



UNIVERSIDADE DE BRASÍLIA

FACULDADE DE CIÊNCIA DA INFORMAÇÃO - FCI

PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA INFORMAÇÃO – PPGCINF

ISABELA PEREIRA CARDOSO

**DADOS CIENTÍFICOS E METADADOS: ESTUDO SOBRE O USO DOS PADRÕES DE
METADADOS NO FLUXO DA INFORMAÇÃO CIENTÍFICA SOBRE BIODIVERSIDADE**

Brasília
2020



UNIVERSIDADE DE BRASÍLIA

FACULDADE DE CIÊNCIA DA INFORMAÇÃO - FCI

PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA INFORMAÇÃO – PPGCINF

ISABELA PEREIRA CARDOSO

DADOS CIENTÍFICOS E METADADOS: ESTUDO SOBRE O USO DOS PADRÕES DE METADADOS NO FLUXO DA INFORMAÇÃO CIENTÍFICA SOBRE BIODIVERSIDADE

Dissertação apresentada ao Programa de Pós-Graduação em Ciência da Informação (PPGCINF) da Universidade de Brasília (UnB) como requisito parcial para obtenção de título de Mestre em Ciência da Informação.

Orientadora: Prof.^a Dra. Fernanda Passini Moreno

Brasília
2020

PC268d Pereira Cardoso, Isabela
Dados científicos e metadados: estudo sobre o uso dos
padrões de metadados no fluxo da informação científica sobre
biodiversidade / Isabela Pereira Cardoso; orientador
Fernanda Passini Moreno. -- Brasília, 2020.
429 p.

Dissertação (Mestrado - Mestrado em Ciência da Informação)
-- Universidade de Brasília, 2020.

1. Uso de metadados. 2. Padrões de metadados. 3. Gestão
de dados. 4. Comunicação Científica. 5. Biodiversidade. I.
Passini Moreno, Fernanda, orient. II. Título.

**FOLHA DE APROVAÇÃO**

Título: "DADOS CIENTÍFICOS E METADADOS: ESTUDO SOBRE O USO DOS PADRÕES DE METADADOS NO FLUXO DA INFORMAÇÃO CIENTÍFICA SOBRE BIODIVERSIDADE"

Autor (a): Isabela Pereira Cardoso

Área de concentração: Gestão da Informação

Linha de pesquisa: Organização da Informação

Dissertação submetida à Comissão Examinadora designada pelo Colegiado do Programa de Pós-graduação em Ciência da Informação da Faculdade em Ciência da Informação da Universidade de Brasília como requisito parcial para obtenção do título de **MESTRE** em Ciência da Informação.

Dissertação aprovada remotamente em: 07 de outubro 2020.

Presidente (UnB/PPGCINF): Dra. Fernanda Passini Moreno

Membro Externo (FCI/UnB): Dra. Michelli Pereira da Costa

Membro Interno (UnB/PPGCINF): Dr. Marcio de Carvalho Victorino

Suplente (UnB/PPGCINF): Dr. Fernando César Lima Leite

Em 02/09/2020.



Documento assinado eletronicamente por **Fernanda Passini Moreno, Professor(a) de Magistério Superior da Faculdade de Ciência da Informação**, em 09/10/2020, às 11:35, conforme horário oficial de Brasília, com fundamento na Instrução da Reitoria 0003/2016 da Universidade de Brasília.



Documento assinado eletronicamente por **Marcio de Carvalho Victorino, Professor(a) de Magistério Superior da Faculdade de Ciência da Informação**, em 15/10/2020, às 15:30, conforme horário oficial de Brasília, com fundamento na Instrução da Reitoria 0003/2016 da Universidade de Brasília.



Documento assinado eletronicamente por **Michelli Pereira da Costa, Professor(a) de Magistério Superior da Faculdade de Ciência da Informação**, em 16/10/2020, às 10:57, conforme horário oficial de Brasília, com fundamento na Instrução da Reitoria 0003/2016 da Universidade de Brasília.



A autenticidade deste documento pode ser conferida no site http://sei.unb.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador 5668253 e o código CRC C7B9A40F.

Ao Miguel e ao João Pedro

Agradecimentos

A minha família, pelo incentivo constante e total apoio.

Especialmente, preciso agradecer a minha mamãe, que cuidou do meu bebê, meu filho Miguel, em todos os momentos que precisei assistir aulas e estudar. Mamãe, você é minha grande inspiração de força de vontade e sempre me ajuda a superar meus limites. Obrigada Mãezinha.

Ao meu marido, Rogério, muito obrigada por ajudar a coletar os dados da pesquisa e por demonstrar sua admiração por mim na forma de apoio.

Agradeço ao meu irmão Gustavo por diversas vezes me acalmar, sempre falando com amor que tudo vai dar certo.

Agradeço ao meu pai, que sempre gostou de estudar e me ensinou o valor e importância do pensamento crítico. Do meu pai herdei o amor pela área ambiental e espero ser capaz de dar continuidade ao legado de contribuições que ele construiu ao longo de sua carreira profissional.

A minha orientadora, Fernanda Passini Moreno, por quem tenho imensa admiração e respeito. Muito obrigada por aceitar o desafio de orientar uma bióloga na descoberta da Ciência da Informação.

Por fim, agradeço o incentivo e apoio do Ibama, instituição que tenho orgulho e o privilégio de trabalhar.

A todos, muito obrigada.

RESUMO

Há uma demanda crescente por acesso aos dados de pesquisa nas comunidades científicas e também em diferentes esferas do governo e segmentos da sociedade, tendo em vista a valorização desse recurso para melhor fundamentação de análises e decisões. Os avanços científicos e tecnológicos facilitam a execução de tarefas, acelerando o processamento, a comunicação e o crescimento exponencial da informação. Esses avanços conduziram à explosão digital da informação com uma expansão acelerada do volume de dados de pesquisa. Com isso aumenta também a necessidade de promover a integração, uso e circulação adequada desses recursos informacionais. Apesar da rapidez em que ocorrem os avanços tecnológicos, os processos tradicionais que organizam o fluxo da informação e conhecimento científicos precisam ser aperfeiçoados para promover maior disponibilidade e legibilidade à pesquisa científica. É cada vez mais necessário o direcionamento de esforços para aperfeiçoar a gestão, comunicação e governança da informação científica, por meio de modelos que valorizem os processos envolvidos no tratamento dos dados pesquisa. Considerando essa tendência, o presente estudo analisa a o uso dos metadados para descrever os dados de pesquisas. Devido à dimensão expressiva desse universo de dados, o escopo dessa dissertação restringe-se aos metadados referentes à biodiversidade. Para tanto, foram identificados os padrões de metadados amplamente recomendados e utilizados pelas comunidades de pesquisa e comparados a evidências de uso dos metadados extraídos de artigos da literatura científica sobre o tema e dos registros de metadados de conjuntos de dados disponíveis no portal da rede *Global Biodiversity Information Facility* (GBIF). Do ponto de vista metodológico, trata-se de uma pesquisa qualitativa e descritiva, com estratégia de coleta de dados bibliográfica e documental. As buscas por artigos e registros dos metadados foram embasadas na adaptação do método de recuperação de informação realizada pouco a pouco (*berrypicking*). A partir dos resultados encontrados delineou-se o panorama uso de metadados no registro dos dados da pesquisa científica sobre biodiversidade. Além disso, foram identificadas tendências voltadas ao aperfeiçoamento na gestão, comunicação e governança dos dados de pesquisa sobre biodiversidade, já adotadas ou em processo de internalização por diversas comunidades de pesquisa envolvidas nessa temática.

Palavras-chave

Dados de pesquisa; Uso de metadados; Padrões de metadados; Gestão de dados; Comunicação Científica; Biodiversidade.

ABSTRACT

Making data accessible allows further research, provides information for decision-making and contributes to transparency in science. Ensuring data is properly managed and shared is the key to a high-quality research environment in terms of scientific breakthroughs, generating world-leading research and enabling more international research collaborations. However, making data accessible, understandable and truly reusable remains a challenge. Publishing datasets is a time-consuming process that is often seen as a courtesy, rather than a necessary step in the research process. Acknowledging such a significant effort involving the management and publication of a dataset remains a flimsy, not well established practice in the scientific community. Over the last decade, many biodiversity informatics initiatives at global, regional and local scales have emerged with a clear goal: to compile and share data, making science opens worldwide. This study presents a review about the scientific communication of the global biodiversity datasets analyzed under the perspective of metadata use throughout the scientific information flow. The proposed methodology aims to establish a baseline by identifying the internationally recommended metadata standards and compare it with existent metadata evidences. This study is grounded in a qualitative and descriptive research of metadata evidences that was guided by the berry picking method. Furthermore, the systematic review and meta-synthetic analysis were defined as methods to collect and analyze metadata evidences extracted from scientific papers and metadata records of Global Biodiversity Information Facility (GBIF) datasets. By investigating literature papers and metadata records of biodiversity datasets the metadata use was outlined in a comparative analysis view to recommended metadata standards. It was also identified best practices that have been adopted in the management, communication and governance of biodiversity research data.

Keywords: *Research data. Metadata use. Metadata standards. Data Management. Scientific Communication. Biodiversity.*

LISTA DE FIGURAS

Figura 1. Estrutura dos dados científicos.....	24
Figura 2. Características dos metadados.....	46
Figura 3. Bases da Ciência de Dados.....	57
Figura 4. Modelo simplificado da web semântica.....	64
Figura 5. Representação de uma tripla.....	65
Figura 6. Delineamento do escopo e etapas da pesquisa.....	72
Figura 7. Relação entre estratégias metodológicas, métodos e instrumentos.....	73
Figura 8. Relação entre estratégias metodológicas e etapas da pesquisa.....	81
Figura 9. Modelo conceitual da pesquisa.....	85
Figura 10. Estrutura do <i>Darwin Core Archive</i> (DwC-A).....	98
Figura 11. Países integrantes da rede GBIF.....	102
Figura 12. Processo de seleção e coleta de artigos científicos.....	120
Figura 13. Quantidade de publicações por década.....	123
Figura 14. Quantidade de publicações por assunto associado ao uso de metadados.....	169
Figura 15. Quantidade de publicações por década e por assunto associado ao uso de metadados.....	170
Figura 16. Quantidade de artigos por área do conhecimento do periódico.....	173
Figura 17. Frequência em que uso de metadados' é assunto principal ou parte da metodologia.....	174
Figura 18. Quantidade de artigo com referências a cada padrão de metadados.....	175
Figura 19. Quantidade de artigos com referência a cada padrão por década.....	176
Figura 20. Outras referências de organização de dados identificadas nos artigos.....	177
Figura 21. Quantidade de conjuntos de dados por continente.....	180
Figura 22. Quantidade de instituições por continente.....	186
Figura 23. Distribuição dos registros de ocorrências do conjunto de dados EOD em dezembro de 2019.....	187
Figura 24. Abrangência dos dados depositados na GBIF nos meses de junho e julho de 2020.....	187
Figura 25. Quantidade de conjuntos de dados depositados por ano na GBIF.....	188
Figura 26. Quantidade de conjuntos de dados (por classe e total) agrupados por década considerando a data de depósito.....	189
Figura 27. Distribuição espacial dos dados de ocorrência do ano 1600.....	190
Figura 28. Distribuição espacial dos dados de ocorrência do período entre os anos 1600 a 1700.....	191
Figura 29. Distribuição espacial dos dados de ocorrência do ano de 2020.....	191
Figura 30. Quantidade de elementos de metadados em uso por percentual de <i>datasets</i> (classes evento e ocorrência) – decis.....	202
Figura 31. Quantidade de elementos de metadados em uso por percentual de <i>datasets</i> (classes evento e ocorrência) – quartis.....	203
Figura 32. Quantidade de <i>datasets</i> por percentual de elementos de metadados em uso (classes evento e ocorrência) – decis.....	205

Figura 33. Quantidade de <i>datasets</i> por percentual de elementos de metadados em uso (classes evento e ocorrência) – quartis.	206
Figura 34. Quantidade de elementos de metadados em uso por percentual de <i>datasets</i> (classe <i>checklist</i>) – decis.	209
Figura 35. Quantidade de elementos de metadados em uso por percentual de <i>datasets</i> (classe <i>checklist</i>) – decis.	210
Figura 36. Quantidade de <i>datasets</i> por quantidade percentual de elementos de metadados em uso (classe <i>checklist</i>) - decis.	212
Figura 37. Quantidade de <i>datasets</i> por quantidade percentual de elementos de metadados em uso (classe <i>checklist</i>) – quartis.	213
Figura 38. Panorama na América do Sul: quantidade de <i>datasets</i> e instituições por país.	215
Figura 39. Brasil: quantidade de elementos de metadados em uso por percentual de <i>datasets</i> (classes evento e ocorrência) – decis.	216
Figura 40. Brasil: quantidade de <i>datasets</i> por percentual de elementos de metadados em uso (classes evento e ocorrência) – decis.	217

LISTA DE QUADROS

Quadro 1. Sistematização da história da Catalogação em períodos	37
Quadro 2. Tipologia de Padrões de Dados	39
Quadro 3. Tipos de Metadados.	47
Quadro 4. Diferentes categorias de metadados e suas funções.	49
Quadro 5. Tipos de metadados	50
Quadro 6. Etapa I - Definição da linha de base – levantamento dos padrões de metadados para biodiversidade: relação entre os objetivos específicos e as estratégias metodológicas	82
Quadro 7. Etapa II - Coleta e análise das evidências do uso dos padrões de metadados: relação entre os objetivos específicos e as estratégias metodológicas.	83
Quadro 8. Etapa III: Análise Comparativa e identificação de Tendências: relação entre os objetivos específicos e as estratégias metodológicas.	84
Quadro 9. Padrões de Metadados para Biologia.....	87
Quadro 10. Perfis e extensões de metadados para Biologia.	88
Quadro 11. Ferramentas de registro /recuperação de metadados	90
Quadro 12. Padrões de metadados na temática Biodiversidade	91
Quadro 13. Padrão de metadados ABCD	94
Quadro 14. Padrão de metadados EML	96
Quadro 15. Padrão de metadados <i>Darwin Core</i>	97
Quadro 16. Síntese de informações sobre o perfil de metadados GMP.	100
Quadro 17. Campos obrigatórios e recomendados para a classe metadados de recursos da GBIF.	107
Quadro 18. Metadados obrigatórios e recomendados para conjuntos de dados da classe <i>checklist</i>	111
Quadro 19. Metadados obrigatórios e recomendados para conjuntos de dados da classe dados de ocorrência.....	114

Quadro 20. Metadados obrigatórios e recomendados para conjuntos de dados da classe evento	117
Quadro 21. Estrutura das buscas por artigos científicos: a relação entre os termos abrangentes e específicos utilizados.	119
Quadro 22. Dados de pesquisa sobre biodiversidade: necessidades, possíveis soluções e iniciativas existentes	136
Quadro 23. Quantidade de instituições depositárias e quantidade de conjuntos de dados (<i>datasets</i>) depositados por país na rede GBIF.....	180
Quadro 24. Instituições de abrangência global e sem referência a um único país.	184
Quadro 25. Elemento de metadados utilizados na descrição dos conjuntos de dados.	192
Quadro 26. Frequência do uso do elementos/campos de metadados para descrição dos dados em <i>datasets</i> da classe ocorrência e da classe eventos no portal GBIF.....	195
Quadro 27. Intervalos estabelecidos para análise tendências do uso de metadados nos <i>datasets</i>	201
Quadro 28. Frequência do uso do elementos/campos de metadados para descrição dos dados em <i>datasets</i> da classe <i>checklist</i> no portal GBIF.	207
Quadro 29. Identificação dos padrões relacionados aos elementos de metadados em uso nas descrição dos <i>datasets</i>	219
Quadro 30. Identificação dos padrões relacionados aos elementos de metadados em uso para descrição dos dados de pesquisa das classes eventos e ocorrência	222
Quadro 31. Identificação dos padrões relacionados aos elementos de metadados em uso para descrição dos dados de pesquisa da classe <i>checklist</i>	231
Quadro 32. Comparação dos assuntos associados ao uso de metadados identificados nos artigos científicos e as iniciativas globais das comunidades de pesquisa	233

SUMÁRIO

1	INTRODUÇÃO	14
1.1	AVANÇO DA TECNOLOGIA E A GÊNESE DIGITAL DA INFORMAÇÃO	17
1.2	FLUXO DE INFORMAÇÃO CIENTÍFICA E A EXPLOSÃO DIGITAL DE DADOS	21
1.3	O QUE É BIODIVERSIDADE E QUAIS SÃO AS DEMANDAS POR INFORMAÇÃO?	25
1.4	DEFINIÇÃO DO PROBLEMA DA PESQUISA	28
1.5	OBJETIVOS	29
1.5.1	OBJETIVO GERAL	29
1.5.2	OBJETIVOS ESPECÍFICOS	29
1.6	JUSTIFICATIVA	29
2	REVISÃO DE LITERATURA	31
2.1	GESTÃO E COMUNICAÇÃO DA INFORMAÇÃO CIENTÍFICA	32
2.1.1	COMUNICAÇÃO CIENTÍFICA	32
2.1.2	GESTÃO DA INFORMAÇÃO	35
2.2	ORGANIZAÇÃO DA INFORMAÇÃO	35
2.2.1	O QUE SÃO METADADOS?	41
2.2.2	TIPOS DE METADADOS	47
2.2.3	PADRÕES DE METADADOS	51
2.3	BIODIVERSIDADE E A GOVERNANÇA DOS DADOS CIENTÍFICOS	54
2.3.1	GOVERNANÇA DE RECURSOS INFORMACIONAIS HETEROGÊNEOS	58
2.3.2	MODELO DE GOVERNANÇA DOS COMUNS	59
2.3.3	WEB SEMÂNTICA	61
2.3.3.1	Modelo simplificado da web semântica	63
2.3.3.2	Gestão de dados na web semântica: aspectos sociais e tecnológicos.....	67
3	PROCEDIMENTOS METODOLÓGICOS	71
3.1	CARACTERIZAÇÃO DA PESQUISA	71
3.2	UNIVERSO DA PESQUISA E HORIZONTE TEMPORAL	73
3.3	ESTRATÉGIAS METODOLÓGICAS	73
3.3.1	ESTRATÉGIA DE COLETA DO MATERIAL BIBLIOGRÁFICO E DOCUMENTAL	74
3.3.1.1	Documentação dos padrões e perfis de metadados para biodiversidade.....	75
3.3.1.2	Artigos científicos sobre o uso de metadados	77
3.3.1.3	Documentação dos metadados dos conjuntos de dados	77
3.3.2	ESTRATÉGIA DE COLETA E ESTRUTURAÇÃO DE DADOS	78
3.3.2.1	Roteiro de estudo dirigido para coleta de dados nos artigos científicos.....	78
3.3.2.2	Lista de verificação para coleta de metadados dos conjuntos de dados.....	78
3.3.3	ESTRATÉGIA DE ANÁLISE DOS DADOS	78
3.4	ETAPAS DA PESQUISA	81
3.4.1	ETAPA I – IDENTIFICAÇÃO E ANÁLISE DOS PADRÕES DE METADADOS	82
3.4.2	ETAPA II – COLETA E ANÁLISE DE EVIDÊNCIAS DO USO DE METADADOS	82
3.4.3	ETAPA III – ANÁLISE DE TENDÊNCIAS NA GESTÃO DE DADOS DE PESQUISA	83
3.5	MODELO CONCEITUAL DA PESQUISA	84
4	RESULTADOS E DISCUSSÃO	87
4.1	ETAPA I – PADRÕES DE METADADOS PARA BIODIVERSIDADE	87

4.1.1	ACCESS TO BIOLOGICAL COLLECTION DATA	92
4.1.2	ECOLOGICAL METADATA LANGUAGE	94
4.1.3	<i>DARWIN CORE</i>	96
4.1.3.1	<i>Darwin Core Archive (DwC-A)</i>	97
4.1.4	GBIF METADATA PROFILE (GMP)	99
4.1.4.1	Rede GBIF: características, abrangência e principais iniciativas	100
4.1.4.2	Ferramentas disponibilizadas pela GBIF	102
4.1.4.3	<i>Data papers</i>	104
4.1.5	LINHA DE BASE: PARÂMETROS PARA AVALIAR O USO DOS METADADOS	104
4.1.5.1	Classes de conjuntos de dados	105
4.1.5.2	Metadados de recursos – <i>resources metadata</i>	106
4.1.5.3	Lista de verificação de espécies – <i>checklist data</i>	109
4.1.5.4	Dados de Ocorrência.....	113
4.1.5.5	Dados de eventos de amostragem	115
4.2	ETAPA II – EVIDÊNCIAS DE USO DE METADADOS	118
4.2.1	COLETA DE DADOS NOS ARTIGOS CIENTÍFICOS	118
4.2.2	ANÁLISE DOS DADOS COLETADOS NOS ARTIGOS CIENTÍFICOS	122
4.2.2.1	1990 - 1999	123
4.2.2.2	2000 - 2009	128
4.2.2.3	2010 - 2019	140
4.2.2.4	Síntese integrativa dos dados coletados nos artigos	169
4.2.3	COLETA DOS METADADOS DOS CONJUNTOS DE DADOS DA BIODIVERSIDADE	178
4.2.4	ANÁLISE DOS METADADOS DE CONJUNTOS DE DADOS	179
4.2.4.1	Abrangência espacial	179
4.2.4.2	Abrangência temporal	188
4.2.4.3	Campos de metadados usados para a descrição dos conjuntos de dados	192
4.2.4.4	Campos de metadados usados para a descrição dos dados da biodiversidade	194
4.2.4.5	Contribuições do Brasil	214
4.3	ETAPA III – IDENTIFICAÇÃO DE TENDÊNCIAS	218
5	CONCLUSÕES	236
6	REFERÊNCIAS BIBLIOGRÁFICAS	239
	APÊNDICES	256
	ANEXOS	292

1 INTRODUÇÃO

Acessar a informação científica é uma demanda crescente, não apenas para a comunidade acadêmica, mas também como suporte às análises e aos processos decisórios nas diferentes esferas do governo e nos mais diversos segmentos da sociedade. Entretanto, o fornecimento ou ampliação desse acesso depende da superação dos desafios referentes à organização, gestão, comunicação e governança da informação, assuntos já discutidos pela Ciência da Informação (CI) há pelo menos meio século. Abordada em diferentes perspectivas, a preocupação com essas questões tem sido amplamente difundida por vários autores desde o fenômeno denominado explosão da informação, ocorrida no período após a 2ª Guerra Mundial. Com os avanços tecnológicos das últimas décadas, surgiu também o fenômeno da explosão digital da informação, que intensificou a necessidade de superar esses e outros obstáculos de forma a evitar a duplicação de esforços e a desaceleração do progresso na ciência.

A explosão digital da informação conduziu a um maior reconhecimento do “potencial informacional dos dados de pesquisa para a ciência contemporânea e transformou a visão que os caracterizava como simples subprodutos dos processos de pesquisa” (SAYÃO e SALES, 2015, p. 5). Essa valorização se deu principalmente devido ao aumento da demanda por transparência nas pesquisas como meio cada vez mais necessário para impulsionar novos estudos, gerar conhecimento e subsidiar decisões que demandam análises complexas envolvendo grande quantidade de dados. Em diversas áreas há o registro de iniciativas voltadas ao aperfeiçoamento do fluxo de informação científica e melhor comunicação dos dados de pesquisa, como o surgimento de políticas institucionais de gestão de dados ou políticas de depósito de dados para publicação em periódicos científicos.

Diante dessa tendência de valorização dos dados, este estudo visa analisar como se dá o uso dos metadados para a descrição dos dados de pesquisa ao longo do fluxo de informação científica. Devido à dimensão expressiva do universo dos dados pesquisa decorrente de diversas áreas do conhecimento, o escopo foi delimitado ao uso dos metadados para conjuntos de dados de pesquisa da biodiversidade.

A ampliação da demanda por acesso aos dados de pesquisa da biodiversidade teve sua discussão intensificada diante da necessidade de criar instrumentos de análise e

monitoramento capazes de auxiliar no estabelecimento de políticas globais voltadas a reduzir os impactos ambientais, desequilíbrio ambiental, extinção de espécies, perda de *habitats* e outras ações necessárias ao desenvolvimento sustentável. Apesar da rapidez em que ocorrem os avanços tecnológicos, diversos autores apontam que há uma inércia do sistema tradicional de comunicação científica que limita a disponibilidade, acesso e descoberta da maior parte da pesquisa relacionada ao tema (HAIGEN XU, ZHENNING GAO, *et al.*, 1999; KWANG-TSAO SHAO, CHING-I PENG, *et al.*, 2007; BAIRD, 2010; BERENDSOHN, CHAVAN e MACKLIN, 2010; VOLLMAR, MACKLIN e FORD, 2010; BENDIX, NIESCHULZE e MICHENER, 2012; DAVIS, TENOPIR, *et al.*, 2014). As tecnologias que conduziram à digitalização da informação também têm facilitado à expansão acelerada do volume de dados científicos produzidos. Consequentemente, aumenta a necessidade de promover a integração, uso e circulação adequada desses recursos.

Existe amplo consenso nas comunidades de pesquisa da biodiversidade sobre a necessidade de promover o acesso livre, aberto e qualificado aos dados de pesquisa (MORITZ, KRISHNAN, *et al.*, 2011; BERENDSOHN, CHAVAN e MACKLIN, 2010; ROCHE, KRUIK, *et al.*, 2015; SILVA, CORRÊA, *et al.*, 2015; WILKINSON, DUMONTIER, *et al.*, 2016; BRASIL. MCTIC. PPBIO, 2016; WILKINSON, VERBORGH, *et al.*, 2017; GBIF, 2018b). Entretanto, o acesso efetivo a esses recursos ainda demanda superação de muitos desafios, que envolvem desde questões comportamentais dos pesquisadores a problemas relacionados à gestão dos dados.

A publicação de conjuntos de dados é um processo que consome tempo e, muitas vezes, sua disponibilização é vista no ambiente acadêmico como uma cortesia, em vez de um passo necessário ao processo de pesquisa. O tratamento dos dados precisa ser devidamente valorizado, pois há necessidade crescente de direcionamento de esforços significativos para aperfeiçoar a gestão e comunicação dos dados de pesquisa. Estudos sugerem que a retenção de dados pelos pesquisadores pode ser atribuída à falta de conhecimento, à falta de capacidade técnica, à falta de mecanismos adequados para reconhecimento dos esforços dos pesquisadores e às preocupações com retenção por razões de interesse pessoal ou organizacional (MORITZ, KRISHNAN, *et al.*, 2011; BENDIX, NIESCHULZE e MICHENER, 2012; DAVIS, TENOPIR, *et al.*, 2014).

Já os desafios mais comuns relacionados à gestão, organização e comunicação dos dados de pesquisa da biodiversidade referem-se às discrepâncias entre as descrições dos

diversos conjuntos de dados, dificuldades em obter ou sustentar financiamento e a ausência de padrões na estruturação das soluções computacionais desenvolvidas para esses fins (BERENDSOHN e SELTMANN, 2010; BAIRD, 2010; VOLLMAR, MACKLIN e FORD, 2010; HAIGEN XU, ZHENNING GAO, *et al.*, 1999; KWANG-TSAO SHAO, CHING-I PENG, *et al.*, 2007). De forma geral, são problemas que dificultam a descoberta, interoperabilidade, compartilhamento e reuso dos dados de pesquisa (DAVIS, TENOPIR, *et al.*, 2014; BENDIX, NIESCHULZE e MICHENER, 2012; WILKINSON, DUMONTIER, *et al.*, 2016).

Nesse sentido, é importante compreender que a disseminação da informação e do conhecimento em uma comunidade científica depende da rede de comunicação que nela se estabelece, ou seja, como se organizam os seus fluxos de informação e do conhecimento (MORENO, LEITE e ARELLANO, 2006; LEITE, 2011; SAYÃO e SALES, 2015). Assim, aperfeiçoar a compreensão a respeito dos impactos das tecnologias sobre a organização, gestão, comunicação e governança dos diversos produtos gerados no fluxo de informação científica é fundamental para promover melhor uso e circulação do conhecimento. A seguir, é apresentada a estrutura de organização das seções que integram este estudo.

A primeira seção detalha o contexto do avanço tecnológico que conduziu a gênese digital da informação. Decorrente desse cenário, em seguida, apresenta-se os desafios decorrentes da obstrução do fluxo de informação científica potencializados pela explosão digital da informação. Considerando também a temática delimitadora deste estudo, a contextualização é concluída apresentando a relevância dos dados das pesquisas sobre biodiversidade em âmbito mundial. Finalizando essa seção, apresenta-se a delimitação do problema, questão, objetivos e a justificativa da pesquisa.

A segunda seção traz a revisão de literatura, iniciada com a discussão acerca das relações entre a gestão e a comunicação da informação científica. De forma a apresentar os conceitos referentes à temática central, a revisão da literatura também trata da organização da informação até alcançar os conceitos relacionados aos metadados. Por fim, são descritos os desafios referentes à governança de dados sobre biodiversidade, abordando aspectos dos dados heterogêneos e questões relevantes apresentadas no modelo de governança dos comuns conforme proposto por Elinor Ostrom (OSTROM, 1990; OSTROM e HESS, 2007).

A terceira seção apresenta os procedimentos metodológicos deste estudo, incluindo a caracterização da pesquisa e estratégias adotadas na busca, coleta, estruturação e análise dos dados. A quarta seção apresenta os resultados alcançados em cada etapa da pesquisa e

a discussão baseada na análise integrativa dos achados e identificação das tendências relacionadas ao uso de metadados nos dados das pesquisas sobre a biodiversidade. Por fim, a quinta seção traz a conclusão deste estudo.

1.1 AVANÇO DA TECNOLOGIA E A GÊNESE DIGITAL DA INFORMAÇÃO

Os avanços científicos e tecnológicos facilitam a execução de tarefas, acelerando o processamento, a comunicação e o crescimento exponencial da informação. Segundo Le Coadic (2004), a implosão do tempo e a explosão informacional são fenômenos que conduzem a um fluxo de consideráveis quantidades de informação em curto espaço de tempo, cuja necessidade de controle suscita novos desafios à sociedade.

Dentre eventos históricos marcantes, após a Segunda Guerra Mundial, nos primeiros anos da Guerra Fria, os investimentos intensivos em ciência e tecnologia ocasionaram o crescimento exponencial da produção científica e tecnológica, que ficou conhecido como explosão informacional. Desse fato, decorreu a necessidade de buscar soluções para “a tarefa massiva de tornar mais acessível, um acervo crescente de conhecimento” (BUSH, 1995, p. 1). Para Saracevic (1996), esses problemas informacionais existem há um longo tempo e sempre estiveram mais ou menos presentes, mas sua importância real ou percebida modificou completamente sua relevância para a sociedade naquele período e essa mudança foi responsável pelo surgimento da Ciência da Informação.

Le Coadic (2004, p. 19) defende que “toda ciência é uma atividade social determinada por condições históricas e socioeconômicas”. Especificamente, o autor entende que a pesquisa em ciência da informação responde a uma necessidade social, uma vez que a sociedade necessita de uma ciência que estude as propriedades da informação e os processos de sua construção, comunicação e uso. Também Wersig e Neveling (1975) afirmam que “o problema da transferência do conhecimento para aqueles que dele necessitam é uma responsabilidade social e esta responsabilidade social parece ser o motivo real da ciência da informação”.

Dessa forma, a CI possui duas raízes: uma no campo tecnológico e outra no campo social. A raiz tecnológica engloba o conjunto de meios – métodos, técnicas e instrumentos inspirados nas ciências exatas para controlar e aperfeiçoar o sistema de informação. Já a raiz social representa a finalidade dessa ciência, preocupada em tratar a informação para as pessoas dentro de necessidades e contextos. Sobre essa ótica, Capurro (2003) entende que

uma raiz está ligada a todos os aspectos sociais e culturais próprios do mundo humano; e outra é de caráter tecnológico recente.

A primeira raiz nos leva às próprias origens, certamente obscuras, da sociedade humana entendida como um entrelaçamento ou uma rede de relações baseadas na linguagem, isto é, em âmbito hermenêutico aberto, onde os entrecruzamentos metafóricos e metonímicos permitem não apenas manter fluido o mundo das convenções e fixações que tornam possível uma sociedade humana relativamente estável, como também nos permitem gerar a capacidade de perguntar pelo que não sabemos a partir do que cremos que sabemos.

(...)

A outra raiz é de caráter tecnológico recente e se refere ao impacto da computação nos processos de produção, coleta, organização, interpretação, armazenagem, recuperação, disseminação, transformação e uso da informação, e em especial da informação científica registrada em documentos impressos (CAPURRO, 2003).

Segundo Araújo (2003), a CI não surgiu como uma ciência social, pois inicialmente estava mais voltada à tecnologia e à necessidade de recuperação automática da informação e, apenas nos anos 70, houve sua inscrição efetiva nas ciências sociais. O autor ainda explica que, logo no início, a ciência da informação aproxima-se exatamente do modelo positivista e funcionalista e, como já vinha de uma formalização muito próxima à das ciências exatas, a CI encontrou, no conjunto de ciências sociais também próximas às ciências exatas, maior identidade de propostas e métodos.

é por pretender se aproximar da ‘ciência pós-moderna’, superando os limites do modelo até então dominante, buscando superar seus impasses metodológicos simplificadores e abarcar um pensamento pautado pela complexidade, que a ciência da informação evolui para novas etapas de diálogo e inserção nas ciências sociais (ARAÚJO, 2003, p. 26).

Capurro (2003) entende que essas “(...) duas raízes: uma é a biblioteconomia clássica ou, em termos mais gerais, o estudo dos problemas relacionados com a transmissão de mensagens, sendo a outra a computação digital”. Essas duas facetas ou ‘raízes’ da CI – social e tecnológica – são complementares e necessárias à compreensão dos fenômenos informacionais. As demandas sociais referentes à informação recebem suporte das tecnologias que surgem e, também, são geradas ou potencializadas devido aos novos cenários proporcionados pelas inovações. Exemplo disso pode ser observado no fenômeno da digitalização, que representa um suporte fundamental a disseminação, mas ao mesmo tempo, por facilitar o processo de geração, o volume de informação cresce exponencialmente e traz desafios referentes ao seu controle e recuperação. Nesse sentido, Araújo (2003) apresenta como ponto para discussão:

A reflexão sobre a evolução da ciência da informação, suas relações com as ciências sociais e com o modelo moderno da ciência como um todo, é fundamental para a

realização de pesquisas nessa área que efetivamente incorporem todo o avanço acumulado nesse processo – e que não apenas fiquem no discurso de uma mudança paradigmática, sem concretamente realizá-la (ARAÚJO, 2003, p. 26-27).

Dessa forma, os mesmos avanços tecnológicos que permitem o encurtamento acentuado do tempo para a execução de tarefas e o crescimento exponencial de dados, também geram cada vez mais inquietações relacionadas aos dilemas da sociedade diante da necessidade de extrair informações e conhecimento, bem como de lidar com novos fenômenos, tais como a exclusão informacional, baixa democratização da informação, más condições de acesso, obstrução do fluxo de informação científica e outros. Assim, é relevante o desenvolvimento de estudos que promovam a valorização e aproximação dessas duas raízes da CI, entendendo os aspectos sociais como os objetivos e a tecnologia como meio de alcançá-los.

No contexto contemporâneo, verifica-se que o caráter tecnológico apontado por Capurro (2003) têm cada vez mais ganhado espaço na CI, pois a geração massiva de dados em todas as áreas do conhecimento vem aumentando e, assim, se mantém atual a necessidade de encontrar formas para acessar e controlar as informações e gerar conhecimento por meio de recursos tecnológicos. Conforme aponta Leite (2011, p. 8), “o fenômeno da gênese digital da informação além de reproduzir problemas similares ao da explosão informacional no pós-guerra, como a recuperação da informação, traz à tona novos desafios, como o da preservação digital e das transformações provocadas pela Internet”. Para Silva e Tomaél (2009), nas “três últimas décadas, a exemplo dos demais campos da ciência, o campo da política da informação atravessou cenários repletos de transformações provocadas pela emergência e profusão das tecnologias que abarcam todas as esferas da vida social”.

Hilbert e Lopez (2012) apresentam a evolução da comunicação e armazenamento de dados ao longo dos anos e, segundo esse estudo, no ano 1986 apenas 1% da informação produzida no mundo era comunicada, incluindo todos os recursos disponíveis tais como papel, filmes, vinis, fitas magnéticas. O percentual de informação comunicada cresceu para 5% no ano 2000 e para 16% em 2007 (HILBERT e LOPEZ, 2012, p. 959). Esse estudo também aponta que em 2000, do total de informação comunicada, apenas 25% era digital.

Adicionalmente, Costa e Cunha (2014, p. 191) apontam que 2002 foi o primeiro ano em que os dados digitais passaram a representar a maior parte da informação armazenada e

comunicada, marcando o início da era da informação digital. Ainda segundo os autores, em 2007 apenas 7% dos dados armazenados eram analógicos e a previsão para 2013 era de que o volume de informação armazenada no mundo equivaleria a 1200 *exabytes*, sendo que destes, apenas 2% seria analógico.

Superando essa estimativa apresentada para 2013, o relatório da *International Data Corporation* – IDC (2012) aponta que já em 2012, 2.8 *Zettabytes*¹ de dados digitais foram criados, dos quais menos de 1% foi analisado e menos de 20% são protegidos. Também segundo esse relatório, o universo digital apresenta a tendência de dobrar a cada dois anos até 2020. A IDC ainda ressalta que um fator importante da expansão do universo digital é o crescimento de dados gerados por ‘máquinas’, que aumentarão de 11% do universo digital em 2005 para mais de 40% até 2020.

Destaca-se que esse aumento de dados resultantes de instrumentos – ‘máquinas’, telescópios, satélites, sensores especializados dentre outros – é cada vez mais percebido, devido ao aperfeiçoamento constante e a capacidade que esses equipamentos possuem para gerar e armazenar dados em uma escala sem precedentes, que vai muito além da capacidade humana de análise. Esse fenômeno fez surgir o termo *e-Science*, que se refere ao conjunto de instrumentos e tecnologias necessárias para apoiar a pesquisa científica do Século XXI, considerando a massa crescente de dados e sua intrínseca natureza colaborativa e multidisciplinar. Segundo a revisão realizada por Costa e Cunha (2014, p. 192), dentre as denominações utilizadas para *e-Science*, destacam-se na literatura os termos: ciência orientada a dados, computação fortemente orientada a dados, ciberinfraestrutura ou quarto paradigma. Ainda segundo Costa e Cunha (2014, p. 193), a *e-Science* traz implicações relevantes sobre a comunicação científica, afinal dela resultam de fato dados científicos primários, que geram efeitos nos serviços e produtos de informação, afetando diretamente as bibliotecas digitais e exigindo reflexões sobre preservação digital e o planejamento das bases de dados.

¹ Um byte (*binary term*) ou octeto é um dos tipos de dados integrais em computação usado com frequência para especificar o tamanho da memória ou da capacidade de armazenamento de certo dispositivo, independentemente do tipo de dados. **Zettabyte (ZB)** ou Zebibyte é a medida de armazenamento que corresponde a 1.000.000.000.000.000.000 (10²¹) ou 1180591620717411303424 (2⁷⁰) Bytes, dependendo do contexto. Para não haver confusão, foi criada uma nova nomenclatura para diferenciar valores em base 10 e os em base 2, sendo esta última Zebibyte. Para fins de compreensão da dimensão que representa essa grandeza, segue a comparação: 1 ZB (2⁷⁰ Bytes) também equivale a 1.024 Exabytes, 1.048.576 Petabytes, ou 1.073.741.800 Terabytes.

Entretanto, considerando essa massa de dados em constante crescimento, Isotani e Bittencourt (2015, p. 16) destacam:

Contudo, grande parte destes dados não está disponível ao público. Estes dados tampouco estão estruturados para facilitar sua compreensão mesmo por aqueles que podem acessá-los e manipulá-los. Como resultado, a extração de informações e a produção de conhecimentos que poderiam ser úteis para a sociedade não acontecem com a agilidade e a eficácia necessárias para lidar com as questões sociais e econômicas do século 21.

A Fundação Americana de Ciência (EUA.NSF, 2007) aponta que a liderança em ciência dependerá cada vez mais de capacidade de alavancar os reservatórios de dados capturados em formato digital e transformar em informação e conhecimento, auxiliados por ferramentas sofisticadas de mineração, integração, análise e visualização de dados. Assim, considerando a produção de informação científica, bem como o compartilhamento e o uso do conhecimento como fatores determinantes para o desenvolvimento científico, também é relevante compreender o contexto da explosão digital de dados e seus impactos no fluxo da informação científica.

1.2 FLUXO DE INFORMAÇÃO CIENTÍFICA E A EXPLOSÃO DIGITAL DE DADOS

Os desafios referentes à circulação da informação científica são abordados por diversos autores na CI. Em 1895, Paul Otlet e Henri la Fontaine tinham por objetivo o desenvolvimento de um grande movimento cooperativo, em nível planetário, para que fosse estabelecida uma espécie de inventário de toda a produção humana de conhecimento registrado (ARAÚJO, 2014, p. 100). Segundo Capurro (2003), na bibliografia universal de Paul Otlet e Henri Lafontaine está explícita a intenção distinguir o conhecimento e o seu registro em documentos.

Em 1968, Borko (1968) demonstrou sua preocupação com os impactos dos avanços tecnológicos para a CI, segundo o autor:

A ciência da informação não acompanhou os outros avanços científicos, e agora é preciso concentrar esforços nesse campo e recuperar o atraso. Se os procedimentos de comunicação e troca de informações não forem melhorados, todos os outros trabalhos científicos serão impedidos; a falta de comunicação resultará em uma duplicação de esforços e uma desaceleração do progresso (BORKO, 1968, p. 2 tradução livre).

Ainda segundo Borko (1968) isso pode ser percebido por meio da observação das seguintes pressões:

1. O tremendo crescimento da ciência e tecnologia e o ritmo acelerado com que novos conhecimentos tornam-se disponíveis e os conhecimentos antigos se tornam obsoletos;
2. A rápida taxa de obsolescência do conhecimento técnico, de modo que o antigo graduado deve voltar para a escola e atualizar suas habilidades;
3. O grande número de cientistas que trabalham e o grande número de revistas científicas e técnicas que existem;
4. O aumento da especialização que torna a comunicação e troca de informações entre disciplinas muito difíceis;
5. O curto intervalo de tempo entre pesquisa e aplicação, que torna a necessidade de informações mais premente e mais imediata. (BORKO, 1968, p. 2 tradução livre).

No contexto contemporâneo, esses problemas, já apontados por Borko, são cada vez mais percebidos e a eles foram somadas novas questões, tais como a necessidade de análise e extração de informações da gigantesca massa digital de dados gerada. Segundo Barreto (1998, p. 125), a comunicação eletrônica modifica estruturalmente o fluxo de informação e conhecimento, atuando basicamente nos seguintes pontos:

- A interação do receptor com a informação: o receptor da informação deixa a sua posição de distanciamento alienante em relação ao fluxo de informação e passa a participar de sua fluidez como se estivesse posicionado em seu interior. Sua interação com a informação é direta, conversacional e sem intermediários;
- Tempo de interação: o receptor conectado on-line está desenhando a sua própria interação com o fluxo de informação em tempo real, isto é, com uma velocidade que reduz o tempo de contato ao entorno de zero. Essa velocidade de acesso e uso o coloca em nova dimensão para o julgamento de valor da informação; o receptor passa a ser o julgador de relevância da informação acessada em tempo real, no momento de sua interação e não mais em uma condição ex-post de retroalimentação intermediada;
- A estrutura da mensagem: em um mesmo documento, o receptor pode elaborar a informação em diversas linguagens, combinando texto, imagem e som. Não está mais preso a uma estrutura linear da informação, que passa a ser associativa em condições de um hipertexto “As we may think” (como pensamos), diria Vanevar Bush em artigo histórico. Cada receptor interage com o texto da mensagem circularmente, e cria o seu próprio documento com a intencionalidade de uma percepção orientada por sua decisão;
- A facilidade de ir e vir: a dimensão de seu espaço de comunicação é ampliada por uma conexão em rede, o receptor passeia por diferentes memórias ou estoques de informação no momento de sua vontade.

Capurro (2003) entende que o caráter tecnológico tem criado novos desafios, pois a geração massiva de dados em todas as áreas do conhecimento continua aumentando e, assim, se mantém atual a necessidade de encontrar formas para comunicar e gerar conhecimento. Capurro (2003) também aponta que a “documentação e, em seguida, a Ciência da Informação têm a ver, aparentemente, em primeiro lugar com os suportes físicos do conhecimento, mas na realidade sua finalidade é a recuperação da própria informação, ou seja, o conteúdo de tais suportes”.

Nesse sentido, Leite (2011, p. 1) explicita que há uma relação cíclica, pois o avanço científico e tecnológico de um país depende da infraestrutura de informação e comunicação científica nele estabelecida e, tal infraestrutura, por sua vez, contribui diretamente para que pesquisadores sejam capazes de alcançar resultados de pesquisa. Adicionalmente, Leite (2011, p. 3) aponta que há uma obstrução do fluxo de informação que, dentre outros fatores, decorre da indisponibilidade (pesquisadores retêm os dados) e da baixa rastreabilidade dos dados científicos. Isso interfere na confiabilidade dos dados e resultados, conforme exposto por Spang-Hanssen (2001):

nem a pessoa que solicita informações nem a pessoa que as entrega devem ignorar a confiabilidade dos dados e esquecer a configuração geral na qual os dados são adquiridos. As informações sobre alguma propriedade física de um material estão incompletas sem informações sobre a precisão dos dados e sobre as condições sob as quais esses dados foram obtidos. Além disso, várias investigações de uma propriedade muitas vezes levaram a resultados diferentes que não podem ser comparados e avaliados separadamente das informações sobre seus antecedentes.

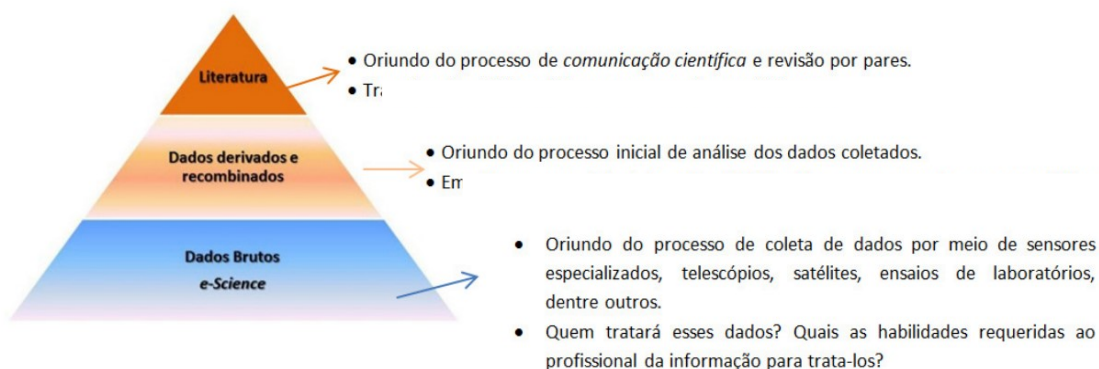
Leite (2011, p. 1) explica que o sistema de comunicação da informação científica é responsável pelos fluxos que alimentam e que resultam das atividades de pesquisa, os quais se constituem, concomitantemente, entrada e saída dos processos de produção do conhecimento. Dessa forma, a informação científica tem seu fluxo promovido por processos de comunicação científica. Por isso, ainda segundo Leite (2011, p. 2) existem três questões basilares que têm exercido influência sobre essa dinâmica cíclica do sistema de comunicação da informação científica e sua efetividade em escala global:

1. Volume crescente e exponencial de informação digital em rede e à falta de controle decorrente disso;
2. O sistema de comunicação científica atual limita, mais do que expande, a disponibilidade e legibilidade de maior parte dos resultados da pesquisa, obstruindo o fluxo da informação científica;
3. As transformações das atividades de produção do conhecimento científico decorrentes, sobretudo, do uso crescente de tecnologias de informação e comunicação, além de impactarem na maneira como a pesquisa científica é conduzida, requerem mudanças nos modos como seus resultados são disseminados, acessados e utilizados.

De forma mais ampla, “a explosão bibliográfica, fenômeno comum a todas as áreas do conhecimento e talvez a característica mais visível das literaturas científicas, pode ser definida como a quantidade crescente de documentos científicos produzidos e a rapidez com que esse número aumenta” (MUELLER, 2000, p. 24). Restringindo essa perspectiva para compreender os impactos no universo dos dados científicos, é relevante a proposição apresentada por Costa e Cunha (2014, p. 194) para a categorização em três grupos: 1) os

dados analisados e comunicados; 2) os dados coletados em processamento; e 3) os dados brutos (coletados e não analisados, ou coletados automaticamente por equipamentos), conforme apresentado na figura 1.

Figura 1. Estrutura dos dados científicos.



Fonte: Adaptado de Costa e Cunha (2014, p. 194)

Considerando a necessidade de gestão e governança dessa massa de dados, ainda não há consenso sobre quais são os atores responsáveis, papéis, etapas e melhores procedimentos para explorar o expressivo e crescente universo de dados produzidos. Para Bell (2011, p. 11) os dados científicos devem “permanecer para sempre num estado submetido à curadoria e acessível para o público para análise contínua”. Tudo isso conduz a um novo capítulo de pesquisa científica, caracterizada por uma quantidade esmagadora de dados, trazendo como paradigma o fato de que os dados de pesquisa passaram a ocupar uma posição central (LI, GREENBERG e DUNIC, 2020).

Conforme apresentado até aqui, desde o surgimento da CI, diversos autores demonstraram preocupação com o fenômeno da explosão da informação e, mais recentemente, com a explosão digital de dados e fluxo da informação científica. Nesse sentido, a presente pesquisa visa analisar os padrões e o uso de metadados ao longo do fluxo de informação científica. Para tanto, considera conceitos e aspectos referentes à governança, gestão e comunicação de dados científicos, detalhados na revisão de literatura deste estudo. Além disso, diante da diversidade e abrangência dos campos do conhecimento que geram dados científicos, optou-se por restringir o estudo à análise do uso dos metadados na pesquisa científica sobre biodiversidade, cujo contexto e relevância são detalhados a seguir.

1.3 O QUE É BIODIVERSIDADE E QUAIS SÃO AS DEMANDAS POR INFORMAÇÃO?

A preocupação com o uso adequado e a conservação dos recursos naturais vem aumentando em todo o mundo desde o século passado. Por isso, diferentes comunidades científicas e segmentos da sociedade estão envolvidos em pesquisas e discussões sobre o conhecimento, a conservação e o uso sustentável da biodiversidade e do ambiente onde vivem esses organismos.

Documentos oriundos de diferentes convenções e reuniões internacionais têm expressado preocupação com a persistência e a velocidade dos impactos negativos sobre espécies e sistemas ecológicos responsáveis pelo fornecimento de bens e serviços indispensáveis para a humanidade (BRASIL. MCTIC. PPBIO, 2016, p. 29-30). Os padrões atuais de produção, consumo e ocupação do solo geram impactos negativos sobre as espécies e os ambientes em que vivem, por isso, o futuro das nações depende, cada vez mais, de soluções que considerem a saúde e a riqueza dos ecossistemas como variáveis importantes.

Nesse contexto, o documento norteador sobre a biodiversidade é denominado Convenção sobre Diversidade Biológica – CDB, que consiste em um tratado da Organização das Nações Unidas estabelecido durante a Conferência das Nações Unidas sobre Meio Ambiente e Desenvolvimento – CNUMAD (também conhecida como ECO-92), realizada no Rio de Janeiro em junho de 1992. Em vigor desde dezembro de 1993 e assinado por mais de 160 países, esse tratado é considerado um dos mais importantes instrumentos internacionais relacionados ao meio ambiente.

A CDB está estruturada sobre três bases principais: a conservação da diversidade biológica, o uso sustentável da biodiversidade e a repartição justa e equitativa dos benefícios provenientes da utilização dos recursos genéticos. Além disso, considera biodiversidade em três níveis: ecossistemas, espécies e recursos genéticos (CDB, 1992, p. Art 1.).

A CDB abarca tudo o que se refere direta ou indiretamente à biodiversidade e funciona como arcabouço legal e político para diversas outras convenções e acordos ambientais mais específicos, tais como Protocolo de Cartagena sobre Biossegurança; o Tratado Internacional sobre Recursos Fitogenéticos para a Alimentação e a Agricultura; as Diretrizes de Bonn; as Diretrizes para o Turismo Sustentável e a Biodiversidade; os Princípios de Addis Abeba para a Utilização Sustentável da Biodiversidade; as Diretrizes para a Prevenção, Controle e Erradicação das Espécies Exóticas Invasoras; e os Princípios e

Diretrizes da Abordagem Ecosistêmica para a Gestão da Biodiversidade (MINISTÉRIO DO MEIO AMBIENTE - MMA, 2019; SCDB, 2019; ONU, 2019).

O conceito de biodiversidade ou diversidade biológica, segundo a CDB, contempla:

a variabilidade entre os organismos vivos de todas as origens, incluindo, entre outros, os ecossistemas terrestres, marinhos e outros ecossistemas aquáticos e os complexos ecológicos dos quais fazem parte; isso inclui diversidade dentro das espécies, entre espécies e ecossistemas (CDB, 1992, p. Art. 2).

A crise da biodiversidade levou a um aumento do interesse na teoria e prática do monitoramento da biodiversidade. Embora as diretrizes para monitoramento tenham sido publicadas desde a década de 1920, sabemos pouco sobre as práticas atuais nos esquemas de monitoramento existentes (LENGYEL, KOSZTYI, *et al.*, 2018). Por isso, um dos compromissos assumidos pelos países signatários da CDB é desenvolver o conhecimento sobre a biodiversidade como um precursor para a sua conservação e uso sustentável. E um dos maiores desafios refere-se ao fato de que o conhecimento da diversidade biológica é bastante heterogêneo e suas pesquisas tratam de diferentes escalas, variando de moléculas a biomas, bem como, envolvendo diferentes metodologias e abordagens (BRASIL. MCTIC. PPBIO, 2016, p. 19).

“O Brasil está entre os países que detém alta diversidade biológica (chamados de países megadiversos), abrigando em seu território cerca de 20% das espécies vivas conhecidas em todo o mundo” (BRASIL. MCTIC. PPBIO, 2016, p. 29). O país tem, portanto, papel decisivo e de vanguarda na geração de conhecimento sobre as espécies e os ecossistemas que possibilitem o uso sustentável e a conservação dessa imensa riqueza.

As decisões sobre intervenções humanas no meio ambiente envolvem questões para as quais não há, na maioria das vezes, respostas imediatas e fáceis, tampouco certeza de que a solução escolhida esteja correta ou dê os resultados esperados (BRASIL. MCTIC. PPBIO, 2016). O desenvolvimento da pesquisa e a gestão da biodiversidade requerem ações variadas e relacionadas entre si. Juntas, essas ações formam um complexo sistema que inclui desde levantamentos biológicos e estudos de variabilidade genética, até a formação de coleções biológicas, a avaliação de impactos ambientais e o planejamento do uso da terra, entre outros. Por isso, à procura de soluções que preencham as lacunas de conhecimento nesse complexo sistema, em 2004 foi criado o Programa de Pesquisa em Biodiversidade (PPBIO). Esse programa tem por missão “desenvolver estratégias de investimentos que agreguem as diversas competências em pesquisa e transferência de conhecimento sobre

biodiversidade, gerando, integrando e disseminando informações que possam ser utilizadas pela sociedade” (BRASIL. MCTIC. PPBIO, 2016, p. 17).

Em situações complexas, como é o caso da gestão em biomas e ecossistemas tropicais, algumas decisões podem ter melhor qualidade se seus efeitos forem monitorados de modo a permitir, por um lado, ajustes de decisões já tomadas, e por outro, a implantação de correções e mitigações por danos eventuais e não previstos (BRASIL. MCTIC. PPBIO, 2016). Dessa forma, o monitoramento promove um ciclo de aprendizado que diminui incertezas. Nesse sentido, a ciência teve papel central na definição das quatro diretrizes voltadas à criação e ao desenvolvimento do conhecimento no âmbito do PPBIO (BRASIL. MCTIC. PPBIO, 2016), que são:

- Realizar pesquisa científica com protocolos adequados para gerar informação e conhecimento que possam ser integrados e que, portanto, permitam alimentar modelos preditivos. Esses modelos são essenciais para apoiar o planejamento e monitoramento de intervenções, bem como a avaliação de seus impactos. Essa experiência consolida a pesquisa colaborativa de grande escala que permite apoiar o desenvolvimento e implantação de decisões em acordos internacionais, com apoio central dos membros da academia que contribuem em escalas regionais e globais.
- Assegurar que os dados primários de cada pesquisa sejam armazenados de forma permanente e institucionalizada, contendo adequada política de acesso, com a qual cada pesquisador integrante esteja compromissado.
- Realizar reuniões periódicas com um conselho de governança que possua representatividade equilibrada de diferentes segmentos, entre membros da academia e tomadores de decisão tanto do setor privado quanto governamental. Esses conselhos são essenciais para aprovarem os planos do programa, monitorarem seus resultados, articularem iniciativas para catalisar ações conjuntas e analisarem relatórios de avaliações independentes.
- Disseminar os resultados de forma orientada a diferentes públicos, dentre eles o científico, o público leigo, e o público técnico.

A partir dessas diretrizes o PPBIO buscou ampliar o uso dos dados e das informações primárias geradas por cada pesquisador da área de biodiversidade, evitando que dados similares viessem a ser coletados em duplicidade por ocasião do início de novas linhas de

pesquisa. As fontes de recursos para o PPBIO são providas por diferentes instituições, que são majoritariamente governamentais, mas também entidades internacionais, como é o caso do *Global Environment Facility* (GEF) que apoiou a implantação do Sistema de Informação para a Biodiversidade Brasileira (SiBBr).

Diversas iniciativas e esforços internacionais têm sido direcionados nas últimas décadas para aperfeiçoar a gestão dos dados de pesquisa sobre biodiversidade. O Brasil tem um papel importante nessas iniciativas, pois além de ser um país megadiverso, é também signatário dos tratados internacionais mais relevantes nessa temática.

1.4 DEFINIÇÃO DO PROBLEMA DA PESQUISA

A obstrução do fluxo de informação científica possui diversas causas, que reunidas representam prejuízos ao processo de criação, compartilhamento e uso do conhecimento científico, limitando, por esta razão, a velocidade das descobertas, o impacto dos resultados das pesquisas e o progresso da ciência (LEITE, 2011). Dentre os problemas identificados na literatura, a ausência, insuficiência ou inadequado registro dos metadados representam um dos obstáculos para promover efetividade e eficiência aos processos de gestão e comunicação de dados científicos.

Frente a esses desafios, busca-se compreender como a comunidade científica utiliza os metadados no tratamento dos dados sobre biodiversidade, tomando por referência os padrões de metadados estabelecidos na literatura internacional para essa temática. Assim, de forma a contribuir para compreensão do uso dos metadados no fluxo da informação científica e até mesmo para o desenvolvimento teórico referente à contribuição dos padrões de metadados para a gestão e comunicação dos dados científicos, visando o aperfeiçoamento das atividades de pesquisa sobre biodiversidade, a proposta de estudo aqui apresentada pretende responder a seguinte pergunta:

Como os padrões de metadados estão incorporados ao fluxo de informação científica sobre biodiversidade?

1.5 OBJETIVOS

1.5.1 OBJETIVO GERAL

Avaliar o uso dos padrões de metadados no fluxo de informação científica sobre biodiversidade.

1.5.2 OBJETIVOS ESPECÍFICOS

- Identificar os padrões de metadados recomendados internacionalmente para a temática biodiversidade;
- Identificar o uso dos padrões de metadados nas pesquisas sobre biodiversidade em artigos da literatura científica;
- Identificar o uso dos padrões de metadados nos conjuntos de biodiversidade em repositórios;
- Comparar os padrões de metadados recomendados internacionalmente com evidências de uso efetivo no registro dos dados sobre biodiversidade.

1.6 JUSTIFICATIVA

Desenvolver o conhecimento visando avanço científico e tecnológico demanda o aperfeiçoamento do sistema de comunicação científica, com destaque às necessidades de redução de custos, compartilhamento da informação e aumento da produção científica. Segundo Leite (2011), quanto mais rápida e completamente pesquisadores receberem a informação científica necessária às suas atividades, mais produtos científicos eles gerarão a custos menores.

A obstrução do fluxo de informação científica decorre, dentre outros fatores, da indisponibilidade ou baixa rastreabilidade dos dados científicos. Além disso, o fenômeno da explosão digital da informação, também impõe a necessidade de superar problemas referentes ao acesso e integração dos dados digitais. Nesse sentido, Isotani e Bittencourt (2015, p. 17) afirmam que para modificar esta situação, diversas empresas, governos e institutos de pesquisa têm realizado esforços para disponibilizar dados e produzir tecnologias que permitam criar um ecossistema de produção e consumo de dados com o

objetivo de agilizar a descoberta de novos conhecimentos e agregar valor a qualquer informação.

Compreender como se dá o uso dos padrões de metadados na descrição dos dados de pesquisa da biodiversidade visa auxiliar na identificação dos obstáculos que impedem ou dificultam o acesso à informação científica. Também tem a função de reunir insumos para proposição de soluções futuras voltadas a melhorar a comunicação científica, sob a perspectiva da incorporação dos padrões de metadados nos processos de produção, disseminação e uso da informação científica entre grupos de pesquisa, indústrias e organizações. Dessa forma, nesta pesquisa, a análise do uso de metadados refere-se a identificação das evidências de incorporação dos padrões de metadados nas pesquisas sobre biodiversidade, utilizando como fontes os artigos científicos e documentação dos conjuntos de dados resultantes de pesquisas sobre a biodiversidade.

Portanto, a presente pesquisa é relevante para a CI como contribuição a compreensão do real papel dos dados de pesquisa como estoques de informação no fluxo da informação científica, bem como na evolução da comunicação desses recursos. Adicionalmente, considerando a diversidade e abrangência dos campos do conhecimento que geram dados científicos, a análise do uso dos metadados é restrita a pesquisa sobre biodiversidade.

A escolha da temática está embasada em três motivos relevantes. Primeiramente, a abrangência do tema, tendo em vista que engloba diferentes áreas do conhecimento – saúde, biotecnologia, biologia, biomedicina, ecologia e outras. Em segundo lugar o conhecimento já acumulado, considerando que já existem padrões de metadados existentes para tratar especificamente dessa temática. Por fim, a terceira motivação se deve a existência de um número relevante de publicações científicas sobre o uso dos metadados nessa área e a disponibilidade de conjuntos de dados resultantes de pesquisas sobre biodiversidade, que garantem uma amostra significativa de dados a serem analisados. Além desses aspectos, cabe também destacar que o Brasil é um dos países com maior biodiversidade e a indisponibilidade dos dados científicos sobre este tema dificulta o estabelecimento de parcerias internacionais para alavancar as pesquisas científicas, bem como fragiliza a capacidade de planejamento para o uso adequado, a preservação e o combate à biopirataria desses recursos.

2 REVISÃO DE LITERATURA

“A ciência e a produção do conhecimento científico têm sido objeto de estudo de várias áreas do conhecimento, entre elas a Filosofia da Ciência, a Sociologia da Ciência e a Ciência da Informação – CI” (MUELLER e PASSOS, 2000, p. 13). A CI tem estudado, desde a sua gênese, questões relacionadas à comunicação científica como fenômeno central de interesse, uma vez que a disciplina surgiu em razão da preocupação de cientistas, tecnólogos e documentalistas de meados do século passado com os fluxos da informação em ciência e tecnologia (COSTA, LEITE e TAVARES, 2018, p. 17).

Apesar de cada área da ciência apresentar especificidades para estruturação do conhecimento há naturalmente pontos de interesse convergentes, entre os quais, segundo Muller (2000, p. 21), a confiabilidade é uma das características mais importantes, pois distingue o conhecimento científico do popular, não científico. Ainda segundo Muller,

para obter confiabilidade, além da utilização de uma rigorosa metodologia científica para a geração do conhecimento, é importante que os resultados obtidos pelas pesquisas de um cientista sejam divulgados e submetidos ao julgamento de outros cientistas, seus pares. A ampla exposição dos resultados de pesquisa ao julgamento da comunidade científica e sua aprovação por ela propicia confiança nesses resultados (MUELLER, 2000, p. 21).

Nesse sentido, todo o conhecimento científico está sujeito à validação da comunidade científica, que apresenta um papel ativo no ciclo de comunicação, conforme também é explicitado por Ziman (1968, p. 9), “aplaudindo ou vaiando, jogando flores ou tijolos, exercem controle sobre a substância das comunicações que recebem”. Referente a esse ciclo, Meadows (1999, p. vii) entende que a “comunicação eficiente e eficaz constitui parte essencial do sistema de pesquisa científica”. Também para Leite (2011, p. 25-26), a ciência é altamente dependente da circulação de informação que ocorre devido à existência de um complexo sistema de comunicação científica.

Sob essa perspectiva, a revisão de literatura apresentada a seguir tem foco em identificar os conceitos fundamentais à compreensão da aplicação dos padrões de metadados na pesquisa científica e na disseminação de seus resultados, com ênfase no aperfeiçoamento do uso e circulação dos dados no fluxo da informação científica. Tais conceitos estão inseridos nos seguintes campos da CI: Gestão, Comunicação e Organização da Informação.

2.1 GESTÃO E COMUNICAÇÃO DA INFORMAÇÃO CIENTÍFICA

A comunicação da informação é responsável pelos fluxos de informação, e fenômenos associados, que viabilizam a atividade humana em diferentes contextos; já a gestão da informação e a gestão do conhecimento são as responsáveis por sua sistematização e dão conta de sua complexidade na contemporaneidade (COSTA, LEITE e TAVARES, 2018, p. 15). Tais definições induzem à consideração de que há relação de complementaridade e interdependência entre as duas práticas. Isso porque, segundo os autores, por um lado, a gestão do conhecimento disciplina, sistematiza e torna mais efetivos os processos de comunicação. Por outro, a comunicação permite que a gestão do conhecimento seja viabilizada, na medida em que possibilita, entre outros processos, a interação entre indivíduos e, por consequência, a criação, o compartilhamento e o uso do conhecimento.

Além disso, Costa, Leite e Tavares (2018) entendem que o pressuposto básico da argumentação que sustenta a relação entre a comunicação e a gestão do conhecimento é o fato de que “a primeira, em seu aspecto da interação humana, e não reduzida, mas apoiada nas tecnologias, é um dos processos essenciais para o sucesso da gestão do conhecimento”. Ainda segundo os autores, essa relação,

Trata-se de uma mão de via dupla. Primeiro porque a comunicação serve à gestão da informação e do conhecimento, uma vez que a produção, compartilhamento e usos constituem, essencialmente, processos de comunicação humana. Segundo porque a gestão da informação e do conhecimento promovem, em última análise, a interação entre sujeitos que alternam em seus papéis de produtores e usuários em diferentes contextos (COSTA, LEITE e TAVARES, 2018, p. 15).

A presente pesquisa busca, portanto, compreender os conceitos nos campos da Gestão e Comunicação da Informação para analisar o papel dos metadados no uso e compartilhamento dos dados no fluxo da informação científica sobre biodiversidade. Por isso, a seguir apresenta-se uma breve revisão voltada a situar este estudo nos referidos temas.

2.1.1 COMUNICAÇÃO CIENTÍFICA

O processo de Comunicação Científica é um ciclo contínuo e regenerativo (LANCASTER e SMITH, 1978, p. 369), que inclui atividades associadas com a produção, disseminação e uso da informação, desde a hora em que o cientista teve a ideia da pesquisa

até o momento em que os resultados de seu trabalho são aceitos formalmente como parte integrante do conhecimento científico (GARVEY, 1979).

“Enquanto de alguma forma os canais informais de comunicação são tão antigos quanto à própria ciência, outros canais têm origem muito mais recente” (LANCASTER e SMITH, 1978, p. 370), devido aos avanços tecnológicos e o crescimento da comunicação informal. O sistema de comunicação na ciência apresenta dois tipos de canais de comunicação dotados de diferentes funções (GARVEY, 1979). O canal informal de comunicação é caracterizado por contatos pessoais, conversas telefônicas, correspondências, cartas, pré-prints e assemelhados. O canal formal é representado pela informação publicada em forma de artigos de periódicos, livros, comunicações escritas em encontros científicos, etc.

Nos canais informais o processo de comunicação é ágil e seletivo. A informação circulada tende a ser mais atual e ter maior probabilidade de relevância, porque é obtida pela interação efetiva entre os pesquisadores. Já nos canais formais o processo de comunicação é lento, mas necessário para a memória e a difusão de informações para o público em geral. Os canais formais destinam-se a transferir informações a uma comunidade, não a um indivíduo.

Barreto (1998, p. 123) explica que a publicidade do conhecimento produzido é uma condição necessária para sua validação e socialização, construindo, também, um ciclo constante e auto regenerativo: conhecimento; publicidade; opinião pública; novo conhecimento. Para o autor, a rapidez e a qualidade desse ciclo e da informação que está sendo legitimada dependerão da velocidade com que se processa a cadeia de eventos da publicidade do conhecimento, pois a opinião do público a quem se direciona o conhecimento é que vai lhe conferir ou não legitimidade e aceitação, definindo as condições de reingresso no fluxo de informação e conhecimento. Segundo Muller (2000, p. 22), todo trabalho intelectual depende de um intrincado sistema de comunicação, que compreende canais formais e informais.

Esse complexo sistema de comunicação é também influenciado pela mediação, que segundo Ruben (1993, p. 228) tem como função ampliar o processamento e comunicação da informação humana, no que se refere a sua produção, distribuição, recepção, armazenamento e recuperação. Ruben também explica que a mediação pode ocorrer por meio de:

1. sistemas de informação formais (exemplos: buscar informação em bases de dados ou assistir TV a cabo) ou sistemas de informação informais (exemplos: perguntar algo por telefone ou tomar emprestado anotações de uma palestra);
2. meios “antigos” (caneta e papel) ou novos meios (computadores e dispositivos móveis);
3. meios usados para a comunicação de massa (exemplos: televisão ou livros), meios usados para a comunicação organizacional (exemplos: boletins de notícias, mensagens para múltiplos usuários ou mala-direta), meios usados para a comunicação interpessoal (exemplos: telefone e secretária eletrônica), meios usados para comunicação científica (exemplos jornais e relatórios científicos);
4. Interfaces tecnológicas (exemplos: teclado do computador ou secretária eletrônica) ou interfaces humanas (exemplos: bibliotecário ou editor de um jornal) (RUBEN, 1993, p. 227).

Assim, também a mediação é fator relevante para compreender como se dá a circulação da informação. Por isso, Ruben (1993, p. 223 e 230) recomenda que os estudos dos fenômenos da informação considerem a análise integrada dos seguintes conceitos: comunicação, mediação, informação e instituições.

De forma geral, a comunicação científica constitui tópico bastante explorado e discutido na CI, gerando vários modelos teóricos, abordagens e contextos encontrados na literatura, que evoluíram para absorver as mudanças tecnológicas, conforme aponta Costa (2005, p. 165).

Comunicação científica constitui um dos tópicos que têm sido muito explorados e discutidos na ciência da informação, ao longo das últimas quatro décadas. Por esse motivo, tem contribuído de forma significativa para a construção de conhecimento na área. Consequentemente, são vários os modelos teóricos, as abordagens e os contextos encontrados na literatura para seu estudo. Isso, por sua vez, reflete uma variedade de aspectos por meio dos quais o processo de comunicação entre pesquisadores tem sido estudado. A partir principalmente da década de 90, estudos sobre os impactos que a introdução de tecnologias da informação no ambiente acadêmico provoca na comunicação científica vêm contribuindo para o enriquecimento do debate a respeito das questões pertinentes ao tópico e que são, por conseguinte, relevantes para estudo.

Apesar da existência de diversos modelos e abordagens voltados à comunicação científica que foram desenvolvidos na CI, ainda é necessário ampliar o conhecimento em busca de soluções voltadas ao uso e compartilhamento dos dados científicos. Tendo em vista o fenômeno mais recente da explosão digital de dados, verifica-se a necessidade de desenvolver estudos voltados a aperfeiçoar o fluxo da informação científica para adequá-lo a esse novo contexto. Muitos dos desafios existentes na comunicação dos dados científicos estão relacionados à necessidade de aperfeiçoamento da gestão desses recursos, cujos principais conceitos são apresentados a seguir.

2.1.2 GESTÃO DA INFORMAÇÃO

O entendimento de diferentes autores sobre gestão da informação e gestão do conhecimento é esclarecido na revisão apresentada por Leite (2006, p. 216):

A gestão do conhecimento engloba as práticas e metodologias de gestão da informação, especialmente no que concerne aos processos relacionados à coleta, armazenamento, recuperação de parte do conhecimento tácito, o qual é reduzido a estruturas de informação. Entretanto, a gestão do conhecimento não pode ser reduzida a, nem confundida com a gestão da informação. Isso porque quando o tipo de compartilhamento do conhecimento exige que ele seja veiculado por meio de sistemas formais que permitem o seu armazenamento e recuperação, o conhecimento é, de fato, reduzido a estruturas de informação. Esta, no entanto, poderá desencadear a criação de conhecimento no indivíduo.

Considerando diversos posicionamentos da Ciência da Informação, Costa e Leite (2018, p. 27) relatam que o entendimento predominante é de que a Gestão da Informação deve se ocupar do conhecimento registrado. Os autores também afirmam que “tanto para a investigação quanto para a aplicação de conhecimentos na solução de problemas reais relacionados com a melhoria da comunicação da informação entre seres humanos, a Ciência da Informação se vale da perspectiva de processos de gestão da informação” (COSTA, LEITE e TAVARES, 2018, p. 32).

Ao encontro dessa perspectiva, Campos e demais autores (2005, p. 58-59) entendem que a gestão se insere como um conceito em construção, contemplando o gerenciamento das informações por meio do tratamento de conteúdos semânticos. As autoras explicam que para a qualidade desse tratamento são fundamentais as investigações relativas a padrões de intercâmbio, controle de linguagem e modelos de representação, utilizando metadados, vocabulários, ontologias, tesouros, taxonomias, entre outras. Por fim, as autoras concluem que a gestão envolve uma série de procedimentos de uniformização e padronização de informações, baseados em uma visão integradora, fundamentada em dois conceitos essenciais: o de organização e o de comunicação.

O uso dos padrões de metadados, portanto, é um assunto que se insere entre os procedimentos de uniformização e padronização de informações. Por isso, também é importante compreender essa temática pela ótica da Organização da Informação.

2.2 ORGANIZAÇÃO DA INFORMAÇÃO

Organizar significa categorizar, incluir, em função de algum critério, determinados elementos em uma classe e, nesse mesmo movimento, excluir outros elementos, pressupõe,

portanto, a identificação de semelhanças e de diferenças (SMIT e KOBASHI, 2002, p. 11-12). Segundo Lima e Cunha (2006, p. 1), “os esforços para a organização da informação geram modelos artificiais que tentam representar o mundo real, criando sistemas e esquemas de conceitos utilizados nos processos de classificação e indexação de recursos informacionais”. No sentido proposto por Bräscher e Café (2008, p. 5):

A organização da informação é, portanto, um processo que envolve a descrição física e de conteúdo dos objetos informacionais. O produto desse processo descritivo é a representação da informação, entendida como um conjunto de elementos descritivos que representam os atributos de um objeto informacional específico.

Para Café e Sales (2010) a organização da informação é um processo de arranjo de acervos tradicionais ou eletrônicos realizado por meio da descrição física e de conteúdo (assunto) de seus objetos informacionais. Um mesmo objeto pode apresentar diversas descrições sucessivas e interdependentes, que se constituem em um enunciado de suas propriedades ou das relações desse objeto com outros que o identificam.

Assim, “os fundamentos da **catalogação** de assuntos, da **classificação**, da **indexação** e da **análise documental** são abordagens do tratamento temático da informação e responsáveis pela formação da organização da informação como a conhecemos hoje” (CAFÉ e SALES, 2010, p. 116, grifo nosso.). A catalogação é dividida em “descrição bibliográfica, pontos de acesso e dados de localização” (MEY e SILVEIRA, 2009, p. 94). A descrição bibliográfica é responsável pela caracterização do recurso bibliográfico. Os pontos de acesso são os elementos do registro bibliográfico pelos quais o usuário pode acessar o recurso. Dados de localização são as informações que permitem ao usuário localizar o item no acervo (ciberespacial ou físico), permitindo assim a individualizar os registros bibliográficos, reuni-los por suas semelhanças, e apontar sua localização (MEY e SILVEIRA, 2009, p. 94-96). Portanto, a descrição física de um objeto informacional se dá pelo processo de catalogação cujo resultado é a representação do suporte físico ou documento, podendo utilizar linguagens específicas, normas e formatos que padronizam este tipo de descrição. Conforme apresenta Alves (2010, p. 41), é possível analisar a evolução dos padrões de catalogação a partir da perspectiva de diferentes períodos históricos – quadro 1.

Quadro 1. Sistematização da história da Catalogação em períodos

PERÍODOS HISTÓRICOS	CONTRIBUIÇÕES À TEORIA DA CATALOCAÇÃO	DESENVOLVIMENTO DE PADRÕES	INTEGRAÇÃO ESTRATÉGICA COM AS TECNOLOGIAS
Períodos Remotos	Catálogo de Saint Martin: dividido em três seções, a última incluía análise das partes (início das entradas analíticas)	Não havia uma forma de organização e sistematização unificada das informações.	Suportes informacionais: tabletes de argila, madeira, papiro, pergaminho, papel
Período Tipográfico	Primeiras iniciativas para construção de um código de catalogação – sem padronização unificada e sem consolidação de princípios	Primeiros catálogos livreiros, sem padronização unificada	Tipografia: catálogos em forma de livros impressos; catálogos em fichas
Período Tradicional	Princípio de Panizzi Princípio de Jewett Princípio de Cutter Princípio de Lubetsky	Regras de Catalogação do Museu Britânico (1839) Regras de Jewett (1853) Regras de Cutter (1876) <i>Anglo-American</i> (1908) Instruções Prussianas Código da Vaticana (1920) LC (1949) ALA (1949)	Tecnologias mecânicas: início dos catálogos automatizados (sistema de processamento Hollerith de cartões perfurados); tecnologias de armazenamento – fitas magnéticas
Período Pré-Mecanizado	Princípio de Paris (1961)	AACR (1967) MARC (1965)	Tecnologias de informática: construção de bancos de dados para processamento da informação; tecnologias de armazenamento – disco rígido
Período Mecanizado	RIEC (1969) CBU (Controle Bibliográfico Universal)	ISBD - Padrão Internacional para Descrição Bibliográfica (1971) AACR2 (1978) AACR2R (1988)	Tecnologias de informática: consolidação de grandes bancos de dados, redes cooperativas, OPACs; desenvolvimento e aprimoramento dos modelos e linguagens de construção de banco de dados
Período de Metadados	Desenvolvimento dos padrões de metadados	Padrão de metadados Dublin Core (1995) AACR2R (1998) AACR2R (2002) MARC 21 e desdobramentos	Tecnologias de informática e TICs: sistemas de recuperação da informação, bancos de dados com acesso à web, aprimoramento das tecnologias para a construção dos bancos de dados

Fonte Alves (2010, p. 41)

No período de metadados, segundo Alves (2010, p. 38), a recuperação da informação torna-se foco de investigação nas áreas de Ciência da Informação e Ciência da Computação, que convergem na busca de soluções, criando uma infraestrutura tecnológica e representacional para diversos ambientes científicos.

Destaca-se que as comunidades científicas têm despendido esforços para criar catálogos de dados. Nessa linha, o *Digital Curation Center* (DCC, 2019), coordena um estudo piloto para testar abordagens visando o desenvolvimento de um serviço de catalogação voltado a agregar metadados.

Dados de pesquisa são valiosos, então vale a pena saber o que você tem e onde está localizado. Um registro de ativos de dados ou catálogo de dados é uma maneira útil de registrar isso.

Algumas instituições estão começando a trabalhar nessa área e o DCC está coordenando um estudo piloto para testar abordagens para um serviço que agregará metadados relativos a coleções de dados ou conjuntos de dados mantidos em instituições de pesquisa do Reino Unido e centros especializados de dados (DCC, 2019).

Segundo o DCC (2019), a iniciativa tem por objetivo ajudar a definir quais metadados devem ser coletados, quais padrões usar e como o catálogo de metadados deve integrar ou explorar sistemas existentes, como repositórios ou de Gestão da Informação de Pesquisa (*Research information management – RIM*).

Já a **descrição do assunto** recebe várias denominações na literatura como Análise documentária, Análise Temática, Análise de Assunto, Descrição de Conteúdo e Tratamento Temático da Informação (CAFÉ e SALES, 2010, p. 118). Essa descrição produz a chamada informação documentária por meio dos processos de classificação, indexação e resumo ou condensação documental (KOBASHI, 2018; CAFÉ e SALES, 2010).

A **classificação** gera uma representação da informação no formato de números ou símbolos de classificação, geralmente descreve o conteúdo do documento de forma mais abrangente. A **indexação** é a operação pela qual se escolhe os termos mais apropriados para descrever o conteúdo de um documento e da pergunta do usuário, resultando em palavras-chave ou descritores que reunidos podem formar índices, complementando a organização da informação realizada pela classificação. Por fim, o **resumo** é o terceiro processo de descrição de conteúdo da organização da informação, cuja função é complementar a informação dada pela classificação e pela indexação (CAFÉ e SALES, 2010, p. 119). De forma geral, os resumos podem indicar apenas os pontos principais do documento, apresentar síntese de partes do documento ou uma análise crítica redigida por especialistas.

Segundo Kobashi (2008, p. 49), “a análise documentária define-se como uma operação com textos”, que são desestruturados em operações sucessivas de análise e síntese, distinguindo informação essencial de acessória. Em seguida as informações

selecionadas são estruturadas formando novos textos, apresentados como resumos ou em símbolos de uma linguagem específica, dita linguagem documentária.

Boughida (2005, p. 49) entende que é importante especificar e diferenciar os padrões de linguagem documentária envolvidos na catalogação, classificação e indexação. Para tanto o autor, propõe o agrupamento em quatro categorias: conteúdo; valor; estrutura e formato - Quadro 2.

Quadro 2. Tipologia de Padrões de Dados

TIPOS	DESCRIÇÃO	EXEMPLOS
Padrões de conteúdo de dados (preenchimento de campos)	São as regras que orientam o preenchimento de um determinado campo de metadados. Padrões de conteúdo de dados (regras de catalogação e códigos). Estas são diretrizes para o formato e a sintaxe dos valores de dados usados para preencher elementos de metadados.	<i>International Standard Bibliographic Description (ISBD), Anglo-American Cataloging Rules 2 (AACR2), e Catalogação de Objetos Culturais (CCO) e outros</i>
Padrões de valor de dados (esquemas baseados em padrões de valores)	Fornecem termos e outros valores com os quais se devem preencher um campo de metadados. Padrões de valor de dados (vocabulários controlados, tesouros, listas controladas). Estes são os termos, nomes e outros valores usados para preencher padrões de estrutura de dados ou conjuntos de elementos de metadados.	<i>Library of Congress Subject Headings (LCSH), Library of Congress Name Authority File (LCNAF), LC Thesaurus for Graphic Materials (TGM), Medical Subject Headings (MeSH), Art & Architecture Thesaurus (AAT), Union List of Artist Names (ULAN), Getty Thesaurus of Geographic Names (TGN), ICONCLASS (sistema de classificação específico para pesquisa iconográfica e documentação de imagens; e outros).</i>
Padrões da estrutura de dados (esquemas transversais de estrutura)	Especificam as categorias e organização de elementos de metadados. Representam uma estrutura de dados padrão, que sustenta as relações e hierarquias entre os elementos. São conjuntos de elementos de metadados, esquemas, representando as “categorias” ou “contêineres” de dados que compõem um registro ou outro objeto de informação.	Campos do tipo <i>MARC (Machine-Readable Cataloging format), Encoded Archival Description (EAD), Dublin Core Metadata Element Set (DCMES), Categories for the Description of Works of Art (CDWA),</i> Categorias principais do <i>Visual Resources Association (VRA)</i> e outros.
Padrões de formato de dados (ou expressões técnicas de dados, proporcionam interoperabilidade)	São a expressão técnica ou codificação de dados em um arquivo ou uma tabela de banco de dados.	Esquemas e documentos XML e as definições do tipo <i>Document Type Definition – DTD</i> , que define a estrutura, os elementos e atributos de um documento XML. <i>MARC21, MARCXML, EAD XML DTD, METS, MODS, CDWA Lite, XML schema, Simple Dublin Core XML schema, Qualified Dublin Core XML schema, VRA Core 4.0 XML</i>

FONTE: Boughida (2005, p. 49-50)

Campos e demais autores (2005, p. 58-59) afirmam que o conceito de organização pressupõe procedimentos classificatórios e tais procedimentos possibilitam o agrupamento e a recuperação de informações de acervos diversos, que podem estar em forma estruturada (bancos de dados, por exemplo) e não estruturada (textos integrais, por exemplo). Para as autoras, tais procedimentos classificatórios, além de evidenciar os contornos de atuação de uma instituição/organização, facilitam os processos de seleção e de tratamento de informações, cujo produto se apresenta como um mapa de conteúdos - taxonomias - das atividades produzidas.

Nesse contexto, para Campos e demais autores (2005, p. 59), o desenvolvimento de taxonomias tem sido um dos pilares da gestão da informação e do conhecimento. As taxonomias são listas de categorias de assunto estruturadas, consideradas como ferramentas com a função de organização lógica de conteúdos informacionais, mas que não incluem a definição de tópicos, somente apresenta as relações hierárquicas.

O uso de taxonomias permite que se estabeleçam padrões de alto nível para a ordenação e classificação de informações através de mecanismos de herança que permitem reduzir esforços na produção e utilização do conhecimento. Ainda segundo Campos e demais autores (2005)

[...] herança é um dos conceitos mais poderosos no desenvolvimento de software. As máquinas podem compreender corretamente relacionamentos de generalização e especialização entre as entidades atribuindo propriedades às classes gerais e então assumindo que as subclasses herdaram estas propriedades. Entretanto, para que isso ocorra, é necessário promover interoperabilidade entre os diversos sistemas de uma instituição. Isto significa promover a capacidade dos sistemas potencializarem oportunidades de intercâmbio e reutilização de informações, interna ou externamente (2005, p. 59).

Harpring (2016, p. 203) explica que a anotação dos valores nos campos ou elementos de metadados dedicados especificamente a certos elementos de conteúdo garante que os dados sejam consistentemente registrados e indexados no mesmo lugar, utilizando as mesmas convenções. A autora ressalta que sem adequados metadados descritivos, a informação permanece inacessível e, por isso, uma solução completa demanda o uso de vocabulários controlados (HARPRING, 2016, p. 22).

Nesse sentido, evidencia-se a relação entre Comunicação e Organização. A primeira como processo que permite a transmissão de conteúdos informacionais a partir de uma visão integrada de conteúdos. Intrinsecamente ligada à Organização, cujas ações referentes à

definição de metadados e à construção de terminologias padronizadas são fundamentais para viabilizar o tratamento e a recuperação das informações.

A partir do exposto, constata-se que o estudo dos metadados está circunscrito às áreas de Organização, Gestão e de Comunicação da Informação, estabelecendo relações integradoras entre esses campos da CI.

2.2.1 O QUE SÃO METADADOS?

Metadados, “dados sobre dados”, é ainda um termo frequentemente entendido de maneiras diferentes por diversas comunidades e profissionais que projetam, criam, descrevem, preservam e usam sistemas e recursos de informação (GILLILAND, 2016). “Os recursos informacionais possuem características que requisitam um tratamento descritivo específico para aperfeiçoar sua busca, acesso e recuperação” (SANTOS, SIMIONATO e ARAKAKI, 2014, p. 148). Então, de forma abrangente, podemos dizer metadados são elementos descritivos que representam as características desses recursos.

“Nos últimos cem anos, pelo menos, a criação e o gerenciamento de metadados foram principalmente de responsabilidade de profissionais da informação envolvidos em catalogação, classificação e indexação” (GILLILAND, 2016). Segundo Sayão (2010, p. 2), os metadados foram usados inicialmente no contexto dos bancos de dados para descrever e controlar a gestão e o uso dos dados, mas é comum associar o termo ao seu uso em bibliotecas, considerando seu papel em um esquema formal para descrição de todo tipo de objetos informacionais, digitais e não digitais. Entretanto, foi sua aplicação nos meios digitais e na Web aumentaram suas interpretações e uso para muito além dos profissionais da informação.

Embora o termo metadados não seja muito familiar entre usuários comuns, criadores e consumidores de conteúdo digital na Web, todos esses indivíduos são cada vez mais estimulados a criar, explorar e avaliar metadados nos ambientes digitais, desde títulos de páginas da Web, *folksonomias*², marcadores sociais, até outros meios voltados à descrição

² *Folksonomia* é um neologismo que se refere a um conjunto de conceitos representados por termos e nomes (chamados tags) que são compilados por meio do processo social de criação de tags. Esse processo corresponde à prática e ao método descentralizado pelos quais indivíduos e grupos criam, gerenciam e compartilham tags (termos, nomes etc.) para anotar e categorizar recursos digitais em um ambiente social on-line. Esse método também é chamado de classificação social, indexação social, indexação e categorização popular (HARPRING, 2016, p. 52).

ou avaliação da qualidade dos recursos informacionais. Entretanto, “esse processo social não é necessariamente colaborativo, porque o esforço geralmente não é organizado; na realidade, os indivíduos não trabalham em conjunto ou de forma coordenada, e não empregam a padronização e um vocabulário comum” (HARPRING, 2016, p. 52).

Assim, considerando a existência de diversas abordagens referentes ao termo metadados, a seguir são apresentadas algumas das definições segundo diferentes autores.

Berners-Lee (1997), criador da Web, define metadados como “informações compreendida por máquinas referentes aos recursos Web ou outras coisas”. Segundo Berners, existem maneiras de representar as coisas e tanto quanto possível, a sintaxe e a semântica devem poder ser obtidas por meio da estrutura de metadados, que deve incluir vários vocabulários dentro em seu escopo. Ainda segundo o autor, a sintaxe e a estrutura devem ser tais que o maior número possível de processamentos possa ser feito sem a necessidade de conhecer a semântica do vocabulário em uso.

Dempsey e Heery (1997, p. 5) conceituaram metadados como “dados que descrevem atributos de um recurso. [...] dados que associados a objetos desoneram os usuários potenciais de ter conhecimento completo antecipado da existência e características desses objetos”. Para os autores está associado às seguintes funções: localização, descoberta, documentação, avaliação, seleção e outras. Também, os autores entendem que essas atividades podem ser realizadas por usuários finais humanos ou por seus agentes (humanos ou automatizados).

Takahashi (2000, p. 59 e 172) conceitua metadados como: “dados a respeito de outros dados, ou seja, qualquer dado usado para auxiliar na identificação, descrição e localização de informações. Trata-se, em outras palavras, de dados estruturados que descrevem as características de um recurso de informação”. Para o autor, o conceito de metadados tem adquirido importância crescente, na medida em que mais e mais conteúdos completos são armazenados em computadores e transmitidos via redes, tais como informações cartográficas, grandes bases de dados em textos livres e em diversas línguas. Por isso, o autor defende que o processamento adequado dessas grandes massas de dados passa pelo processamento dos metadados entremeados aos dados, que impõem estrutura e inteligibilidade aos mesmos.

Orozco García-Mayorca (2001, p. 1) comenta que “a generalização do conceito abrangeu qualquer tipo de informação descritiva sobre recursos [...] Portanto, a

categorização (normalizada) pode ser considerada como um processo de geração de metadados, que pode ser apresentado em um arquivo (não digital) ou em um ambiente digital”. Para a autora, os metadados possuem três funções básicas: a primeira, fornecer uma descrição de um objeto ou entidade de informação juntamente com informações necessárias para gerenciamento e preservação; a segunda para fornecer pontos de acesso àquela descrição, através dos quais um índice será gerado; e a terceira para codificar a descrição visando facilitar seu gerenciamento por meios automatizados (OROZCO GARCÍA-MAYORCA, 2001, p. 2).

Polydoratos e Nicholas (2001, p. 309) consideram que os registros de metadados são uma solução para o problema de compartilhamento de dados e padronização de informações na Internet. Os autores argumentam que, diante da necessidade de uma abordagem padronizada, foi produzida pela Organização Internacional de Padronização (*International Organization for Standardization – ISO*) a norma sobre Especificação e Padronização de Dados em Tecnologia da Informação (ISO/IEC 11179)³.

Eva Mendes Rodrigues (2001, p. 1) argumenta que a necessidade de rotular, catalogar e descrever informação estruturada de forma a permitir que os objetos da Web possam ser armazenados, localizados, processados e compartilhados, sem dúvidas, fez dos metadados protagonistas do novo paradigma dos sistemas de informação deste milênio. Segundo a autora, em vista dessa realidade, os metadados constituem um mecanismo crítico tanto na representação do conhecimento das coleções digitais quanto na exploração e uso dos dados. São elementos ou estruturas de organização da informação que, atribuídas a cada objeto de informação eletrônica, classificam, categorizam ou descrevem (RODRÍGUEZ, 2001, p. 8). Smiraglia (2005, p. 2) aponta que no nível mais básico de sua definição, metadados são descritores estruturados de recursos informacionais desenvolvidos para promover a recuperação da informação.

Dziekaniak (2007, p. 232) considera metadado como a “informação que descreve e explica qualquer dado que, de modo geral, possa vir a aparecer em meio eletrônico”. Mas a autora explica que os metadados não são utilizados apenas em aplicações digitais e lembra que a “Biblioteconomia utiliza metadados desde seus primórdios. Podem ser consideradas metadados as notações oriundas das linguagens documentárias que propiciam a

³ ISO/IEC 11179 International Organization for Standardization – ISO. Information technology - specification and standardisation of data elements. Parts1-6. Disponível em: <http://www.iso.ch>

representação do conteúdo das obras”, tais como notação numérica ou alfanumérica nos sistemas de classificação e os descritores presentes nos tesouros.

Alves e Souza (2007, p. 22) explicam que “os elementos de metadados têm o propósito primário de descrever, identificar e definir um recurso de informação com o objetivo de modelar e filtrar o acesso. Os metadados são importantes na organização, gestão e recuperação da informação digital, principalmente.” Dessa forma, as autoras entendem que os metadados enriquecem ou complementam os objetos ou serviços referenciados, aumentando o potencial informativo, pois “são dados definidores que fornecem informação sobre ou documentação de outros dados dentro de uma aplicação ou de um ambiente” (2007, p. 23).

Ainda segundo Alves e Souza (2007, p. 21) o constante desenvolvimento das tecnologias da informação e de comunicação tem possibilitado o avanço metodológico, exemplo disso são os modelos metadados, que propiciam novas práticas para a organização e tratamento da informação digital, proporcionando diferentes mecanismos de busca e recuperação. Essa “utilização de metadados na organização eletrônica de recursos vem ao encontro da necessidade crescente de descobrir e disponibilizar informações na internet e nas intranets” (2007, p. 22). Assim, as autoras ressaltam que o uso de padrões de metadados possibilita aos sistemas de informação e de gestão do conhecimento a integração e o compartilhamento de recursos e aplicações.

Campos (2007, p. 16) acrescenta que,

Metadados são habitualmente definidos simplesmente como dados descrevendo outros dados. No entanto, cada vez mais, especialmente no meio digital, o conceito tem sido empregado em variados contextos que envolvem diversos propósitos e tecnologias. Com o objetivo primário de identificar as funções de metadados nesses diferentes contextos (...) discerniram-se dez temas, denominados categorias funcionais dos metadados, que ressaltam diferentes ênfases em sua utilização, assim nomeadas: Descritividade, Padronização, Interoperabilidade, Modularidade, Reflexibilidade, Visibilidade, Flexibilidade, Automatização, Administração e Preservação.

Sayão (2007, p. 34) aprofunda a discussão ao afirmar que as funções dos metadados compreendem: **a descoberta de recursos**, permitindo que “recursos sejam identificados, localizados, selecionados por critérios de relevância e distinguidos por diferenças e similaridades; **a organização de recursos; a facilitação da interoperabilidade; a identificação digital; e a preservação digital**”. Mey e Silveira (2009, p. 133) elucidam algumas vantagens obtidas com o uso e aplicação dos metadados: “a) a rapidez do registro

bibliográfico; b) a normalização requerida a um registro bibliográfico; c) a possibilidade de busca por campo”.

Lima (2008) afirma que existe um grande número de iniciativas e padrões de metadados para as mais diversas aplicações. Alves e Café (2010, p. 185) apontam que dentre as ferramentas disponíveis atualmente para a padronização da descrição das informações, em rede eletrônica, encontram-se os metadados, que tornam mais viável a troca de informações entre aplicativos e organizações, bem como facilitam e tornam mais precisa a recuperação de dados e informações. De forma mais ampla, Alves (2010, p. 47) entende que os metadados podem ser descritos como,

[...] atributos que representam uma entidade (objeto do mundo real) em um sistema de informação. Em outras palavras, são elementos descritivos ou atributos referenciais codificados que representam características próprias ou atribuídas às entidades; são ainda dados que descrevem outros dados em um sistema de informação, com o intuito de identificar de forma única uma entidade (recurso informacional) para posterior recuperação.

Segundo Sayão (2010), ao submergir no mundo dos documentos digitais, constatamos que outras dimensões dos metadados, que ultrapassam os limites de ferramenta para a descrição e descoberta de recursos, precisam ser reveladas e exploradas. Pois segundo o autor, “os objetos digitais para serem gerenciados e usados requerem processos de maior amplitude, que implica em identificar informações precisas para instruí-los adequadamente” (SAYÃO, 2010, p. 2).

Siqueira e Modesto (2011, p. 12) consideram que os metadados “são informações estruturadas que descrevem, identificam, localizam ou tornam mais fácil a recuperação, o uso ou o gerenciamento de fontes de informação digital”.

Isotani e Bittencourt (2015) conceituam metadados como “informações adicionais que descrevem os dados contidos nos documentos são chamadas de metadados, ou seja, dados sobre dados”. Segundo os autores, “os metadados em conjunto com métodos de modelagem conceitual da informação [...] são essenciais para publicação de dados conectados” (ISOTANI e BITTENCOURT, 2015).

Caldera-Serrano e Freire-Andino (2016, p. 65) entendem que “os metadados, hoje em dia, estão se tornando cada vez mais importantes como possíveis mecanismos que permitam realizar esta representação do conteúdo dos recursos”, são ferramentas que ajudam a estabelecer uma descrição correta de um recurso eletrônico e que permite torná-lo visível e acessível dentro da rede. Os autores então definem “metadados como informações que

levam o usuário ao conteúdo de um recurso, revelando informações sobre o contexto do recurso, ou seja, indicando aspectos sobre quem, o quê, por quê, quando, onde e como, relacionados à criação do recurso”.

Gilliland (GILLILAND, 2016) apresenta metadados como “a soma total do que se pode dizer em um dado momento sobre qualquer objeto de informação em qualquer nível de agregação”. Para a autora, “um objeto de informação é qualquer coisa que possa ser endereçada e manipulada como uma entidade discreta por um ser humano ou um sistema de informação”. Ela também explica que os metadados possuem fonte interna (normalmente quando há geração pelo criador do objeto informacional) ou externa (geração após a criação do objeto, por outro que não o criador), podendo ser criados automaticamente por computadores ou manualmente por humanos. Por fim, a autora destaca que, “em geral, todos os objetos de informação, independentemente da forma física ou intelectual, possuem três recursos - conteúdo, contexto e estrutura -, os quais podem e devem ser refletidos por meio de metadados”.

Segundo a Organização Americana de Padronização da Informação, “metadado é a informação estruturada que descreve, explica, localiza um recurso informacional, ou possibilita sua recuperação, uso e gerenciamento. O termo metadados frequentemente designa dados sobre dados” (NISO, 2017, p. 1). Frente ao exposto, percebe-se que o termo metadados tem seu conceito cada vez mais expandido para abarcar a caracterização e relações de um objeto de informação – figura 2.

Figura 2. Características dos metadados.

Compreensão mais ampla de metadados é a soma total do que se pode dizer sobre qualquer Objeto de Informação (*Information Object*)

OBJETO DE INFORMAÇÃO

- Item ou grupo de itens digitais, não importando tipo ou formato, que pode ser dirigido ou manipulado como um objeto único por um computador (imagens digitais de obras de arte, PDFs, etc)
- Pode ser usado para “substitutos” digitais de documentos (DLO) e para descrever registros relacionados a objetos e/ou coleções (catálogo de registros, *finding aids*, etc)

3 características principais que devem (deveriam) refletir nos metadados:

Conteúdo - relativo ao que o objeto contém ou fala sobre. Intrínseco ao objeto de informação;

Contexto - indica os aspectos de quem, o quê, porquê, onde e como são associados ao objeto. Extrínsecos ao objeto

Estrutura - relativo ao conjunto formal de associações por dentro de, ou entre objetos individuais de informação. Podem ser intrínsecos, extrínsecos ou ambos.

Fonte: Barbosa (2018)

Diante da expansão da abordagem e uso, tem sido cada vez mais importante que, além dos profissionais de informação, pesquisadores, criadores e utilizadores de conteúdo digital, entendam os principais papéis dos diferentes tipos de metadados, de forma a assegurar acessibilidade, autoria, interoperação, granulação e preservação dos dados científicos e seus sistemas de registro.

2.2.2 TIPOS DE METADADOS

Kenney, Rieger e Entlich (2003) identificam três tipos de metadados: os **metadados descritivos**, que descrevem o conteúdo dos recursos informacionais; os **metadados estruturais**, que fornecem insumos sobre a estrutura de armazenamento das fontes de dados e os **metadados administrativos**, que controlam o acesso a cada um dos recursos informacionais identificados.

Refletindo o ponto de vista de Kenney, Rieger e Entlich (2003), percebe-se que, ao identificar tipos de padrões de metadados são estabelecidas as funções básicas de um metadado dentro de um sistema informacional. Para os autores, em uma biblioteca digital, os padrões de metadados incluem metadados com estas três funções distintas: os descritivos, que descrevem o recurso informacional (autor, título, assunto e outros); os estruturais utilizados para estruturação visual da biblioteca digital (basicamente *tags* de marcação de HTML); e os administrativos, que controlam datas de acesso ao recurso informacional, de direitos autorais etc. (senhas, data de digitalização, legislação sobre direitos de reprodução etc.) - ver quadro 3.

Quadro 3. Tipos de Metadados.

TIPO	OBJETIVO	ELEMENTOS	EXEMPLOS
Metadados Descritivos	<p>Descrição e identificação de recursos de informação: - No nível local (sistema) para permitir busca e recuperação (por exemplo, busca por uma coleção de imagens para encontrar pinturas com ilustrações de animais);</p> <p>- No nível da Web, permite que os usuários descubram recursos (por exemplo, pesquisando na Web para encontrar coleções digitalizadas sobre poesia).</p>	<p>- Identificadores exclusivos (PURL, Handle);</p> <p>- atributos físicos (mídia, características das dimensões),</p> <p>- atributos bibliográficos (título, autor / criador, idioma, palavras-chave).</p>	<p>- Localizador de Recursos (Localizador Uniforme e Contínuo de Recursos); Dublin Core; MARC; Meta HTML <i>Tags</i> (Meta <i>Tags</i> HTML).</p> <p>- vocabulários controlados, tais como: Thesaurus em Arte e Arquitetura; Categorias para a descrição de obras de arte.</p>

TIPO	OBJETIVO	ELEMENTOS	EXEMPLOS
Metadados Estruturais	<p>Facilitar a navegação e apresentação de recursos eletrônicos:</p> <ul style="list-style-type: none"> - Fornecem informações sobre a estrutura interna de recursos, incluindo página, seção, capítulo, numeração, índices e índice; - Descrevem a relação entre os materiais (por exemplo, a fotografia B foi incluída no manuscrito A); - Unem os arquivos e os textos relacionados (por exemplo, FileA é o formato JPEG da imagem do arquivo de FileB) 	<p>Estruturação de rótulos, como página de título, sumário, capítulos, partes, errata, índice, relação com um subobjeto (por exemplo, fotografia de um jornal).</p>	<p>SGML; XML; Descrição arquivística codificada, EAD (descrição do arquivo codificado); MOA2, Elementos Metadados Estruturais; União Eletrônica, Encadernação Eletrônica, Ebind).</p>
Metadados Administrativos	<p>Facilitar a gestão e processamento de coleções digitais a curto e longo prazo:</p> <ul style="list-style-type: none"> - Incluem dados técnicos sobre criação e controle de qualidade; - Incluem os requisitos de gerenciamento de direitos e controle de acesso e uso; - Informações sobre ação de preservação. 	<p>Dados técnicos, como tipo e modelo do scanner, resolução, profundidade de bits, espaço de cores, formato de arquivo, compactação, fonte de luz, proprietário, data de registro de direitos autorais, limitações de cópia e distribuição, informações sobre licença, atividades de preservação (ciclos de atualização, migração, etc.).</p>	<p>MOA2, Elementos de Metadados Administrativos; Biblioteca Nacional da Austrália, Metadados de Preservação de Coleções Digitais (Biblioteca da Universidade de Cornell / Departamento de Preservação e Conservação para coleções digitais)</p>

Fonte: Kenney, Rieger e Entlich (2003)

Nessa mesma linha, segundo Lourenço (2017),

Os metadados descritivos são utilizados para descrever um objeto digital, identificando por meio de etiquetas colocadas antes de dados replicantes visando uma melhor recuperação deste recurso informacional [...] Os metadados estruturais tem por função estruturar a apresentação dos objetos digitais contidos nas páginas web, de maneira que estes possam interagir entre si, proporcionando uma melhor recuperação da informação [...] E finalmente, os metadados administrativos que irão identificar dados que servirão, não para descrição dos objetos digitais, mas para sua preservação, para controle de usos deste objeto digital.

Gilliland (2016) identifica os tipos de metadados de acordo com a função destes no ambiente de uma biblioteca digital. Percebe-se que os tipos de metadados identificados por Gilliland (2016) e Kenney, Rieger e Entlich (2003) seguem uma estrutura lógica semelhante. A diferença entre as tipologias propostas é que Gilliland (2016) chama os metadados estruturais de “técnicos” e subdivide o tipo administrativo de Kenney em “administrativo”, “de preservação” e “de uso”, conforme quadro 4.

Quadro 4. Diferentes categorias de metadados e suas funções.

CATEGORIA	DEFINIÇÃO	EXEMPLO
Administrativo	Metadados usados no gerenciamento e administração de coleções e recursos de informações	<ul style="list-style-type: none"> • Informação de aquisição e avaliação • Direitos e rastreamento de reprodução • Documentação de requisitos e protocolos legais, culturais e de acesso à comunidade • Informação de localização • Critérios de seleção para digitalização • Documentação de repatriação digital
Descritivo	Metadados usados para identificar, autenticar e descrever coleções e recursos de informações confiáveis relacionados	<ul style="list-style-type: none"> • Metadados gerados pelo criador e sistema original • Pacote de informação de submissão • Registros de catalogação • Encontrar ajudas • Controle de versão • Índices especializados • informação curatorial • Relacionamentos interligados entre recursos • Descrições, anotações e emendas de criadores e outros usuários
Preservação	Metadados relacionados à gestão de preservação de coleções e recursos de informação	<ul style="list-style-type: none"> • Documentação do estado físico dos recursos • Documentação de ações tomadas para preservar versões físicas e digitais de recursos (por exemplo, atualização de dados e migração) • Documentação de quaisquer alterações que ocorram durante a digitalização ou preservação
Técnico	Metadados técnicos relacionados a como um sistema funciona ou metadados se comporta	<ul style="list-style-type: none"> • Documentação de hardware e software • Informações processuais geradas pelo sistema (por exemplo, metadados de roteamento e de evento) • Informações técnicas de digitalização (por exemplo, formatos, taxas de compactação, rotinas de dimensionamento) • Acompanhamento dos tempos de resposta do sistema • Autenticação e dados de segurança (por exemplo, chaves de criptografia, senhas)
Uso	Metadados relacionados ao nível e tipo de uso de coleções e recursos de informação	<ul style="list-style-type: none"> • registros de circulação • registros de exposições físicas e digitais • Uso e rastreamento de usuários • Informações sobre reutilização e multiversão de conteúdo • Logs de pesquisa • Metadados de direitos

Fonte: Gilliland (2016)

Para Gilliland (2016), os metadados administrativos são especificamente as informações a respeito da forma de aquisição do recurso informacional pela biblioteca digital, os direitos de reprodução relativos à permissão ou não de cópias e aos critérios de seleção para digitalização referentes a acervos físicos que foram digitalizados para formarem uma biblioteca digital. Já os metadados descritivos são os metadados utilizados para o tratamento dos recursos informacionais, utilizando para isso o conjunto de técnicas de catalogação, classificação e indexação da Ciência da Informação.

Os metadados de preservação, para Gilliland (2016), constituem os metadados utilizados para informar o usuário da biblioteca digital sobre o original físico daquele recurso digitalizado: seu estado de conservação, sua localização física, etc. Com relação aos recursos informacionais gerados na própria web, as informações de preservação serão os dados relativos ao organismo e/ou pesquisador responsável pelo seu conteúdo e outras formas de aquisição, no caso deste recurso informacional vir a não estar disponível na biblioteca digital.

A classificação apresentada pela Organização Americana de Padronização da Informação (NISO, 2017) abrange as duas visões já citadas e acrescenta de forma destacada às linguagens de marcação – Quadro 5.

Quadro 5. Tipos de metadados

TIPO	DESCRIÇÃO
Metadados Descritivos	Encontrar ou compreender recurso
Metadados Administrativos - Metadados Técnicos - Metadados de Preservação - Metadados de Direitos	Para decodificar e renderizar arquivos; Gestão de longo prazo de arquivos Direitos de Propriedade Intelectual referente ao conteúdo
Metadados Estruturais	Relações entre partes dos recursos
Linguagens de Marcação (<i>Markup Languages</i>)	Integra metadados e <i>flags</i> para outras feições estruturais ou semânticas com conteúdo

Fonte: (NISO, 2017)

Segundo Campos (2007), os metadados referentes à administração, acesso, preservação e uso das coleções, não apenas descrevem e identificam um objeto informacional, mas apontam as condições corretas ou ideais de seu gerenciamento, as relações do objeto com outros na coleção, sua função, utilização, comportamento, contexto de criação e condições de preservação.

Sayão (2010) entende que as discussões a respeito dos tipos e dos contornos das informações necessárias para instruir corretamente os processos de preservação digital permitiram estabelecer um consenso em torno de cinco grandes categorias de informação.

Essas categorias são materializadas por uma descrição aprofundada e ampla dos aspectos técnicos, custodiais e legais dos recursos digitais que devem ser traduzidos por metadados de preservação. Resumidamente, são as seguintes: **1) proveniência** – os metadados de preservação devem registrar informações sobre a história do objeto desde sua origem, traçando a sua cadeia de custódia e de propriedade; **2) autenticidade** – os metadados de preservação devem incluir informações suficientes para validar que o objeto é de fato o que diz ser e que não

sofreu alterações intencionais ou não documentadas; **3) atividades de preservação** – os metadados de preservação devem documentar as ações tomadas ao longo do tempo para preservar o objeto digital e as consequências dessas ações sobre aparência, usabilidade e funcionalidades do objeto; **4) ambiente técnico** - os metadados de preservação devem descrever as dependências técnicas necessárias para a apresentação e uso dos objetos digitais, tais como hardware, sistema operacional e software de aplicação; **5) gestão de direitos** – os metadados de preservação devem registrar todos os itens relacionados às questões de propriedade intelectual que limitem as ações de preservação, de disseminação e uso por parte de usuários de hoje e do future (SAYÃO, 2010, p. 11-12, grifo nosso).

Quanto à estruturação das informações necessárias para preservação digital na forma de esquemas de metadados, Sayão (2010, p. 12) afirma que muitos fatores devem ser levados em consideração, mas destaca três que considera mais importantes:

abrangência – o esquema deve ter uma amplitude tal, em termos de escopo e de profundidade, que considere as necessidades presentes e futuras de preservação do sistema de repositório considerado; **orientação para a implementação** – o esquema deve ser projetado tendo como perspectiva os níveis práticos de implementação e a possibilidade de adaptação a sistemas automatizados voltados para gerir e assinalar metadados; **interoperáveis** – os esquemas devem ser pensados para promover e facilitar as transações entre diversos fatores que envolvam o objeto digital e os seus diversos metadados ao longo do seu ciclo de vida, por exemplo, submissão a um repositório, disseminação para um usuário ou transferência para outro repositório (SAYÃO, 2010, p. 12, grifo nosso).

As diferentes abordagens na classificação dos metadados apresentam similaridades e são orientadas a caracterizar com base na função da descrição. A partir dessas divisões com base nas funções são estabelecidos os padrões de metadados, que podem ser genéricos ou mais específicos, conforme a granularidade do recurso informacional e a descrição desejada.

2.2.3 PADRÕES DE METADADOS

A obstrução do fluxo de informação científica apontada por Leite (2011, p. 3), considerando especificamente os dados científicos, dentre outros fatores, decorre principalmente da indisponibilidade (pesquisadores retêm os dados) e da baixa rastreabilidade dos dados científicos. Isso interfere na confiabilidade dos dados e resultados, conforme exposto por Spang-Hanssen (SPANG-HANSEN, 2001):

[...] nem a pessoa que solicita informações nem a pessoa que as entrega devem ignorar a confiabilidade dos dados e esquecer a configuração geral na qual os dados são adquiridos. As informações sobre alguma propriedade física de um material estão incompletas sem informações sobre a precisão dos dados e sobre as condições sob as quais esses dados foram obtidos. Além disso, várias investigações de uma propriedade muitas vezes levaram a resultados diferentes que não podem ser comparados e avaliados separadamente das informações sobre seus antecedentes.

Segundo a IDC (2012), o universo digital compreende dados de todos os tipos, no entanto, a grande maioria dos novos dados gerados não é estruturada e, na maioria das vezes, sabemos pouco sobre os dados, a menos que sejam de alguma forma caracterizados ou marcados, uma prática que resulta no registro de metadados. A IDC (2012) afirma que os metadados são um dos subsegmentos que mais crescem no universo digital, embora os próprios metadados sejam uma pequena parte desse universo em geral, pois a estimativa é de que apenas 3% dos dados potencialmente úteis possuem metadados. Para a Organização Americana de Normatização da Informação (NISO, 2017), os metadados são sempre estruturados para cumprir a uma finalidade útil e auxiliam o gerenciamento de objetos digitais, bem como fornecem os insumos necessários para processar o conteúdo digital de maneira apropriada para atender às necessidades do usuário.

Os metadados influenciam tanto no acesso quanto na recuperação da informação, mas para que tenham aceitação internacional e auxiliem na interoperabilidade eles devem possuir uma padronização (SANTOS, SIMIONATO e ARAKAKI, 2014, p. 149). De acordo com o estudo elaborado por Polydoratos e Nicholas (2001, p. 316-317), as denominações mais populares utilizadas para se referir aos padrões de metadados são: esquemas e padrões (20%), conjuntos de elementos (19%) e formatos (16%), embora outras denominações como os sistemas (12%), catálogos (11%) e outros (2%) sejam também utilizadas. Entretanto, os autores ressaltam que há distinções na aplicação das expressões formato de metadados e esquema de metadados. Sendo a primeira atribuída ao conjunto de regras que governam a estrutura e o conteúdo dos dados, e a segunda referente ao conjunto estruturado de atributos com nomes dos elementos e uma semântica associada (POLYDORATOU e NICHOLAS, 2001, p. 315).

Já definição da *Dublin Core Metadata Initiative* (DCMI) de esquema / esquema justifica a assimilação de ambas as denominações (formato e esquema): combinação sistemática e ordenada de elementos. Um conjunto de regras, mantido por uma comunidade específica de usuários, para codificar informações.

Rodríguez (2001, p. 787) entende o padrão de metadados como um conjunto de elementos que compõem uma especificação (geralmente associada a nomes de um domínio de informações) dos rótulos que devem ser usados com um valor dêitico para um objeto de informação (recursos de informação ou *Document Like-Object – DLO*). Segundo a autora, “um objeto de informação é, em termos gerais, um grupo de elementos digitais que um

computador pode manipular como se fosse um único objeto, independentemente do formato que eles têm” (RODRÍGUEZ, 2001, p. 95), sendo DLO uma expressão amplamente reconhecida na literatura sobre metadados para se referir a documentos da internet, que podem ser formados por texto, imagem, áudio ou vídeo e hipertexto. Rodríguez ainda ressalta que a abstração do DLO é uma questão chave no campo da informação distribuída na Web e a base para definir os metadados necessários para caracterizá-lo e recuperá-lo.

Segundo Méndez Rodríguez, mesmo sendo criados para o contexto tecnológico, muitos dos padrões de metadados desenvolvidos não atendem satisfatoriamente às necessidades de representação informacional em um domínio específico, pois seus esquemas de descrição são amplos e gerais e por isso não apresentam especificidade do domínio. Deste modo, presencia-se atualmente o delinear de um novo panorama, pautado no uso intensivo de tecnologias de informática e no estabelecimento de princípios mais consistentes para a determinação de metadados para que representem um recurso informacional com unicidade

Alves (2010, p. 47) entende que os padrões de metadados podem ser descritos como,

[...] estruturas de descrição constituídas por um conjunto predeterminado de metadados (atributos codificados ou identificadores de uma entidade) metodologicamente construídos e padronizados. O objetivo do padrão de metadados é descrever uma entidade gerando uma representação unívoca e padronizada que possa ser utilizada para recuperação da mesma.

Um padrão de metadados define parâmetros que devem ser utilizados de forma global, tornando mais prático e seguro o compartilhamento dos conjuntos de dados. Portanto, estabelecem regras consistentes e vocabulários padronizados, que possibilitam aperfeiçoar e tornar eficiente a comunicação entre os geradores e consumidores de dados (SILVA, CORRÊA, *et al.*, 2015).

Dessa forma, quanto mais adequado o registro de metadados melhores serão as condições para que os dados sejam validados, registrados, preservados, disseminados e reutilizados por outros pesquisadores em novos processos de uso do conhecimento, permitindo aumentar a efetividade do sistema de produção e comunicação da informação científica.

2.3 BIODIVERSIDADE E A GOVERNANÇA DOS DADOS CIENTÍFICOS

Os modelos e processos informacionais podem ser analisados por diferentes perspectivas, que variam conforme a demanda por conhecimento, os contextos sociais e tipos de abordagens. Além disso, cada área do conhecimento apresenta formas próprias para gerir e comunicar seus dados. Portanto, existe uma estrutura de governança que se estabelece em cada contexto, gerando desafios para generalização de modelos voltados a tratar os fenômenos informacionais nos mais diversos campos do conhecimento.

Durante as últimas décadas, há registro de muitas iniciativas de pesquisas referentes ao tratamento dos dados sobre biodiversidade em escalas global, regional e local com um objetivo claro: compilar e compartilhar dados, tornando-os abertos à ciência em todo o mundo (GILMAN, KING e PETERSON, 2009). Em uma perspectiva ideal, os dados disponíveis gratuitamente na *web* oferecem uma vasta gama de usos potenciais que podem promover avanços na pesquisa da biodiversidade e melhor apoio às decisões.

As atuais práticas de financiamento e publicação exigem cada vez mais o compartilhamento dos conjuntos de dados resultantes das pesquisas e a maioria dos pesquisadores entende que a publicação dos dados é, portanto, não apenas desejável, mas necessária. De fato, roteiros para eventualmente institucionalizar a publicação de conjuntos de dados foram propostos (MORITZ, KRISHNAN, *et al.*, 2011) e muitos desafios para tal mudança cultural já foram identificados (CHAVAN, PENEV e HOBERN, 2013). No entanto, tornar os dados acessíveis, compreensíveis e verdadeiramente reutilizáveis continua a ser um desafio (COSTELLO e WIECZOREK, 2014), mas esses dados não podem ser compartilhados e usados sem critérios e precisam ser mobilizados de acordo com padrões estabelecidos, amplamente conhecidos e aceitos (PETERS, 2010).

Nesse sentido, é importante discutir caminhos para a governança de dados biodiversidade, abordando os aspectos relacionados à gestão de dados heterogêneos, bem como a possibilidade de aplicação do modelo de governança dos comuns conforme proposto por Elinor Ostrom (OSTROM, 1990; OSTROM e HESS, 2007), que considera a informação como recurso. Segundo Glushko (2013),

O recurso possui um sentido comum de 'qualquer coisa de valor que pode apoiar a atividade orientada a um objetivo. Esta definição significa que um recurso pode ser uma coisa física, ou uma coisa não física, informações sobre coisas físicas, ou informações sobre coisas não físicas, ou qualquer coisa que você quer organizar. Em outras palavras, o que direciona este amplo escopo são entidade, objeto, item,

e instancia. O termo documento é frequentemente utilizado para um recurso informacional, seja ele em formato digital ou analógico (GLUSHKO, 2013, p. 8).

Conforme explicam Santos e demais autores (2014, p. 148), entende-se que o termo recurso informacional refere-se à informação objetivada no contexto de um campo do conhecimento podendo ser apresentado em uma estrutura analógica e/ou digital, com valor informacional que caracteriza a sua concepção intelectual expressa na corporificação de manifestações estruturadas na forma de itens.

A governança refere-se a padrões de articulação e cooperação entre atores e arranjos institucionais que coordenam e regulam transações dentro e através das fronteiras dos diferentes setores da sociedade. Incluindo não apenas os mecanismos tradicionais de agregação e articulação de interesses, tais como os partidos políticos e grupos de pressão, mas também redes sociais informais – de fornecedores, prestadores de serviços e outros –, hierarquias e associações de diversos tipos. Segundo Calame (2001), governança engloba o conjunto de processos, costumes, políticas, leis, regulamentos e instituições que regulam a maneira como algo – uma empresa, um projeto ou um sistema – é dirigido, administrado ou controlado.

O termo inclui também o estudo sobre as relações entre os diversos atores envolvidos – *stakeholders* – e os objetivos pelos quais a gestão se orienta. Isto incide sobre o aspecto formal e informal dos atores envolvidos na gestão dos processos de execução das atribuições, considerando as estruturas formais e informais estabelecidas para implantar os objetivos estratégicos.

Assim, a eficiência de um modelo de gestão depende da capacidade de definir os domínios de governança para cada atribuição e, também, de adaptar-se a novas situações. Além disso, a qualidade do modelo de gestão apresenta relação direta com a habilidade em adotar sistemas de representação dos processos e de organização dos recursos em busca de uma autogestão que permita a boa governança.

Há diferentes perspectivas em diversas áreas de estudos que abordam questões referentes à gestão e comunicação de dados, informação e conhecimento. Considerando que a gestão de dados e metadado é assunto central desta pesquisa, cabe compreender a perspectiva da Ciência da Informação e da Ciência de Dados para lidar com diferentes contextos, ou seja, diferentes estruturas de governança.

A abordagem da CI para essa questão, segundo Costa, Leite e Tavares (2018, p. 14) é

Entre as características que desde sempre distinguem a Ciência da Informação das áreas como a Biblioteconomia e a Documentação e a Informática, duas merecem destaque. A primeira é a preocupação com a comunicação da informação entre seres humanos que, em sentido amplo, envolve questões relacionadas com a produção, compartilhamento e uso da informação nas sociedades. Este talvez seja o fenômeno nuclear e fundamental de interesse do campo, a partir do qual irradiam tópicos de grande relevância como é o caso do comportamento e práticas informacionais e organização da informação. A segunda, decorrente da primeira, é a noção de contexto, que diz respeito ao conjunto de forças – sociais, tecnológicas, culturais, econômicas e legais – que em qualquer medida exercem influência sobre as práticas de produção, compartilhamento e uso da informação. Subjacente e derivada de tais preocupações, a gestão da informação e, mais recentemente, a gestão do conhecimento – subáreas dedicadas às questões relacionadas com a sistematização dos processos de produção, compartilhamento e usos da informação e do conhecimento – têm sido fortemente desenvolvidas no campo da Ciência da Informação.

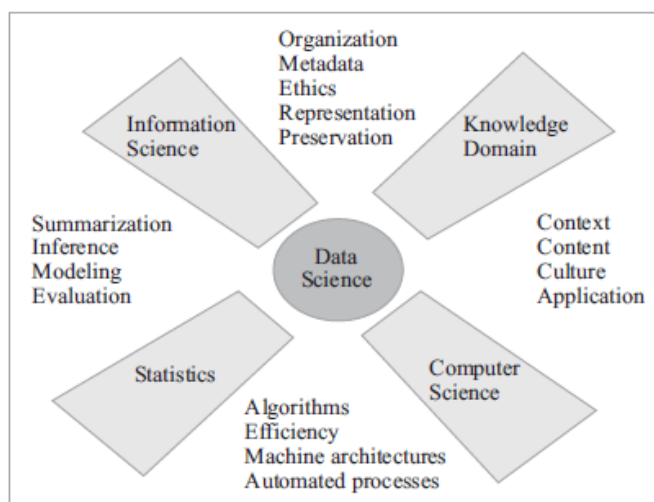
Conforme apontam os autores, há uma crescente preocupação com a estrutura de governança que se impõe aos objetos de estudo da CI, ou seja, o conjunto de forças – sociais, tecnológicas, culturais, econômicas e legais – que exercem influência sobre os processos informacionais.

Já na perspectiva da Ciência de Dados houve uma evolução de caráter mais prático, voltada à aplicação de soluções para resolver demandas emergentes, conforme explica Marchionini (2016),

Há muito se discute as distinções da biblioteconomia, da ciência da informação e da informática, e como essas áreas diferem e se sobrepõem à ciência da computação. Hoje, o termo ciência de dados está surgindo, gerando entusiasmo e questionamentos sobre como ele se relaciona e difere dessas outras áreas de estudo. Para os nossos propósitos aqui, considero a ciência da informação como o termo geral que inclui a biblioteconomia e a informática e foca-se em distinções e semelhanças entre essas disciplinas que cada uma informa a ciência de dados. Nos níveis mais gerais, a ciência da informação lida com a gênese, fluxo, uso e preservação da informação; A ciência da computação lida com algoritmos e técnicas para processos computacionais. A ciência de dados como um conceito emerge das aplicações de estudos existentes de medição, representação, interpretação e gerenciamento para problemas nas áreas do comércio, saúde, meio ambiente, governo e outros domínios. Cada uma dessas áreas aproveita cada vez mais a melhoria rápida das habilidades de capturar, compartilhar e analisar fluxos de dados. Em última análise, a ciência de dados é importante porque as pessoas são capazes de tomar melhores decisões sobre o mundo - isto é, como as pessoas na medicina, nos negócios, no governo ou na pesquisa aplicam seus princípios e técnicas aos problemas específicos do domínio (MARCHIONINI, 2016, p. 1).

Marchionini (2016, p. 2) também explica que a Ciência de dados, como qualquer campo emergente, extrai teoria e prática das bases de conhecimento existentes e propõe um modelo para o surgimento dessa ciência, apoiado em quatro principais alicerces: ciência da informação; estatística; ciência da computação; e um setor geral representando diferentes domínios do conhecimento – Figura 3.

Figura 3. Bases da Ciência de Dados.



Fonte: Marchionini (2016, p. 2)

Portanto, a Ciência de Dados tem origem no atendimento de demandas práticas, que tendem a utilizar as melhores ferramentas e métodos para solucionar problemas específicos em cada contexto. Enquanto a Ciência da Informação procura compreender os fenômenos informacionais de forma mais ampla, considerando o papel de cada elemento em cada contexto. Em comum, ambas tendem a tratar em seus estudos de contextos específicos, persistindo os desafios impostos pelas diferentes estruturas de governança encontradas em cada caso.

O foco na especificidade dos contextos estudados é um dos fatores limitantes para transpor as barreiras que levam a obstrução dos fluxos de informação entre diferentes áreas do conhecimento, dificultando seu uso e circulação. Isto ocorre porque não é possível prever todos os possíveis usos a que um dado e uma informação podem vir a ser destinados. Nessa perspectiva, os modelos e processos propostos são limitados aos contextos considerados. Assim, independente da abordagem, seja por meio da Ciência de Dados ou na Ciência da Informação, para reduzir as barreiras que dificultam o uso e a circulação dos dados, é relevante compreender como superar as limitações de impostas pelas estruturas de governança existentes.

2.3.1 GOVERNANÇA DE RECURSOS INFORMACIONAIS HETEROGÊNEOS

“Os desafios que os profissionais da informação enfrentam não são novos: de algum modo, foram percebidos por estudiosos e teóricos, mas precisaram se ajustar às contribuições tecnológicas do momento” (GOMES, 2018, p. 17). Segundo Gomes (2018), o maior desafio, na atualidade, são os arquivos correntes e heterogêneos em meio eletrônico, conforme exemplifica:

Trazendo à tona a problemática dos arquivos heterogêneos, alguns exemplos mostram a complexidade do problema. Também afetado pela introdução da computação, ressalta em tal contexto a necessidade de tratamento terminológico voltado não somente para o tratamento da documentação textual, mas para documentos de natureza vária em diversificados suportes. Este ambiente requer a inclusão do tratamento do conteúdo do documento.

Alguns exemplos: seja a documentação de um projeto de pesquisa em Antropologia: ela pode conter, além de textos – cadernos de anotação, relatórios e assemelhados –, filmes e fotos sobre o comportamento e valores de uma comunidade – dança, iniciação, entrevistas e assemelhados –, e assim por diante. Ou a documentação de um instituto de pesquisa agrícola que pretende desenvolver melhoramento de uma espécie, cuja documentação inclui análises diversas que resultam em imagens. Ou, ainda, um departamento de pesquisa em patologia que precisa organizar e recuperar algumas centenas de lâminas.

Em tais ambientes de pesquisa, a documentação é heterogênea e requer tratamento de seu conteúdo. Este é um aspecto relevante para o profissional de informação. E o instrumento de controle é a taxonomia com seus termos controlados do ponto de vista semântico (GOMES, 2018, p. 22).

Para Gomes (2018), a recuperação da informação traz questões relacionadas à classificação e à terminologia e desafia o profissional para tratar informações em várias áreas do saber. Nesse contexto, o autor ressalta que a web semântica revolucionou a Ciência da Informação a partir da introdução do conceito de metadado.

Ao propiciar que cada dado – nome próprio ou não –, seja registrado em um metadado que indique sua classe (pessoa, objeto, tamanho ou qualquer outra característica considerada relevante no momento do planejamento do serviço), torna-se possível recuperar informação com mais precisão. Assim, um nome de pessoa pode estar caracterizado em suas diferentes funções em um documento: ou autor, ou assunto, ou revisor, ou tradutor e por aí vai. Do mesmo modo, na descrição do conteúdo, o metadado pode indicar todo ou qualquer aspecto da descrição do objeto: tamanho, velocidade, parte ou assemelhado. Como indica o nome, o metadado implica um termo amplo, uma classe, cujos dados são representados por termos da mesma classe (GOMES, 2018, p. 22).

Ainda segundo Gomes (2018), a introdução do metadado requer a elaboração dos dados/termos em taxonomias, mas o vocabulário não deve ser estabelecido a priori, pois a taxonomia é criada no momento que surge a necessidade de um novo termo/conceito, cuja posição é estabelecida a partir de princípios teóricos, portanto, a posteriori. Assim, essa estrutura sistemática possibilita a correta escolha do termo para um metadado. Entretanto,

para o autor essas limitações deixam de existir na recuperação da informação, pois todos os elementos de descrição considerados relevantes em um dado contexto podem potencialmente produzir produtos informacionais diversos.

Conforme apontam Cavalcante e Kobashi (2018, p. 142), “atualmente, a discussão sobre a interoperabilidade e a integração de acervos de diferentes instituições como arquivos, museus e bibliotecas está centrada nas tecnologias da web semântica e de dados abertos interligados, com uso de anotações semânticas, metadados e ontologias”. Dessa forma, a web semântica representa um caminho promissor para integrar grande quantidade de dados provenientes de diferentes contextos, mesmo existindo divergências de estruturas e significados. Assim, considerando suas tecnologias e seus processos ainda em amadurecimento, a web semântica pode ser considerada um modelo capaz de estimular a integração das áreas da CI, englobando as raízes social e tecnológica, bem como exercendo influência na governança e estimulando a integração da gestão e comunicação de dados.

A possibilidade de coexistência de contextos diferentes e até divergentes para a classificação de objetos na web semântica permite lidar com os principais desafios de governança enfrentados por grande parte dos modelos de gestão e comunicação de dados, que são vinculados a áreas específicas. Por isso, é importante analisar a evolução da governança de dados considerando que os dados podem ser utilizados de diversas formas, com diferentes significados, por diferentes atores e em diferentes contextos que se somam e não se anulam. Nesse sentido, o modelo de governança dos comuns pode contribuir adicionalmente para o estabelecimento de regras de boa governança.

2.3.2 MODELO DE GOVERNANÇA DOS COMUNS

Segundo Ostrom (2007, p. 4-5), Comuns (*Commons*) é um termo geral que se refere a um recurso compartilhado por um grupo de pessoas, que pode ser: pequeno e servir um pequeno grupo (computador da família); em nível comunitário (bibliotecas); ou pode se estender a níveis internacionais e globais (internet e conhecimento científico). Em relação aos limites, os Comuns podem ser bem limitados (biblioteca comunitária); transfronteiriços (internet); ou sem limites claros (conhecimento).

O modelo de Ostrom aponta oito princípios para a boa governança dos Comuns:

1. Limites claramente definidos;

2. Coerência das regras com as necessidades e condições do contexto;
3. Arranjos de decisão coletiva. Indivíduos afetados por essas regras devem ter acesso para participar do controle e evolução das regras.
4. Reconhecimento de direitos de organização. O direito dos membros da comunidade de elaborar suas próprias regras é respeitado por autoridades externas.
5. Um sistema de monitoramento.
6. Sanções graduadas.
7. Arranjos de resolução de conflitos. Os membros da comunidade devem ter acesso a mecanismos de resolução de conflitos de baixo custo;
8. Alinhamento e articulação da gestão entre as partes envolvidas, permitindo que a apropriação, provisão, monitoramento e sancionamento, resolução de conflitos e outras atividades de governança sejam organizadas em uma estrutura aninhada com várias camadas de atividades.

“A análise de qualquer Comum deve envolver as regras, decisões e comportamentos realizados em grupo em relação ao recurso compartilhado” (OSTROM e HESS, 2007, p. 7). Sob essa ótica, entre os desafios encontrados ao lidar com os fenômenos informacionais, podemos analisar questões referentes à gestão dos dados científicos abertos, ao aperfeiçoamento do fluxo de informação científica e aos repositórios de dados científicos, bem como outros assuntos similares. Todas essas questões podem ser analisadas sob a perspectiva de construir um modelo de governança voltado ao uso de dados, informações e conhecimentos científicos como bens comuns.

É relevante ressaltar que para Ostrom (2007, p. 7), o conhecimento abrange todas as ideias, informações e dados inteligíveis em qualquer forma na qual são expressos ou obtidos. Assim, especificamente, o Conhecimento Comum pode referir-se a vários tipos de bens e regimes, bem como apresentar muitas características similares aos demais recursos comuns – aquíferos, terras e recursos pesqueiros. Entretanto, o Conhecimento Comum não obedece ao princípio da subtrabilidade do modelo original de Ostrom (1990), pois o seu uso não reduz a disponibilidade para o uso por outras pessoas (OSTROM e HESS, 2007, p. 48). Para Ostrom (2007, p. 11), “a tragédia dos anticomuns na arena do

conhecimento está na potencial subutilização de recursos científicos causada pelo excesso de direitos de propriedade intelectual e patentes na pesquisa”.

A temática do Conhecimento Comum aplicada ao modelo de Ostrom pode apresentar ainda mais benefícios. Exemplo disso pode ser observado no acesso à informação, que é muito diferente do acesso à terra ou à água. Neste último caso, o acesso pode significar uso irrestrito, levando ao consumo excessivo e ao esgotamento. No que se refere ao Conhecimento Comum, em vez de ter efeitos negativos, o acesso à informação proporciona um bem público universal: quanto mais informação de qualidade, maior o bem público.

Assim, como qualquer outro recurso comum, o Conhecimento Comum é um recurso compartilhado vulnerável a dilemas sociais, pois os resultados das interações de pessoas e recursos podem ser positivos ou negativos. Frequentemente, no que se refere ao conhecimento, o conceito de Comuns é facilmente ligado à liberdade de expressão, acesso universal aberto e auto governança.

A aplicação do modelo dos Comuns para avaliar aspectos de gestão e comunicação de dados permite considerar a coexistência de diversos modelos de governança existentes, os quais podem ser articulados para alcançar objetivos comuns. Nesse sentido, isso permite analisar os aprimoramentos que a utilização da web semântica representa para superar os desafios na governança de recursos informacionais heterogêneos.

2.3.3 WEB SEMÂNTICA

Semântica é um termo usado para descrever o estudo do significado. Explicando de forma simplificada, o estudo da semântica “é o campo explora como os conceitos são codificados na linguagem através de um sistema de símbolos que facilitam a comunicação entre as entidades” (CALARESU e SHIRI, 2015). Assim, é importante considerar que a linguagem humana é extremamente complexa, especialmente quando considerado os diversos conceitos que uma palavra pode assumir em diferentes contextos, bem como um conceito pode ser representado por diferentes palavras ou conjunto de palavras.

A Web é a maior e mais diversificada fonte de conhecimento da humanidade. Em junho de 2018, o número estimado de usuários ativos da Internet aumentou para mais de

4,2 bilhões, representando cerca de 55% da população da Terra (INTERNET WORLD STATS, 2018). Diante da grandiosidade e da expressiva taxa de crescimento da circulação de informações na Web, conduzir a sua organização de uma forma a permitir o uso efetivo de seus conteúdos é um dos desafios mais prementes deste século. “A Web se desenvolveu mais rapidamente como um meio de transmissão de documentos para as pessoas, ao invés de dados e informações que possam ser processados automaticamente” (LAUFER, 2015, p. 20).

De acordo com Calaresu e Shiri (2015, p. 84), o ponto crucial está no paradoxo de que, embora o volume de dados codificados e enviados para a Web só possa ser realisticamente ordenado por máquinas, a linguagem humana em si é complexa, o que torna difícil para que as máquinas atuais possam interpretar e extrair informações e conhecimento. Segundo Jackendoff (2012),

Avanços recentes em psicologia e neurociência contribuíram muito para melhorar nossa compreensão de como o cérebro desempenha tarefas. A pesquisa sobre o funcionamento interno da estrutura da linguagem já nos deu a análise mais detalhada e precisa conhecida sobre a realização de qualquer tarefa pelo cérebro, mas ainda há muito a aprender. O que está claro é que, quando se trata de linguagem humana, nem mesmo o computador mais avançado é compatível com as habilidades do cérebro humano (JACKENDOFF, 2012).

A web atual liga documentos e será incrementada pela web semântica, que permite ligar dados. O termo web semântica foi apresentado por Berners-Lee (1997), o criador da web, na primeira *World Wild Web Conference* realizada em 1994. Entretanto, o potencial da web semântica só foi mais bem explicado em 2001, quando Berners-Lee e demais autores publicaram um artigo na revista *Scientific American* (LAUFER, 2015, p. 33), intitulado *The Semantic Web - A new form of Web content that is meaningful to computers will unleash a revolution of new possibilities* (Web Semântica - um novo formato de conteúdo para a Web que tem significado para computadores vai iniciar uma revolução de novas possibilidades). Esse artigo descreveu uma evolução esperada da web existente para uma web semântica. Assim, a ideia de agregar semântica aos dados visa facilitar o entendimento e a interoperabilidade do universo de informações heterogêneas, publicadas em diferentes formatos e com diferentes protocolos de acesso na web.

Segundo Sayão (2005), alguns autores definem a “web semântica como uma extensão da Web tradicional, onde, a partir do uso intensivo de metadados, espera-se obter o acesso automatizado às informações, com base no processamento semântico de dados e

heurísticas feitos por máquinas”. Porém, Sayão (2005) ressalta que desde 1998 a equipe da W3C trabalha arduamente no “desenvolvimento de tecnologias avançadas, que visam à representação estrutural e semântica dos recursos na Web” (SAYÃO, 2005, p. 283). Ainda segundo o autor, essas tecnologias, aliadas à teoria de domínios ou ontologias, permitem oferecer um serviço com um nível maior de qualidade, tornando a Web capaz de tecer uma rede extensa de conhecimento humano, podendo ainda, por meio do processamento via máquina, inferir novos conhecimentos.

A web semântica segue os padrões do *World Wide Web Consortium* (W3C), que definem formatos de dados e protocolos na web. De acordo com o W3C (2013), "A Web Semântica fornece uma estrutura comum que permite que os dados sejam compartilhados e reutilizados em toda a aplicação, empresa e limites da comunidade". Portanto, é considerada como um integrador em de diferentes conteúdos, aplicações de informação e sistemas.

Segundo Laufer (2015, p. 7), “Web Semântica é uma teia de informações construída de forma a ser facilmente processável por máquinas em uma escala global. A ideia geral é criar uma maneira eficiente para representar dados na *Web* de forma a construir um banco global de dados conectados”. Assim, a web semântica tem por objetivo tornar a web um sistema global, distribuído e baseado em semântica.

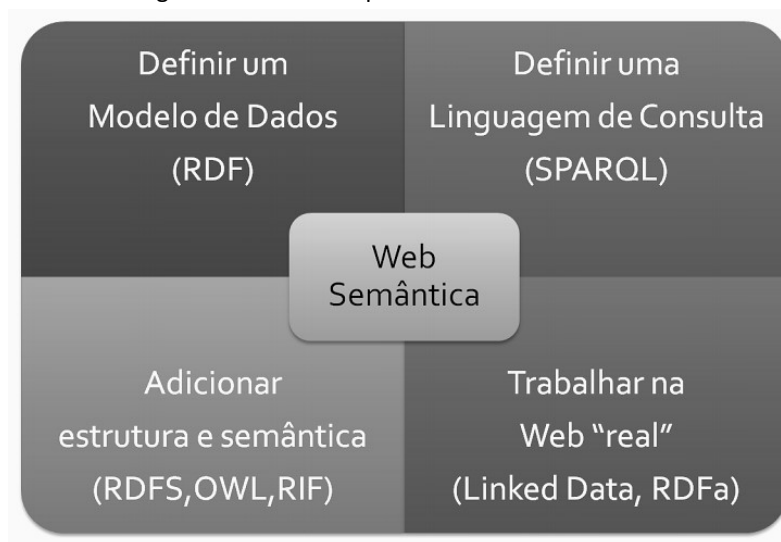
A web semântica não apresenta uma definição formal e existem diferentes abordagens e aplicações, tais como a integração de dados, representação e análise de conhecimento, serviços de catalogação, aprimoramento de algoritmos e métodos de pesquisa, redes sociais e outras (W3C, 2013). Mas para compreender como ocorre esse aprimoramento, é preciso ter uma noção das tecnologias e processos que integram o modelo da web semântica.

2.3.3.1 Modelo simplificado da web semântica

Considerando suas tecnologias e seus processos ainda em amadurecimento, a seguir é apresentado um modelo simplificado da web semântica – figura 4 – contendo quatro elementos fundamentais: 1) representação conceitual dos termos e ligação entre eles, estabelecendo uma estrutura linguística similar a humana com sujeito, predicado e objeto (grafos – modelo de dados em RDF); 2) vocabulário controlado – estruturas de conhecimento de taxonomias, tesouros e ontologias como elementos correlacionados a

arquitetura proposta permitindo atribuir significado a cada objeto (estrutura e semântica – RDFS, OWL e RIF); 3) integração dinâmica e real de dados (ligação dos dados na web – URIs); 4) meios para a pesquisa de dados (linguagem de consulta – SPARQL).

Figura 4. Modelo simplificado da web semântica.



Fonte: Souza (2013)

A web semântica é baseada no uso do *Resource Description Framework* – RDF (Estrutura de Descrição de Recursos), que é o modelo conceitual padrão definido pela W3C para ligar dados na Web, cujo ponto forte refere-se à facilidade em promover a interoperabilidade de estruturas diferentes, suportando a evolução dos esquemas ao longo do tempo sem exigir que as fontes e os consumidores de dados sejam alterados. Isotani e Bittencourt (2015, p. 104), esclarecem

Resource Description Framework (RDF) é a especificação proposta pelo *World Wide Web Consortium* (W3C) para descrever metadados. Ela permite criar triplas que contêm um nó sujeito, uma relação chamada de predicado e o nó objeto (sujeito, predicado, objeto). Mediante essa tripla, é possível indicar a relação entre dados e usá-la para representar a semântica contida neles.

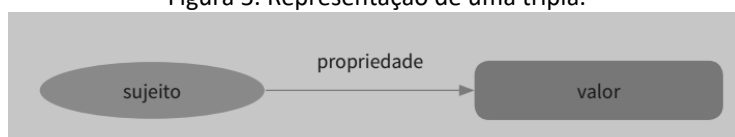
“Além de auxiliar na descrição dos recursos disponíveis na Web de maneira mais precisa (descrição sintática), o RDF também oferece a possibilidade de descrever a relação e os significados entre diversos recursos (descrição semântica)” (ISOTANI e BITTENCOURT, 2015, p. 59). Segundo Laufer (2015, p. 34),

O RDF permite fazer afirmações sobre recursos. Recursos são quaisquer coisas, tanto concretas quanto abstratas. Uma determinada empresa, uma pessoa, uma página Web são considerados recursos. Um sentimento, uma cor, também são recursos.

O formato das afirmações é simples. Uma afirmação RDF consiste de três elementos (uma tripla) e tem a seguinte estrutura: <sujeito> <predicado> <objeto>. Uma afirmação RDF expressa uma relação entre dois recursos. O sujeito e o objeto representam os dois recursos sendo relacionados; o predicado representa a natureza desta relação, que é formulada de modo direcional (do sujeito para o objeto) e é chamada em RDF de propriedade. Um objeto pode também ser um literal, definindo uma propriedade para um recurso.

Essa estrutura de vinculação de várias triplas – figura 5 – forma grafos, no qual as setas representam o predicado de ligação entre dois recursos, sendo os recursos representados pelos nós. A visualização na forma de grafo é o um modelo mental simplificado frequentemente usado para facilitar a compreensão das conexões entre os dados (W3C, 2013). Assim cada sujeito e objeto têm URI próprio e a ligação entre eles também tem significado e se dá por meio de um predicado, representado por uma seta.

Figura 5. Representação de uma tripla.



Fonte: Adaptado de Laufer (2015).

O RDF estende a estrutura de vinculação da Web utilizando *Uniform Resource Identifier* – URI (identificador uniforme de recursos) para nomear cada um dos recursos (W3C, 2013). Segundo Hondros (2010),

O RDF é uma estrutura útil para integração de dados. Ao usar URIs para identificar entidades e modelar relacionamentos entre elas, simplifica a integração de dados de fontes heterogêneas que implicitamente compartilham essas entidades. (...) todas as unidades de publicação mantêm vocabulários controlados (...) que se sobrepõem a outras unidades de publicação; no entanto, como todos usam diferentes formatos e identificadores, os dados não podem ser combinados. O RDF garante que os dados de uma unidade de publicação sejam compatíveis com os dados de outro, referindo-se ao mesmo URI que pode ser usado para identificá-lo (HONDROS, 2010).

Conforme apontam Isotani e Bittencourt (2015, p. 105), “para expressar a semântica por meio de triplas, também conhecidas como vocabulário RDF, é necessária a definição de *tags*. *RDF-Schema* (ou RDF-S) é utilizada para tal fim. RDF-S é a especificação que define classes, propriedades e seus relacionamentos”.

Além disso, SPARQL (*SPARQL Protocol and Query Language*) é a linguagem de consulta da Web Semântica. Podemos fazer uma analogia entre SPARQL e a linguagem SQL de consulta a bancos de dados relacionais, considerando que SPARQL tem uma sintaxe

adequada e admite uma série de filtros e operadores que permitem fazer consultas complexas aos dados representados como um conjunto de triplas RDF (LAUFER, 2015, p. 55 - 57).

Diversos bancos de dados de triplas oferecem pontos de acesso via Web (URLs), que aceitam o protocolo SPARQL e sua linguagem de consulta. Esses pontos de acesso são denominados SPARQL *endpoints*. Um SPARQL *endpoint* aceita consultas e retorna os resultados via HTTP. *Endpoints* genéricos executam consultas em qualquer dado RDF com acesso possível pela Web (especificado como parâmetro). *Endpoints* específicos executam consultas apenas em conjuntos de dados particulares, fixados pela aplicação (LAUFER, 2015, p. 58).

As tecnologias da web semântica são relacionadas à forma como sintaticamente os metadados podem ser agregados às informações, mas para que o significado dos dados possa ser igualmente compreendido por todos é necessário o uso de vocabulários controlados que possuam uma semântica bem definida, em especial a utilização de vocabulários de referência, ou vocabulários de uso mais comum (LAUFER, 2015, p. 92).

Segundo Kobashi (2018, p. 92), vocabulário controlado

é um sistema autônomo, estável, coeso, que requer processos contínuos de interação com o contexto: o que será organizado; para quem será organizado. Portanto, a construção de um vocabulário controlado (VC) requer procedimentos sistemáticos, entre os quais:

- a) reconhecimento do contexto de uso;
- b) modelagem do domínio de aplicação;
- c) escolha de um método de coleta de termos;
- d) definição dos termos;
- e) estabelecimento de relações lógico-semânticas entre os termos;
- f) padronização sintática dos termos;
- g) teste de aplicação;
- h) correção de inconsistências;
- i) elaboração da versão final.

Conforme aponta Almeida (2018, p. 108), “recursos informais não são capazes de prover interoperabilidade entre sistemas, como tanto se divulga. A interoperabilidade semântica, que permita a comunicação automática entre máquinas, e enseja o tipo de sistema e ambiente previsto para o futuro, só é possível via ontologias.” Ainda segundo o autor, isso acontece porque são construídas a partir de princípios filosóficos, que lhes dão o caráter de verdade, como correspondência à realidade, e especificadas em linguagens lógicas, onde o rigor e as restrições estabelecem sem ambiguidade o significado pretendido para o termo.

Nesse sentido, a Web Semântica se torna um modelo diferenciado à medida que viabiliza a coexistência de diversos vocabulários controlados para um mesmo conjunto de

dados, permitindo que possam ser organizados e analisados por perspectivas diferentes e até divergentes.

2.3.3.2 Gestão de dados na web semântica: aspectos sociais e tecnológicos

Nos modelos tradicionais da CI, define-se primeiramente o contexto de uso dos dados (a governança), para em seguida traçar os procedimentos de gestão e comunicação. É fato que não é possível conhecer, mesmo que para um universo de dados restrito, todos os possíveis contextos que podem ser atendidos, pois sempre poderão surgir novas perspectivas de análise dos dados. Nesse sentido, a web semântica se coloca de forma diferenciada, pois sua abordagem facilita o estabelecimento de diversos modelos conceituais para um mesmo conjunto de dados, permitindo análises em diferentes perspectivas. Dessa forma, a web semântica representa um caminho capaz de gerar melhorias na governança de dados e maior integração entre as raízes social e tecnológica da Ciência da Informação, facilitando a integração, comunicação, uso e governança de dados.

Devido a seu potencial, verifica-se uma tendência de esforço para automação das atividades envolvidas no processo de implantação das tecnologias web semântica. Nesse processo, os profissionais da CI são atores com papéis relevantes em diversas etapas, em especial na integração dos metadados, na indexação dos vocabulários controlados e definição das triplas, podendo também atuar como mediadores para a compatibilização de diferentes estruturas de governança. “Os metadados em conjunto com métodos de modelagem conceitual da informação [...] são essenciais para publicação de dados conectados” (ISOTANI e BITTENCOURT, 2015, p. 43).

Conforme aponta o projeto *Creating Knowledge out of Interlinked Data* (AUER, BRYL e TRAMP, 2014, p. 1), a web semântica tem ganhado impulso com a publicação de dados estruturados em RDF. Isso tem permitido que o paradigma de dados ligados evolua de uma ideia de pesquisa para um caminho promissor voltado a enfrentar os desafios na área de gerenciamento inteligente de informações, como um poderoso facilitador para a transição da atual web orientada a documentos para uma web de dados interligados e, finalmente, para a web semântica. Entretanto, “a ideia de agregar semântica aos dados não é suficiente para a criação de um banco de dados global, pois é preciso que esses dados sejam conectados” (LAUFER, 2015, p. 71), por isso em 2006, Berners-Lee definiu uma lista de quatro princípios para a conexão desses dados:

1. usar URIs para identificar as coisas (recursos);
2. usar HTTP URIs, de forma a possibilitar que as pessoas possam procurar essas coisas na Web;
3. fornecer informações relevantes utilizando os padrões (RDF, SPARQL) para as pessoas que buscam URIs;
4. incluir links para outras URIs, de forma a possibilitar que mais coisas possam ser descobertas.

Como forma de orientar a publicação de dados abertos na Web, nessa nova visão semântica, Berners-Lee também sugeriu um esquema de classificação de cinco estrelas, para uma Web de Dados onde todas as informações estariam conectadas de acordo com os quatro princípios descritos acima (LAUFER, 2015, p. 71). Cada um desses níveis apresenta custos e benefícios, indicando a maturidade para caminhar na direção de um conjunto de dados semanticamente melhor descritos, e conectados com outros dados. Essa escala serve para qualificar uma publicação quanto a sua situação em relação aos de dados abertos e é composta de cinco níveis:

1. Publicar a informação na Web (qualquer que seja o formato) sob um tipo de licença de dados abertos;
2. Publicar a informação na forma de dados estruturados (por exemplo, uma planilha Excel ao invés de uma imagem digital de uma tabela);
3. Utilizar formatos não proprietários (por exemplo, uma tabela CSV ao invés de uma planilha Excel) para publicar a informação;
4. Utilizar URIs para identificar as recursos, de tal forma que as pessoas possam apontar para as informações publicadas;
5. Criar a conexão entre dados publicados e outros dados, de forma a prover um contexto maior para as informações.

Com a associação dos princípios para ligar dados e os níveis para qualificar a publicação, verifica-se que o modelo da web semântica é orientado a promover boas

práticas de governança de dados que convergem e podem ser aperfeiçoados a partir das recomendações apresentadas no Modelo de Governança de Comuns.

Em uma publicação pioneira conectar dados com web semântica, Fink (2018, p. 41) apresenta boas práticas para desenvolvimento de projetos de dados abertos ligados (*Linked Open Data – LOD*) e resume as seguintes recomendações:

1. Estabeleça políticas de dados;
2. Escolha as licenças de dados facilmente compreendidas;
3. Planeje sua seleção de dados;
4. Reconhecer que as normas são necessárias para tornar mais eficaz uso do LOD;
5. Escolha ontologias com colaboração;
6. Defina um modelo de referência para orientar o processo de mapeamento;
7. Crie uma identidade institucional para domínios raiz de URI
8. Prepare seus dados e certifique-se de incluir identificadores exclusivos
9. Esteja ciente dos desafios ao exportar dados do seu CIS; Desenvolva um script de extração ou API;
10. Se terceirizar o mapeamento e a conversão de seus dados para LOD, não assuma que a contratada entende como funcionam o seus dados ou saiba o que você pretende fazer com eles;
11. Aceite que você não pode alcançar 100% de precisão, 100% de cobertura, 100% Completitude: Comece em algum lugar, aprenda, corrija;
12. Operacionalize o LOD.

Além disso, Cavalcante e Kobashi (2018, p. 143-144) apontam que os resultados dos projetos de web semântica realizados até o momento indicam que o foco da investigação para os problemas de integração de dados não deve centrar apenas em questões técnicas, como as tecnologias e linguagens, mas também no potencial usuário, em particular os aspectos sobre a interação humano-máquina. Por isso as pesquisas se encaminham no sentido de definir metodologias para o desenho de uma estrutura coerente de metadado, voltada a permitir a representação descritiva e temática e descrever os atributos técnicos. Segundo Kobashi (2018, p. 91), “a mineração de dados será mais eficaz se os sistemas forem dotados de dicionários semânticos. Não por acaso, as pesquisas sobre a Web Semântica preveem uma camada semântica (ontologias) para buscas.”

Assim, a web semântica tem por objetivo estender a orientação da *web* de documentos para dados, permitindo maior capacidade de leitura por máquinas e melhor identificação de significado pelos usuários, sem exigir mudanças nas fontes de origem para que essa integração ocorra. Isso permite reduzir as limitações dos usuários, que passam a ser mais ativos ao expressar raciocínios complexos nas buscas por informações. Já os computadores tornam-se capazes de respostas mais precisas, pois interpretam em uma

linguagem estruturada de forma mais próxima à expressão humana, permitindo uma melhor experiência e a exploração de novas interações e interpretações dos dados.

Sob a perspectiva das máquinas, a estrutura da web semântica é uma camada de dados consumidos por uma camada de aplicações. Os dados gerados por aplicações podem ser utilizados como fontes de dados por outras aplicações, constituindo uma nova camada que pode também ser consumida por novas aplicações e assim por diante. Formando “um universo de camadas de dados e de aplicações que podem construir novas camadas de forma infinita, com cada aplicação oferecendo um conjunto específico de tarefas” (LAUFER, 2015, p. 30). Sob essa ótica, as interpretações dos dados e de seus possíveis significados são de responsabilidade das aplicações, resultantes das relações que estas fazem a partir do consumo do universo de dados dispostos nas diversas camadas. Sob a perspectiva das pessoas, as aplicações devem prover interfaces onde os dados resultantes das tarefas requisitadas sejam apresentados de forma a ter a sua compreensão facilitada (LAUFER, 2015, p. 30). “Os Dados Conectados podem ser considerados um subconjunto da Web Semântica” (ISOTANI e BITTENCOURT, 2015, p. 45). “*Resource Description Framework (RDF)* é a especificação proposta pelo *World Wide Web Consortium (W3C)* para descrever metadado” (ISOTANI e BITTENCOURT, 2015, p. 104). Isotani e Bittencourt (2015, p. 45 e 46) enfatizam “que não há como ter Dados Conectados sem fazer a utilização do modelo de dados RDF”, mas que também há uma série de exigências que precisam ser atendidas para que os dados sejam estruturados no modelo RDF. Ainda segundo Isotani e Bittencourt (2015, p. 49)

O W3C criou um grupo de trabalho (*CSV on the Web*) com o objetivo de definir um vocabulário para metadados em CSV e métodos padronizados para:

- (a) encontrar os metadados via protocolo HTTP;
- (b) encontrar os dados descritos pelos metadados por meio de uma API; e
- (c) mecanismos de mapeamento de CSV para outros formatos como RDF e JSON.

Isotani e Bittencourt (2015, p. 49)

a Web apresenta uma arquitetura cliente-servidor. A comunicação na Web acontece por meio de protocolos-padrões que permitem a troca de mensagens entre um Servidor web que implementa este protocolo e um Browser Cliente que envia a solicitação para o servidor. na comunicação cliente-servidor, o servidor irá retornar para o cliente (navegador) uma representação em um determinado formato. Cada formato de representação retornado para o cliente conterá informação de Metadados e Dados. Os metadados são os campos de cabeçalhos utilizados para identificar o formato de representação.

3 PROCEDIMENTOS METODOLÓGICOS

Os aspectos metodológicos deste estudo, incluindo procedimentos e caracterização da pesquisa, estão baseados nas definições apresentadas por Creswell (2010); Kothari (2004); Kauark, Manhães e Medeiros (2010); Appolinário (2006) e Bates (1989).

3.1 CARACTERIZAÇÃO DA PESQUISA

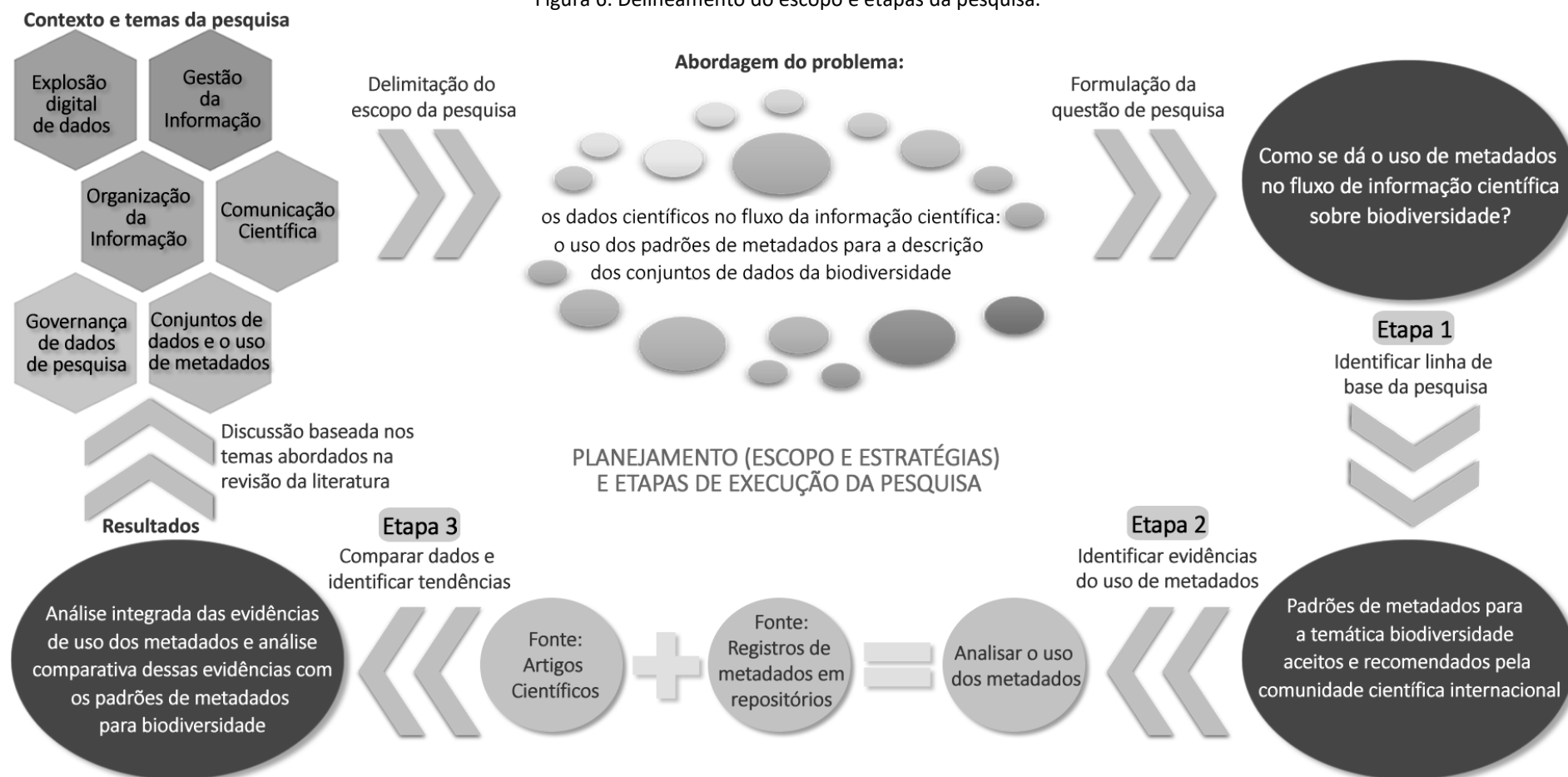
Esta pesquisa tem propósito descritivo (KOTHARI, 2004, p. 37), pois será realizada por meio da observação e descrição da realidade, sem interferência do pesquisador, reunindo estudos de forma a preencher uma lacuna do conhecimento. Também é considerada uma pesquisa de finalidade básica, pois apresenta objetivo científico não comercial (APPOLINÁRIO, 2006), assim, contribuindo para qualquer pesquisa sobre uso dos metadados para conjuntos de dados da biodiversidade.

A estratégia dessa pesquisa em relação à coleta de dados se caracteriza como bibliográfica e documental, pois as fontes usadas para a coleta de informações são artigos disponíveis na literatura científica e evidências documentais extraídas dos repositórios de dados analisados.

Devido ao fato de não empregar coleta ou análise de dados quantitativos, nem combiná-los com qualitativos e considerando os objetivos propostos somados ao quadro teórico esboçado neste projeto de estudos, esta pesquisa caracteriza-se por ser de natureza qualitativa. Nesse sentido, a abordagem adotada para esta pesquisa é qualitativa (KAUARK, MANHÃES e MEDEIROS, 2010, p. 26) e indutiva (CRESWELL, 2010, p. 26 e 92) por ser mais adequada para estudo de fenômenos, partindo de questões específicas para temas mais amplos. Sendo a compressão do uso de metadados a questão mais específica e a comunicação científica dos dados de pesquisa o tema mais amplo.

O escopo desse estudo foi delimitado a partir da identificação dos temas considerados relevantes para compreensão do contexto atual da comunicação científica dos dados de pesquisas. Analisando as interseções entre esses temas, verificou-se a relevância de avaliar o uso dos padrões de metadados nos conjuntos de dados de pesquisa. Diante da abrangência do tema, optou-se por delimitar o universo amostral ao uso dos metadados em conjuntos de dados de biodiversidade. Esse processo de delineamento está ilustrado de forma sintética na figura 6.

Figura 6. Delineamento do escopo e etapas da pesquisa.



Fonte: Elaboração própria.

3.2 UNIVERSO DA PESQUISA E HORIZONTE TEMPORAL

De forma a explorar e entender o significado atribuído ao problema, este estudo analisa os dados da biodiversidade e respectivos metadados em ambiente natural de atividades de pesquisa. Espera-se com isso alcançar o tema mais amplo, a comunicação científica dos dados de pesquisa no fluxo da informação científica e aspectos intrínsecos, como gestão, comunicação e governança.

Portanto, os dados coletados para subsidiar a pesquisa têm duas fontes de coleta:

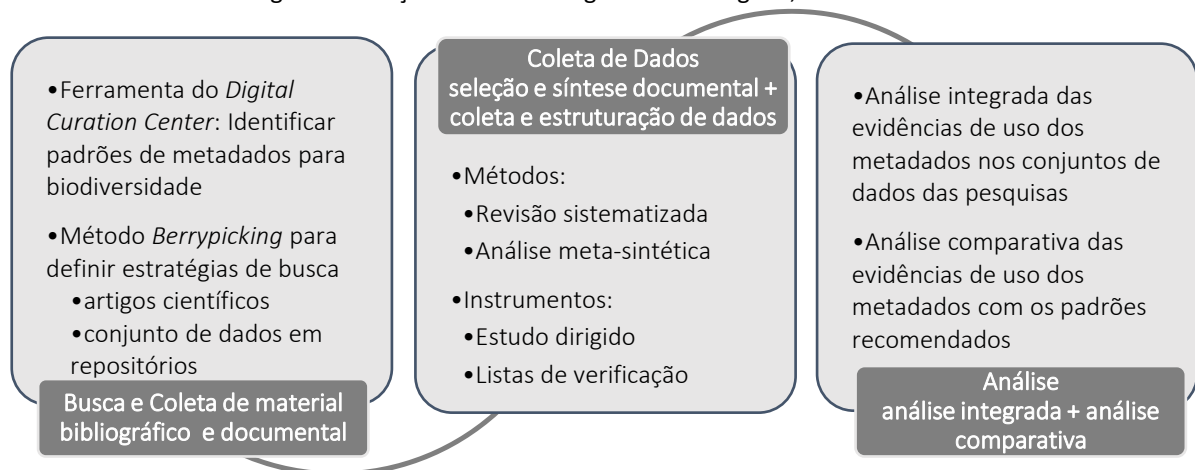
- Artigos revisados por pares e publicados na literatura científica, que tratam do uso de metadados para biodiversidade;
- Registros de metadados relacionados aos conjuntos de dados resultantes de pesquisa sobre a biodiversidade.

Optou-se por não restringir o horizonte temporal, de forma a permitir a análise das tendências do objeto da pesquisa e sua evolução ao longo do tempo.

3.3 ESTRATÉGIAS METODOLÓGICAS

A figura 7 apresenta esquematicamente a relação entre essas estratégias adotadas, considerando principais métodos e instrumentos. De forma a facilitar a compreensão, as estratégias metodológicas foram agrupadas e descritas em três categorias: 1) busca e coleta de material bibliográfico e documental; 2) coleta e estruturação de dados; e 3) análise dos dados.

Figura 7. Relação entre estratégias metodológicas, métodos e instrumentos



Fonte: Elaboração própria.

3.3.1 ESTRATÉGIA DE COLETA DO MATERIAL BIBLIOGRÁFICO E DOCUMENTAL

A escolha das fontes e a seleção dos documentos mais relevantes e aderentes ao tema estudado constitui fator crítico para o desempenho da pesquisa. Por esse motivo, o delineamento das estratégias de pesquisa utilizadas neste trabalho tem embasamento nos modelos de busca da informação.

A busca da informação refere-se a “(...) variedade de métodos que as pessoas utilizam para descobrir e ganhar acesso às fontes de informação”. Já o comportamento de busca em sistemas de informação estaria “(...) relacionado com as interações entre o usuário em um sistema de informação computadorizada.” (WILSON, 1999, p. 249).

Tradicionalmente, os autores relacionam a busca informacional a uma necessidade pessoal de determinado indivíduo. Entretanto, ampliando esse entendimento para abranger a pesquisa científica, podemos dizer que o pesquisador realiza a busca por informação visando responder o problema da pesquisa à comunidade científica, contribuindo para o desenvolvimento da ciência.

O planejamento e a execução de uma estratégia de pesquisa demandam a escolha de onde, quando e com que investigar cuidadosamente uma fonte de informação para alcançar objetivos específicos. A partir desses aspectos, é possível considerar os modelos de buscas existentes e adaptá-los, visando explicar as estratégias desenhadas para explorar as fontes de informação utilizadas em uma pesquisa científica.

Nesse sentido, a estratégia definida para pesquisa por artigos e registros de metadados neste trabalho segue uma adaptação do modelo proposto por Bates (1989). Segundo a autora, o modelo trata da recuperação de informação realizada pouco a pouco, a cada instante (*bit-at-a-time* ou *berrypicking*). Bates (1989) também afirma que o processo de amostragem e seleção da informação funciona como uma “escolha de frutas”, por isso o termo “*berrypicking*” é usado em analogia à coleta de frutas silvestres, que estão espalhadas nos arbustos e devem ser colhidas uma a uma.

No modelo de buscas proposto por Bates (1989), a busca de informação pode ser iniciada a partir de um tópico amplo ou com uma referência relevante e, em seguida, movimentar-se por diversas fontes. Cada nova informação encontrada oferece novas ideias e orientações e, conseqüentemente, nova concepção da consulta que pode levar a modificação dos termos de pesquisa.

A própria consulta e os termos de pesquisa utilizados são continuamente refinados em parte ou na totalidade. Assim, as consultas não são estáticas, mas em constante evolução. A coleta das informações é realizada em partes até que seja atendido o objetivo da pesquisa. Por isso, a definição dos parâmetros de busca e triagem de fontes bibliográficas e documentais consiste também em um resultado dessa pesquisa.

3.3.1.1 Documentação dos padrões e perfis de metadados para biodiversidade

Nesse estudo, o *Digital Curation Center* (DCC) é adotado como referência para a identificação dos padrões de metadados aceitos e recomendados internacionalmente por oferecer:

- Acesso gratuito a ferramentas e recursos de gestão de dados;
- Treinamento e eventos de amplitude internacional, com validação de experiências e ampla divulgação dos casos de sucesso;
- Identificação e compartilhamento das melhores práticas de gestão de dados;
- Consolidação e expansão dos esforços de gestão de dados existentes;
- Contribuição para o desenvolvimento de padrões, ferramentas e práticas inovadoras na gestão de dados;
- Suporte contínuo por meio da oferta de recursos de aprendizado para atualizar continuamente as habilidades e conhecimentos de gestão de dados

O DCC foi criado para apoiar a superação dos desafios referentes ao tratamento dos dados e informações digitais, de forma a aperfeiçoar os serviços de pesquisa, desenvolvimento, aconselhamento e suporte para instituições de ensino superior no Reino Unido. Tendo em vista a expressiva adesão da comunidade científica, com a participação de mais de 20 universidades, e o sucesso em promover a adoção dos padrões de metadados e das práticas de gestão de dados internacionalmente aceitos, o DCC tornou-se uma organização de referência internacional.

O DCC também dialoga com a comunidade científica internacional realizando de eventos e participando da *Research Data Alliance* – RDA⁴, que é uma organização global voltada a construir a infraestrutura social e técnica necessária ao compartilhamento aberto de dados. A RDA é composta por mais de 44 organizações membro, tais como o DCC, o Conselho de Ciência e Tecnologia do Reino Unido (*Science and Technology Facilities Council* –

⁴ Portal da RDA na internet - <https://www.rd-alliance.org>

STFC)⁵, a Associação Europeia de Pesquisa em Biblioteconomia (*Ligue des Bibliothèques Européennes de Recherche – LIBER*)⁶ e instituições privadas de pesquisa e educação (exemplo: WILEY⁷).

Portanto, neste estudo, a busca e coleta de informações sobre os padrões de metadados para a biodiversidade foi iniciada a partir do Portal do DCC com o uso da ferramenta de busca por padrões de metadado por áreas temáticas – *Disciplinary Metadata*⁸. Esse recurso é suportado pela rede criada pelo DCC que permite o compartilhamento e evolução das experiências, fornecendo informações sobre os padrões de metadados em diversas áreas temáticas, incluindo perfis e ferramentas para implantar os padrões, bem como divulgando os casos de repositórios de dados que atualmente que já adotam as ações recomendadas. Além disso, a ferramenta se concentra em padrões descritivos que auxiliam na descoberta e reutilização de dados; bem como na definição da informação a ser mantida com os conjuntos de dados científicos. O foco inicial do DCC tem sido nos padrões de metadados para dados tabulares.

Segundo o DCC (2019), “embora os gestores de dados, e cada vez mais pesquisadores, saibam que bons metadados são essenciais para o acesso e reutilização de dados de pesquisa, descobrir precisamente quais metadados capturar e como capturá-los é uma tarefa complexa”. Por isso, muitas áreas da comunidade acadêmica têm cada vez mais apoiado iniciativas para formalizar as especificações de metadados necessárias ao uso e compartilhamento dos dados.

Adicionalmente a triagem realizada no Portal do DCC, levantou-se a documentação dos padrões de metadados nas entidades responsáveis por sua manutenção. Dessa forma, à medida que os resultados são obtidos é feita a retroalimentação para estabelecer a linha de base, como forma para corroborar ou qualificar os achados iniciais.

Essa estratégia metodológica foi delineada para alcançar o primeiro objetivo específico da pesquisa, que trata da identificação das referências de metadados para biodiversidade.

⁵ Portal do STFC na internet - <https://stfc.ukri.org/>

⁶ Portal da LIBER na internet - <https://libereurope.eu>

⁷ Portal da WILEY na internet - <https://www.wiley.com>

⁸ Ferramenta de busca por padrões de metadados do DCC - <http://www.dcc.ac.uk/resources/standards>

3.3.1.2 Artigos científicos sobre o uso de metadados

A busca por artigos científicos foi realizada na ferramenta de “Pesquisa por Assunto” disponível no Portal de Periódicos da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) ⁹. Trata-se de um agregador indexado que “reúne e disponibiliza a instituições de ensino e pesquisa no Brasil [...] um acervo de mais de 45 mil periódicos com texto completo, 130 bases referenciais [...], além de livros, enciclopédias e obras de referência, normas técnicas, estatísticas e conteúdo audiovisual” (BRASIL. MCTIC.CAPES, 2019).

A opção por utilizar essa ferramenta foi motivada, principalmente, pela possibilidade de englobar publicações de diferentes campos do conhecimento e oriundas de diversas bases oficiais da literatura científica. Além disso, esse agregador apresenta comportamento estável, que permite a reprodução das consultas, promovendo reprodutibilidade, transparência e confiabilidade a esta pesquisa.

Conforme descrito inicialmente, as buscas seguem o modelo de Bates (1989), partindo de termos mais abrangentes até alcançar os mais específicos e atingir o objetivo da pesquisa. Dessa forma, optou-se por iniciar as buscas utilizando como termos mais abrangentes *biodiversity* e *metadata*. Os termos mais específicos a serem utilizados são os nomes dos padrões e perfis de metadados identificados na etapa anterior. O uso dos termos em língua inglesa visa abranger as pesquisas em âmbito global. Além disso, foram considerados apenas artigos científicos revisados por pares.

3.3.1.3 Documentação dos metadados dos conjuntos de dados

Inicialmente, seriam coletados metadados de conjuntos de dados de pesquisa sobre biodiversidade obtidos a partir de dois repositórios, preferencialmente um nacional e outro externo ao Brasil, visando comparar experiências. Entretanto, ao longo do desenvolvimento da pesquisa, foi identificado que a rede *Global Biodiversity Information Facility* – GBIF (GBIF, 2017b) disponibiliza ferramenta de busca por conjuntos de dados de biodiversidade publicados por instituições de pesquisa em todo o mundo.

⁹ Portal de Periódicos da Capes - <http://www.periodicos.capes.gov.br>

Dessa forma, optou-se por utilizar essa ferramenta para coletar a documentação dos conjuntos de dados provenientes de instituições em diversos países do mundo e realizar uma análise mais ampla sobre o uso dos padrões de metadados da biodiversidade.

3.3.2 ESTRATÉGIA DE COLETA E ESTRUTURAÇÃO DE DADOS

A coleta e estruturação das informações foram realizadas em duas partes. A primeira com roteiro de estudo dirigido para extrair os dados dos artigos científicos. A segunda com a lista de verificação para extrair dados da documentação dos conjuntos de dados.

A seguir são apresentados os instrumentos utilizados para coleta, estruturação e síntese dos dados coletados.

3.3.2.1 Roteiro de estudo dirigido para coleta de dados nos artigos científicos

A coleta de dados dos artigos científicos foi orientada pelo roteiro de estudo apresentado no **APÊNDICE A**. As questões apresentadas nesse roteiro foram formuladas para permitir a identificação do uso dos padrões de metadados em conjuntos de dados de pesquisas sobre a temática biodiversidade. Considerando a estratégia de projeção dos temas mais amplos até alcançar mais específicos (BATES, 1989), houve o refinamento desse instrumento ao longo da pesquisa de forma a abranger questões amadurecidas a partir do conhecimento construído.

3.3.2.2 Lista de verificação para coleta de metadados dos conjuntos de dados

A lista de verificação foi utilizada para coleta, estruturação e análise dos metadados referente aos conjuntos de dados da biodiversidade – **APÊNDICE B**. Assim como ocorreu com o roteiro de estudo, essa lista de verificação foi aperfeiçoada ao longo da pesquisa.

3.3.3 ESTRATÉGIA DE ANÁLISE DOS DADOS

As estratégias adotadas nesta pesquisa para análise das informações coletadas incluem a revisão sistematizada e meta-análise (ou análise meta-sintética), que permitem resumir e sintetizar evidências sobre o tema estudado. O uso dessas técnicas contribui ao avanço na produção do conhecimento a partir de procedimentos que orientam e integram

os resultados dos estudos produzidos em uma determinada área (FILHO, PARANHOS, *et al.*, 2014, p. 207).

A revisão sistematizada (ou revisão integrativa) surgiu como alternativa à revisão sistemática, permitindo combinar e revisar rigorosamente estudos de diversas áreas e diferentes metodologias, integrando seus resultados. Tem o potencial manter o rigor metodológico das revisões sistemáticas, porém estabelece estratégias de busca diferenciadas, sem a obrigatoriedade de garantir a busca exaustiva. O propósito deste método de pesquisa é obter um profundo entendimento de um determinado fenômeno baseando-se em estudos anteriores.

Este método permite a combinação de dados de literatura teórica e empírica. Isso permite elaborar uma revisão com diferentes finalidades, tais como a definição de conceitos, a revisão de teorias ou a análise metodológica dos estudos incluídos de um tópico particular (MENDES, SILVEIRA e GALVAO, 2008).

A revisão sistematizada com meta-análise consiste em integrar diferentes estudos para utilizar metodologia analítica e estatísticas em busca de explicar a variância dos resultados utilizando fatores comuns aos estudos (ROSCOE e JENKINS, 2005, p. 54), sendo um procedimento metodológico que “sintetiza uma determinada quantidade de conclusões num campo específico” (FILHO, PARANHOS, *et al.*, 2014, p. 209). Segundo Glass (1976, p. 3), a meta-análise refere-se à análise de análises, considerando uma grande coleção de resultados de estudos individuais com a finalidade de integrar os resultados.

A revisão sistematizada ou integrativa surgiu como alternativa para revisar rigorosamente e combinar estudos com diversas metodologias, por exemplo, delineamento experimental e não experimental, e integrar os resultados, ainda que os estudos sejam de caráter qualitativo. Tem o potencial de promover os estudos de revisão em diversas áreas do conhecimento, mantendo o rigor metodológico das revisões sistemáticas.

Ainda que este estudo seja de caráter qualitativo, o método de revisão integrativa permite a combinação de dados da literatura empírica e teórica que podem ser direcionados à definição de conceitos, identificação de lacunas nas áreas de estudos, revisão de teorias e análise metodológica dos estudos sobre um determinado tópico. A combinação de pesquisas com diferentes métodos combinados na revisão sistematizada amplia as possibilidades de análise da literatura.

Roscoe e Jenkins (2005, p. 54) afirmam que “meta-análise consiste em colocar diferentes estudos juntos em um mesmo banco de dados e utilizar metodologias analíticas e estatísticas para explicar a variância dos resultados utilizando fatores comuns aos estudos”. Segundo Figueiredo (2014, p. 209), a meta-análise é um procedimento metodológico que sintetiza uma determinada quantidade de conclusões num campo de pesquisa específico, elevando a objetividade das revisões de literatura, minimizando possíveis vieses e aumentando a quantidade de estudos analisados.

Especificamente na área da Ciência da Informação, a revisão sistematizada pode contribuir para a concentração de resultados de várias outras pesquisas em um mesmo trabalho, aumentando a confiabilidade e revelando o status de um problema de pesquisa, visto que a maioria das revisões de literatura é de cunho narrativo e sem a adoção de roteiro específico de análise. Esse tipo de revisão requer uma pergunta clara, a definição de uma estratégia de busca, o estabelecimento de critérios de inclusão e exclusão dos artigos e, acima de tudo, uma análise criteriosa da literatura selecionada. O processo de desenvolvimento desse tipo de revisão inclui caracterizar cada estudo selecionado, avaliar, identificar conceitos importantes, comparar as análises apresentadas e concluir sobre o que a literatura informa em relação ao assunto que delimitado.

Por fim, a revisão sistematizada é utilizada nesta pesquisa para a construção de uma análise ampla da literatura, contribuindo para discussões sobre uso de metadados como e reflexões sobre futuros estudos. A opção por esse método visa permitir uma síntese de evidências qualitativas e o desenho da tendência histórica, com o propósito de obter um profundo entendimento da evolução do uso de metadados para biodiversidade.

. O processo de desenvolvimento desse tipo de revisão inclui caracterizar cada estudo selecionado, avaliar, identificar conceitos importantes, comparar as análises apresentadas e concluir sobre o que a literatura informa em relação ao assunto que delimitado. Esse processo envolve as seguintes atividades:

- Identificação do tema e seleção da hipótese ou questão de pesquisa para a elaboração da revisão;
- Estabelecimento de critérios para inclusão e exclusão de estudos/amostragem ou busca na literatura;
- Definição das informações a serem extraídas dos estudos selecionados/categorização dos estudos;

- Análise dos estudos incluídos na revisão integrativa;
- Interpretação dos resultados;
- Apresentação da revisão/síntese do conhecimento.

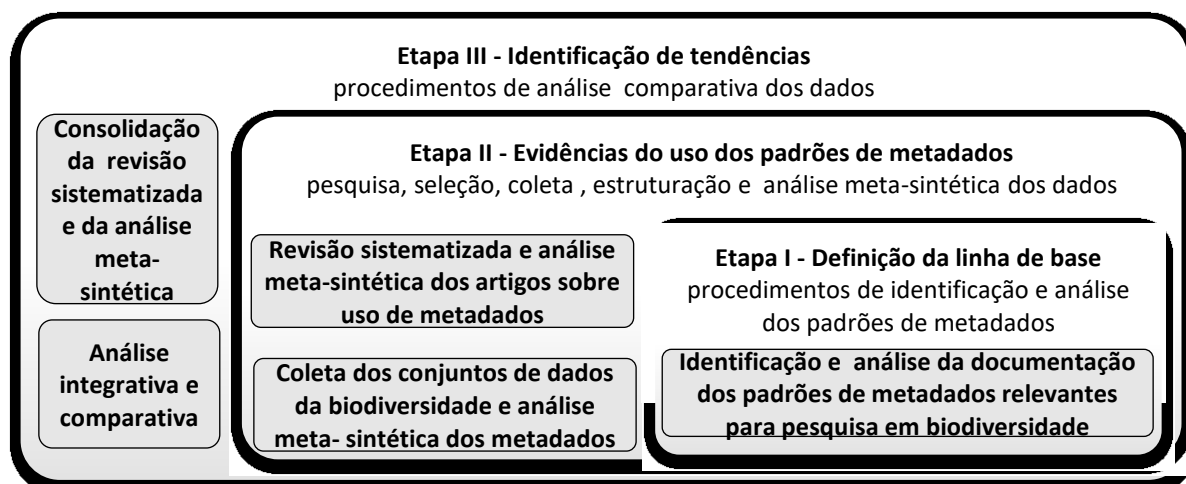
3.4 ETAPAS DA PESQUISA

As estratégias metodológicas descritas na **subseção 3.3** são aplicadas para alcançar diferentes objetivos específicos neste estudo. Para fins didáticos, de forma a facilitar a compreensão das atividades de pesquisa desenvolvidas e a ordem em que foram executadas para alcançar cada objetivo específico proposto, a apresentação e discussão dos resultados deste estudo são organizadas e descritas em três etapas:

- I. **Definição da linha de base da pesquisa** – etapa de identificação e análise dos padrões de metadados existentes para a temática biodiversidade aceitos e utilizados globalmente pelas comunidades de pesquisa;
- II. **Coleta e análise de evidências do uso dos padrões de metadados** – etapa de coleta, estruturação e análise de evidências do uso dos padrões de metadados em artigos científicos e na documentação de conjuntos de dados de biodiversidade;
- III. **Identificação de tendências** – etapa de análise integrada dos resultados das etapas I e II para identificação de tendências relacionadas ao uso dos padrões de metadados para dados de pesquisas da biodiversidade.

A relação de complementaridade entre as etapas da pesquisa está ilustrada de forma sintética na figura 8.

Figura 8. Relação entre estratégias metodológicas e etapas da pesquisa



Fonte: Elaboração própria.

3.4.1 ETAPA I – IDENTIFICAÇÃO E ANÁLISE DOS PADRÕES DE METADADOS

A etapa I consiste na identificação dos padrões e perfis de metadados existentes que são utilizados nas pesquisas sobre biodiversidade. A síntese dos procedimentos metodológicos adotados nessa etapa e sua relação com os objetivos específicos da pesquisa constam no quadro 6.

Quadro 6. Etapa I - Definição da linha de base – levantamento dos padrões de metadados para biodiversidade: relação entre os objetivos específicos e as estratégias metodológicas

OBJETIVOS ESPECÍFICOS	FONTE/BUSCA	COLETA	ANÁLISE
Identificar os padrões de metadados recomendados internacionalmente para a temática biodiversidade.	Publicações oficiais do <i>Digital Curation Center</i> sobre padrões de metadados e das entidades responsáveis pela manutenção desses padrões	<p>Método: Busca de informação realizada pouco a pouco - <i>berrypicking</i> (BATES, 1989).</p> <p>Técnica: Coleta documental, identificando os padrões de metadados recomendados para a temática biodiversidade recomendados pelo DCC.</p> <p>Instrumento: Ferramenta de busca para padrões de metadados por áreas temáticas do DCC. E páginas na web com a documentação oficial de cada padrão.</p>	<p>Método: Análise meta-sintética.</p> <p>Técnica: Integração interpretativa e síntese de dados qualitativos referentes aos padrões de metadados recomendados pelo DCC, estabelecendo uma lógica sistematizada e coerente com os objetivos propostos neste estudo.</p>

Fonte: Elaboração própria.

3.4.2 ETAPA II – COLETA E ANÁLISE DE EVIDÊNCIAS DO USO DE METADADOS

Na etapa II são coletadas e analisadas as evidências de uso dos padrões de metadados a partir de duas fontes principais: 1) artigos científicos disponibilizados por meio do Portal Capes; e 2) registros de metadados relacionados aos conjuntos de dados sobre biodiversidade disponibilizados na rede internacional *Global Biodiversity Information Facility* – *GBIF*. A síntese dos procedimentos a serem adotados nesta etapa e sua relação com os objetivos específicos da pesquisa constam no quadro 7.

Quadro 7. Etapa II - Coleta e análise das evidências do uso dos padrões de metadados: relação entre os objetivos específicos e as estratégias metodológicas.

OBJETIVOS ESPECÍFICOS	FONTE	COLETA	ANÁLISE
Identificar o uso dos padrões de metadados para o registro de dados sobre biodiversidade em artigos da literatura científica	Artigos científicos disponibilizados no Portal de Periódicos da Capes/MEC.	<p>Método: Pesquisa Sistematizada e Descritiva.</p> <p>Técnica: Recuperação de informação realizada pouco a pouco ou berrypicking (Bates, 1989) para levantamento dos artigos que tratam dos padrões de metadados para biodiversidade.</p> <p>Instrumentos: Ferramenta de busca por assunto no Portal de Periódicos da Capes .Estudo Dirigido para coleta preliminar de informações nos artigos.</p>	<p>Método: Pesquisa baseada em evidências, com uso da revisão sistematizada como fonte de evidência científica.</p> <p>Técnica: Revisão sistematizada da literatura científica que trata do uso dos padrões de metadados para a temática biodiversidade.</p> <p>Instrumento: Roteiro de estudo para comparar os principais aspectos, conceitos e informações a serem extraídas de cada artigo.</p>
; Identificar o uso dos padrões de metadados no registro de dados científicos sobre biodiversidade em repositórios	Registros de metadados relacionados aos conjuntos de dados obtidos no Portal da rede GBIF.	<p>Método: Pesquisa sistematizada e descritiva.</p> <p>Técnica: Recuperação de informação realizada pouco a pouco ou berrypicking (Bates, 1989) para levantamento dos conjuntos de dados e respectivos metadados.</p> <p>Instrumento: Lista de Verificação.</p>	<p>Método: Análise qualitativa.</p> <p>Técnica: Descrição sintética dos registros de metadados e estruturação com classificação qualitativa.</p> <p>Instrumento: Lista de Verificação para estruturar, comprar e analisar os principais aspectos, conceitos e informações sobre o uso de metadados.</p>

Fonte: Elaboração própria.

3.4.3 ETAPA III – ANÁLISE DE TENDÊNCIAS NA GESTÃO DE DADOS DE PESQUISA

A etapa III da pesquisa consiste em comparar os padrões de metadados recomendados internacionalmente às evidências de uso identificadas na literatura científica e na documentação dos metadados dos conjuntos de dados. A síntese dos procedimentos a serem adotados nesta etapa e sua relação com os objetivos específicos da pesquisa constam no quadro 8.

Quadro 8. Etapa III: Análise Comparativa e identificação de Tendências: relação entre os objetivos específicos e as estratégias metodológicas.

OBJETIVOS ESPECÍFICOS	FONTE	COLETA	ANÁLISE
Comparar os padrões de metadados recomendados internacionalmente com evidências de uso dos metadados no registro dos dados das pesquisas sobre biodiversidade	Material coletado nas outras etapas desta pesquisa	--- Coleta realizada em outras etapas do estudo	Método: Análise comparativa Técnica: Comparação dos padrões recomendados pelo DCC com os efetivamente identificados na literatura científica e nos repositórios. Instrumento: Lista de Verificação os padrões recomendados pelo DCC e os registros de metadados identificados nos artigos e repositórios selecionados nesta pesquisa.

Fonte: Elaboração própria.

3.5 MODELO CONCEITUAL DA PESQUISA

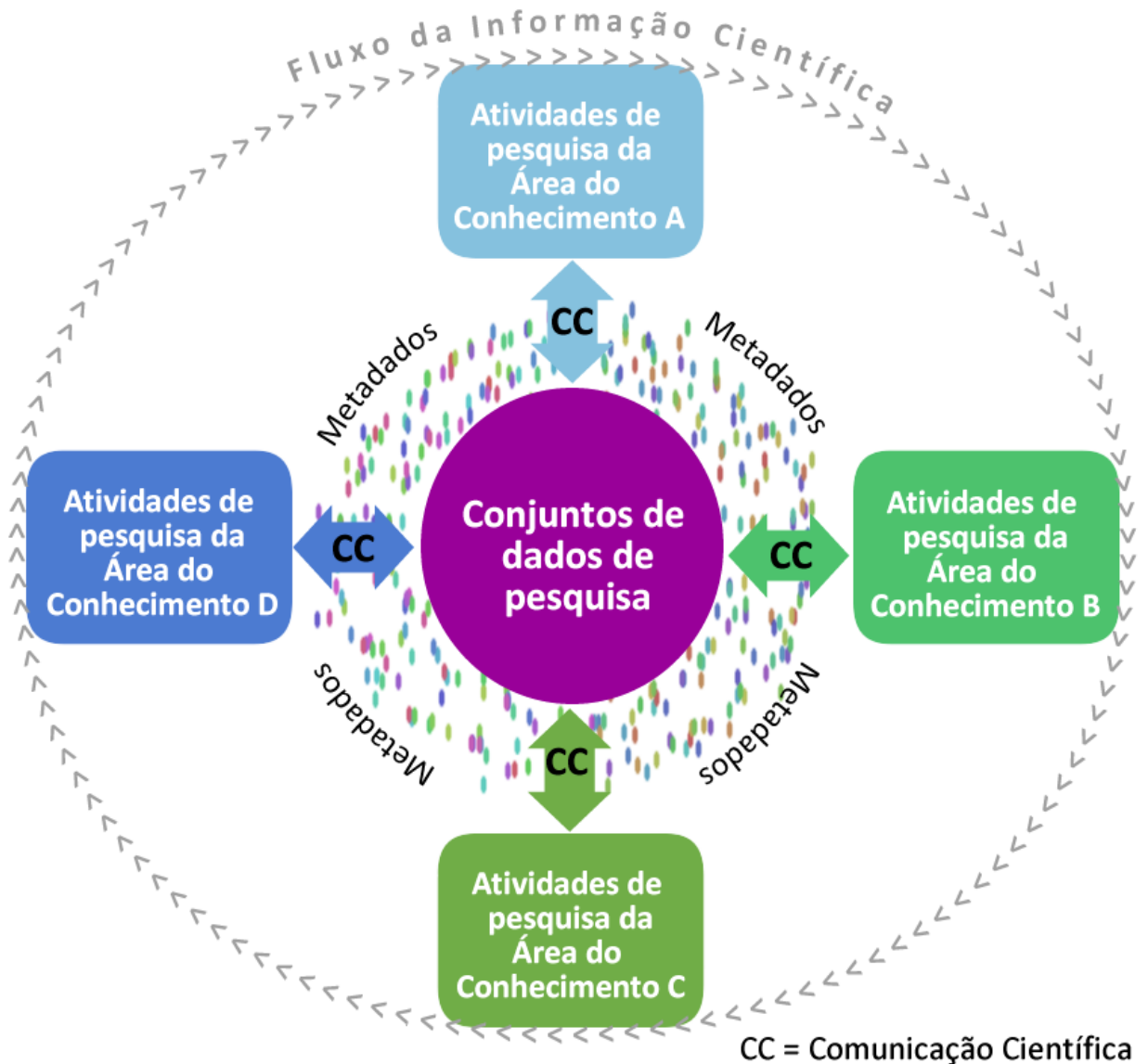
O modelo conceitual proposto considera os elementos essenciais para este estudo, que foram extraídos a partir da revisão de literatura e de suas respectivas inter-relações. Esses elementos são:

- Conjuntos de dados de pesquisa → dados coletados e utilizados para subsidiar atividades de pesquisa.
- Metadados → apresenta diversas definições descritas na revisão da literatura deste estudo. Para fins do modelo conceitual, devem ser compreendidos como as informações que descrevem os conjuntos de dados de pesquisa da biodiversidade.
- Comunicação Científica (CC) → todo o conhecimento científico está sujeito à validação da comunidade científica, que apresenta um papel ativo em um intrincado sistema de comunicação composto por canais formais e informais. A comunicação científica é um ciclo contínuo e regenerativo da informação, que inclui atividades associadas à produção, disseminação e uso da informação científica desde o planejamento até a conclusão da pesquisa, aceitos formalmente como parte integrante do conhecimento. Esse sistema de comunicação científica varia conforme regras estabelecidas pelas comunidades de cada área do conhecimento.

- Fluxo da informação científica → Fluxo da informação que alimenta e que resulta das atividades de pesquisa.

Considerando essa abordagem conceitual, a figura 9 ilustra as relações entre os conceitos elencados acima.

Figura 9. Modelo conceitual da pesquisa



Fonte: Elaboração própria.

Partindo das evidências de uso dos metadados para descrição de dados de pesquisa, busca-se compreender a comunicação desses recursos. O uso dos padrões de metadados é analisado como elemento central no processo de aperfeiçoamento da circulação e

compartilhamento dos dados de cada pesquisa, como elementos fundamentais à descoberta e ao reuso.

Também é relevante compreender como as diversas comunidades científicas de cada área do conhecimento têm absorvido os padrões de metadados para a descrição dos dados que resultam do desenvolvimento de pesquisas relacionadas à biodiversidade. Por isso, observando o fluxo da informação científica, busca-se compreender a comunicação dos dados de pesquisa sob a perspectiva do uso dos metadados.

Os conjuntos de dados de pesquisa são insumos relevantes não apenas às comunidades científicas das quais resultam como produto, mas também fundamentais ao avanço de pesquisa nas demais áreas do conhecimento. Ao mesmo tempo, esses produtos integram o fluxo da informação científica como insumos às pesquisas das diversas áreas do conhecimento. Entretanto, muitas vezes os produtos resultantes das pesquisas de diferentes áreas do conhecimento, mesmo dentro da temática biodiversidade, apresentam parâmetros de comunicação científica específicos, estabelecidos por comunidades diretamente vinculadas ao assunto de cada pesquisa.

O modelo conceitual proposto (figura 9) destaca a relação direta que há entre as exigências impostas pelo processo de comunicação científica estabelecido por cada comunidade científica e o uso dos metadados na descrição dos conjuntos de dados. Por conseguinte, o uso dos metadados não apenas influencia a comunicação dos conjuntos de dados na comunidade que o gerou, mas também apresenta efeitos sobre o fluxo da informação científica entre as diversas comunidades de outras áreas do conhecimento, com impactos sobre o potencial de descoberta e reuso desses dados para retroalimentar as outras atividades de pesquisa.

Por fim, destaca-se a relevância de aperfeiçoar a descrição dos dados de pesquisa, de forma que parâmetros específicos do processo de comunicação científica não limitem a descoberta e reuso por outras áreas do conhecimento.

4 RESULTADOS E DISCUSSÃO

A partir dos procedimentos metodológicos estabelecidos, a apresentação e discussão dos resultados estão organizadas em três etapas detalhadas nas subseções a seguir.

4.1 ETAPA I – PADRÕES DE METADADOS PARA BIODIVERSIDADE

A primeira etapa desta pesquisa refere-se à identificação dos padrões de metadados para biodiversidade, bem como a definição dos parâmetros a serem utilizados para avaliar o uso dos padrões de metadados nas pesquisas dessa temática.

Utilizando a ferramenta *Disciplinary Metadata*¹⁰ disponibilizada pelo DCC realizou-se a consulta dos padrões de metadados organizados por áreas temáticas. Inicialmente, os padrões são apresentados nessa ferramenta em cinco categorias: (1) Biologia; (2) Ciências da Terra; (3) Ciências Físicas; (4) Ciências Sociais e Humanas; e (5) Dados gerais de pesquisa. Essa última categoria vincula-se a informações sobre padrões mais amplos de metadados que foram adaptados para atender às necessidades de dados de pesquisa, englobando soluções para as disciplinas que ainda não estabeleceram um padrão de metadados e para repositórios contendo dados de várias disciplinas.

Considerando o escopo definido para este trabalho, a análise dos padrões recomendados pela DCC teve início com a pesquisa na categoria **Biologia**, que apresenta 11 padrões de metadados (quadro 9).

Quadro 9. Padrões de Metadados para Biologia.

Nome do Padrão	Descrição
1. <i>Access to Biological Collection Data</i> (ABCD)	Norma para o acesso e o intercâmbio de dados primários de biodiversidade, incluindo amostras e observações.
2. Darwin Core	Conjunto de padrões, incluindo um glossário de termos destinados a facilitar o compartilhamento de informações sobre diversidade biológica, fornecendo definições de referência, exemplos e comentários.
3. <i>Ecologic Metadata Language</i> (EML)	Especificação de metadados especialmente desenvolvida para a disciplina de ecologia.
4. Metadados do genoma	Dados descritivos sobre genomas no <i>Pathosystems Resource Integration Center</i> .
5. Tab ISA	Estrutura de propósito geral para capturar e comunicar metadados em arquivos de dados de experimentos, empregando combinações de tecnologias.
6. <i>Minimum Information for Biological and</i>	Estrutura de listas de verificação de informações mínimas

¹⁰ Ferramenta do DCC para pesquisa de Padrões de Metadados por área do conhecimento - < <http://www.dcc.ac.uk/resources/metadata-standards>>

Nome do Padrão	Descrição
<i>Biomedical Investigations</i> (MIBBI)	em quase 40 disciplinas biológicas e biomédicas.
7. Observ-OM	Padrão utilizado para integrar e comparar dados de observação em projetos experimentais, como bancos de dados sobre doenças e biobancos clínicos.
8. <i>Open Microscopy Environment</i> (XML – OME-XML)	Padrão de metadados e de formato de arquivo para dados de microscopia.
9. <i>Protein Data Bank Exchange Dictionary</i> (PDBx / mmCIF)	Dicionário de Intercâmbio do Banco de Dados de Proteínas e das Informações Cristalográficas Macromoleculares é um formato de arquivo padrão usado pelo <i>Protein Data Bank</i> (PDB), que fornece metadados e dados de acordo com as propriedades definidas no <i>PDB Exchange Dictionary</i> e no <i>Macromolecular Crystallographic Information Framework</i> (mmCIF).
10. <i>Protocol Data Element Definitions</i>	<i>Protocol Registration System</i> (PRS) é o sistema de registro de protocolo usado pelos Institutos Nacionais de Saúde dos EUA para registrar estudos clínicos em seres humanos no site <i>ClinicalTrials.gov</i> .
11. <i>Repository-Developed Metadata Schemas</i>	Esquemas de metadados desenvolvidos para repositórios. Em alguns repositórios os padrões de metadados não atendiam às suas necessidades e, portanto, criaram foram criados esquemas próprios.

Fonte: DCC (2019)

Além dos padrões apresentados acima, na área do Conhecimento da Biologia há recomendação de 14 perfil/extensões (quadro 10), que incluem adaptações de padrões ou conjunto de padrões para tipos específicos de repositórios ou para tipos específicos de dados.

Quadro 10. Perfis e extensões de metadados para Biologia.

Nome do Perfil ou Extensão	Descrição
1. ABCDDNA	Extensão do padrão ABCD para dados de DNA.
2. <i>Apple Core</i>	Documentação e recomendações do Darwin Core para herbários.
3. <i>Darwin Core Geospatial Extension</i>	Esquema XML independente de protocolo para uma extensão geoespacial do Darwin Core.
4. <i>DwC Germplasm</i>	Extensão do padrão Darwin Core, inclui termos adicionais necessários para descrever os recursos genéticos das plantas e, em particular, amostras de sementes de germoplasma.
5. <i>EDMED Metadata Profile</i>	Esquema de metadados do Diretório Europeu de Conjuntos de Dados Ambientais Marinhos, que é um perfil baseado na ISO 19115, que contém esquema adotado internacionalmente para descrever informações e serviços geográficos.
6. <i>FGDC/CSDGM Biological Data Profile</i>	Perfil do padrão de metadados <i>Federal Geographic Data Committee Content Standard for Digital Geospatial</i>

Nome do Perfil ou Extensão	Descrição
	<i>Metadata</i> (FGDC / CSDGM), destinado a suportar a coleta e o processamento de dados biológicos.
7. GBIF <i>metadata profile</i>	Estabelecido por uma rede global de países e organizações, o Sistema Global de Informação sobre a Biodiversidade (GBIF) é um portal da web que promove e facilita a mobilização, acesso, descoberta e uso de dados de biodiversidade. O portal usa um perfil de EML e disponibiliza um guia de instruções e um guia de referência para o uso desse perfil.
8. <i>Herbarium Information Standards and Protocols for Interchange of Data</i> (HISPID)	Extensão do ABCD 2.06 projetada para permitir o armazenamento e a transmissão de dados de amostras de plantas de herbário.
9. ISA-TAB Nano	Extensão do ISA-TAB especificando o formato para representar e compartilhar informações sobre nanomateriais, pequenas moléculas e amostras biológicas, juntamente com seus dados de caracterização do ensaio.
10. isaconfig-diXa	Extensão do ISA-TAB para representar e compartilhar metadados sobre experimentos toxicogenômicos.
11. <i>MIBBI Portal</i>	Cerca de 40 projetos de padrões de Informações Mínimas registrados na iniciativa MIBBI
12. <i>Open Microscopy Environment TIFF</i> (OME-TIFF)	Especificação sobre como incorporar metadados OME-XML em um arquivo de imagem TIFF ou BigTIFF para um ambiente de microscopia aberta.
13. SNRNASM ISA-Tab	Padrão baseado em ISA-Tab para relatar os resultados de experimentos de mapeamento de estrutura de ácido nucleico com resolução de nucleotídeo único.
14. VarioML	Conjunto de ferramentas e práticas para a melhoria da disponibilidade, da qualidade e da compreensibilidade das informações de avaliação humana, voltadas a permitir que os pesquisadores, laboratórios de diagnóstico e clínicas compartilhem essas informações com facilidade, clareza e sem ambiguidade.

Fonte: DCC (2019).

Também na área do conhecimento da Biologia, são apresentadas como referências à organização dos dados de pesquisa 13 ferramentas (quadro 11), que são *softwares* desenvolvidos para capturar ou armazenar metadados em conformidade com um padrão específico. Por fim, há uma extensa lista com mais de trinta registros de estudos de caso, incluindo repositórios institucionais e portais de dados que utilizam os padrões da área da Biologia para determinar os metadados coletados para o depósito de dados – <<http://www.dcc.ac.uk/resources/subject-areas/biology>>.

Quadro 11. Ferramentas de registro /recuperação de metadados

Nome da Ferramenta	Descrição
1. Bio-Formats	Leitor de dados e metadados proprietários de imagens de microscopia, que os converte em OME-TIFF, uma combinação dos padrões TIFF e OME-XML.
2. Darwin Core Archive Assistant	Aplicativo web com uma interface fácil para descrever elementos de dados como arquivos de texto básicos, compondo um arquivo descritivo XML Darwin Core voltado aos editores de dados que desejam integrar-se à rede GBIF.
3. Darwin Core Archive Validator	Ferramenta para validar metadados XML com base nas Diretrizes de texto principais do Darwin Core.
4. Fiji	Pacote de processamento de imagens que suporta o modelo de dados OME para imagens
5. Integrated Publishing Toolkit	Plataforma embasada no Darwin Core e EML para facilitar a publicação eficiente de dados de biodiversidade na Internet, usando a rede GBIF.
6. ISA Software Suite	Ferramentas de rastreamento de metadados ISA de código aberto que facilitam a coleta, a curadoria, o gerenciamento local e a reutilização de conjuntos de dados em conformidade com o padrão ISA-TAB em um conjunto cada vez mais diversificado de domínios relacionados às Ciências da vida.
7. MetaCat	Repositório de dados e metadados voltado a auxiliar os cientistas a encontrar, entender e usar efetivamente os conjuntos de dados que gerenciam ou que foram criados por outros.
8. MOLGENIS	Gerador de software para criar rapidamente bancos de dados para genotipagem, fenótipo, QTL e pipelines de análise.
9. Morpho	Aplicativo para acessar e manipular metadados e dados (localmente e na rede), com assistentes criando arquivos de metadados usando um subconjunto da Linguagem de Metadados Ecológicos (EML).
10. OMERO	Software de repositório para organizar, visualizar, analisar e compartilhar imagens de microscopia biológica, que suporta formatos de arquivo proprietários, mas normaliza para OME-TIFF / OME-XML.
11. PATRIC Download Tool	Ferramenta para baixar dados do <i>Pathosystems Resource Integration Center</i> (PATRIC).
12. PDBx/mmCIF Software Resources	Ferramentas e bibliotecas de análise, validação e visualização que suportam PDBx / mmCIF, que é o padrão de dados usado pelo Banco Mundial de Dados de Proteína.
13. ProteoRed Tools	Ferramentas de bioinformática para criar e extrair metadados compatíveis com os requisitos mínimos registrados no MIBBI MIAPE.

Fonte: DCC (2019)

Considerando o escopo desta pesquisa, verificou-se que muitos dos padrões relacionados à área do conhecimento da Biologia não são aplicáveis à temática Biodiversidade. Dessa forma, após análise dos itens listados nos quadros acima, identificou-se a necessidade de selecionar apenas os padrões de metadados, perfis/extensões e

ferramentas diretamente relacionados ao tratamento de dados sobre biodiversidade sob a perspectiva da ecologia.

Nesse sentido, por tratar-se de campos do conhecimento distantes da temática estudada, foram excluídos os elementos relacionados aos assuntos: Genoma, Genética, Microscopia, Biomedicina, Bioengenharia, Biologia Molecular, Análise Laboratoriais, Análises Clínicas e outras. Após esse refinamento, foram mantidos os itens relacionados às temáticas: Ecologia, Botânica, Zoologia, Biogeografia e Biodiversidade. Dessa delimitação de assuntos, resultaram três padrões de metadados diretamente recomendados em todas as áreas temáticas elencadas: 1) *Darwin Core* (DwC); 2) *Ecological Metadata Language* (EML); e 3) *Access to Biological Collection Data* (ABCD). O quadro a seguir apresenta brevemente os padrões de metadados selecionados para integrar a linha de base desta pesquisa.

Quadro 12. Padrões de metadados na temática Biodiversidade

Área do conhecimento: Biodiversidade	
Nome do Padrão	Descrição
1. <i>Access to Biological Collection Data</i> (ABCD)	Norma para o acesso e o intercâmbio de dados primários de biodiversidade, incluindo amostras e observações.
2. Darwin Core	Conjunto de padrões destinados a facilitar o compartilhamento de informações sobre diversidade biológica, fornecendo definições de referência, exemplos e comentários.
3. <i>Ecological Metadata Language</i> (EML)	Especificação de metadados especialmente desenvolvida para a disciplina de ecologia.

Fonte: DCC (2019)

Especificamente na área de ecologia também foi identificado o padrão de metadados *Directory Interchange Format* (DIF). Esse padrão surgiu da iniciativa da comunidade acadêmica relacionada às Ciências da Terra e tem foco em instrumentos de captura das características temporais e espaciais dos dados. Assim, por ser mais relacionado a Geociências e não tratar especificamente de características da biodiversidade, o padrão DIF não foi selecionado para análise neste estudo.

Além dos padrões identificados, as buscas realizadas na ferramenta *Disciplinary Metadata* do DCC também apontam como recomendação para todas as áreas selecionadas, o perfil de metadados chamado *GBIF Metadata Profile* (GMP). Esse perfil foi desenvolvido pela *Global Biodiversity Information Facility* (GBIF), que é uma rede mundial voltada a promover a mobilização, acesso, descoberta e uso dos dados da biodiversidade. Assim, o

GMP também foi incluído na análise para comparação com o panorama do uso dos metadados.

De forma a identificar os padrões e perfis de metadados recomendados internacionalmente para a temática biodiversidade, além do Portal do DCC, foram também considerados os documentos e informações oficiais disponibilizados pelas entidades mantenedoras dos padrões de metadados selecionados e no portal da rede GBIF.

No domínio do conhecimento sobre biodiversidade, conforme aponta a GBIF (2019c), a principal referência de desenvolvimento e manutenção dos padrões de informação sobre biodiversidade é o grupo *Taxonomic Databases Working Group* (TDWG)¹¹, recentemente conhecido como *Biodiversity Information Standards*. Trata-se de uma associação científica e educacional sem fins lucrativos que aceita como membros indivíduos e instituições – incluindo agências governamentais, organizações não governamentais, programas e projetos. “Tudo o que o TDWG produz provém da experiência, colaboração e suporte de seus membros - pessoas e organizações que criam, gerenciam, integram e analisam informações sobre biodiversidade” (TDWG, 2019a). A maioria dos membros são biólogos profissionais ou cientistas da informação.

Assim como o DCC, a GBIF (2019c) aponta que os padrões de informações comumente usados na temática biodiversidade são o *Access to Biological Collections Data* e o padrão *Darwin Core*. Além disso, a GBIF desenvolveu o *GBIF Metadata Profile* (GMP), que é um perfil de metadados predominantemente embasado no padrão de *Ecological Metadata Language*. A seguir são apresentados os resultados da análise meta-sintética da bibliografia sobre os padrões de metadados para biodiversidade, considerando as recomendações de entidades globais representativas que tratam do tema – DCC, GBIF, TDWG e outras.

4.1.1 ACCESS TO BIOLOGICAL COLLECTION DATA

O *Access to Biological Collections Data* (ABCD) é um padrão em evolução, baseado em XML e estabelecido para o acesso e intercâmbio de dados primários sobre a biodiversidade, incluindo espécimes e dados primários (DCC, 2019a; TDWG, 2019b). O esquema do ABCD é abrangente e estruturado para suportar dados de uma ampla variedade de bancos de dados. Isso o torna compatível com vários padrões de dados existentes

¹¹ Portal do grupo *Biodiversity Information Standards* (TDWG) <<https://www.tdwg.org>>.

(TDWG, 2019b). Esquemas secundários também são disponibilizados para viabilizar notações em texto livre.

O desenvolvimento do ABCD começou em 2001 e em 2005 foi ratificado como padrão TDWG. A última versão, o ABCD 3.0, foi publicada em janeiro de 2019 (TDWG, 2019b), contendo novo esquema XML, com recursos focados na reutilização de elementos, implementação de solicitações da comunidade e harmonização com a nova ontologia. Prevalece nesse padrão a disponibilização de recursos detalhados para descrever características de coleções da área da botânica, pois grande parte da equipe responsável por documentar a especificação pertence ao Jardim e Museu Botânico de Berlim – *Botanischer Garten und Botanisches Museum* (BGBM).

A especificação do ABCD 3.0 publicada em janeiro de 2019 (TDWG, 2019b) contém como principais inovações: um novo esquema XML, com recursos focados na reutilização de elementos; a qualificação semântica, pois passou a ser descrito como um esquema XML e uma ontologia. Houve alterações na estrutura do padrão em busca de melhor atender a casos de uso considerados específicos. Como prova de conceito dessas alterações, foi desenvolvido um Esquema de Aplicação ABCD 3.0 para Dados de Coleções Zoológicas¹².

O ABCD também é usado para publicar dados na rede GBIF, na Federação Alemã de Dados Biológicos – GFBio, na Europeana e outras bases. Esse padrão foi estendido para o mapeamento de amostras de DNA e para registros relacionados às Geociências. Também é usado para publicação em portais especializados temáticos.

Entre os portais que utilizam o padrão ABCD, destaca-se o *Biological Collection Access Service*, comumente conhecido como BioCASE¹³. Esse portal é alimentado por uma rede internacional que vincula dados de coleções biológicas de museus de história natural, jardins botânicos, zoológicos e instituições de pesquisa. O BioCASE tem como base o padrão de troca de dados ABCD (GBIF, 2019c), mas tem protocolo próprio para a transferência de dados. O protocolo BioCASE independe de esquemas e pode ser usado em conjunto com qualquer esquema XML definido de forma personalizada, geralmente utilizando o esquema ABCD (BIOCASE, 2019; GBIF, 2019c) ou um de seus esquemas de extensão como ABCD-EFG (*Extension for Geosciences*) ou ABCD-GGBN (*Global Genome Biodiversity Network*).

¹² Exemplo de esquema de aplicação XML envolvendo o ABCD 3.0 para Dados de Coleções Zoológicas <<https://abcd.tdwg.org/xml/#application-schemas>>

¹³ Portal da rede BioCase <<https://www.biocase.org>>

A rede BioCASE também está integrada à GBIF e representa a principal fonte de dados no padrão ABCD. Em 2016, cerca de 35 milhões de registros da BioCASE referente a espécimes e observações alimentavam a rede GBIF. Esse valor tem aumentado significativamente, em 2019 já existem 160 milhões de registros. O portal GBIF exibe estatísticas de uso para informações mediadas pelo BioCASE (BIOCASE, 2016).

O esquema do ABCD possui 64 classes que expressam os principais conceitos que podem ser descritos pelo padrão. Há também oito classes que existem para modelar vocabulários controlados para os termos específicos, usualmente utilizados ao expressar o ABCD no RDF. A lista e descrição das 71 classes constam no ANEXO 1. Além disso, os elementos pertencentes a cada classe são listados no ANEXO 2.

O padrão ABCD também possui em sua ontologia lista de propriedades de objetos, utilizadas para conectar classes; e propriedades de tipos de dados, que permitem coletar valores de tipos específicos de classes. O quadro 13 apresenta sinteticamente informações relevantes sobre o padrão de metadados ABCD.

Quadro 13. Padrão de metadados ABCD

PADRÃO DE METADADOS ABCD	
Vocabulários Relacionados	Vocabulários <i>Life Science Identifiers (LSID)</i> , que são os elementos básicos da ontologia do <i>Biodiversity Information Standards</i> – antigo <i>Taxonomic Databases Working Group (TDWG)</i> < https://www.tdwg.org >
Especificação	http://www.tdwg.org/standards/115/
Website do Padrão	https://abcd.tdwg.org/
Projetos relacionados	BioCASE - Biological Collection Access Service for Europe Global Biodiversity Information Facility GeoCASE - Geosciences Collection Access Service

Fonte: DCC (2019a)

4.1.2 ECOLOGICAL METADATA LANGUAGE

A *Ecological Metadata Language (EML)* é um padrão de metadados que registra informações sobre conjuntos de dados ecológicos em documentos XML modulares e extensíveis. Trata-se de uma especificação de metadados inicialmente desenvolvida para a disciplina de ecologia, baseada em trabalhos realizados pela *Ecological Society of America and Associated Efforts (ECOINFORMATICS, 2019)*.

O padrão EML é administrado e mantido pelo *Knowledge Network for Biocomplexity* (KNB) ¹⁴, que disponibiliza um repositório internacional destinado a facilitar a pesquisa ecológica e ambiental (KNB, 2019). O repositório KNB utiliza o sistema de gerenciamento de dados Metacat e armazena conjuntos de dados descritos no padrão EML, mas também pode armazenar qualquer documento de metadados baseado em XML (KNB, 2019). Conforme apresentado em sua especificação (KNB, 2019), a EML define um vocabulário abrangente e uma sintaxe de marcação XML para documentar dados de pesquisa. Essa linguagem é amplamente utilizada nas ciências da terra e do meio ambiente, e cada vez mais em outras disciplinas.

O esquema do padrão EML foi projetado em módulos para facilitar o aperfeiçoamento da sua linguagem em termos de amplitude e detalhamento (KNB, 2019). Com sua arquitetura extensível, é possível escolher quais dos módulos principais são pertinentes para descrever os dados. Além disso, se o EML ficar aquém das necessidades de uma área específica, poderá ser estendido por meio da criação de um novo módulo para descrever o recurso.

Trata-se, portanto, de um padrão que evolui para atender às necessidades de documentação, preservação e compartilhamento aberto dos dados de pesquisa. O padrão EML inclui diversos módulos que têm como finalidades: (i) identificar e citar conjuntos de dados; (ii) descrever a extensão espacial, temporal, taxonômica e temática dos dados; (iii) descrever métodos e protocolos de pesquisa; (iv) descrever a estrutura e o conteúdo dos dados, mesmo em conjuntos de dados complexos; e (v) notação precisa dos dados com vocabulários semânticos. Os esquemas, descrição e tipo módulos que integram a EML constam no ANEXO 3 e a relação entre esses módulos e os elementos é listada no ANEXO 4.

A estrutura de nível superior do EML foi projetada para ser compatível com a sintaxe Dublin Core (KNB, 2019). Em geral, recursos de conjunto de dados, recursos de literatura, recursos de software e recursos de protocolo compreendem a lista de informações que podem ser descritas em EML. Esse padrão foi amplamente projetado para descrever recursos digitais; no entanto, também pode ser usado para descrever recursos não digitais (KNB, 2019). O quadro 14 apresenta sinteticamente informações relevantes sobre o padrão EML.

¹⁴ Portal do KNB < <https://knb.ecoinformatics.org> >

Quadro 14. Padrão de metadados EML

ECOLOGICAL METADATA LANGUAGE – EML	
Vocabulários relacionados	Vocabulário próprio e expansível
Especificação	< http://knb.ecoinformatics.org/software/eml/eml-2.1.1/index.html >
Página na web	< http://knb.ecoinformatics.org/software/eml/ >
Projetos relacionados	Global Biodiversity Information Facility KNB - The Knowledge Network for Biocomplexity Long Term Ecological Research Network National Center for Ecological Analysis and Synthesis

Fonte: DCC (2019a)

Destaca-se que o padrão EML é amplamente utilizado para classificar e documentar os conjuntos de dados de pesquisa, mas seus atributos não se estendem para descrição detalhada dos dados como ocorre no DwC. Dessa forma, as notações do DwC costumam associar-se a arquivos EML.

4.1.3 DARWIN CORE

O *Darwin Core* (DwC) surgiu como uma extensão do Dublin Core para a biodiversidade, em uma iniciativa apoiada pelo Centro National de Pesquisa dos EUA – *National Science Foundation* – iniciada em 1998. Posteriormente, tornou-se um padrão de metadados revisado e ratificado pelo TDWG em 2009. É considerado o padrão mais utilizado para dados de biodiversidade, pois foi desenvolvido para fornecer uma maneira simples de documentar e compartilhar informações sobre ocorrências de espécies, coletadas em campo ou a partir de uma coleção de museu (GBIF, 2019e). Trata-se de um conjunto de padrões, que inclui um glossário de termos destinado a facilitar o compartilhamento de informações sobre diversidade biológica e que fornece definições de referência, exemplos e comentários (DCC, 2019b).

A organização do padrão é baseada principalmente nos táxons¹⁵ e na ocorrência desses táxons na natureza. Isso é documentado por observações, coleta de espécimes, amostras e informações relacionadas (TDWG, 2015a). A última versão do DwC foi publicada em outubro de 2015 (TDWG, 2015b) e sua gestão é realizada pelo *Darwin Core Maintenance Group*. Esse grupo é composto por membros da comunidade diretamente interessada em dados sobre biodiversidade e faz parte do *Biodiversity Information Standards* (TDWG, 2015c). Diferente do ABCD e EML que têm maior foco na descrição do conjunto de dados e

¹⁵ Taxonomia é a disciplina do campo da biologia que define os grupos de organismos com base em características comuns e dá nomes a esses grupos. **Táxon** é uma unidade taxonomica associada a um sistema de classificação científica da taxonomia.

documentos associados, o DwC apresenta classes que permitem melhor descrever informações sobre os campos de dados.

O DwC oferece uma estrutura estável, direta e flexível para compilar dados de biodiversidade de fontes variadas e envolvendo diferentes variáveis (GBIF, 2019c). A maioria dos conjuntos de dados compartilhados por meio do portal GBIF segue esse padrão. Além disso, os principais projetos que utilizam o DwC são: *Atlas of Living Australia*; GBIF; *Ocean Biogeographic Information System (OBIS)*; Rebioma; e VertNet (DCC, 2019b). O quadro 15 apresenta sinteticamente informações relevantes sobre o padrão de metadados ABCD.

Quadro 15. Padrão de metadados *Darwin Core*

PADRÃO DARWIN CORE	
Vocabulários Relacionados	Audubon Core; GBIF Vocabularies
Especificação	< http://www.tdwg.org/standards/450 >
Página na web	< http://rs.tdwg.org/dwc/index.htm >
Projetos relacionados	<i>Atlas of Living Australia</i> ; <i>Global Biodiversity Information Facility</i> ; OBIS - <i>Ocean Biogeographic Information System</i> ; Rebioma; VertNet

Fonte: DCC (2019b).

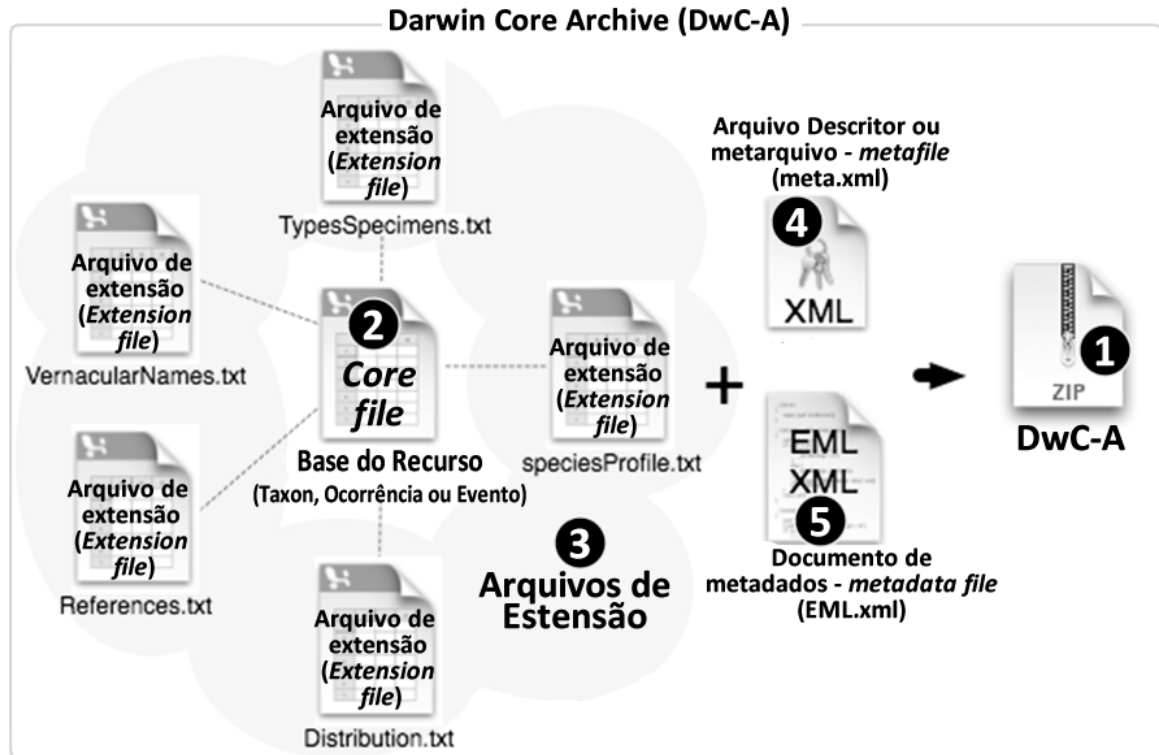
4.1.3.1 *Darwin Core Archive (DwC-A)*

O uso do padrão *Darwin Core* gira em torno de um formato de arquivo padrão, o *Darwin Core Archive (DwC-A)*. Esse formato refere-se a um pacote compactado – em formato ZIP – que contém arquivos de texto interconectados e permite que os pesquisadores ao publicar seus dados utilizem terminologias padronizadas. Esse arquivo utiliza os termos do *Darwin Core* para documentar um conjunto de dados independente e compartilhar dados de listas de verificação de espécies, de ocorrência ou de eventos de amostragem.

O arquivo DwC-A final compactado no formato ZIP – figura 10 – é composto por quatro tipos de arquivos: 1) *Core File* é o arquivo principal que contém um conjunto padrão de termos do *Darwin Core* e pode ser do tipo Taxon, Ocorrência ou Evento; 2) Arquivos de Extensão que são opcionais e correlacionados com o arquivo núcleo pelo arquivo descritor; 3) Arquivo descritor ou *metafile* define a estrutura e relacionamentos entre o arquivo

principal e os arquivos de extensões (meta.xml); 4) Documento de metadados é um arquivo complementar de metadados que descreve os conjuntos de dados contidos no DwC-A.

Figura 10. Estrutura do *Darwin Core Archive* (DwC-A)



1 - DwC-A em formato ZIP. 2 - *Core File* (tipo Taxon, Ocorrência ou Evento). 3 - Arquivos de Extensão (opcionais). 4 - Arquivo descritor (define a estrutura e relacionamentos entre *core file* e extensões). 5 - Arquivo de metadados (descreve os conjuntos de dados). Fonte: Adaptado de OBIS (2013).

Os arquivos extensão (*extension files*) em formato texto (.txt) podem ser incluídos para adicionar informações vinculadas aos registros do arquivo principal. Esses arquivos de extensão são opcionais e permitem que o DwC-A contenha relacionamentos muitos-para-um entre o arquivo principal e quaisquer arquivos com dados mais específicos a serem correlacionados.

Cada DwC-A contém duas peças fundamentais que ajudam as máquinas e os seres humanos a interpretar os dados (GBIF, 2017b; OBIS, 2013). A primeira refere-se ao arquivo descritor (meta.xml), descreve a estrutura e os relacionamentos entre o *core file* e quaisquer extensões (*extensions files*). A segunda é o documento complementar de metadados de recursos que fornece informações sobre o conjunto de dados em si, como um resumo, descrevendo os agentes responsáveis pela autoria, publicação e documentação, bem como informações bibliográficas, métodos de coleta e muito mais informações caso desejado. Esse

documento de metadados complementar normalmente utiliza a *Ecological Metadata Language* (EML) – arquivo EML.xml – mas existem outros padrões suportados (GBIF, 2017b).

Por fim, destaca-se que o *core file* (*taxa.txt*, *occurrence.txt* ou *event.txt*) apresenta um conjunto padrão de termos do Darwin Core conforme o seu tipo (GBIF, 2017b). Considerando a quantidade de atributos relacionados aos dados de origem e o que se deseja compartilhar, opta-se por um dos três tipos do arquivo principal (GBIF, 2018a):

- Taxon (*Taxon core*) - lista um conjunto de espécies, normalmente provenientes da mesma região ou que compartilham características comuns.
- Evento (*Event core*) - lista os estudos de campo e métodos, incluindo os protocolos usados, o tamanho da amostra e o local de cada um.
- Ocorrência (*Occurrence core*) - lista um conjunto de horários e locais em que espécies específicas foram registradas.

As classes e respectivas propriedades (atributos ou elementos) são descritas como extensões ao padrão DwC (GBIF, 2015a; GBIF, 2015b; GBIF, 2015c). A lista completa é apresentada no ANEXO 5.

Além das classes e respectivos elementos estabelecidos para descrição dos dados, há ainda vocabulário controlado associado para orientar o preenchimento correto. Essa padronização não apenas simplifica o processo de publicação, mas também de descoberta dos conjuntos dados. Consequentemente, facilita a avaliação e comparação desses dados em atividades de pesquisa que demandam o uso intensivo de dados da biodiversidade.

4.1.4 GBIF METADATA PROFILE (GMP)

GBIF Metadata Profile (GMP) foi desenvolvido pela *Global Biodiversity Information Facility* (GBIF) e sua documentação é disponibilizada no portal da rede, juntamente com guia de instruções e referência. A GBIF é uma rede internacional de pesquisa financiada pelos governos do mundo, destinada a fornecer a qualquer pessoa, em qualquer lugar, acesso aberto a dados sobre todos os tipos de vida na Terra (GBIF, 2019b). Essa rede promove e facilita a mobilização, o acesso, a descoberta e o uso de dados sobre biodiversidade.

Documentar a proveniência e o escopo dos conjuntos de dados é requisito fundamental para publicá-los na rede GBIF. O GMP foi desenvolvido para padronizar essa descrição e existem várias maneiras de escrever um documento de metadados em

conformidade com esse perfil (GBIF, 2017a; GBIF, 2018b). Um exemplo disso são documentos de metadados gerados no padrão DwC, que devem ter o arquivo de metadados do GMP incluído no arquivo DwC-A (GBIF, 2017a).

O GMP é baseado principalmente no padrão EML, pois utiliza um subconjunto e o estende para incluir requisitos adicionais. Os elementos de metadados que integram o GMP são organizados em categorias. A lista e descrição das categorias e respectivos elementos são apresentadas no ANEXO 6. O quadro 16 apresenta sinteticamente informações relevantes sobre o GMP.

Quadro 16. Síntese de informações sobre o perfil de metadados GMP.

GBIF METADATA PROFILE (GMP)	
Vocabulários relacionados	EML, <i>Darwin Core</i> e ABCD
Especificação	< https://github.com/gbif/ipt/wiki/GMPHowToGuide >
Página na web	https://www.gbif.org/standards >
Projetos relacionados	<i>Global Biodiversity Information Facility</i>

Fonte: DCC (2019)

O GMP resulta da ampla experiência desenvolvida pela rede GBIF, que envolve infraestrutura, entidades e pesquisadores de todo o mundo. Além disso, trata-se de um perfil de metadados que absorveu dos principais padrões de metadados – ABCD, DwC e EML – os esquemas, recomendações e vocabulários considerados mais relevantes para descrever conjuntos de dados resultante de pesquisas sobre a biodiversidade.

Nesse sentido, optou-se por aprofundar o conhecimento sobre as recomendações e exigências da GBIF para publicação de conjuntos de dados.

4.1.4.1 Rede GBIF: características, abrangência e principais iniciativas

A rede *Global Biodiversity Information Facility* (GBIF) – Sistema Global de Informação sobre Biodiversidade – surgiu a partir de um esforço científico internacional coordenado pela Organização para Cooperação e Desenvolvimento Econômico (OCDE)¹⁶ para permitir que usuários de todo o mundo compartilhem e utilizem abertamente grandes quantidades

¹⁶ OCDE entidade internacional que atua juntamente aos governos, formuladores de políticas e cidadãos para formulação de políticas voltadas solucionar problemas sociais, econômicos e ambientais.

de dados sobre a biodiversidade. O principal objetivo da GBIF é tornar a informação sobre biodiversidade disponível de forma livre e universal para a ciência e a sociedade.

As coleções de história natural, os programas de monitoramento ambiental, as sociedades voltadas ao registro de dados biológicos (*recording societies*) e seus esquemas de registros (*recording schemes*), os projetos científicos com ou sem participação da sociedade e outros dados valiosos sobre a biodiversidade são coletados e gerenciados em muitos sistemas e ambientes diferentes. Isso varia amplamente conforme o objetivo da entidade ou do pesquisador, dependendo do tipo de detalhes que são capturados e armazenados os dados para qualquer registro individual. Por isso, integrar e viabilizar o uso desses diversos conjuntos de dados para finalidades distintas é um desafio.

A partir das diferentes abordagens e diversas iniciativas, a GBIF estabeleceu uma infraestrutura robusta e um modelo governança eficaz que a tornaram uma referência mundial no que se refere à integração das informações sobre biodiversidade. No início de 2020, a rede registra mais de 50 mil conjuntos de dados e cerca de 1,4 milhão de dados de ocorrência de espécies (GBIF, 2019b).

A preocupação com o uso adequado e a conservação dos recursos naturais vem aumentando em todo o mundo desde o século passado. Os programas, projetos e iniciativas fomentados pela GBIF auxiliam o desenvolvimento tecnológico sustentável e a pesquisa científica em muitas disciplinas, com foco na conservação da biodiversidade e no compartilhamento equitativo de seus benefícios para melhorar a qualidade de vida da sociedade.

A GBIF é, portanto, uma iniciativa intergovernamental e seus participantes são países, economias e organizações internacionais que colaboram para promover o acesso livre e aberto aos dados da biodiversidade. A rede é constituída de 97 entidades membros e mais de 1500 colaboradores distribuídos em 58 países do mundo. Entretanto, há registro do uso de suas ferramentas em mais 70 países.

Há também os países, organizações ou outras entidades internacionais chamados participantes associados ou observadores, que integram a rede, mas não participam do apoio financeiro. Esses participantes demonstram interesse em compartilhar dados sobre biodiversidade e seguir as atividades da rede. O Brasil é um país do tipo participante associado desde 2012, com 103 publicadores registrados. O mapa a seguir esboça o alcance global do GBIF, distribuído nos cinco continentes (GBIF, 2019b).

A obtenção de dados na rede GBIF é facilitada por ferramentas de pesquisa que disponibilizam diversos parâmetros de busca e oferecem diferentes formas de visualização, bem como outros recursos para auxiliar a consulta e exportação dos dados. É possível pesquisar conjuntos de dados a partir de parâmetros relacionados à ocorrência, à espécie ou seus publicadores. Há também ferramenta de análise de tendências, que permite compreender a mudança na disponibilidade dos registos de ocorrência de espécies ao longo do tempo. Além disso, a GBIF disponibiliza orientações sobre como citar os conjuntos de dados extraídos da rede, bem como utilizar ferramentas mais avançadas de processamento e modelagem.

O volume expressivo de dados da biodiversidade e as ferramentas disponibilizadas, bem como as diversas iniciativas coordenadas pela GBIF são atrativos relevantes aos pesquisadores de todo o mundo. Em contrapartida, para integrar-se à rede e publicar dados é necessário atender aos padrões de informação recomendados pela GBIF, tendo em vista que “são o principal facilitador para reunir as centenas de milhões de registos primários de biodiversidade” (GBIF, 2019c).

Em relação à publicação de dados na GBIF, é necessário documentar a proveniência e o escopo dos conjuntos de dados para publicar nessa rede (GBIF, 2017a). Essa documentação do conjunto de dados visa à descrição dos metadados que permitem aos usuários avaliar a adequação para uso de um conjunto de dados. Nesse sentido, a GBIF apresenta diversas iniciativas voltadas a promover boas práticas de gestão de dados e orientar sobre a importância do uso dos padrões da informação – programas de capacitação, guias e ferramentas.

Existem duas ferramentas principais que orientam a publicação de conjuntos de dados na rede GBIF. A primeira é o validador de dados, um serviço que permite a qualquer pessoa submeter um conjunto de dados sobre biodiversidade para gerar um relatório sobre a correção sintática e a validade do conteúdo contido no conjunto de dados. O relatório gerado contém orientações sobre os procedimentos de validação e interpretação geralmente associados à publicação no GBIF e determina rapidamente se há problemas. Caso o conjunto de dados seja devidamente validado, o colaborador poderá decidir se deseja publicar.

A segunda ferramenta relacionada à publicação de dados é o *Integrated Publishing Toolkit* (IPT), um *software* livre e de código aberto desenvolvido para auxiliar a publicação e

compartilhamento dos conjuntos de dados de biodiversidade que facilita o atendimento aos padrões de informação estabelecidos. Os dados e metadados registrados pelo IPT são integrados e indexados para consulta via rede e portal GBIF, assim são automaticamente disponibilizados para uso público (GBIF, 2019a). Essa aplicação é compatível com vários idiomas e possibilita configurar e agendar ciclos de publicação automática relacionados aos programas de monitoramento em andamento e projetos de amostragem (GBIF, 2019e).

4.1.4.3 *Data papers*

Diante ao significativo esforço necessário para preparar, selecionar e descrever dados, os *data papers* – artigos que descrevem conjuntos de dados ou artigos de dados – têm ganhado espaço no ambiente acadêmico. Esses artigos são documentos revisados por pares que descrevem conjuntos de dados publicados em periódicos científico, em geral especializados em compartilhamento e preservação de dados científicos (GBIF, 2019f). O padrão EML engloba atributos de metadados voltados a detalhar o registro dessas publicações.

A rede GBIF desenvolve iniciativas para promover e facilitar a publicação dos *data papers* visando o reconhecimento a todos os envolvidos na publicação de dados e a publicidade dos conjuntos de dados perante a comunidade científica. No Portal GBIF há ferramentas e guias voltados a incentivar esse tipo de publicação (GBIF, 2019f). Todas as descrições de conjuntos de dados na rede GBIF se baseiam metadados que usam o padrão EML (GBIF, 2019c). Inclusive os arquivos no padrão Darwin Core, pois cada arquivo DwC-A inclui como um de seus componentes um arquivo de referência do padrão EML (escrito em formato XML). Além disso, a ferramenta *Integrated Publishing Toolkit* (IPT), desenvolvida pela rede GBIF baseada no perfil GMP, contém um editor de metadados interno que permite preencher facilmente metadados, validá-los e produzir um arquivo EML no formato XML – EML.xml.

4.1.5 LINHA DE BASE: PARÂMETROS PARA AVALIAR O USO DOS METADADOS

A identificação dos padrões de metadados envolveu a análise da documentação oficial disponibilizada pelas entidades com abrangência global e consideradas referência para esse tema – DCC, TDWG, GBIF, KNB, ECOINFORMATICS. De forma complementar, foi

também analisada a documentação disponibilizada por projetos associados a essas entidades e aos padrões identificados.

Desse conhecimento gerado, destaca-se a identificação do perfil de metadados GMP desenvolvido pela GBIF, que é embasado nos outros três padrões selecionados neste estudo para análise – EML, DwC e ABCD. Também é relevante a compreensão de que a rede GBIF permite o registro de conjuntos de dados descritos em todos esses padrões. Além disso, os dados são depositados nessa rede por pesquisadores e entidades de diversos países localizados nos cinco continentes e seguem as diretrizes estabelecidas por consenso com a comunidade internacional.

Isso permite a integração de conjuntos de dados descritos de diferentes formas e níveis de detalhamento. Portanto, a rede GBIF representa um ambiente com recursos propícios à avaliação comparativa das experiências globais para compreender o uso dos padrões de metadados.

Nesse sentido, esta pesquisa adota como linha de base as diretrizes, recomendações e exigências apresentadas pela GBIF para a publicação dos conjuntos de dados de pesquisa sobre a biodiversidade. Tais referências, portanto, serão utilizadas para avaliar como se dá o uso efetivo de metadados considerando as evidências coletadas na etapa II deste estudo.

4.1.5.1 Classes de conjuntos de dados

Os principais tipos de conjuntos de dados da biodiversidade são os listas de verificação de espécies (descrevem as espécies) e os de ocorrências (identificam os locais de presença), bem como dados de recursos (informações associadas), tais como de provedores de dados, bancos de dados, dados aeroespaciais, dados bibliográficos e arquivos multimídia (SILVA, CORRÊA, *et al.*, 2015, p. 32). A GBIF (2019d) agrupa os conjuntos de dados em classes, que iniciam de recomendações mais simples até progressivamente alcançar estruturas mais detalhadas e complexas. Dessa forma, nem todos os conjuntos de dados na rede GBIF incluem informações no mesmo nível de detalhe.

Existem várias maneiras de escrever um documento de metadados em conformidade com as exigências da GBIF. As formas mais comuns são: (i) o uso do editor de metadados do GBIF Integrated Publishing Toolkit (IPT), (ii) o modelo de metadados do recurso GBIF ou (iii) a geração manual de um documento de metadados.

Os colaboradores que compartilham dados na rede GBIF são constantemente estimulados a publicar os dados nas classes mais estruturadas, de forma a permitir o uso em uma ampla variedade de abordagens e pesquisas. Entretanto, a GBIF (2019d) ressalta que compartilhar dados menos estruturados sobre biodiversidade também é valioso, pois mesmo informações parciais podem auxiliar na resposta a perguntas importantes.

Em todas as classes propostas pela GBIF, os conjuntos de dados são associados a um *Digital Object Identifier – DOI*¹⁷ quando registrados na rede. Segundo a GBIF (GBIF, 2019d; GBIF, 2018b), a rede suporta a publicação, descoberta e uso de quatro classes de dados:

- (i) Metadados dos Recursos – *Resources metadata*;
- (ii) Lista de verificação de espécies – *Checklist data*;
- (iii) Dados de ocorrência (dados primários) – *Occurrence data*; e
- (iv) Dados do evento de amostragem – *Sampling-event data*.

4.1.5.2 Metadados de recursos – *resources metadata*

Os metadados de recursos – *resources metadata* – permitem documentar as coleções de dados. Incluem a localização, proveniência, proprietário e qualidade dos dados das coleções, podendo também abranger aspectos técnicos do conjunto de dados que permitem avaliar sua adequação para uso. Considerado o nível mais simples, essa classe engloba informações básicas para que as instituições criem conjuntos de dados descrevendo recursos não digitalizados, como coleções da história natural e outras (GBIF, 2018b).

Todos os demais tipos de classes de conjuntos de dados incluem esta informação de base. Essas informações também facilitam a avaliação da relevância e do valor de coleções não digitalizadas para definir prioridades de digitalizações futuras (GBIF, 2019c).

A descrição dessa classe de conjunto de dados é registrada em um documento de metadados XML, que é baseado na EML e deve estar em conformidade com o GMP. A GBIF não disponibiliza um modelo de arquivo tabular para metadados do recurso. Por isso, sugere o uso do editor de metadados no IPT, que garante que o preenchimento no formato XML válido e apropriado. Se os metadados já estiverem descritos em EML ou DwC é possível enviá-los nesses formatos para o IPT.

¹⁷ DOI foi concebido como uma estrutura genérica para gerenciar a identificação de conteúdo em redes digitais, reconhecendo a tendência de convergência digital e disponibilidade de multimídia **Fonte bibliográfica inválida especificada..**

A documentação desses metadados é uma ferramenta valiosa para os pesquisadores descobrirem recursos que ainda não estão disponíveis online. Também é uma maneira útil de avaliar a importância e o valor de recursos não digitais para planejar uma digitalização futura. O quadro a seguir apresenta os campos obrigatórios e os recomendados para essa classe de conjunto de dados, que derivam principalmente do padrão EML.

Quadro 17. Campos obrigatórios e recomendados para a classe metadados de recursos da GBIF.

Elementos obrigatórios:	Descrição:
<i>Title</i>	<i>Title for the resource. This title will appear as the Name of the resource throughout. Recommendation: a brief, but descriptive title that characterizes the dataset in an international context and distinguishes it from similar datasets in other institutions.</i>
<i>Description</i>	<p><i>A brief overview of the resource broken into paragraphs. This should provide enough information to help potential users of the data to understand if it may be of interest.</i></p> <p><i>This may include a longer version of title, a description of geographic, temporal and taxonomic scope of the checklist, methodology and purpose of the underlying data compilation (e.g. red list, invasive species, freshwater taxa, regional flora), relevant literature references, and any other information relevant to characterize the dataset</i></p>
<i>Publishing Organization</i>	<p><i>Organization responsible for publishing (producing, releasing, holding) this resource. It will be used as the resource's rights holder and publishing organization when registering this resource and when submitting metadata during DOI registrations. It will also be used to auto-generate the citation for the resource (if auto-generation is turned on), so consider the prominence of the role.</i></p> <p><i>The publishing organization is the institution which holds or owns the dataset and is in charge of its contents and maintenance. The title given should be the official title of the organization as registered with relevant authorities, listed on websites, and, if applicable, as stated in the project contract.</i></p>
<i>Type</i>	<p><i>Dataset metadata EML. Type of resource. The value of this field depends on the core mapping of the resource and is no longer editable if the Darwin Core mapping has already been made.</i></p> <p><i>The record type describes the main focus of all records contained in the dataset (core records).</i></p>
Elementos obrigatórios:	Descrição:
<i>License</i>	<i>The license applied to the resource. The license provides a standardized way to define appropriate uses. GBIF encourages publishers to adopt the least restrictive license possible from among three (default) machine-readable options (CCO 1.0, CC-BY 4.0 or CC-BY-NC 4.0) to encourage the widest possible use and application of data.</i>

<i>contact(s)</i>	<p>List of people and organizations that should be contacted to get more information about the resource, that curate the resource or to whom putative problems with the resource or its data should be addressed.</p> <p>Contact data (minimum: name and email) for at least one administrative contact for the dataset.</p>
<i>creator(s)</i>	<p>One or more individuals, groups, or institutions responsible for the creation of the dataset. All contributors to the dataset should be listed - see data citation principle #2. Creators should be aware, however, that the full list can be truncated by the Journal during typesetting (e.g. According to Nature's guidelines, they will truncate at 6 creators). Creators should be listed last name first, followed by initials of given names. Creators are listed according the importance of the role they played in the creation of the dataset, with the most important creator appearing first. Multiple creators are separated by commas.</p>
<i>metadata provider(s)</i>	<p>People and organizations responsible for producing the resource metadata. If this person or organization is the same as the first resource contact, all of the details of the latter can be copied into the equivalent fields for the resource creator by clicking on the link labeled "copy details from resource contact". The metadata provider has all of the same fields and requirements as the resource contact. Refer to the field explanations under Resource Contacts, above.</p> <p>Contact data (minimum: name and email) for the author of the dataset metadata</p>
Elementos recomendados: Descrição:	
<i>sampling methodology - in situations where data comes from a sampling event</i>	<p>Contains information about sampling methods used for the data represented by the resource. A text-based/human readable description of the sampling procedures used in the research project. The content of this element would be similar to a description of sampling procedures found in the methods section of a journal article.</p>
<i>citation</i>	<p>Contains information about how to cite the resource as well as a bibliography of citations related to the data set, such as publications that were used in or resulted from the production of the data. Each Citation, whether for the resource or in the bibliography, consists of an optional unique Citation Identifier allowing the citation to be found among digital sources and a traditional textual citation.</p>
<i>additional metadata</i>	<p>Enrich the description and improve the fitness of use of the dataset.</p>

Fonte: GBIF (2018b); GBIF (020b) e KNB (2019)

Conforme boas práticas recomendadas pela GBIF (GBIF, 2017a), além dos elementos obrigatórios e recomendados citados no quadro acima, deve-se realizar a descrição do maior número de informações recomendadas para recursos de metadados no GMP. As outras três classes de conjuntos de dados propostas pela GBIF incluem essas informações básicas da classe metadados de recursos.

4.1.5.3 Lista de verificação de espécies – *checklist data*

Os conjuntos de dados dessa classe fornecem um catálogo ou lista de organismos ou táxons. Embora possam incluir detalhes adicionais, como nomes locais de espécies ou citações de espécimes, a classe de dados lista de verificação de espécies – *checklist data* – normalmente categoriza as informações de acordo com linhas taxonômicas, geográficas e temáticas, ou alguma combinação.

Os termos "lista de verificação de espécies" e "catálogo taxonômico" podem se referir a diversos recursos. Todos esses produtos contêm conjuntos de nomes científicos que se referem implícita ou explicitamente a táxons. Conforme aponta o manual de boas práticas para publicação de listas de verificação de espécies da GBIF (2018c), em ordem crescente de detalhes, esse tipo de conjunto de dados podem incluir os seguintes tipos de recursos:

1. Listas de nomes - Listas simples de nomes de espécies sem indicação explícita do status taxonômico, mas geralmente focadas em consistir nomes aceitos de táxons. Essas listas geralmente identificam um conjunto de táxons incluídos em algum contexto regional ou temático.
2. Listas nomenclaturais (Nomencladores) - Listas de nomes, incluindo os táxons nominais, ou seja, o registro de usos publicados de nomes científicos representando atos nomenclaturais, conforme regidos pelos respectivos códigos de nomenclatura. A maioria desses atos contém 'descrições originais' de novos nomes científicos, mas outros podem incluir emendas, tipificações ou outros atos regidos pelos Códigos. A sinonímia não é incluída nessas listas como conceito taxonômico, mas apenas como uma combinação recém-estabelecida (para botânicos) vinculada a um basiônimo, fornecendo assim um sinônimo nomenclatural.
3. Lista de verificação taxonômica - Essas listas estendem as listas nomenclaturais adicionando opinião taxonômica na forma de informações explícitas sobre o status taxonômico e a inclusão de nomes colocados na sinonímia. As listas taxonômicas simples dessa categoria não fornecem detalhes específicos da circunscrição sobre a base da sinonímia. As listas são frequentemente organizadas em classificações. O termo "catálogo taxonômico" também pode ser usado para se referir a instâncias deste tipo e

às categorias restantes, principalmente se o recurso incluir publicação verificada e detalhes do status taxonômico.

4. Listas de verificação anotadas - Esta categoria estende as listas de verificação taxonômicas adicionando outros tipos de dados (anotações) à lista de verificação principal e com sinônimos, como nomes comuns, status de ameaças, distribuição e informações descritivas básicas. Quando os tipos de anotação fornecem detalhes suficientes para definir ou circunscrever efetivamente um táxon, como descrições e ilustrações detalhadas de diagnóstico, dados moleculares, amostras, etc., a lista anotada pode se enquadrar em uma das duas categorias definidas abaixo.
5. Listas de flora ou fauna - normalmente fornecem informações detalhadas de espécies para uma região específica. Os detalhes podem incluir muitos dos tipos de dados incluídos nas listas de verificação anotadas, mas incluem tipos de dados específicos, como descrições e ilustrações detalhadas, referências de amostras e outros detalhes que definem explicitamente (circunscreva) o táxon no escopo da região que não é necessariamente global.
6. Monografias - são relatos detalhados de espécies para um grupo de táxons em escala global. Contém informações detalhadas da nomenclatura, da sinonímia e da circunscrição do táxon, que incluem descrições e ilustrações textuais, detalhes dos espécimes examinados e incluídos na definição e uma bibliografia da literatura examinada.

O arquivo no formato DwC-A associado às extensões de lista de verificação de espécies suporta a troca de elementos de dados importantes em todos esses tipos de listas de verificação (GBIF, 2018c). O nível específico de cobertura depende muito do recurso a ser publicado.

A GBIF disponibiliza modelo de estrutura tabular para facilitar o preenchimento de uma lista ou taxonomia de espécies, indicando os campos obrigatórios e os recomendados (GBIF, 2018b; GBIF, 2019d) – quadro 18. Colunas adicionais também podem ser adicionadas utilizando os termos estabelecidos no DwC.

Quadro 18. Metadados obrigatórios e recomendados para conjuntos de dados da classe *checklist*

Elementos obrigatórios:		Descrição
<i>TaxonID</i>		<i>An identifier for the set of taxon information (data associated with the Taxon class). May be a global unique identifier or an identifier specific to the data set.</i>
<i>ScientificName</i>		<i>An identifier for the nomenclatural (not taxonomic) details of a scientific name.</i>
<i>TaxonRank</i>		<i>The taxonomic rank of the most specific name in the scientificName.</i>
Elementos fortemente recomendados:		Descrição
<i>ParentNameUsageID</i>		<i>An identifier for the name usage (documented meaning of the name according to a source) of the direct, most proximate higher-rank parent taxon (in a classification) of the most specific element of the scientificName.</i>
<i>AcceptedNameUsageID</i>		<i>An identifier for the name usage (documented meaning of the name according to a source) of the currently valid (zoological) or accepted (botanical) taxon.</i>
<i>Kingdom</i>		<i>The full scientific name of the kingdom in which the taxon is classified.</i>
Elementos recomendados (se disponíveis):		Descrição
<i>vernacularName</i>		<i>A common or vernacular name.</i>

Fonte: GBIF (2018b); GBIF 2019d; TDWG (2015a)

Na descrição de conjuntos de dados da classe lista de verificação de espécies devem ser adicionados os metadados de recursos. Apesar da referência de elementos obrigatórios e recomendados, a GBIF orienta incluir o máximo de propriedades considerando o guia de boas práticas para publicação dessa classe de dados e as extensões relacionadas (GBIF, 2018b; GBIF, 2018c). A seguir é apresentada a lista completa de extensões que podem ser usadas para fornecer informações adicionais sobre um registro de táxon (GBIF, 2018c):

1. Extensão Taxon (Core) - conjunto de termos para fornecer as informações fundamentais para uma lista de verificação de espécies, incluindo classificação, sinonímia e outros elementos-chave. Os termos desta classe suportam métodos diferentes para representar informações de classificação.

2. Extensão de nomes vernaculares - fornece os meios para compartilhar informações relacionadas a nomes comuns (vernáculos) vinculados a táxons no arquivo de dados principal. Vários nomes vernáculos podem ser vinculados ao mesmo táxon via taxonID.
3. Extensão de Referências – permite descrever uma ou mais referências bibliográficas relacionadas a um táxon no arquivo de dados principal. É usada para qualificar as referências. Esta extensão suporta o compartilhamento de listas de verificação de sinônimos referenciadas.
4. Extensão de Distribuição de Espécies – permite compartilhar informações sobre uma ou mais referências de distribuição para um táxon. Um ou mais registros de localidade podem estar vinculados ao mesmo táxon. Por exemplo, várias localidades, regiões ou países podem estar listados. Esta extensão auxiliar a descrição do status da ameaça para um táxon, alterações sazonais na distribuição e outras propriedades vinculadas a um táxon em uma região específica.
5. Extensão da descrição da espécie – fornece espaço para inclusão de texto descritivo para um táxon. Isso geralmente ocorre na forma de um único parágrafo por registro, como seria normalmente armazenado em um banco de dados. As descrições podem ser qualificadas por um tipo para indicar, por exemplo, que está relacionada à conservação, reprodução morfológica, etc. Várias descrições equivalem a vários registros em um arquivo de descrições.
6. Identificadores alternativos – permite o registro de mais de um identificador ou link para informações sobre o táxon. Um banco de dados de origem pode, por exemplo, fornecer acesso aos registros de dados de origem por meio de uma página da web, um serviço da web e um identificador resolvível, como LSID, DOI ou outros meios.
7. Extensão de tipos e espécimes – permite compartilhar dados relacionados a uma ou mais amostras ou referências de tipo vinculadas ao táxon principal.
8. Extensão de Relação de Recursos – é usada para descrever um ou mais relacionamentos que o táxon principal possui com outros táxons, na lista de fontes ou incluídos no registro. Essa extensão pode ser usada, por exemplo,

para fornecer uma lista de espécies de plantas (um registro por espécie) polinizadas por uma espécie de abelha listada na lista de espécies principais.

A classe lista de verificação de espécies pode funcionar, portanto, como um resumo rápido ou um inventário de táxons em um determinado contexto. O primeiro passo para a publicação na rede GBIF refere-se a transformar o conjunto de dados em uma estrutura tabular, cujos atributos são nomeados com termos do padrão DwC.

Os conjuntos de dados do tipo lista de verificação podem ser publicados utilizando o arquivo do tipo arquivo DwC-A. Cada arquivo pode conter muitos registros de extensão relacionados a um registro de núcleo táxon – *Core taxon* – permitindo a inserção de informações adicionais. De forma a auxiliar a preparação desses arquivos, a GBIF disponibiliza modelos para diversas extensões e recomendações das melhores práticas para essa classe de conjuntos de dados (GBIF, 2018c).

Na documentação dos conjuntos de dados da classe lista de verificação, os metadados podem descrever o escopo e a função pretendida da lista, metodologias e recursos utilizados para sua compilação e os indivíduos e organizações envolvidos em sua criação e gerenciamento. Entretanto, se houver informações suficientes no conjunto de dados de origem ou que se aplicam a todas as espécies da lista de verificação é recomendável que o conjunto de dados seja apresentado como do tipo dados de ocorrência.

4.1.5.4 Dados de Ocorrência

Os conjuntos de dados dessa classe apresentam detalhes e contribuem com informações sobre a localização de organismos individuais no tempo e no espaço, pois oferecem evidências da ocorrência de uma espécie (ou outro táxon) em um local específico e normalmente em uma data especificada (GBIF, 2018b). Esses conjuntos de dados expandem a maioria dos dados da lista de verificação de espécies, pois contribuem para o mapeamento da distribuição histórica ou atual de uma espécie.

A maior parte dos dados publicados na rede GBIF pertence à classe dados de ocorrência (GBIF, 2019d). Os exemplos podem variar de espécimes e fósseis em coleções de história natural, observações de pesquisadores de campo e cientistas cidadãos e dados coletados de armadilhas fotográficas ou satélites de sensoriamento remoto.

Os registros de ocorrência nesses conjuntos de dados às vezes fornecem apenas informações gerais da localidade ou simplesmente identificam o país (GBIF, 2018b).

Preferencialmente devem incluir coordenadas e uma precisão de coordenadas para suportar o mapeamento e definição de escala. Também podem registrar separadamente vários indivíduos da mesma espécie. Em muitos casos informam descrição dos locais, coordenadas geográficas, métodos, tipo de coletor e condições de coleta. Assim, permitem análises e mapeamentos em grande escala sobre a distribuição de espécies (GBIF, 2019d; GBIF, 2018b).

Exemplos de tais conjuntos de dados incluem bancos de dados de espécimes em coleções de história natural, dados de observações da ciência cidadã, dados de projetos de atlas de espécies, etc. Se existir informação suficiente no conjunto de dados de origem (ou se aplicar a todas as ocorrências no conjunto de dados) é recomendável que esses conjuntos incluam os dados de eventos de amostragem (GBIF, 2018b).

Quadro 19. Metadados obrigatórios e recomendados para conjuntos de dados da classe dados de ocorrência

Elementos obrigatórios:	Descrição
<i>occurrenceID</i>	<i>A unique identifier for the occurrence, allowing the same occurrence to be recognized across dataset versions as well as through data downloads and use. Ideally, the occurrenceID is a persistent global unique identifier. As a minimum requirement, it has to be unique within the published dataset. It allows recognizing the same occurrence over time when the dataset indexing is refreshed.</i>
<i>basisOfRecord</i>	<i>The type of the individual record, e.g. observation, physical specimen, fossil, living ex-situ, culture collection specimen. Even if the basis of record is supplied for the dataset as a whole, it is recommended practice to add it to each individual record as well to make the information explicitly available</i>
<i>scientificName</i>	<i>The full scientific name of the organism, to the lowest level taxonomic rank that is possible to supply, and including authorship and year of the name where applicable. The scientific name is one of the most important and frequently used search criteria for data access, and is indispensable for identifying the biological organism that is at the center of any occurrence record.</i>
<i>eventDate</i>	<i>The date or date interval during which the occurrence record was collected, following ISO 8601 date-time standard.</i>
Elementos fortemente recomendados:	Descrição
<i>countryCode</i>	<i>A two-letter standard abbreviation for the country of the occurrence locality.</i>

taxonRank	<i>The taxonomic rank of the most specific name in the scientificName.</i>
Kingdom	<i>The full scientific name of the kingdom in which the taxon is classified.</i>
decimalLatitude / decimalLongitude	<i>The geographic latitude or longitude, resp., in decimal degrees. Where coordinate values are available, both fields need to be supplied.</i>
geodeticDatum	<i>The coordinate system and set of reference points upon which the geographic coordinates are based</i>
coordinateUncertaintyInMeters	<i>The horizontal distance from the given decimalLatitude and decimalLongitude in meters, describing the smallest circle containing the whole of the Location</i>
individualCount / organismQuantity / organismQuantityType	<i>To record the quantity of a species occurrence, e.g. as the number of individuals, percentage of vegetation coverage, or the biomass. In many cases, a single occurrence record represents more than one individual, and for uses like population studies, this information is interesting to users of data.</i>
Elementos recomendados (se disponíveis):	Descrição
informationWithheld	<i>Additional information that exists, but that has not been shared in the given record.</i>
dataGeneralizations	<i>Actions taken to make the shared data less specific or complete than in its original form. Suggests that alternative data of higher quality may be available on request.</i>
eventTime	<i>The time or interval during which an Event occurred.</i>
Country	<i>The name of the country or major administrative unit in which the Location occurs.</i>

Fonte: GBIF (2018b); GBIF (2019d); TDWG (2015a); GBIF (2020b)

Na descrição de conjuntos de dados da classe dados de ocorrência também devem ser adicionados os metadados de recursos. Apesar da referência de elementos obrigatórios e recomendados, a GBIF orienta incluir o máximo de propriedades considerando as extensões do Dwc e boas práticas para publicação dessa classe de dados (GBIF, 2018b; GBIF, 2020b; GBIF, 2015b).

4.1.5.5 Dados de eventos de amostragem

Os conjuntos de dados de eventos de amostragem muitas vezes fornecem mais detalhes, além das evidências de ocorrência de espécies em um determinado local e data, que possibilitam avaliar a composição das comunidades de grupos taxonômicos mais amplos ou mesmo a abundância de espécies em vários momentos e locais (GBIF, 2019d). Dados

apresentados nessa classe geralmente derivam de protocolos para medir e monitorar a biodiversidade, tais como os transectos de vegetação, os censos de aves e as amostragens de água doce ou marinha (GBIF, 2018b).

Ao indicar os métodos, eventos e abundância relativa de espécies registradas em uma amostra, esses conjuntos de dados melhoram as comparações com os dados coletados usando os mesmos protocolos em diferentes momentos e locais – em alguns casos, até levando os pesquisadores a inferir a ausência de espécies específicas em locais específicos (GBIF, 2019d; GBIF, 2018b). Os esforços para rastrear mudanças nos padrões de biodiversidade no espaço e no tempo aumentaram a quantidade de informações disponíveis sobre espécies coletadas por meio de programas de amostragem e monitoramento (GBIF, 2019e). Os conjuntos de dados de eventos, com base em amostras, capturam detalhes mais complexos sobre quantidades e frequência de espécies. A adição do núcleo do tipo evento ao padrão Darwin Core incluiu vários novos termos aplicáveis aos dados de monitoramento e baseados em amostragem (GBIF, 2019e).

Com a inclusão frequente de medições realizadas nos mesmos locais, os dados de eventos de amostragem de investigações ecológicas e ambientais são melhores para detectar mudanças e tendências nas populações de espécies, bem como críticos para definir o escopo e a velocidade das mudanças globais (GBIF, 2018b). Assim, quando se adota no DwC-A um núcleo do tipo eventos – *event core*, um dos arquivos de extensão geralmente contém os elementos exibidos no núcleo de ocorrência – *occurrence core*, permitindo a inclusão de muitos registros de observação como parte de um único estudo de campo.

O núcleo do tipo evento permite que o evento de amostragem ocupe uma posição central do conjunto de dados, simplificando a vinculação do esforço e das medições das ocorrências de espécies. Como resultado, os pesquisadores podem acessar registros mais complexos e quantitativamente mais detalhados. Isso também viabiliza análises combinadas com outros arquivos focados em organismos únicos ou em táxons individuais.

A GBIF (2019e) entende que essa classe viabiliza melhorias na qualidade e utilidade dos conjuntos de dados, até mesmo os já publicados e que derivam de pesquisas e censos mais complexos. Pois a combinação dessas fontes variadas de dados pode facilitar a descoberta e reutilização dos dados, podendo permitir até a revelação de relacionamentos que não seriam aparentes ao examinar registros individuais.

Quadro 20. Metadados obrigatórios e recomendados para conjuntos de dados da classe evento

Elementos obrigatórios:	Descrição
eventID	<i>The date or date interval during which the sampling event as a whole took place, following ISO 8601 date-time standard</i>
eventDate	<i>The date or date interval during which the sampling event as a whole took place, following ISO 8601 date-time standard</i>
samplingProtocol	<i>The name of, reference to, or description of the method or protocol used during a sample event. Sample events typically use specific methods or follow certain protocols that standardize the sampling effort to a certain degree. Knowledge about the sampling protocol gives users additional information that is helpful for the interpretation of the attached occurrence records</i>
samplingSizeValue / samplingSizeUnit	<i>A numeric value and the corresponding unit for the value, specifying the size of an individual sample in the sampling event. The two sampleSize fields always go together, and specify the size of an individual sample within a sample event. The sample size can relate to time duration, a spatial lengths (e.g. of a trawl), an area or a volume.</i>
Elementos fortemente recomendados:	Descrição
countryCode	<i>A two-letter standard abbreviation for the country of the occurrence locality.</i>
parentEventID	<i>A cross-reference to the eventID of a broader event, e.g. a long-term monitoring project that the specific event is a part of, or a general vegetation survey of a larger area that is comprised of a number of sub-plots. In order to be able to reference a parent event, this event needs to be specified as a separate entry, typically within the same dataset, carrying its own eventID. Refer to the eventID of the parent event in the sample event record to specify the relationship between the two entries.</i>
samplingEffort	<i>The measure for the amount of effort that was expended during a sampling event</i>
locationID	<i>An internal or external reference that links to a set of data describing the sample event location, if available.</i>
decimalLatitude / decimalLongitude	<i>The geographic latitude or longitude, resp., in decimal degrees. Where coordinate values are available, both fields need to be supplied.</i>
geodeticDatum	<i>The coordinate system and set of reference points upon which the geographic coordinates are based</i>
coordinateUncertaintyInMeters	<i>The horizontal distance from the given decimalLatitude and decimalLongitude in meters, describing the smallest circle containing the whole of the Location</i>
footprintWKT	<i>An alternative area description, specifying the location of the sample event in Well-known text (WKT) markup language.</i>
occurrenceStatus	<i>This applies to associated occurrence data, not to the sample event itself. A qualifier for individual occurrence records, marking a taxon as either present or absent at a location during the sampling event. Since sample datasets document the sampling effort exerted during the event, it can often be valuable to not only document taxa as being present (observed, collected) at the location at the time, but also to record negative occurrences (absences) for taxa that could be reasonably expected, but were not encountered in the event.</i>

Fonte: GBIF (2018b; GBIF (2019d); TDWG (2015a) e GBIF (2020b)

Na descrição de conjuntos de dados da classe eventos devem ser adicionados os metadados de recursos. Apesar da referência de elementos obrigatórios e recomendados, a GBIF orienta incluir o máximo de propriedades considerando o guia de boas práticas para publicação dessa classe de dados e as extensões relacionadas (GBIF, 2018b; GBIF, 2015c).

4.2 ETAPA II – EVIDÊNCIAS DE USO DE METADADOS

A busca e coleta bibliográfica e documental das evidências de uso dos padrões de metadados nas pesquisas foram realizadas em duas fontes: 1) artigos científicos revisados por pares obtidos a partir do Portal de Periódicos da CAPES; e 2) documentação de metadados referente a conjuntos de dados de pesquisa disponíveis na rede GBIF. A seguir são apresentados os resultados encontrados a partir dessas duas estratégias. Além disso, as evidências coletadas representam insumo para a última etapa da pesquisa – Etapa III.

4.2.1 COLETA DE DADOS NOS ARTIGOS CIENTÍFICOS

As buscas por artigos científicos iniciaram a partir de termos mais abrangentes, que foram gradativamente correlacionados a termos específicos. Os termos mais abrangentes utilizados foram: *biodiversity* (biodiversidade) e *metadata* (metadado). Em seguida, esses foram correlacionados aos mais específicos, compostos pelos nomes dos três padrões e um perfil de metadados identificados para a temática biodiversidade: *Darwin Core* (DwC); *Ecological Metadata Language* (EML); e *Access to Biological Collection Data* (ABCD); GBIF metadata profile (GMP). As siglas foram utilizadas nas buscas. Apenas para *Darwin Core* as pesquisas foram realizadas com o uso do nome completo, pois a sigla não retornava resultados ou apresentava assuntos divergentes do estudado. O quadro a seguir exhibe a combinação de termos e as diversas buscas realizadas – quadro 21.

Quadro 21. Estrutura das buscas por artigos científicos: a relação entre os termos abrangentes e específicos utilizados.

Estrutura da consulta	Nome dos padrões de metadados estudados associados às palavras-chave principais que delimitam a pesquisa (<i>biodiversity e metadata</i>)				Buscas apenas com <i>Biodiversity e Metadata</i>
	ABCD	Darwin Core	EML	GBIF <i>profile</i>	
P (ANY) + M (ANY)	ABCD (ANY) + M (ANY)= 195	DW(ANY) + M (ANY)= 347	EML (ANY) + M (ANY) = 612	GBIF (ANY) + M (ANY) = 526	---
P (SUB) + M (SUB)	ABCD (SUB) + M (SUB)= 01	DW(SUB) + M (SUB)= 02	EML (SUB) + M (SUB) = 06	GBIF (SUB) + M(SUB) = 06	---
P (ANY) + M (SUB)	ABCD (ANY) + M (SUB)= 25	DW(ANY) + M (SUB)= 53	EML (ANY) + M (SUB) = 75	GBIF (ANY) + M (SUB) = 52	---
P (SUB) + M (ANY)	ABCD (SUB) + M (ANY)= 07	DW(SUB) + M (ANY)= 13	EML (SUB) + M (ANY) = 09	GBIF(SUB) + M (ANY) = 16	---
P (ANY) + B e M(ANY)	ABCD (ANY) + B e M (ANY)= 61	DW(ANY) + B e M (ANY)= 300	EML (ANY) + B e M (ANY) = 118	GBIF (ANY) + B e M (ANY) = 500	---
P (ANY) + B e M (SUB)	ABCD (ANY) + B e M (SUB)= 03	DW(ANY) + B e M (SUB)= 08	EML(ANY) + B e M (SUB) =06	GBIF(ANY) + B e M (SUB) =22	---
P (SUB) + B e M (ANY)	ABCD (SUB) + B e M (ANY)= 03	DW (SUB) + B e M (ANY)= 12	EML(SUB) + B e M (ANY) =1	GBIF (SUB) + B e M (ANY) =15	---
P (SUB) + B e M (SUB)	ABCD (SUB) + B e M (SUB) = 0	DW (SUB) + B e M (SUB) = 0	EML (SUB) + B e M (SUB) = 1	GBIF (SUB) + B e M (SUB) = 04	---
P e B (ANY) + M(ANY)	ABCD e B (ANY) + M (ANY)= 64	DW e B (ANY) + M (ANY) =300	EML e B (ANY) + M (ANY) =118	GBIF e B (ANY) + M (ANY) =500	---
P e B (ANY) + M(SUB)	ABCD e B (ANY) + M (SUB)= 12	DW e B (ANY) + M (SUB) =35	EML e B (ANY) + M (SUB) = 36	GBIF e B (ANY) + M (SUB) = 51	---
P e B (SUB) + M (SUB)	ABCD e B (SUB) + M (SUB) = 0	DW e B (SUB) + M (SUB) =0	EML e B (SUB) + M (SUB) =01	GBIF e B (SUB) + M (SUB) =04	---
P e B (SUB) + M (ANY)	ABCD e B (SUB) + M (ANY)= 03	DW e B (SUB) + M (ANY) =05	EML e B (SUB) + M (ANY) = 01	GBIF e B (SUB) + M (ANY) = 14	---
B (ANY) + M (ANY)					B (ANY) + M (ANY) = 4903
B (ANY) + M (SUB)					B (ANY) + M (SUB) =246
B (SUB) + M (ANY)					B (SUB) + M (ANY) =971
B (SUB) + M (SUB)					B (SUB) + M (SUB) = 68
"B e M" (ANY) 1 campo					"B e M" (ANY) = 15
"B e M" (SUB) 1 campo					"B e M" (SUB) = 03
B e M (ANY) 2 campos					B e M (ANY) = 4903
B e M (SUB) 2 campos					B e M (SUB) = 68

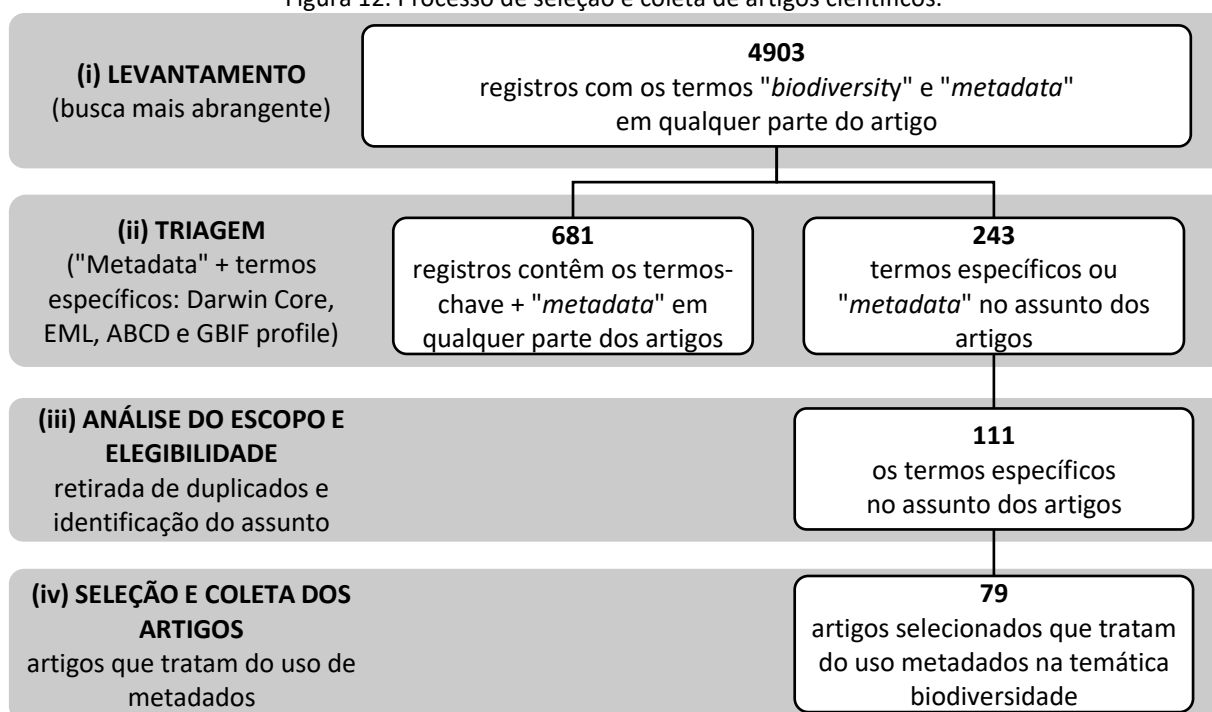
Consultas com resultados úteis.

Consultas com a maior parte dos resultados fora do escopo da pesquisa e resultados úteis já inclusos nas demais consultas.

B = *biodiversity*. **M** = *metadata*. **B e M** = "*metadata biodiversity*" (expressão em um único campo de pesquisa). **P** = indica uso de um dos padrões de metadados na busca (*Darwin Core*, *EML*, *ABCD* ou *GBIF profile*). **"e"** = indica uso de dois termos em um único campo de pesquisa. **ANY** = busca em qualquer parte do artigo. **SUB** = busca no assunto do artigo. + = indica o uso de dois campos de pesquisa associados. Fonte: Elaboração própria.

A partir de diversas correlações de termos nas pesquisas no Portal da Capes – quadro 21 – foi possível identificar os artigos científicos que tratam do uso de metadados em dados de pesquisas relacionadas à biodiversidade. Considerando a adaptação do modelo *berrypicking* (BATES, 1989), o processo de seleção dos artigos científicos para coleta de dados partiu de buscas com termos abrangentes até mais específicos, de forma a alcançar o escopo desejado neste estudo. Portanto, esse processo foi realizado em quatro fases principais: (i) levantamento; (ii) triagem; (iii) análise de escopo e elegibilidade; e (iv) seleção e coleta, conforme sintetizado na figura 12.

Figura 12. Processo de seleção e coleta de artigos científicos.



Fonte: Elaboração própria.

Inicialmente, na **fase de levantamento**, as pesquisas utilizaram os termos mais abrangentes, *metadata* e *biodiversity* em campos distintos e buscados em qualquer parte dos artigos. Com isso foram identificadas 4903 publicações – data do levantamento 21 de dezembro de 2019. Apesar do expressivo volume desse universo de artigos, analisando os assuntos correlacionados, verificou-se que mais de 80% apenas mencionavam os termos buscados, mas não tratavam efetivamente do uso de metadados para biodiversidade.

Os resultados mais efetivos foram obtidos na **fase de triagem** utilizando termos mais específicos – Darwin Core, EML, ABCD, GBIF profile. Nessa segunda fase resultaram 681

artigos, entretanto, muitas das pesquisas apenas mencionavam os termos buscados e tratavam de assuntos divergentes – e.g. métodos de amostragem, modelos preditivos de análise da biodiversidade ou equipamentos e técnicas de coleta de fauna e flora. Constatou-se também que apenas o termo *metadata* buscado no assunto não era efetivo para identificar a literatura científica dentro do escopo desejado, mesmo quando correlacionado aos termos mais específicos buscados em qualquer parte do artigo. Por isso, ainda na fase de triagem ao restringir as buscas desses termos específicos ao assunto dos artigos foram obtidos 243 registros, mas muitas das correlações de termos utilizadas apontaram para publicações semelhantes, acarretando em duplicidade de registros.

Considerando o resultado obtido na triagem, foi iniciada a **fase de análise do escopo e elegibilidade** dos artigos. Nessa terceira fase, foram retirados os registros duplicados e os artigos relacionados às temáticas relacionadas à biomedicina – e.g. genoma, genética e assuntos correlatos – pois se referem a outro grupo de padrões de metadados mais específicos que não fazem parte do objeto deste estudo. Isso resultou em uma lista de 111 artigos considerados elegíveis – APÊNDICE G.

Identificou-se que os registros retornados das buscas com os termos abrangentes e pesquisados no campo assunto dos artigos – *biodiversity* e *metadata* em campos distintos (68 artigos) e com a expressão “*biodiversity metadata*” (15 artigos) – estão englobados na lista de artigos elegíveis selecionados para análise. Isso também comprova a relevância das buscas realizadas e a aderência dos artigos ao escopo deste estudo.

Por fim, a partir da lista de 111 registros considerados elegíveis, realizou-se a **fase seleção e coleta dos artigos**. Nessa etapa os artigos foram lidos em sua completude, visando extrair o assunto principal e verificar a aderência ao objeto deste estudo. A partir dessa análise foram descartados 32 artigos, por apresentarem assunto não aderente ao escopo deste estudo – a lista dos artigos excluídos consta no APÊNDICE H. Com isso, foram **selecionados 79 artigos científicos** para a coleta de dados sobre uso de metadados em dados de pesquisa da biodiversidade – APÊNDICE I.

Pesquisas adicionais foram realizadas considerando a expressão “uso de metadados”, com a seguinte estruturação: 1) expressão “*metadata use*”, associada ao termo “*biodiversity*” e à expressão “*biodiversity data*”; 2) a expressão “*metadata standard use*” associada ao termo “*biodiversity*”. As expressões “*metadata use*” e “*metadata standard use*”

buscadas em qualquer parte dos artigos; e o termo “*biodiversity*” ou a expressão “*biodiversity data*” buscadas no assunto. Ao restringir “*metadata use*” ao campo assunto, a pesquisa não retornou nenhum artigo. Resultaram dessas buscas **sete novos artigos** - ANEXO 7.

Dessa forma, somados esses novos artigos aos demais já descritos, este estudo selecionou o total de 86 artigos para coleta e análise dos dados sobre o uso de metadados nas pesquisas sobre biodiversidade.

4.2.2 ANÁLISE DOS DADOS COLETADOS NOS ARTIGOS CIENTÍFICOS

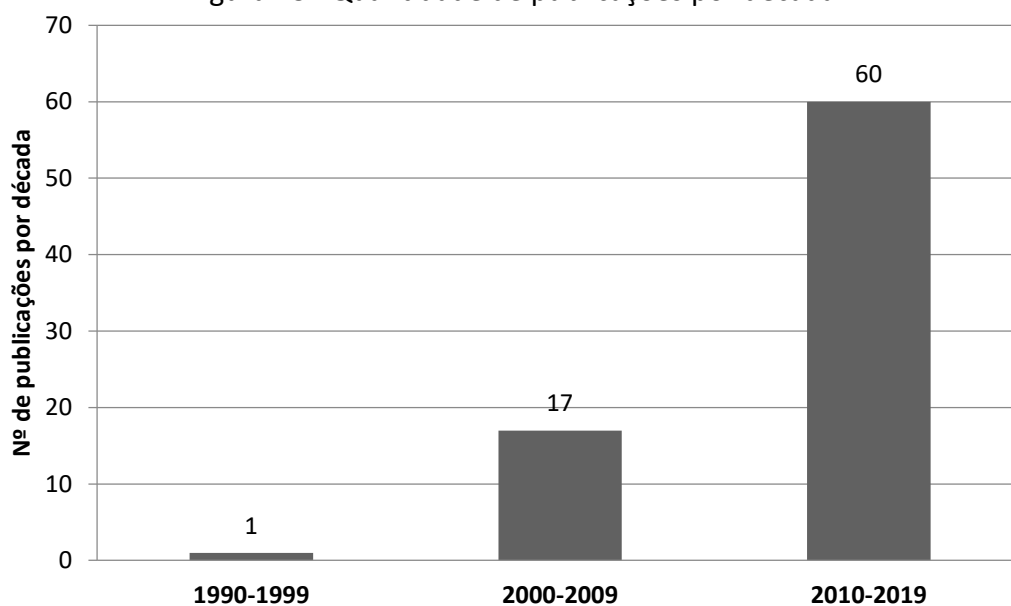
Os 86 artigos selecionados foram submetidos à coleta e análise de dados utilizando o roteiro de estudo apresentado no APÊNDICE A. Após esse procedimento foram excluídos 8 artigos – 6 por tratar de *data papers*; 1 por não ser artigo científico e 1 por não tratar do uso de metadados.

Conforme metodologia proposta, a última etapa deste estudo abrange a coleta direta dos conjuntos de dados e metadados para análise. Os *data papers* são documentos que descrevem conjuntos de dados. Dessa forma, a exclusão das publicações desse tipo se deu para não haver coleta e análise redundante de conjuntos de dados. A lista dos 8 artigos excluídos consta no APÊNDICE J.

Os resultados apresentados a seguir resultam, portanto, da análise dos dados coletados a partir dos 78 artigos listados no ANEXO 8. O registro dos dados coletados de cada um desses artigos é apresentado em quatro partes: APÊNDICE C; APÊNDICE D; APÊNDICE E; APÊNDICE F.

Os artigos coletados distribuem-se entre os anos de 1999 e 2019. A análise do contexto histórico considerando o ano de publicação permite compreender a evolução das pesquisas relacionadas ao uso de metadados na temática biodiversidade. Dessa forma, optou-se por apresentar fatos históricos relevantes e contextualizar os artigos dividindo a análise por décadas: 1990-1999; 2000-2009; e 2010 a 2019 – figura 13.

Figura 13. Quantidade de publicações por década



Fonte: Elaboração própria.

Apresenta-se a seguir a análise detalhada dos artigos relacionada ao contexto histórico de cada uma das três décadas. Para concluir essa análise, ao final é apresentada a síntese integrativa dos resultados obtidos considerando os três períodos – subseção 4.2.2.4.

4.2.2.1 1990 - 1999

Marcos referentes à gestão de dados de pesquisa – 1990 a 1999

O final da década anterior e a década de 1990 são períodos marcados por avanços na gestão de dados de pesquisa e na comunicação científica, em especial na área ligada a dados geográficos. A temática biodiversidade tem seus dados intrinsecamente relacionados às referências de dados de localização geográfica, por isso, o desenvolvimento da gestão de dados nessa temática decorre diretamente do avanço das discussões relativas às chamadas Ciências da Terra e Geociências.

Em 1984, em resposta a recomendações do grupo de trabalho sobre Crescimento, Tecnologia e Emprego da Cúpula Econômica das Nações Industriais (G7), foi criado o Comitê de Observação da Terra por Satélites – *Committee on Earth Observation Satellites* (CEOS) – (CEOS, 2020) – voltado à coordenação dos esforços internacionais de observação da Terra em benefício da sociedade. Em 1987, foi realizado pelo CEOS o Workshop de Sistemas de Dados de Ciências da Terra – *Earth Science and Applications Data Systems Workshop*

(ESADS). Nesse evento o Grupo de Trabalho de Interoperabilidade de Catálogos (composto por várias agências federais e internacionais dos EUA) apresentou o protótipo do diretório da NASA – *NASA's Master Directory* (NMD) – que utiliza o *Directory Interchange Format* (DIF). A partir de 1988 o padrão DIF passou a ser adotado formalmente pelo CEOS.

Em 1990, o grupo de trabalho interagências de dados referentes à mudança global – *Interagency Working Group on Data Management for Global Change* (IWGDMGC) – adotou como diretório o *Global Change Master Directory* (GCMD), antes chamado NMD. Desde então o padrão DIF tornou-se referência e dele decorreram diversas iniciativas para organização de dados geográficos e dados temáticos relacionados a dados de localização.

O DIF não concorre com outros padrões de metadados, pois é simplesmente o "contêiner" para os elementos de metadados que são mantidos no banco de dados da rede criada pelo CEOS – *International Directory Network* (IDN). O DIF é usado para criar entradas de diretório que descrevem um grupo de dados, portanto, consiste em uma coleção de campos que detalham informações específicas sobre os dados voltados a permitir a validação de campos obrigatórios, identificar palavras-chave, responsáveis e outros.

Considerando os aspectos relacionados à organização da informação, essa década também foi permeada por marcos regulatórios voltados à criação das entidades responsáveis pela gestão de dados, com destaques neste estudo aos dados geográficos (de forma simples, são os dados relacionados a uma localização). Também nesse período surgiram iniciativas voltadas a fomentar o compartilhamento de dados de pesquisa – *Clearing-house Mechanisms*.

Originalmente o termo *clearing-house* se referia a um estabelecimento financeiro no qual eram trocados cheques e notas promissórias, mas o significado foi aos poucos estendido para incluir qualquer agência que reúna buscadores e fornecedores de bens, serviços ou informações, combinando, assim, demanda e oferta (BCH, 2016). O termo *Clearing-house Mechanisms* (CHM) foi estendido em significado para referir-se a mecanismos de compartilhamento de informação (não apenas financeira) por meio de atores independentes visando benefícios a todas as partes interessadas (CRAIN, 1994, p. 5; EUA. OMB. FGDC., 2010). Na temática biodiversidade, consolidou-se como termo relacionado a facilitar a cooperação técnica e científica por meio de infraestrutura para

depósito e livre compartilhamento de informações sobre recursos biológicos, incluindo coleções de dados, pesquisas e tecnologia (CRAIN, 1994, p. 5).

Em 1990 os Estados Unidos criaram o Comitê Federal de Dados Geográficos – *Federal Geographic Data Committee* (FGDC) – e a Infraestrutura Nacional de Dados Espaciais – *National Spatial Data Infrastructure* (NSDI), publicando também norma com orientações direcionadas às agências federais responsáveis por criar, manter ou usar dados espaciais direta ou indiretamente (EUA. FGDC., 2020; EUA. OMB. FGDC., 2010). A partir dessas iniciativas, surgiram programas de incentivo ao desenvolvimento do *National Spatial Data Clearinghouse*, que se refere ao serviço eletrônico de acesso a dados espaciais documentados e metadados de fontes de dados distribuídas com abrangência em diversas temáticas (EUA. OMB. FGDC., 2010). E em 1992 a FDGC (2020) criou a *Geospatial Clearinghouse Network Initiative*, uma rede entre as agências federais dos EUA voltada a minimizar a duplicação de esforços na coleta de dados espaciais digitais e promover atividades cooperativas de coleta de dados digitais.

Em 1993 o Conselho Nacional de Pesquisa dos EUA – *National Research Council* – publicou o relatório "*Biological Survey for the Nation*" que recomendou o desenvolvimento de um sistema nacional de informações sobre recursos da biodiversidade, considerado vital para uso de diversos campos científicos, educacionais e governamentais (EUA. NATIONAL RESEARCH COUNCIL, 1994). Decorrente dessa recomendação, em 1994 foi criada a *National Biological Information Infrastructure* (NBII) – *NBII Clearinghouse*. Em 1999 foi publicada a documentação do perfil de metadados associado ao sistema NBII – *Content Standard for Geospatial Metadata and the Biological Data Profile* (EUA. FGDC. USGS, 1999). Ainda nessa década também surgiram importantes padrões de metadados, o Dublin Core em 1995 (DCMI, 2012) e a EML em 1997 (KNB, 2019).

Outra linha de eventos relevantes na década de 1990 refere-se ao surgimento da web (*world wild web*). A web facilitou e tornou mais amigável a identificação de recursos na internet (rede mundial de computadores) por meio do uso de hipertexto e técnicas multimídia. Também se destaca a criação do *World Wide Web Consortium* (W3C), consórcio internacional voltado à criação de padrões para a web. Esse conjunto de fatos históricos revolucionou a visão sobre o potencial de soluções que poderiam ser criadas para aperfeiçoar a gestão de informações.

Marcos referentes à temática biodiversidade – 1990 a 1999

O movimento ambientalista ganhou novo impulso em 1962 com a publicação do livro de Rachel Carson, “A Primavera Silenciosa”, que fez um alerta sobre o uso agrícola de pesticidas químicos sintéticos (ONU, 2020). A preocupação universal sobre o uso saudável e sustentável do planeta e de seus recursos continuou a crescer. Em 1972, a ONU convocou a Conferência das Nações Unidas sobre o Ambiente Humano, na cidade de Estocolmo na Suécia, que resultou em Manifesto Ambiental com 19 princípios.

A história é permeada por fatos relevantes em que a humanidade precisou enfrentar desastres ambientais, bem como crises sanitárias e de saúde de grande dimensão. Esses eventos, sejam eles causados por fenômenos naturais ou por ações antrópicas, tornaram a discussão global das temáticas ambientais ainda mais relevante. Alguns exemplos de eventos de grande impacto ambiental são apresentados a seguir.

Em 1986, o desastre de Chernobyl devido a um acidente nuclear ocorrido no reator nuclear da Usina Nuclear de Chernobyl no norte da Ucrânia. Em 1987, na cidade de Goiânia no Brasil, houve a exposição indevida de pessoas ao Césio 137, decorrente do descarte inadequado de resíduos radioativos do lixo hospitalar. Também existem diversos registros de derramamentos de petróleo, como o desastre de Cubatão no Brasil em 1984 ou no final da década de 1980, o ocorrido com o navio Exxon Valdez, ambos os eventos com desastres ambientais de grandes proporções e que ganharam atenção mundial. Em relação a desastres naturais, há temporadas de furacões e enchentes anuais que causam muitas fatalidades e prejuízos, bem como episódios de terremotos e maremotos.

Esses desastres ambientais enfrentados pela humanidade conduziram ao fortalecimento da cooperação global. Em 1987, o Relatório de Brundtland, também conhecido como “Nosso Futuro Comum”, apresenta o conceito de desenvolvimento sustentável para o discurso público. (ONU, 2020). Em 1988, o Programa das Nações Unidas para o Meio Ambiente (PNUMA) criado pela ONU e a Organização Meteorológica Mundial (OMM), se uniram para criar o Painel Intergovernamental para as Mudanças Climáticas (IPCC), que se tornou a fonte proeminente para a informação científica relacionada às mudanças climáticas. Em 1992 foi estabelecida a Convenção sobre Diversidade Biológica (CDB, 1992). Considerada o documento norteador sobre a biodiversidade, a CDB consiste em

um tratado da Organização das Nações Unidas decorrente da conferência mundial sobre meio ambiente realizada no Rio de Janeiro naquele ano – também conhecida como ECO-92. Mais de 160 países são signatários.

A CDB (1992), entre diversas outras recomendações, estabelece a necessidade de desenvolver mecanismos de compartilhamento – *Clearing-house Mechanisms* (CHM) – para promover e facilitar a cooperação técnica e científica. Em relação a mudanças climáticas, em 1992 foi adotado o principal instrumento internacional, a Convenção-Quadro das Nações Unidas sobre Mudanças Climáticas – *United Nations Framework Convention on Climate Change* (UNFCCC). Essa convenção criou o Painel Intergovernamental sobre Mudanças Climáticas (IPCC) que reúne cientistas independentes de todo o mundo, incluindo pesquisadores brasileiros. Em 1997, foi estabelecido o Protocolo de Kyoto, com metas obrigatórias para 37 países industrializados e para a comunidade europeia referente à redução das emissões de gases de efeito estufa.

Também em 1992, a OCDE realizou o primeiro Fórum de Megaciência¹⁸, voltado a fornecer espaço para diálogos intergovernamentais entre cientistas e formuladores de políticas e fortalecer a cooperação científica em projetos de grande complexidade e amplo impacto. Na edição realizada em 1996, foi criado o Grupo de Trabalho de Biologia Computacional (GTBC/OECD) com base nas diretrizes estabelecidas pela CDB.

Entre os diversos objetivos traçados para o grupo GTBC/OECD destacam-se: (i) promover a cooperação internacional no desenvolvimento e implantação de bancos de dados interoperáveis e outros recursos de informática relacionados à diversidade biológica; (ii) promover o rápido desenvolvimento e distribuição geral de ferramentas de informática para o campo de sistemática da biologia; (iii) identificar as questões de direitos de propriedade intelectual relacionadas aos bancos de dados de biodiversidade, bem como as condições necessárias para manter uma acessibilidade aberta a esses recursos universais (OECD, 1999, p. 5).

O GTBC/OECD publicou relatório contendo recomendações referentes à integração dos dados sobre biodiversidade (OECD, 1999). Nesse documento, destaca-se a indicação para criação de um mecanismo internacional capaz de tornar os dados e informações sobre biodiversidade acessíveis em todo o mundo para produzir benefícios econômicos e sociais

¹⁸ Conhecido como Fórum Global de Ciências a partir de 1999.

visando o desenvolvimento sustentável (OECD, 1999, p. 2). Especificamente, foi recomendada a criação da rede GBIF (OECD, 1999, p. 14, griffo nosso):

recomenda que os governos dos países membros da OCDE estabeleçam e apoiem um sistema distribuído de módulos interligados e interoperáveis (bancos de dados, software e ferramentas de rede, mecanismos de pesquisa, algoritmos analíticos etc.) que juntos formarão um Sistema Global de Informações sobre Biodiversidade – **Global Biodiversity Information Facility (GBIF)**.

Após endossada pela OCDE em 1999, foi estabelecida formalmente em 2001 a *Global Biodiversity Information Facility – GBIF* como uma rede de colaboração internacional, cuja infraestrutura de pesquisa é financiada por governos de todo o mundo com o objetivo de “dar a qualquer pessoa, em qualquer lugar, acesso aberto aos dados sobre toda a vida na Terra” (GBIF, 2019g). A OCDE (2010) também publicou um Memorando de Entendimento entre os governos participantes que estabelece as diretrizes para adesão aberta à rede GBIF (GBIF, 2019b).

Análise dos artigos – 1990 a 1999

Apesar dos marcos histórico apontados nesse período, identificou-se apenas um artigo relacionado ao uso de metadados, que foi publicado no final dessa década. Isso provavelmente se deve ao fato de que as iniciativas começaram a se estruturar no âmbito governamental a partir de tratados internacionais e seus efeitos repercutiram nas pesquisas na década seguinte.

O artigo identificado para essa década trata das iniciativas ocorridas na China para criação de CHM como ações visando adequação à CDB. Os autores apresentam ações desenvolvidas na CHINA para identificar as bases de dados de biodiversidade existentes por meio do mapeamento dos metadados e realizar levantamento de desafios e necessidades visando à integração de dados e o desenvolvimento de um CHM para o compartilhamento (HAIGEN XU, ZHENNING GAO, *et al.*, 1999).

4.2.2.2 2000 - 2009

Marcos referentes à gestão de dados de pesquisa – 2000 a 2009

Essa década é marcada pela discussão sobre o acesso livre aos dados de pesquisa. Entretanto, muito antes de ser um objeto de discussão técnica ou movimento, a ideia de

dados abertos estava enraizada na comunidade científica, que foi a primeira a perceber o benefício da abertura e do compartilhamento de dados (CHIGNARD, 2013).

Essa ideia foi popularizada por Robert King Merton no início da década de 1940, indicando que a pesquisa (que produz dados) deve ser compartilhada livremente para o bem comum (CHIGNARD, 2013). Em 1966, foi promulgada a Lei de Liberdade de Informação dos EUA – *Freedom of Information Act* (FOIA) – que representou um marco sobre o direito ao acesso público aos resultados da pesquisa. Já no século XXI, Elinor Ostrom mostrou em sua pesquisa a especificidade dos bens comuns da informação, explicando que são muito semelhantes aos bens públicos, pois seu uso por uma pessoa não impede o uso por outras (OSTROM e HESS, 2007). A pesquisadora ressaltou, no entanto, que são bens públicos de um novo tipo: pois seu uso, além de não esgotar o estoque comum, também o enriquece.

Em 2000, o periódico científico *Journal of Cognitive Neuroscience* passou a exigir que dados primários dos artigos publicados na revista fossem disponibilizados de forma pública e sem custo, isso fez emergir um debate considerável sobre o tema na comunidade científica (CHULIN MENG, 2004). Em 2004, a OECD promoveu a discussão com instituições produtoras de dados dos países membros sobre o tema. Em 2007 a OECD publicou o documento “Princípios e Diretrizes da OCDE para Acesso a Dados de Pesquisa a partir de Financiamento Público”, recomendando que todos os dados resultantes de projetos e pesquisas financiados com recursos públicos deveriam ser disponibilizados de forma aberta e sem custos.

Diversas iniciativas e discussões sobre dados abertos permearam o período entre 2000 e 2009. De forma geral, surgiu um movimento que considera os dados de pesquisa uma propriedade comum e os meios para viabilizar isso passam, principalmente, pelo compartilhamento e uso desse bem comum construído sobre três conceitos: abertura, participação e colaboração (CHIGNARD, 2013).

Nesse cenário, surgiram entidades internacionais temáticas voltadas a criar diretrizes e padrões de dados abertos. Exemplo disso na área de pesquisa relacionada aos dados espaciais é o *Open Geospatial Consortium* (OGC), consórcio internacional que utiliza processos de consenso entre os membros para criar padrões geoespaciais livres de royalties e publicamente disponíveis desde 2005. Entre as temáticas relacionadas a dados geográficos, a OGC também trata do tema biodiversidade.

Além disso, apesar do termo web semântica ter sido apresentado por Berners-Lee, o criador da web, na primeira *World Wide Web Conference* realizada em 1994, só foi mais bem explicado em 2001, em artigo publicado na revista *Scientific American* intitulado *The Semantic Web - A new form of Web content that is meaningful to computers will unleash a revolution of new possibilities* – Web Semântica: um novo formato de conteúdo para a Web que tem significado para computadores vai iniciar uma revolução de novas possibilidades (LAUFER, 2015, p. 33). Isso estimulou o aumento de pesquisas relacionadas ao desenvolvimento de ontologias.

Considerando iniciativas sobre o uso de metadados, também é marcante nesse período a publicação de normas internacionais voltadas à organização de dados. Em 2001 o padrão ABCD começou a ser desenvolvido e foi ratificado pela TDWG em 2005. Além disso, uma extensão do Dublin Core para a biodiversidade, o Darwin Core surgiu em uma iniciativa apoiada pelo Centro Nacional de Pesquisa dos EUA – *National Science Foundation* – iniciada em 1998, que foi revisada e ratificada pelo TDWG em 2009. Destacam-se ainda duas normas ISO publicadas em 2003, a primeira é a ISO 19115:2003 *Geographic Information: Metadata*, que define o esquema necessário para descrever informações e serviços geográficos. A segunda é a ISO 15836:2003, que apresenta a documentação do padrão Dublin Core (DCMI, 2012).

Conforme apontado, desde o início da década de 1990 surgiu nos EUA o desenvolvimento das chamadas Infraestruturas de Dados Espaciais (IDEs), que desencadeou diversas iniciativas referentes à organização de dados relacionados às referências de localização. Em 2007 começou a ser desenvolvida a infraestrutura europeia de dados espaciais – *Infrastructure for Spatial Information in the European* (INSPIRE) e em 2008, o Brasil instituiu a Infraestrutura Nacional de Dados Espaciais (INDE).

Marcos referentes à temática biodiversidade – 2000 a 2009

Os avanços ocorridos em termos de normatização na área de ciências da informação e na área ambiental na década anterior começaram a refletir mudanças nas organizações que lidam com pesquisa e dados sobre biodiversidade a partir de 2000. Houve crescimento dos esforços relacionados ao monitoramento ambiental em busca de responder a

numerosos e complexos problemas associados a temas como: sustentabilidade ambiental, mudanças climáticas, biossegurança e uso da terra (HALE e HOLLISTER, 2009).

Dez anos depois da conferência que gerou a CDB, em 2002, foi realizada nova Conferência das Nações Unidas sobre Desenvolvimento Sustentável (Rio+10) em Johannesburgo na África do Sul, que gerou a “Declaração de Johannesburgo”, reafirmando compromissos da CDB e recomendando iniciativas para redução da perda de biodiversidade.

Surgiram também outros protocolos, programas e projetos decorrentes dessas e outras conferências e reuniões globais. Em 2003 entrou em vigor o Protocolo de Cartagena sobre Biossegurança, tratado internacional que rege questões relacionadas a organismos vivos modificados (OVMs). Esse documento estabeleceu a infraestrutura voltada a facilitar o intercâmbio de informações sobre organismos vivos modificados e auxiliar os países na implantação do protocolo – *Biosafety Clearing-House*. Também em 2003, a descoberta do vírus Ebola é um evento que merece destaque. Por representar um potencial risco de pandemia conduziu à cooperação internacional e investimentos em pesquisas para unir conhecimento de epidemiologia, genética e modos de transmissão, bem como seus contextos ecológicos (JONES, SCHILDHAUER, *et al.*, 2006, p. 520).

Em 2004, o então Ministério da Ciência e Tecnologia, atual Ministério da Ciência Tecnologia e Inovação (MCTI), criou o Programa de Pesquisa em Biodiversidade (PPBio) voltado a “desenvolver estratégias de investimentos que agreguem as diversas competências em pesquisa e transferência de conhecimento sobre biodiversidade, gerando, integrando e disseminando informações que possam ser utilizadas pela sociedade” (BRASIL. MCTIC. PPBIO, 2016, p. 17).

Em 2005, a ONU formalizou a Estratégia de Maurício (ONU, 2020), que aborda questões como mudanças climáticas e a elevação do nível do mar; desastres naturais e ambientais; gestão de resíduos; produção e consumo sustentável; saúde; gestão do conhecimento e da informação para tomada de decisão e outras. A ONU declarou o período entre 2005 e 2014 como a Década das Nações Unidas da Educação para o Desenvolvimento Sustentável. Nesse período a ONU priorizou ajudar as populações a desenvolverem atitudes, habilidades e conhecimento para tomarem decisões sobre sustentabilidade (ONU, 2020).

Assim, a partir de diferentes demandas das comunidades científicas e governamentais surgiram projetos voltados à gestão digital dos dados de pesquisa sobre

biodiversidade. Uma grande variedade de conjuntos de dados produzidos por pesquisadores individuais passou a ser sintetizada para abordar questões ecológicas abrangendo uma grande variedade de escalas espaciais e temporais (ELLISON, OSTERWEIL, *et al.*, 2006). Houve o crescimento das iniciativas voltadas aperfeiçoar a catalogação, digitalização, estruturação e compartilhamento das informações sobre os recursos da biodiversidade, conforme apontam os resultados da análise meta-sintética dos artigos a seguir.

Análise dos artigos – 2000 a 2009

No período de 2000 a 2009 houve o aumento das publicações relacionadas ao uso de metadados na temática biodiversidade em comparação com a década anterior. A revisão sistematizada permitiu identificar 17 artigos científicos publicados nesse período que fazem referência ao uso de metadados associado aos seguintes assuntos: **(i) proposta de padrão de metadados** – 1 artigo; **(ii) revisão de padrões de metadados existentes** – 2 artigos; **(iii) soluções computacionais (descobrir, integrar, compartilhar e reutilizar coleções de dados digitais)** – 12 artigos; e **(iv) ontologia** – 2 artigos. A análise meta-sintética dos artigos referentes a esse período é apresentada a seguir.

Referente ao primeiro assunto identificado, **proposta de padrão de metadado (i)**, nessa década há apenas um artigo que apresenta o padrão de metadados para biodiversidade da China (HAIGEN XU, YIMIN LI e HUI DING, 2004). O desenvolvimento desse padrão emergiu da demanda por desenvolver um CHM decorrente das recomendações estabelecidas na CDB. O padrão chinês tem como principais referências o perfil de metadados NBII da FGDC e o padrão *Directory Interchange Format* (DIF) desenvolvido pela *National Aeronautics and Space Administration* (NASA), ambos de entidades dos EUA.

Há dois artigos associados **revisão dos padrões de metadados mais utilizados nesse período (ii)**. O primeiro publicado em 2003 apresenta a discussão dos padrões de metadados mais utilizados para catalogação de dados de história natural e destaca o uso do formato MARC para descrever campos de metadados taxonômicos e geoespaciais (HAASA, HENJUMB, *et al.*, 2003). Nesse estudo, os autores destacam o surgimento dos esquemas de metadados que permitiram o compartilhamento de informações complexas e evidenciaram a necessidade de esforços para promover a interoperabilidade, diante do desafio de integrar dados heterogêneos, descritos em diferentes padrões e em diferentes formatos. Segundo

Haasa e demais autores (2003), para a descrição dos dados de história natural eram utilizados: os esquemas de metadados incluíam o MARC e o Dublin Core no âmbito das bibliotecas; a EML desenvolvida para a disciplina de ecologia; o esquema da Infraestrutura Nacional de Informação Biológica (NBII) criado como um perfil geográfico para biodiversidade pela FGDC; o Darwin Core adotado pelos museus na catalogação de coleções. Além disso, o estudo aponta que a NSF iniciava a construção da *National Science Digital Library* considerando os mecanismos para descoberta de metadados em desenvolvimento pela *Open Archives Initiative (OAI)* e *IEEE Learning Object Metadata (LOM)*.

A segunda revisão padrões de metadados usados para dados da biodiversidade nesse período foi publicada em 2004. Chulin Meng (2004) inicia o artigo apontando os padrões decorrentes de iniciativas da Ciência da Informação, tais como *Anglo-American Cataloguing Rules*, MARC, Dublin core e Z39.50 (CHULIN MENG, 2004). Em seguida o autor descreve as iniciativas mais próximas aos museus de história natural, como o guia de boas práticas do *Consortium for the Computer Interchange of Museum Information (CIMI)*, que apresenta recomendações de como utilizar o Dublin Core para compartilhar dados das coleções de história natural entre museus. Por fim, o autor aponta a evolução em complexidade e detalhamento, bem como os potenciais ganhos semânticos com padrões mais específicos para a biodiversidade, o ABCD, EML e o Darwin Core.

A maior parte dos artigos selecionados nesse período de 2000 a 2009 trata do uso de metadados associado ao assunto **soluções computacionais desenvolvidas para integrar, compartilhar e reutilizar dados (ii)**. A primeira publicação referente a esse assunto foi publicada já no início da década de 2000 e apresenta a primeira versão do CHM de biodiversidade da China (HAIGEN XU, DEHUI WANG e XUEFENG SUN, 2000). Os autores também relatam os principais desafios enfrentados, tais como financiamento insuficiente, inexistência de entidades responsáveis e informações de baixa qualidade.

Em 2003, outro estudo destaca o papel relevante da estrutura dos dados geográficos para o desenvolvimento de CHM na Finlândia em atendimento à CDB. Nesse estudo, Laihonon, Ronka e demais autores (2003) alertam sobre o fato de que muitos dados de biodiversidade acabam sendo descartados por não apresentar registro de localização associados. Além disso, os autores identificam quatro categorias principais de recursos de informação da biodiversidade: (1) coleções biológicas de museus de herbários e história

natural, (2) dados descritivos da biodiversidade (por exemplo, medições ou observações) coletados por organizações ou pessoas individuais, (3) informações sintéticas (por exemplo, informações científicas publicações) e (4) metadados (informações descritivas nas categorias acima).

Schucka e colaboradores (2005) descrevem em seu artigo o sistema de informação florestal europeu – *European Forest Information System (EFIS)* – também estabelecido em atendimento à CDB e desenvolvido orientado pelo padrão Dublin Core, padrões OGC, recomendações do W3C e normativas que instituem a infraestrutura europeia de dados espaciais – *Infrastructure for Spatial Information in Europe (INSPIRE)*. O estudo também cita outras iniciativas que emergiram na web voltados a identificar metadados relacionados aos recursos florestais e integrar diferentes catálogos existentes, tais como “*WWW Virtual Library for Forestry*”, “*BIOME/AgriFor*” e “*Global Forest Information Service (GFIS)*”.

Torres, Medeiro e demais autores (2006) apresentam uma experiência brasileira no desenvolvimento de dois *Biodiversity Information Systems (BISs)*, que se referem aos sistemas voltados a integrar grande volume de coleções de dados heterogêneos – ecológicos e geográficos. Os dois BISs apresentados no artigo foram desenvolvidos para agregar dados referentes a peixes e utilizaram técnicas avançadas aplicadas às bibliotecas digitais. Os autores afirmam que esses sistemas permitiram aperfeiçoar e facilitaram o processo de identificação de peixes. Em termos de arquitetura, para promover interoperabilidade entre as bases de dados, foi utilizado o protocolo *Open Archives Initiative (OAI)*, tanto para publicar dados e metadados estruturados, como para desenvolver os serviços de coleta e processamento de metadados. Mas a principal inovação apontada está na associação do sistema de imagens aos BISs, que utilizou um *metadata-based search component (MBSC)*, o mecanismo de pesquisa vetorial ESSEX, associado a um *content-based image search component (CBISC)* desenvolvido pelos autores.

Ellison, Ostrerwell e demais autores (2006) apontam que a documentação necessária para garantir a confiabilidade e a validação dos conjuntos de dados inclui tanto os metadados descritivos como a documentação formal dos processos científicos usados (isto é, metadados do processo) para tratar os dados brutos. Segundo os autores, essa documentação é complexa e difícil de construir, portanto, é importante ajudar os “produtores” (pesquisadores que coletam e editam os dados) a criar conjuntos de dados

confiáveis e facilitar a criação dos metadados necessários (ELLISON, OSTERWEIL, *et al.*, 2006). Dessa forma, os autores apresentam o software SciWalker e como essa ferramenta foi utilizada em processos científicos para criar redes analíticas e realizar o processamento dos conjuntos de dados brutos e derivados, criando trilha de auditoria e servindo de suporte à síntese e à validação. Como estudo de caso, o artigo apresenta a síntese de medições de longo prazo da troca de carbono em um ecossistema. Por fim, os autores destacam que os metadados do processo criados pelo SciWalker são facilmente adaptados para inclusão em arquivos no padrão EML.

Kwang-Tsao Shao e colaboradores (2007) apresentam o as ações para integração das bases de dados de biodiversidade de mais de 10 instituições e museus em Taiwan em parceria com a rede GBIF, utilizando o padrão Darwin Core e técnicas de gestão de dados espaciais. Wong, Bloom e demais (2007) descrevem o sistema criado pelo governo canadense para apoiar decisões relacionadas à proteção da vida silvestre – *WILDSPACE Decision Support System* (WILDSPACE DSS) – que integrou diversas informações sobre biodiversidade utilizando esquema de metadados de sistemas pré-existentes e seguindo recomendações da OGC para tratamento de dados geográficos. Outra experiência canadense refere-se ao desenvolvimento do SFMN GeoSearch (GONZALES, CARDILLE, *et al.*, 2009), um aplicativo de compartilhamento e visualização de dados criado para a rede de manejo florestal sustentável – *Sustainable Forest Management Network* (SFMN). Esse aplicativo também apresenta esquema próprio de metadados e utiliza *Keyhole Markup Language*, uma linguagem baseada em XML empregada para expressar anotações geográficas e visualização espacial de conteúdos (GONZALES, CARDILLE, *et al.*, 2009).

Bertzky e Stoll-Kleemann (2009) apresentam o levantamento e análise de 27 soluções existentes na web que disponibilizam dados de monitoramento da biodiversidade de forma livre. Os autores analisam a possibilidade de uso desses dados para avaliações sobre a conservação de áreas protegidas. Destacam-se como contribuição relevante deste estudo os resultados decorrentes da análise das de dados disponíveis, bem como do universo de dados não publicados e dados inexistentes (lacunas), identificando as principais necessidades, possíveis soluções e iniciativas existentes, conforme apresentado a seguir - quadro 22.

Quadro 22. Dados de pesquisa sobre biodiversidade: necessidades, possíveis soluções e iniciativas existentes

NECESSIDADES	POSSÍVEIS SOLUÇÕES	INICIATIVAS EXISTENTES
Reduzir a inconsistência conceitual e a complexidade no tratamento de dados disponíveis	Padronização de protocolos de coleta de dados	protocolos IFRI e condições de coleta e manuseio de dados
	Padronização de abordagens para o desenvolvimento de sistemas de monitoramento (abordando apenas parcialmente o problema)	Aplicação de ferramentas simples existentes. Exemplos: BRIM, CAP e CMP's <i>Adaptive</i>
		Ferramenta de gerenciamento de <i>software</i>
	Facilitação de sistemas para manipulação de dados	Miradi
Promover a redução da sensibilidade de determinados dados ou manipular dados sensíveis com cuidado especial (dados não publicados)	Aumento da fiscalização e controle das atividades do mercado negro	Linguagem de Metadados Ecológicos (EML)
	Avaliação da sensibilidade e aprovação dos usuários de dados para evitar uso indevido -	Convenção sobre o Comércio Internacional das Espécies da Fauna e da Flora Silvestres Ameaçadas de Extinção (CITES)
	Promover uma " cultura de segurança " no manuseio e distribuição de dados	Princípios de troca de dados do <i>National Biodiversity Network Trust</i> (rede colaborativa do Reino Unido)
Promover iniciativas de acesso aberto e o compartilhamento de dados, que incluem a redução da competitividade científica e administrativa, bem como a relutância geral em relação aos processos de avaliação (dados não publicados e dados inexistentes)	Obrigação de os titulares fornecerem dados publicamente	Diretrizes da OCDE para a segurança de sistemas e redes de informação
		FOIA dos EUA
		Declarações de Budapeste, Bethesda e Berlim. Documentos publicados entre 2001 e 2003 sobre acesso aberto nas mais variadas áreas de conhecimento
	Garantia de direitos de propriedade intelectual e oferta de incentivos para o fornecimento de dados	Regulamentos de Informações Ambientais do Reino Unido
	Padronização dos procedimentos de avaliação como elementos essenciais no gerenciamento de áreas protegidas	licença <i>Creative Commons</i>
Promover atividades de capacitação financeira e / ou técnica para coleta, análise e uso de dados de monitoramento em áreas protegidas (dados inexistentes)	Exploração total de informações on-line existentes	sistemas de certificação
		Sistema padrão ISO de relatórios de suporte a decisão
		PALnet
	Levantar fundos externos, taxas de mercado de bens e serviços, etc. -	FOS on-line
CMP		
		Vários exemplos podem ser encontrados (Financiamento sustentável de áreas protegidas: uma revisão global de desafios e opções. IUCN, Gland, Suíça / Cambridge, Reino Unido)

Fonte: Adaptado de Bertzky e Stoll-Kleemann (2009)

O estudo de Bertzky e Stoll-Kleemann (2009) é focado em análises voltadas a reduzir a perda da biodiversidade, decorrente das recomendações da CDB (ECO 92) e Declaração de Johannesburgo (Rio+10). Os autores apontam a existência de um grande volume de dados, mas também a grande escassez de metadados, que acarretam em desafios para o compartilhamento e uso desses dados como subsídios às avaliações e decisões relacionadas a áreas protegidas.

Barbosa e demais autores (2008) apresentam outra experiência brasileira no desenvolvimento do BDiG-PELD, um ambiente georreferenciado baseado em biblioteca digital criado para apoiar a integração, disseminação e exploração de dados ecológicos coletados no programa Pesquisa Ecológica de Longa Duração – PELD. O desenvolvimento do BDiG-PELD seguiu os padrões *Open Archives Initiative Protocol for Metadata Harvesting* (OAI-PMH) e *Open Digital Library* (ODL) para integrar as bases de dados e estabelecer uma rede de informações. Além disso, a solução adotou de planilhas padronizadas com campos controlados para preenchimento pelos pesquisadores baseadas no padrão EML.

Por fim, entre os artigos que apresentam desenvolvimento de soluções para integrar e compartilhar dados digitais, dois destacam-se por apresentar uma abordagem voltada às disciplinas **bioinformática** ou **ecoinformática**. Hale e Hollister (2009) definem a ecoinformática como “a ciência da informação em ecologia, ou o estudo da estrutura inerente à informação ecológica, a fim de criar e aplicar a tecnologia da computação para seu gerenciamento e análise”. Segundo os autores, a ecoinformática mudou a maneira como os dados de monitoramento ambiental passaram a ser gerenciados, sintetizados e analisados. Os autores apontam que o surgimento de diversos bancos de dados ecológicos nesse período ocorreu devido aos rápidos avanços em tecnologias da informação, ao surgimento de padrões e protocolos que facilitaram o compartilhamento de dados, bem como pelo o crescente interesse na sustentabilidade da biodiversidade e dos ecossistemas com a ampliação de debates sobre temas como impactos ambientais, mudanças climáticas globais e invasões de espécies (HALE e HOLLISTER, 2009).

Hale e Hollister (2009) apresentam estudo de caso sobre a integração de dados do programa nacional de avaliação costeira – *National Coastal Assessment* (NCA) – da agência de proteção ambiental dos EUA – *Environmental Protection Agency* (EPA). No caso da NCA apresentado nesse artigo, Hale e Hollister (2009) informam que um volume expressivo de

dados foi coletado em milhares de estações na costa dos EUA na década de 1990 utilizando o padrão DIF da NASA e posteriormente o padrão EML. Também foram adotadas as ferramentas também desenvolvidas pela rede KNB: Morpho, gerenciador de dados que facilita a criação e descoberta de dados de EML; e Metacat, um banco de dados de metadados. Por fim, os autores apontam que a adoção desses padrões e dessas ferramentas permitiu analisar a massa de dados da NCA de forma integrada e utilizando diferentes abordagens ecológicas.

Jones, Shildhauer e demais autores (2006) esclarecem que a bioinformática trata da aplicação de ferramentas computacionais no gerenciamento e análise de dados biológicos. Segundo os autores, essa abordagem estimulou rápidos avanços na pesquisa sobre genoma e progressos semelhantes estavam apenas iniciando na ecologia. “Uma razão para a adoção tardia de abordagens de informática em ecologia é a amplitude de dados ecologicamente pertinentes (dos genes à biosfera) e sua natureza altamente heterogênea” (JONES, SCHILDHAUER, *et al.*, 2006, p. 519). Os autores explicam que a variedade de formatos, estruturas lógicas e métodos de amostragem em ecologia criam desafios significativos à integração das coleções de dados e seu compartilhamento. Os autores descrevem ainda soluções de informática para a ecologia, percorrendo desde *data warehouses* específicos a coleções genéricas de dados que usam descrições detalhadas de metadados e ontologias formais para catalogar e fazer referências. Segundo os autores, a combinação dessas abordagens com técnicas automatizadas de integração de dados e sistemas científicos de fluxo de trabalho reduzem custos e abrem novas fronteiras para a pesquisa em ecologia.

Concluindo a análise das publicações desse período, verifica-se uma evolução do entendimento sobre a importância do uso de padrões e o entendimento avançado sobre a importância de investir na qualificação semântica dos dados e das relações entre eles. Já no final da década, há dois artigos identificados referentes ao uso de metadados relacionado às **ontologias (iv)**. O primeiro é focado em apresentar diversos esclarecimentos conceituais relacionados ao uso de ontologias. Madin, Bowers e demais autores (2008) dedicam boa parte do artigo para esclarecer o que são ontologias e a sua relevância para aperfeiçoar a descrição de dados e aperfeiçoar as buscas. Resumidamente, os autores entendem que a ontologia é um “modelo formal que usa lógica matemática para esclarecer e definir conceitos e relacionamentos em um domínio de interesse” (MADIN, BOWERS, *et al.*, 2008).

Madin, Bowers e demais autores (2008) apontam que falta de formalização de conceitos levou a inúmeras controvérsias em ecologia. Para os autores, o uso de termos ambíguos pode até estimular a discussão e levar a várias definições operacionais, mas a ambiguidade terminológica retarda o progresso científico, leva a esforços redundantes de pesquisa e, principalmente, impede os avanços em direção a uma base unificada para a pesquisa relacionada à biodiversidade.

Os autores entendem que “as tecnologias modernas, como a Internet e catálogos de dados eletrônicos, permitem que os pesquisadores troquem ideias e informações com maior rapidez e facilidade do que nunca” (MADIN, BOWERS, *et al.*, 2008). Mas eles destacam que a maioria das abordagens para descrição de dados científicos se baseia no uso *ad hoc* de palavras-chave fornecidas pelo usuário e não garante que esses termos sejam definidos e usados de forma consistente. Por isso, os autores entendem que os sistemas e bases de dados disponíveis não são eficazes para encontrar dados relevantes (baixa precisão e reuso) ou para determinar semelhanças e diferenças entre conjuntos de dados (para integração).

Os autores também apresentam as principais ontologias disponíveis e aplicações – *Extensible Observation Ontology* (OBOE); *Classes for Environmental Data Exchange* (CEDEX); e *Observations Data Model* (ODM); *Evolutionary Trees and Natural History* (ETHAN); *Semantic Web for Earth and Environment Technology* (SWEET). O artigo destaca que a OBOE é considerada a mais compatível com a EML e permiti que outras ontologias sejam associadas ou derivem dela.

Por fim, Madin, Bowers e demais autores (2008) recomendam uma arquitetura para gerenciamento de dados ecológicos e ambientais, consistindo em três níveis principais. Os dados armazenados nos repositórios de dados distribuídos (1) são mediados por metadados e ontologias padrão (2) para fornecer ferramentas de software usadas por cientistas e gerentes de dados (3). Os autores recomentam que os aplicativos usem ontologias e padrões de metadados aprovados pela comunidade para fornecer ferramentas mais eficazes para publicar, consultar, integrar e analisar dados.

No segundo artigo sobre o uso de metadados relacionado à ontologia, Bosh, Lendvai e demais autores (2009) apontam que grande parte dos metadados na temática biodiversidade é composta por texto em linguagem natural. Por isso, os autores propõem o uso associado de dois métodos voltados descoberta automatizada de metadados. O

primeiro método trata da detecção automática de novos metadados nas colunas existentes do banco de dados de texto livre e, segundo, refere-se à descoberta de novas relações ontológicas entre os elementos de metadados.

De forma geral, no período de 2000 a 2009, surgiram diversos projetos voltados a integrar dados, muitas das quais ainda buscavam estabelecer esquemas próprios de metadados ou identificar padrões existentes nos metadados disponíveis. As publicações também apontam que os pesquisadores já identificavam que grande parte dos desafios para integrar e reutilizar as coleções de dados decorria de problemas relacionados ao uso de metadados, desde a ausência de descrição de metadados relevantes à falta de padrão de metadados de uso comum.

4.2.2.3 2010 - 2019

Marcos referentes à gestão de dados de pesquisa – 2010 a 2019

O movimento de acesso aberto à informação se fortaleceu ao longo das últimas décadas. Enquanto na década de 1990 apenas 13 países adotavam legislação para regular essa matéria, já em 2008, de acordo com estudo patrocinado pela UNESCO, mais de 70 países já adotavam leis relacionada ao acesso aberto (MENDEL, 2009). Em 2011, o Brasil publicou a Lei de Acesso à Informação – Lei nº 12.527/2011.

Decorrentes desse movimento de acesso aberto, ao longo das últimas décadas, surgiram e foram ampliadas diversas iniciativas de colaboração global relacionadas à gestão de dados de pesquisa em diferentes temáticas. Em 2012 foi publicada a ISO 19139:2012, que fornece o esquema de implantação XML da ISO 19115, especificando o formato do registro de metadados para descrever, validar e trocar metadados geoespaciais preparados em XML. Em 2014, as diretrizes de implantação da ISO 19115 foram aprovadas para uso nos sistemas de dados das Ciências da Terra da NASA – *Earth Science Data Systems* (ESDS). O documento atual da ISO 19115 (Requisito e orientação para implantação) foi aprovado em julho de 2018.

Desde 2015, o padrão DIF, o modelo de metadados ECHO e a ISO19115 formam a base do modelo unificado de metadados da NASA – *Unified Metadata Model* (UMM). A última atualização do UMM é de 2018. Esse modelo é utilizado no *Common Metadata*

Repository (CMR) que abriga uma das iniciativas precursoras, o *Global Change Master Directory* (GCMD) criado na década de 1990 e que passou a ser mantido pela NASA.

Apesar do expressivo e crescente volume de dados disponíveis, as pesquisas realizadas na década anterior já apontavam lacunas e discrepâncias na descrição dos conjuntos de dados, gerando dificuldades para a integração e o reuso desses recursos em novas pesquisas. Diversos periódicos científicos buscaram adotar políticas exigindo a publicação dos dados de pesquisa como pré-requisito para publicar artigos científicos. Diante das demandas das comunidades científicas e dos periódicos, surgiram grandes repositórios voltados a viabilizar o depósito de dados de pesquisa, soluções públicas de arquivamento de dados – *Public Data Archiving* (PDA).

Para certos tipos de objetos digitais com finalidades especiais, foram criados repositórios bem organizados e profundamente integrados – e.g. o banco de genoma (Genbank); o banco mundial de dados de proteína (*Worldwide Protein Data Bank - wwPDB4*); o banco de dados de ciências da vida relacionado à biomedicina (UniProt); repositório de dados espaciais (*Space Physics Data Facility - SPDF*) e banco de pesquisas de dados da astronomia (*Set of Identifications, Measurements and Bibliography for Astronomical Data - SIMBAD6*). Nesses exemplos, há continuamente ajustes para aprimorar o fluxo da informação científica e a gestão desses conjuntos de dados, considerados de referência e de alto valor (WILKINSON, DUMONTIER, *et al.*, 2016).

No entanto, nem todas temáticas têm seus conjuntos de dados compartilhados em repositórios especificamente estruturados. Segundo Wilkinson e demais autores (2016), em resposta à demanda das demais áreas de pesquisa, surgiram vários repositórios de dados de uso geral, em escalas que variam de institucional (e.g. uma única universidade), a repositórios de escopo global (e.g. Dataverse, FigShare, Dryad, Mendeley Data, Zenodo, DataHub, DANS e EUDat). Esses repositórios “aceitam uma ampla variedade de tipos de dados em uma ampla variedade de formatos, geralmente não tentam integrar ou harmonizar os dados depositados e impõem poucas restrições (ou requisitos) aos descritores da deposição de dados” (WILKINSON, DUMONTIER, *et al.*, 2016). Os autores apontam ainda que o ecossistema de dados digitais resultante dessas soluções mais genéricas é cada vez mais “diversificado e menos integrado, exacerbando o problema de descoberta e reutilização” (WILKINSON, DUMONTIER, *et al.*, 2016).

Diante desse cenário, em 2014, foi realizado um *workshop* na cidade de Leiden na Holanda e um grupo representativo de cientistas e organizações buscou definir requisitos para aperfeiçoar a descoberta e reuso dos dados de pesquisa. A partir desse evento foi criado um grupo de trabalho – FORCE11 *community* – e definidas linhas gerais a serem desenvolvidas. Em março de 2016, o FORCE11 publicou artigo contendo os princípios orientadores para a gestão de dados de pesquisa, conhecidos como princípios FAIR – do acrônimo para *Findable, Accessible, Interoperable, and Reusable*. Esses princípios visam o desenvolvimento de soluções interoperáveis, voltados a facilitar a descoberta, acesso e reuso aos dados (WILKINSON, DUMONTIER, *et al.*, 2016). Em setembro de 2016, durante a reunião da cúpula do G-20 realizada na cidade de Hangzhou na China, foi publicada declaração que recomenda o uso dos princípios FAIR para pesquisa (EUROPEAN COMMISSION. G20, 2016).

Essa década é, portanto, marcada pelo aumento do compartilhamento dos dados de pesquisa e consequente crescimento das soluções genéricas de armazenamento desses recursos. Mas também houve o fortalecimento de iniciativas voltadas a aperfeiçoar a gestão de dados da biodiversidade. Entre outras iniciativas, destaca-se que rede GBIF aderiu ao movimento de aperfeiçoamento da gestão de dados de pesquisa e adequação aos princípios FAIR a partir de 2017 (GBIF, 2016).

Marcos referentes à temática biodiversidade – 2010 a 2019

A década de 2010 a 2019 também é marcada pela ampliação dos mecanismos de cooperação global frente aos problemas ambientais. De forma a dar maior visibilidade e realizar ações efetivas em relação a temas mais específicos, a ONU estabeleceu 2010 como “Ano Internacional da Biodiversidade” e 2011 como o “Ano Internacional das Florestas”.

Em 2010, durante a convenção na cidade de Nagoya no Japão, foi discutido o “Protocolo de Nagoya”, que trata do acesso aos recursos genéticos e a partilha justa e equitativa de benefícios resultantes de sua utilização. Esse acordo internacional foi ratificado por 126 países e entrou em vigor em 2014. Também durante a conferência mundial em Nagoya, a ONU declarou o período de 2011 a 2020 como a década da biodiversidade e estabeleceu o Plano Estratégico para a Biodiversidade (ONU. CDB, 2010). Esse plano tem

foco no desenvolvimento dos CHMs para promover serviços efetivos de informação e outros meios apropriados para promover e facilitar a cooperação científica e técnica.

Em 2012, foi realizada no Rio de Janeiro, nova Conferência das Nações Unidas sobre Desenvolvimento Sustentável (conhecida também como Rio + 20). Resultou dessa conferência um documento contendo medidas claras e práticas para o desenvolvimento sustentável – conjunto de Objetivos de Desenvolvimento Sustentável (ODS). Esses objetivos são embasados nos Objetivos de Desenvolvimento do Milênio (ODM) da ONU e convergem com a agenda de desenvolvimento planejada para o período pós-2015 (ONU, 2015).

Além disso, desde 2014, a ONU passou a contar com a Assembleia Ambiental das Nações Unidas – United Nations Environment Assembly (UNEA). Trata-se da mais importante plataforma representada pelo órgão do mais alto nível de tomada de decisões do mundo sobre meio ambiente. A criação da UNEA marcou o início de um período em que os impactos ao meio ambiente são considerados problemas de todos os países e, pela primeira vez, colocou as preocupações ambientais no mesmo âmbito da paz, segurança, finanças, saúde e comércio (UN, 2020; ONU, 2020). Em 2015, ocorreu em Nova York, na sede da ONU, a Cúpula de Desenvolvimento Sustentável. Nesse encontro, foi definida a Agenda 2030 para o Desenvolvimento Sustentável (ONU, 2015).

A OCDE (2010) publicou o Memorando Entendimento entre os governos participantes estabelecendo as diretrizes para adesão aberta à GBIF. Em 2019, completou 20 anos desde a recomendação em 1999 para a criação da GBIF, em decorrência das diretrizes estabelecidas pela CDB (1992) para a integração de dados da biodiversidade e desenvolvimento de CHMs. A GBIF tornou-se uma plataforma global de acesso aberto e distribuído, que representa um avanço ao acesso aberto, abrangente e de fácil uso aos conjuntos de dados digitais de ocorrência de espécies (CODATA, PFEIFFENBERGER e UHLIR, 2020). A maior parte dos artigos identificados sobre o uso de metadados apresentam referências à GBIF, seja diretamente como parte do assunto ou como citações bibliográficas.

Análise dos artigos – 2010 a 2019

O período de 2010 a 2019 concentra a maior parte dos artigos científicos coletados – 60 do total de 78 artigos. Houve um aumento expressivo de publicações relacionadas ao uso

de metadados na temática biodiversidade, representando um volume três vezes maior comparado com a soma do quantitativo coletado nas duas décadas anteriores.

Os 60 artigos científicos publicados apresentam o uso de metadados associado aos seguintes assuntos: **(i) revisão sobre os desafios na gestão de dados da biodiversidade** – 2 artigos; **(ii) desenvolvimento de indicadores da biodiversidade** – 4 artigos; **(iii) catálogos e ontologia** – 3 artigos; **(iv) proposta de padrão de metadados ou extensão** – 2 artigos; **(v) digitalização de dados** – 10 artigos; **(vi) procedimentos de gestão de dados** – 16 artigos e **(vii) soluções computacionais para descobrir, integrar, compartilhar e reutilizar coleções de dados digitais** – 23 artigos. A análise meta-sintética dos artigos referentes a esse período é apresentada a seguir.

Referente ao primeiro assunto, **revisão sobre os desafios na gestão de dados da biodiversidade (i)**, há dois artigos especialmente voltados a identificar desafios e avaliar as necessidades de informações sobre biodiversidade apontadas por pesquisadores e tomadores de decisão relacionada ao meio ambiente. A primeira publicação aborda necessidades de informação, comportamentos de busca e atributos mais relevantes dos conjuntos de dados sobre biodiversidade. Daves, Tenopir e demais autores (2014) sustentam conclusões anteriores de que informações úteis sobre biodiversidade devem ser facilmente e rapidamente acessíveis, bem como disponibilizadas em soluções que permitam integração e visualização espacial, de forma a corresponder adequadamente às necessidades dos usuários. Os autores concluem que há necessidade de investir esforços para aperfeiçoar o gerenciamento, descoberta e descrição de dados.

Na segunda publicação desse tópico, Bendix, Nieschulze e Michener (2012) apontam diversos desafios relacionados à gestão de dados da biodiversidade (e.g. heterogeneidade dos objetos de informação e dos vocabulários utilizados para descrevê-los; multidisciplinaridade dos dados requerido nas análises; necessidade de dados acumulados em longa escala de tempo; repositórios criados para fins muito específicos ou muito genéricos e que muitas vezes não têm sustentabilidade institucional ou financeira). Diante dos desafios identificados, os autores discutem três questões que consideram fundamentais para promover acesso aberto aos dados globais da biodiversidade. Primeiramente, do ponto de vista técnico, a padronização de dados e metadados, a interoperabilidade de dados e as interfaces amigáveis ao usuário são apontadas como necessidades prioritárias pelos autores.

Em segundo lugar, eles apontam a necessidade de estabelecer estratégias técnicas, institucionais e de financiamento sustentáveis. Por último, os autores ponderam a influência dos direitos de propriedade e do comportamento do usuário nas soluções técnicas e administrativas.

O segundo assunto associado ao uso de metadados refere-se ao **desenvolvimento de indicadores da biodiversidade (ii)**, com quatro artigos identificados, dos quais três tratam do desenvolvimento de variáveis essenciais à análise de dados da biodiversidade – *Essential Biodiversity Variable* (EBV). O conceito de Variáveis Essenciais da Biodiversidade (EBVs) foi introduzido para estruturar o monitoramento da biodiversidade globalmente de forma a harmonizar e padronizar dados de biodiversidade de fontes diferentes capturando um conjunto mínimo de variáveis críticas necessárias a análise e gerenciamento das mudanças na biodiversidade (CHANDLER, SEE, *et al.*, 2017; KISSLING, AHUMADA, *et al.*, 2018)

EBVs tais como a composição do ecossistema por tipo funcional, retenção de nutrientes e extensão ou fragmentação do ecossistema, podem ser monitorados por sensores em larga escala ou sistemas de Observação da Terra (CHANDLER, SEE, *et al.*, 2017). Esses sistemas referem-se ao sensoriamento remoto por satélite (e.g. Landsat, Sentinel), imagens aéreas, além de em como plataformas de observação terrestre (e.g. *National Ecological Observatory Network*, rede de câmeras e outros).

Além disso, observações humanas precisam complementar dados detectados remotamente (e.g. modelos de distribuição de espécies, modelos de associação de habitat de espécies, modelos sazonais de produtividade) fornecendo informações necessárias à calibração e validação dos EBVs (CHANDLER, SEE, *et al.*, 2017). Essa demanda da participação humana tem como obstáculo a ausência de profissionais suficientes (ou financiamento para contratação suficiente de profissionais) para monitorar EBVs em larga escala. Chandler, See e demais autores (2017) entendem que essas lacunas de informações podem ser preenchidas utilizando soluções abertas aos cidadãos – *Citizen Science* (CS) – na qual voluntários participam de alguns ou de todos os aspectos das avaliações ambientais. Para tanto, os autores propõem esquemas de metadados que devem ser priorizados. Por sua vez, Kissling, Ahumada e demais autores (2018) avaliam os desafios do *Big Data* para a criação de produtos globais de dados EBV com abrangência de táxons e escalas espaço-temporais visando à distribuição e abundância de espécies. Por fim, também em abordagem

relacionada ao *Big Data*, Ziliolo, Oggioni e demais autores (2019) propõem métodos para extrair e analisar EBVs, por meio da mineração de dados no núcleo italiano da rede europeia de monitoramento ecológico de longo prazo – *International Long-Term Ecological Research* (ILTER).

Em uma abordagem para tratamento de contextos específicos, Lengyel, Kosztyi e demais autores (2018) desenvolveram indicadores para o desenho amostral, esforço amostral e análise dos dados da biodiversidade para apoiar a avaliação das práticas de monitoramento. Esses indicadores foram desenvolvidos a partir da análise dos metadados de 646 espécies e respectivos esquemas de monitoramento de habitat em 35 países europeus,. Os autores também avaliaram como os indicadores propostos são afetados por fatores socioeconômicos (e.g. ano de início, fonte de financiamento, motivação e abrangência geográfica do monitoramento). Segundo os autores os indicadores propostos fornecem parâmetros de referência para auxiliar na identificação dos pontos fortes e fracos dos esquemas de monitoramento individuais (e.g. esforço de amostragem e análise de dados no monitoramento da biodiversidade) em relação à média de outros esquemas e para melhorar as práticas atuais, formular as melhores práticas, padronizar o desempenho e integrar os resultados do monitoramento.

O assunto **catálogos e ontologias (iii)** associado ao uso de metadados apresenta três publicações identificadas. Walls, Guralnicl e demais autores (2014) descrevem os avanços no desenvolvimento das ontologias *Biological Collections Ontology* (BCO) e *Population and Community Ontology* (PCO), bem como recomendam ferramentas para descrever conjuntos de dados utilizando essas e outras ontologias relacionadas à biodiversidade. Lucarini, Gigante e demais autores (2014) apresentam e recomendam o uso do catálogo taxonômico do projeto *anArchive Checklist* para prover informações semânticas e sintáticas aos conjuntos de dados florísticos, organizando os nomes de forma a manter a integridade taxonômica. Relativo a esse assunto, por fim Guralnik, Cellinese e demais autores (2015) alertam sobre a necessidade de criar perfis de metadados de identificadores únicos aplicados aos registros de conjuntos de dados sobre biodiversidade, pois as “práticas usualmente adotadas para a descrição não preservam as ligações entre os dados, impedindo a interoperação, o rastreamento de proveniência e integração em conjuntos de dados maiores”. Guralnik, Cellinese e demais autores (2015) também apresentam uma visão geral

dos esquemas e práticas disponíveis, bem como de soluções de identificadores transversais que fazem a ponte entre os diversos domínios que integram objetos de informação da biodiversidade.

O assunto **(iv) proposta de padrão de metadados** associado ao uso de metadados apresenta duas publicações nessa última década. No primeiro artigo, Rasaiah, Bellman e demais atores (2015) propõe uma abordagem para o desenvolvimento de padrões de metadados para aplicações específicas, visando promover confiabilidade e interoperabilidade para os conjuntos de dados de espectroscopia. Os autores analisam os desafios para desenvolver um padrão de metadados que atenda aos requisitos exclusivos de comunidades específicas (e.g. que incluam a medição *in situ* da refletância subaquática no monitoramento de corais). Os autores também analisaram os esquemas de metadados de padrões de metadados existentes (e.g. Dublin Core; ABCD; DwC; EML; CSDGM; FGDC NBII) e concluíram que não atendem às necessidades dos cientistas de espectroscopia de campo para aplicações gerais e específicas (e.g. métodos de medição de campo; calibração de instrumentos; representatividade de dados para campanhas de espectroscopia de campo marinho). Por fim os autores apresentam um propões um conjunto de metadados que consideram relevantes e apontam as implicações da semântica e da sintaxe associadas à necessidade de desenvolver um padrão de metadados.

No segundo artigo relacionado à proposta de padrão de metadados, Pooter, Appeltans e demais autores (2017) apresentam o *Ocean Biogeographic Information System* (OBIS), que é o banco de dados de acesso aberto que reúne os dados mais abrangente do mundo relacionados à distribuição de espécies marinhas. O OBIS cresce com milhões de novas observações de espécies todos os anos a partir das contribuições de uma rede de instituições, projetos e indivíduos, unidos pelo objetivos comum de construir uma base de conhecimento científico aberta ao público para descoberta e exploração científica, de forma a detectar tendências e mudanças que informam a sociedade elementos essenciais ao gerenciamento, conservação e desenvolvimento sustentável. Segundo os autores, o OBIS se concentrava apenas na coleta de dados biogeográficos (presença de espécies marinhas no espaço e no tempo) e operava com fluxos de dados otimizados, procedimentos de controle de qualidade e padrões de dados direcionados especificamente para esses dados. Com base nas demandas da comunidade, foi criado OBIS-ENV-DATA para gerenciar conjuntos de dados

que combinam medições biológicas, físicas e químicas. Os autores apresentam neste artigo uma proposta voltada a aperfeiçoar o gerenciamento dos dados referentes à metodologia de amostragem, rastreamento de animais e dados de telemetria, medições biológicas (e.g. comprimento do corpo, porcentagem de cobertura ao vivo), bem como medições ambientais (e.g. concentrações de nutrientes, características de sedimentos ou outras substâncias abióticas¹⁹; parâmetros medidos durante a amostragem para caracterizar o ambiente a partir do qual os dados biogeográficos foram coletados). A prática recomendada baseia-se no padrão Darwin Core Archive (DwC-A) e nas práticas adotadas pelo Global Biodiversity Information Facility (GBIF). Consiste no estabelecimento de um Núcleo de Eventos no arquivo DwC (DwC-A) em combinação com a Extensão de Ocorrência, bem como o aprimoramento para a Extensão *MeasurementOrFact* do DwC. Segundo os autores, essa nova estrutura permite a ligação de medidas ou fatos - propriedades quantitativas e qualitativas - a eventos de amostragem e ocorrências de espécies, e inclui campos adicionais para padronização de propriedades. Os autores também adotaram o uso do novo termo *parentEventID* para o DwC, que permite a criação de uma hierarquia de eventos de amostragem. Os autores acreditam que a adoção desta prática recomendada como um novo padrão de dados para gerenciar e compartilhar conjuntos de dados ambientais biológicos pela comunidade científica internacional é essencial para melhorar a eficácia da base de conhecimento e a integração necessária ao gerenciamento de dados críticos para compreender os processos ecológicos e biológicos no oceano e em terra.

A **digitalização de dados (v)** é o terceiro mais frequente entre os assuntos associados ao uso de metadados no período de 2010 a 2019, com 10 publicações identificadas. A digitalização das informações associadas a espécimes²⁰ em coleções de história natural é considerada fundamental para fornecer informações sobre a biodiversidade em escala global e representa recursos de interesse a uma variedade de campos de pesquisa (BERENDSOHN e SELTMANN, 2010; VOLLMAR, MACKLIN e FORD, 2010; BAIRD, 2010; OTEGUI, ARIÑO, *et al.*, 2013; SEREGIN, 2017; SKEVAKIS, MAKRIS, *et al.*, 2014). Muitos dos esforços de digitalização realizados foram orientados por necessidades institucionais individuais, sem relação com iniciativas de coordenação em escala global, conduzindo a

¹⁹ Substâncias derivadas de aspectos físicos, químicos ou físico-químicos do meio ambiente, tais como a luz e a radiação solar, a temperatura, o vento, a água, a composição do solo e outros.

²⁰ Cada indivíduo representa um espécime coletado. Uma coleção pode conter vários espécimes de cada espécie catalogada.

resultados distintos, com informações irregulares referentes à cobertura taxonômica e geográfica, bem como à qualidade (BERENDSOHN e SELTMANN, 2010).

Vollmar, Macklin e Ford (2010) apresentam os resultados de pesquisa realizada em 2009 sobre os desafios e preocupações envolvidos na digitalização de espécimes de história natural realizada com mais de 200 participantes, integrados por os curadores, gerentes de coleções e administradores da comunidade de história natural. O principal problema apontado foi à falta de financiamento ou questões diretamente relacionadas ao financiamento. Em segundo lugar os participantes apontaram a digitalização desigual que leva a um acúmulo irregular de registros com qualidades variadas e com prioridades diferentes, influenciando a adequação dos dados ao uso. Por fim, outra questão relevante apontada refere-se à heterogeneidade dos recursos informacionais a serem digitalizados. Entretanto, apontaram que muitas disciplinas têm problemas similares e que deveriam investir em ações para promover o conhecimento sobre digitalização, com treinamento para disseminar melhores práticas.

Há um grande volume de coleções a serem digitalizadas, por isso muitas iniciativas buscam definir quais recursos devem passar prioritariamente pelos processos de digitalização (BERENDSOHN e SELTMANN, 2010; BERENDSOHN, CHAVAN e MACKLIN, 2010; VOLLMAR, MACKLIN e FORD, 2010; BAIRD, 2010; SEREGIN, 2017). Berendsohn e Seltmann (2010) apontam que a maioria dos estudos sobre biodiversidade é taxonômica e geograficamente restrita, mas o acesso a informações de coleta não digitalizadas é quase exclusivamente pelo nome do taxon. Por isso, utilizando dados da rede GBIF como estudo de caso, os autores propõem a criação de um "Índice Geotaxonômico" para definir prioridades de digitalização, visando fornecer metadados sobre o número de amostras de uma região geográfica específica pertencente a uma categoria taxonômica mais alta que possa representar um meio de atrair a atenção de pesquisadores e governos para exploração de coleções não digitalizadas relevantes.

Após analisar vários projetos nacionais e internacionais considerados referências, Baird (2010) argumenta que uma abordagem coordenada globalmente para a digitalização de uma massa crítica de dados de pesquisa é altamente desejável e recomenda a elaboração de um plano de ação global. Nessa linha, Berendsohn, Chavan e Mackil (2010) apresentam a estratégia global e plano de ação para a mobilização de dados de história natural elaborada

por grupo de trabalho instituído pela GBIF. Esse grupo definiu três ações principais a serem adotadas pela GBIF para auxiliar a identificação de coleções não digitalizadas e a priorização para a digitalização: (a) facilitar o acesso a informações sobre recursos de coleta não digitalizados, divulgando o potencial de pesquisa de coleções por meio de metadados e avaliando o número de espécimes não digitalizados; (b) aperfeiçoar os procedimentos de digitalização em progresso para aumentar a eficiência da captura de dados de amostras e melhorar a qualidade dos dados por meio de medidas técnicas, garantindo a atribuição e o crédito profissional e influenciando as prioridades das instituições parceiras; (c) melhorar e promover a infraestrutura global para mobilizar dados de coleta digitalizados por meio de medidas técnicas, atividades de extensão e medidas políticas. O artigo apresenta o detalhamento de cada uma dessas ações. A definição dessa estratégia global de digitalização pela GBIF foi fundamental para estimular e qualificar novas iniciativas de digitalização e disponibilização das coleções de história natural.

Decorrente das iniciativas apoiadas pela GBIF, Holetschek, Dröge e demais autores (2012) apresentam o Serviço de Acesso a Coleções Biológicas (BioCAsE), criado para fomentar o fornecimento de dados de história natural utilizando esquemas complexos do padrão ABCD e promover processos de digitalização aperfeiçoados. Essa iniciativa disponibilizou ferramentas para auxiliar a descrição e compartilhamento desses recursos de informação, incluindo o Protocolo BioCAsE e o software PyWrapper, serviço Web para acessar dados de história natural. Os autores também explicam que o padrão ABCD está em constante aperfeiçoamento e fornece uma cobertura abrangente de dados da história natural e foi estendido para cobrir coleções de DNA (ABCD-DNA) e geociências (ABCD-EFG, a extensão para geociências). Também alinhados à GBIF, Blagoderov, Kitching e demais autores (2012) propõem um fluxo de trabalho para a digitalização em massa de coleções de museus de história natural, considerando padrão Darwin Core e identificadores únicos persistentes.

Também considerando o padrão Darwin Core, os autores Thomer, Vaidya e demais (2012), propõem metodologia para extração de dados a partir de notações textuais registradas por pesquisadores e integração com outras fontes dados. Os autores destacam que essas notações de pesquisa geralmente contêm detalhes necessários para entender a

localização e as condições ambientais existentes durante eventos de coleta (amostragem) e representam valor para estudos de mudança global.

Otegui, Ariño e demais autores (2013) avaliam a quantidade e a qualidade das informações disponibilizadas por processos de digitalização nos registros do nó espanhol (GBIF.ES) da rede GBIF. Os autores identificaram padrões referentes à disponibilidade e ao acúmulo de dados e acreditam que isso pode ajudar a decidir quando e como os dados digitalizados podem ser usados. No entanto, os autores apontam a existência de grandes faixas de registros que carecem de elementos de dados, como georreferenciamento ou níveis taxonômicos. Por fim, destacam que embora as informações em geral sejam amplas e adequadas para muitos usos, melhorar a integridade dos registros provavelmente aumentaria o tempo de usabilidade desses dados.

Seregin (2017) apresenta os resultados do processo de digitalização de 77,7% da coleção total (785887 espécimes) realizado entre o período de 2015 a 2016 em um herbário em Moscow – *Herbarium of Moscow State University* (MW). Os autores destacam que os metadados do herbário digital incluem o identificador de cada amostra, o nome da espécie, referência geográfica e está integrado ao banco internacional *Catalogo of Life*.

Minicante, Birello e demais autores (2017) apresentam o *Archivio di Studi Adriatici* (ASA), um repositório do Instituto de Ciências Marinhas (ISMAR-CNR) que resulta da digitalização e compilação de espécimes marinhas e respectivos metadados usando os padrões *Dublin Core* e *Simple Darwin Core*.

Entre os assuntos associados ao uso de metadados, o **procedimentos de gestão de dados (vi)** é o segundo mais frequente no período de 2010 a 2019, com 16 publicações identificadas. Já no início da década, uma das publicações mais relevantes refere-se ao estudo de Moritz, Krishnan e demais autores (2011) que aponta a necessidade de institucionalização de uma estrutura de publicação de dados capaz de acolher restrições socioculturais, técnicas de infraestrutura, políticas e legais, ao mesmo tempo garantir a sustentabilidade e suporte financeiro. Nesse sentido, os autores discutem as 24 recomendações para incentivar o acesso livre e aberto aos dados de biodiversidade do mundo, que foram desenvolvidas por grupo de trabalho instituído pela GBIF. Por fim, os autores destacam o papel fundamental da GBIF em fomentar a divulgação e aplicação dessas recomendações.

Resultando dessas recomendações da GBIF, Remsen, Knapp e demais autores (2012) propõem procedimentos para converter uma lista de verificação florística convencional, escrita em um programa padrão de processamento de texto, em dados estruturados no formato *Darwin Core Archive*. Os autores apontam que os dados também foram publicados e indexados por meio da aplicação *Integrated Publishing Toolkit (IPT)* da GBIF.

King, Rajvanshi e demais autores (2012) afirmam que as informações de biodiversidade obtidas durante o processo de Avaliação de Impacto Ambiental (AIA) raramente são acessíveis para outros usos após a conclusão dos estudos. As análises realizadas na AIA subsidiam o Licenciamento Ambiental (LA) de atividades e empreendimentos potencialmente poluidores ou que degradam o meio ambiente. Por isso, os autores destacam que os dados resultantes da AIA precisam ser facilmente acessíveis e reutilizáveis para subsidiar novas pesquisas da comunidade científica, bem como para o planejamento das ações de conservação e, conseqüentemente, decisões futuras baseadas na AIA. Portanto, os autores propõem uma 'Estrutura de Publicação de Dados de Biodiversidade na AIA', com base nos padrões globais recomendados pela GBIF. Esse estudo destaca a necessidade urgente de adotar uma variedade de ferramentas de publicação de dados e melhores práticas para tornar os dados de biodiversidade de Estudos de Impacto Ambiental (EIA) reutilizáveis, utilizando como estudo de caso projetos-piloto na Índia e na África do Sul.

Güntsch, Fichtmüller e demais autores (2012) afirmam que os dados de biodiversidade carecem de uma estratégia para preservação e disponibilidade em longo prazo, pois correm o risco de ficar desatualizados e perdidos. Os autores apresentam as soluções proposta no projeto reBiND para desenvolver um fluxo de trabalho eficiente e bem documentado visando aperfeiçoar a gestão dos conjuntos de dados. O fluxo de trabalho consiste em uma primeira fase para adequação dos dados aos padrões contemporâneos, seguida da fase de validação de dados, depois armazenamento em um banco de dados XML nativo e, por fim, a publicação de dados em redes internacionais de biodiversidade.

Skevakis, Makris e demais autores (2014) afirmam que uma impressionante abundância de conteúdo científico de alta qualidade disponível em museus de história natural da Europa permanece praticamente inexplorada devido a uma série de barreiras (e.g. falta de interconexão e interoperabilidade entre os sistemas de gerenciamento usados

pelos museus; falta de acesso centralizado por meio de um ponto de referência europeu. como Europeana; e inadequação dos atuais metadados e organização de conteúdo). Nesse sentido, os autores apontam o projeto Europa Natural como uma solução coordenada para superar essas barreiras na Europa. Os autores descrevem como os curadores podem publicar, descrever e gerenciar semanticamente os objetos de patrimônio cultural dos museus, além de divulgá-los para Europeana.eu e BioCASE/GBIF. Além disso, discutem metodologia para a transição para a Web Semântica e para a publicação dos metadados do patrimônio cultural dos NHMs como dados vinculados (*Linked Data*), baseada o Modelo de Dados da Europeana.

Os brasileiros Pôças, Gonçalves e demais autores (2014) propõem uma estrutura metodológica para avaliar a adequação ao uso de conjuntos de dados espaciais para aplicações ambientais e ecológicas. A metodologia é baseada no uso de metadados para analisar a semelhança entre as características dos dados e as necessidades ou expectativas do usuário em relação a indicadores de qualidade. Os autores utilizam o conceito de 'fatores críticos' como os indicadores de qualidade que têm maior importância para os usuários, dados os requisitos ou expectativas nos contextos de aplicação especificados. A metodologia proposta integra-se a metodologia de avaliação da qualidade dos dados espaciais – *spatial data quality* (SDQ) – utilizada nos portais de informações geográficas e nas plataformas WebGIS. Os autores entendem que a metodologia proposta pode incentivar a implementação de rotinas de avaliação SDQ em programas de avaliação e monitoramento ecológicos, promovendo um uso mais adequado dos dados geoespaciais e, finalmente, contribuindo para decisões bem fundamentadas de políticas e gerenciamento.

Bonet, Pérez-Pérez e demais autores (2014) apontam que muitas das melhores práticas relativas ao desenvolvimento de modelos ecológicos ou técnicas analíticas não estão totalmente disponíveis, pois ficam restritas ao conhecimento dos pesquisadores. Dessa forma, os autores propõem uma estrutura conceitual voltada a promover a cooperação entre cientistas, aprimorando a criação de conhecimento relevante que pode ser transferido para pesquisas e análises futuras. A metodologia desenvolvida é voltada a documentar, compartilhar e executar mais de 200 modelos e processos analíticos associados a um programa global de monitoramento de mudanças que está sendo realizado nas montanhas de Sierra Nevada (no sul da Espanha). Os autores criaram um fluxo de trabalho

científico para conectar e executar diferentes tipos de modelos e processos analíticos. Por fim, recomendam a criação de repositórios de modelos para depósito dos modelos documentados visando reuso por outros pesquisadores.

Roche, Kruuk e demais autores (2015) tratam da expansão das políticas de depósito de dados de pesquisa – Public Data Archiving (PDA) – e decorrente ampliação da acessibilidade aos dados relacionados às publicações científicas. Diante desse cenário, os autores analisam se a qualidade dos dados é suficiente para permitir sua reutilização. Foram analisados 100 conjuntos de dados associados a periódicos de pesquisas ecológicas e evolutivas com política de depósito já estabelecida. Desses conjuntos de dados, 56% estavam incompletos e 64% foram armazenados em formato que impede parcial ou totalmente a reutilização. Por fim, os autores apresentam recomendações que passam por mudanças culturais capazes de conscientizar os pesquisadores sobre os benefícios de depositar conjuntos de dados com formatos e descrições adequadas. Além disso, apontam diretrizes para ajudar os autores a aumentar o potencial de reutilização de seus dados e a conformidade com as políticas de dados de periódicos.

Vanderbilt, Lin e demais autores (2015) analisam a rede internacional de pesquisa ecológica de longo prazo – *International Long Term Ecological Research* (ILTER) – que foi criada em 1993 e contém uma diversidade de conjuntos de dados referente a ecossistemas em todo o mundo. Apesar de relatar obstáculos considerados significativos (e.g. publicação de dados ILTER incluem distribuição desigual entre redes de que contribuem com informações; dificuldades no uso de ferramentas e recursos; bem como compatibilização entre idiomas na tradução de termos), os autores apontam que a rede ILTER formou parcerias e colaborou para fornecer treinamento em gerenciamento de informações, adotou um padrão comum de metadados e desenvolveu ferramentas de gerenciamento de informações úteis que incentivam os pesquisadores a compartilhar dados. Por fim, os autores compartilham as lições aprendidas com os sucessos nas parcerias internacionais criadas para capacitar pessoas e torná-las multiplicadoras de conhecimento para projetos que envolvam gestão de dados de pesquisa e apresentem desafios semelhantes aos identificados.

Niu (2016a; 2016b) propõe procedimentos para o controle e integração de dados científicos. O autor afirma que as comunidades científicas e instituições tradicionais ligadas a

gestão de dados de pesquisa usam amplamente as mesmas abordagens de controle para facilitar o acesso e a descoberta de recursos informacionais. O autor identifica as semelhanças e diferenças entre os métodos de organização do conhecimento dos arquivos nacionais e dos repositórios de conjuntos de dados e aponta que os repositórios de conjuntos de dados usam padrões de metadados mais especializados, mas com suporte a organização do conhecimento mais flexível visando facilitar o acesso e uso. Por fim, o autor discute algumas abordagens possíveis que as instituições podem adotar para melhorar a descrição dos dados e oferecer suporte a uma melhor descoberta e uso de dados.

Tessarolo, Ladle e demais autores (2017) apontam que aspectos espaciais ou temporais dos dados da biodiversidade podem influenciar diretamente a utilidade, comparabilidade e confiabilidade dos estudos ecológicos e evolutivos. Os autores destacam que embora os efeitos dos problemas na cobertura espacial dos dados da biodiversidade sejam bem conhecidos pelos pesquisadores, a variação temporal na qualidade dos dados (isto é, a congruência entre informações registradas e reais) é menos conhecida. Nesse sentido, os autores desenvolveram uma estrutura conceitual para entender a influência do tempo na qualidade dos dados da biodiversidade, com base em três processos principais: (1) a dinâmica natural dos sistemas ecológicos - como rotatividade de espécies ou extinção local; (2) revisões taxonômicas periódicas e; (3) a perda de dados físicos e metadados devido à curadoria ineficiente, acidentes ou déficits de financiamento. Apesar de entender que é inevitável a degradação de alguns dados devido a variações naturais nos sistemas ecológicos, os autores apontam que a deterioração dos dados pode ser parcialmente melhorada com a adoção de protocolos padrão para geração, armazenamento e compartilhamento de dados e metadados.

Diante dos desafios enfrentados para aperfeiçoar a gestão dos dados científicos, em 2017 destaca-se a publicação de artigo que propõe a adoção dos Princípios FAIR (WILKINSON, VERBORGH, *et al.*, 2017). Segundo Wilkinson e demais autores (2017), os dados nas ciências da vida são extremamente diversos e armazenados em um amplo espectro de repositórios. Os autores explicam que em resposta à demanda de diversas áreas de pesquisa surgiram vários repositórios de dados de uso geral, em escalas que variam de institucional (e.g. uma única universidade), a repositórios de escopo global (e.g. Dataverse, FigShare, Dryad, Mendeley Data, Zenodo, DataHub, DANS e EUDat). Os autores afirmam que

esses repositórios “aceitam uma ampla variedade de tipos de dados em uma ampla variedade de formatos, geralmente não tentam integrar ou harmonizar os dados depositados e impõem poucas restrições (ou requisitos) aos descritores da deposição de dados” e apontam que o ecossistema de dados digitais resultante dessas soluções mais genéricas é cada vez mais “diversificado e menos integrado, exacerbando o problema de descoberta e reutilização”. Os autores apontam também que a falta de uniformidade nos modelos de dados dos repositórios, bem como na riqueza e disponibilidade das descrições de metadados tornam a integração e a análise dos dados uma tarefa manual, demorada e sem escalabilidade. Diante desse cenário, os autores propõem princípios orientadores para a gestão de dados de pesquisa, conhecidos como princípios FAIR – do acrônimo para *Findable, Accessible, Interoperable, and Reusable*. Esses princípios exploram um conjunto de padrões de design da Web orientados a recursos para descoberta, acessibilidade, transformação e integração de dados que podem ser implementados por qualquer repositório de propósito geral ou especial como um meio para ajudar os usuários a encontrar e reutilizar seus dados.

Weatherdon, Appeltans e demais autores (2017) relatam que geralmente a coleta dos dados de biodiversidade demanda recursos financeiros elevados, principalmente em ambientes marinhos. Portanto, o reuso desses dados é extremamente importante para novas pesquisas, bem como para a implementação de políticas ou outros tipos de tomada de decisão. Os autores apontam que, embora a qualidade e acessibilidade dos dados sobre biodiversidade e conservação marinha tenham melhorado na última década, existem vários conjuntos de dados que não são acessíveis. Por isso, os autores entendem ser fundamental transformar esses conjuntos de dados de biodiversidade marinha em produtos de conhecimento que transmitam informações aos decisores é um passo importante no fortalecimento do intercâmbio de conhecimentos na interface ciência-política. Portanto, os autores propõem aspectos para a seleção de produtos de conhecimento sobre biodiversidade e conservação que contribuem para sua capacidade de apoiar a tomada de decisões e elaborar políticas no domínio marinho.

Huang, Bowman e demais autores (2018) apesar da riqueza de informações contida nas coleções de biodiversidade, relacionadas aos espécimes e respectivos metadados, os repositórios em que estão armazenadas podem representar um desafio para o acesso e reuso dos conjuntos de dados, que raramente são estruturadas observando as necessidades

dos pesquisadores. Os autores apresentam proposta de estruturação dos conjuntos de dados a partir de coleções que não foram estruturadas explicitamente para tratar de questões ecológicas. Os autores destacam a relevância da biogeografia, clima, hospedeiros e separação geográfica, bem como relatam sobre a importância dos metadados para gerar subsídios às pesquisas futuras e responder a questões ecológicas, em especial considerando um estudo de caso na área da micologia.

Referente ao assunto procedimentos de gestão de dados associado ao uso de metadados, a última publicação analisada é de Gordon e Habermann (2018), que apontam que muitas comunidades científicas já usam metadados para possibilitar a descoberta, acesso, uso e entendimento de conjuntos de dados científicos. Além disso, instituições e comunidades científicas também desenvolveram recomendações para controlar a descrição dos metadados, tipicamente para representações específicas. Os autores analisam o conteúdo conceitual das recomendações de forma a compreender a completude dos registros de metadados EML e CSDGM do DataONE em comparação com as recomendações do projeto LTER. Por meio da análise quantitativa e da integridade dos registros de metadados, os autores buscam determinar se a estrutura de metadados do LTER são mais completos em relação às recomendações de outras coleções descritas em EML e no CSDGM. Por fim, os autores concluem que os registros LTER são, em geral, mais completos que as outras coleções EML, mas grande similaridade às coleções CSDGM.

O último assunto associado ao uso de metadados refere-se às **soluções para descobrir, integrar, compartilhar e reutilizar coleções de dados digitais (vii)**. Esse assunto é o mais frequente e representa mais de um terço de todas as publicações analisadas no período de 2010 a 2019, com 23 artigos identificados.

No primeiro artigo nesse assunto, Deng (2010) trata do da reutilização e do processamento de metadados em lote como recurso para melhorar o fluxo de trabalho de catalogação em bibliotecas. Estudos de casos abordados referem-se à transferência de planilha de inventário interno em Coleções especiais para registros *MAchine-Readable Cataloging* (MARC) no *Online Computer Library Center* (OCLC), a importação de dados do Excel para o *Shocker Open Access Repository* (SOAR) baseado em *DSpace* (e.g. dados antropológicos, coleções de imagens biológicas e outros documentos de pesquisa). O autor discute a implementação e customização de ferramentas e programas externo para

aprimorar o fluxo de trabalho, incluindo a preparação de metadados, a normalização e a transformação. O artigo explora as possíveis opções e decisões necessárias para a construção de uma coleção de herbário no SOAR usando uma combinação dos elementos dos padrões *Dublin Core* (DC) e *Darwin Core* (DwC), como meio para promover o compartilhamento e a reutilização de recursos de metadados e melhorar o fluxo de trabalho de catalogação. O autor destaca a relevância da colaboração entre atores de diferentes áreas do conhecimento para aperfeiçoar esse trabalho.

Devarakonda, Palanisamy e demais autores (2010) analisam as funcionalidades e potenciais usos do Mercury, uma ferramenta de coleta, pesquisa e recuperação de metadados, baseada em código, originalmente desenvolvida pelo consórcio entre a NASA, o USGS e o Departamento de Energia nos EUA. O Mercury é um aplicativo que usa a abordagem da arquitetura orientada a serviços – *Service-Oriented Architecture* (SOA) – para capturar e gerenciar metadados relacionados a projetos de pesquisa. Também suporta a reutilização de metadados, permitindo pesquisas em vários padrões e especificações de metadados, incluindo XML, Z39.50 (projeto que deu origem ao Darwin Core), FGDC, Dublin Core, Darwin Core, EML e ISO-19115. Essa aplicação coleta metadados e dados importantes dos servidores de projetos parceiros distribuídos em todo o mundo e cria um índice centralizado, que permite pesquisar mais de 50.000 registros de metadados por meio de várias interfaces de usuário específicas. Trata-se, portanto, de um repositório centralizado de metadados com fontes de dados distribuídas que fornece resultados de pesquisa rápidos, pois permite que os provedores de dados publicizem a disponibilidade de seus dados, mas mantenham o controle e a propriedade completos desses dados.

Aguilar, Pan e demais autores (2010) desenvolveram um sistema de edição e gerenciamento de metadados voltado inicialmente para as ciências ambientais, mas apontam que há potencial de uso em vários domínios do conhecimento. Segundo os autores, o sistema emprega tecnologias XML gerando um design modular e distribuído que permite escalabilidade, flexibilidade, personalizações, bem como a possibilidade de adicionar novas funcionalidades. Trata-se de uma ferramenta que oferece opções ao editor para combinar campos de entrada em formulários online com tags amigáveis e determinar o formato de entrada dos dados, gerando código XForm, baseado no mecanismo Orbeon

XForms. Esse sistema pode gerar aplicativos de entrada de dados para qualquer esquema XML válido sem depender de informações personalizadas no esquema.

Dengler, Jansen e demais autores (2011) compilaram o índice global de bancos de dados de parcelas de vegetação – *Global Index of Vegetation-Plot Databases (GIVD)* – e analisaram os bancos de dados registrados até 30 de dezembro de 2010, considerando a estrutura e a qualidade dos dados. Foram identificadas cerca de 2,4 milhões de parcelas não sobrepostas. Os autores explicam que os dados de parcelas de vegetação no GIVD constituem um recurso importante para a pesquisa da biodiversidade, tanto pelo grande número de registros de ocorrência de espécies quanto pelo armazenamento de informações de coocorrência de espécies em pequena escala, combinados com dados ambientais estruturais e baseados em parcelas. Por fim, concluem que o GIVD oferece oportunidades significativas de descoberta e reuso dos dados para estudos em larga escala em ecologia, macroecologia e pesquisa de mudanças globais.

Baker, Johnson e Young (2011) apontam que as comunidades científicas relacionadas às ciências biológicas e paleontológicas incorporaram os avanços da informática separadamente, criando uma divisão de natureza tecnológica e sociológica. Nesse sentido, os autores propõem expandir a plataforma de depósito de dados da biodiversidade – *Scratchpads* – disponibilizada pelo Museu de História Natural de Londres para também incluir parâmetros relacionados ao tempo geológico, criando um serviço da web utilizando vocabulários da GBIF. Os autores também buscam tornar a solução flexível de forma a possibilitar adaptações às mudanças na definição de intervalos de tempo geológicos, bem como para facilitar consultas aos conjuntos de dados independentemente do formato dos dados de idade geológica utilizados.

Bachman, Moat e demais autores (2011) propõe o uso da GeoCAT, uma ferramenta de código-fonte aberto que permite realizar análises geoespaciais rápidas, para dar suporte as avaliações da lista de espécies ameaçadas de extinção. Segundo os autores, essa ferramenta foi desenvolvida para estruturar dados primários de ocorrência com referência espacial, considerando dois aspectos principais da faixa geográfica de um táxon: a extensão da ocorrência e a área de ocupação. Essas métricas fazem parte das categorias e critérios da Lista Vermelha da União Internacional para a Conservação da Natureza e dos Recursos Naturais – *International Union for Conservation of Nature's Red List of Threatened Species*

(IUCN). Os autores apontam que no ambiente familiar do *Google Maps*, os usuários do GeoCAT podem combinar e visualizar rápida e facilmente dados de várias fontes, como GBIF, Flickr e Scratchpads, além de dados de ocorrência gerados por usuários. As saídas, incluindo os resultados, dados e parâmetros usados para análise, são armazenadas em um arquivo GeoCAT que pode ser facilmente recarregado ou compartilhado com os colaboradores. Por fim, os autores entendem que o uso do GeoCAT é o primeiro passo para automatizar o processo de manipulação de dados da avaliação da Lista Vermelha.

Huettmann, Artukhin e demais autores (2011) apresentam as ferramentas e processos realizados para a compilação, quantificação e resumo dos dados de presença de espécies de aves marinhas para o norte do círculo ártico (> 66 graus de latitude norte) e o período sem gelo (verão). Para o cruzamento taxonômico de espécies compatíveis e aplicações subsequentes (e.g. filogenias), a normalização dos nomes de espécies foi embasada em vários bancos de dados taxonomicamente válidos (e.g. *Integrated Taxonomic Information System* (ITIS), *AviBase*, código de 4 letras do *American Ornithological Union* (AOU), *The British List 2000*, *taxonomic serial numbers* (TSNs), *World Register of Marine Species* (WORMS) e APHIA ID). Com base no algoritmo *RandomForest* de mineração de dados e *machine learning*, os autores identificaram 26 camadas de Sistemas de Informações Geográficas (GIS) disponíveis ao público, que permitiram o desenvolvimento de 27 modelos preditivos de aves marinhas com base em arquivos de dados públicos de acesso aberto, como da GBIF, *North Pacific Pelagic Seabird Database* (NPPSD) e a base de dados do *Programme Intégré des Recherches sur les Oiseaux Pélagiques* (PIROP) integrada ao *Ocean Biodiversity Information System* (OBIS-Seamap). Segundo os autores, esses modelos preditivos apresentaram um desempenho razoável – ruim apenas algumas espécies costeiras com amostras baixas e ótimo para muitas espécies pelágicas. Em conformidade com as políticas de dados do *International Polar Year* (IPY)²¹ e iniciativas semelhantes, os autores documentaram os dados e os modelos com metadados do FGDC NBII e disponibilizaram publicamente, pois entendem que têm potencial para avaliações da sustentabilidade, sinergia e explorações científicas diante da crise global da biodiversidade, dos oceanos e do Ártico.

²¹ Esforço internacional e colaborativo tendo como tema central as regiões polares que realiza eventos anuais com representantes de todo o mundo.

Devarakonda, Palanisamy e demais autores (2011) apontam que há um consenso quanto à necessidade de armazenar e arquivar dados de pesquisa de forma digital e pública, mas que a preservação em longo prazo desses recursos geralmente requer apoio institucional, como aqueles direcionados a comunidades científicas de prática específicas. Os autores consideram que os conjuntos de dados de pesquisa geralmente são úteis para várias comunidades, mas em geral apresentam normas diferentes para estrutura e semântica de dados e metadados, tornando relevante a adoção de padrões eficazes de compartilhamento para que os usuários possam descobrir e recuperar os dados. Nesse sentido, os autores recomendam o uso do Mercury como ferramenta, devido à apresentação coerente de informações em vários padrões. Os autores apontam que o Mercury foi ampliado para permitir a coleta e distribuição de metadados usando o *Open Archive Initiative Protocol for Metadata Harvesting* (OAI-PMH). Por fim, os autores descrevem lições aprendidas ao utilizar diversos recursos do Mercury, considerando vários formatos de metadados e implementações OAI-PMH.

Veen, Reenen e demais autores (2012) destacam que na última década, catalisadas pelo crescimento da internet, várias tecnologias para disseminação e integração de dados foram desenvolvidas e aplicadas em projetos como GBIF, KNB, BioCASE e a *British National Biodiversity Network* (NBN). Os autores analisam o banco de dados nacional de flora e fauna da Holanda – National Database of Flora and Fauna (NDFF) – com aproximadamente 40 milhões de registros de observação e cobre uma ampla variedade de espécies. O NDFF usa um modelo de dados padronizado e semanticamente integrado para combinar efetivamente dados de observação de espécies de vários tipos. Os autores avaliam o modelo de dados do NDFF em comparação com os esquemas do Darwin Core, ABCD e o modelo *Recorder 2000* usado pelo NBN.

Schmidt-Kloiber, Moe e demais autores (2013) apontam que o papel dos metadados é cada vez mais valorizado para aperfeiçoar a descoberta e reuso dos conjuntos de dados nas ciências ecológicas e se tornou cada vez mais importante que pesquisadores. Em seu estudo, os autores avaliam se os padrões de metadados geralmente usados (e.g. Darwin Core, ISO 19115, CSDGM, EML) atendem às necessidades de do banco de metadados – *metadatabase* – criado para o projeto *Water bodies in Europe: Integrative Systems to assess Ecological status and Recovery* (WISER). O *metadatabase* WISER contém informações sobre

114 conjuntos de dados (referente a 22 rios, 71 lagos, 1 curso de água doce geral e 20 em zonas costeiras/de transição). Os autores também avaliam se a documentação simples de metadados é suficiente permitir o reuso de um conjunto de dados, pois entendem que ainda há relutância dos pesquisadores em publicar metadados e dados primários de pesquisa por diversos motivos (e.g. restrições de tempo; restrições financeiras; preocupação com uso indevido de dados; necessidade de abandono dos direitos de propriedade intelectual). Os autores concluem recomendando a publicação de metadados e destacam que isso traz grandes vantagens aos pesquisadores, pois torna possível a descoberta dos conjuntos de dados por outros cientistas, tornando o trabalho mais visível e com maior alcance.

Graham e Kennedy (2014) apresentam a ferramenta *Visual Exploration of SPECies-referenced Repositories* (VESPER), que permite explorar os conjuntos de dados do *Darwin Core Archive* (DwC-A). Essa ferramenta foi desenvolvida para apoiar pesquisadores na redução do tempo e esforço despendidos para verificar a qualidade dos conjuntos de dados. Segundo os autores, embora já existam ferramentas que examinam minuciosamente a presença de dados e a correção de estrutura de dados em relação ao esquema DwC-A, em geral essas soluções não fornecem nenhuma percepção da qualidade do conjuntos de dados. Já o Vesper analisa e exibe os conjuntos de dados DwC-A nas três dimensões fundamentais (taxonômica, geográfica e temporal) e permite a visualização detalhada de cada um desses aspectos dos dados. Isso permite aos pesquisador avaliar se um conjunto de dados é adequado para as tarefas ou análises de cada nova pesquisa. Além disso, um provedor de dados pode identificar em que pontos há problemas de qualidade nos dados (e.g. se contém dados obviamente incorretos, como a inversão clássica da longitude), permitindo descobrir se estão restritos a subpartes identificáveis do conjunto de dados. O Vesper pode manipular conjuntos de dados do com um milhão de entidades devido ao uso criterioso da filtragem, pois muitos dos tipos de dados em registros individuais não são necessários para julgar a distribuição geográfica, temporal ou taxonômica e a extensão de um conjunto de dados. Assim, muitos dos campos mais detalhados do arquivo podem simplesmente ser ignorados durante um estágio inicial de descompactação de dados.

Robertson, Döring e demais autores (2014) apontam que diante da crise global de biodiversidade, para medir a magnitude e a taxa em que as mudanças ocorrem é fundamental ter acesso a dados de biodiversidade de todo o mundo que estejam

organizados em formato digital em uma estrutura que permita a fácil descoberta. Os autores destacam que o registro da representação digital coerente da biodiversidade requer a integração de dados que historicamente são analógicos, dispersos e heterogêneos. Nesse sentido, os autores apresentam a ferramenta *Integrated Publishing Toolkit* (IPT), desenvolvida com apoio da GBIF para apoiar a publicação padronizada de conjuntos de dados de biodiversidade. Os autores destacam duas funções principais do IPT: (1) aprimorar a interoperabilidade utilizando o padrão Darwin Core para codificar os conjuntos de dados (e.g. listas de verificação de ocorrência de espécies existentes (*checklists*); registros de coleções de história natural ou de observações de campo); e (2) publicar e arquivar dados e metadados para uso amplo em um Darwin Core Archive. Os autores discutem as necessidades que levaram ao desenvolvimento do IPT em resposta à contribuição da comunidade, bem como a constante evolução da ferramenta de forma a aperfeiçoar a interoperabilidade, descoberta e mobilização de novos tipos de dados. Por fim, concluem apresentando como o IPT impactou positivamente a comunidade de pesquisa em biodiversidade, permitindo aprimorar a publicação de conjuntos de dados em periódicos tradicionais como é feito com artigos convencionais e *data papers*.

Burkhardt, Russel e demais autores (2014) apresentam o Edaphobase, um *data warehouse* não comercial desenvolvido na Alemanha e integrado a GBIF, que contém dados referentes aos organismos do solo. O Edaphobase combina dados sobre taxonomia, zoogeografia e ecologia desses organismos de maneira abrangente. Os dados são derivados de publicações científicas (coletados das publicações de pesquisas alemãs nos 60 anos anteriores), resultados de estudos de campo não publicados e coleções de museus ou instituições de pesquisa. Segundo os autores, essa solução oferece uma ampla gama de ferramentas para inclusão de dados e exploração de dados (e.g. ferramenta GIS e análise semiautomática de literatura), que permitem desde consultas simples até análises mais sofisticadas para diferentes grupos de dados. De forma a exemplificar o potencial de uso, os autores apresentam exemplos específicos de análises de conjuntos de dados de várias fontes (e.g. preferências de habitat por espécies; padrões de distribuição; ou influências ambientais nas densidades populacionais).

Hock, Payn e Heaphy (2014) apontam que o setor de florestas plantadas da Nova Zelândia é o terceiro maior produtor de exportações do país e, portanto, há uma crescente

demanda por informações científicas sobre a sustentabilidade dos produtos florestais. Os mecanismos internacionais de controle (e.g. Avaliação Global de Recursos Florestais; Processo Montreal, que é um acordo internacional sobre manejo florestal sustentável; e os esquemas de certificação florestal) exigem relatórios sobre o estado e as tendências nas florestas plantadas. Segundo os autores, a disponibilidade de dados para esses relatórios é variável, pois há diversos problemas encontrados ao avaliar a aplicabilidade e a utilidade das fontes de dados ambientais nacionais existentes, incluindo a compatibilidade de informações sobre atributos, resolução, cobertura de dados espaciais e a extensão dos dados existentes. Diante desses problemas, os autores propõem melhorias na estruturação dos dados a serem utilizadas no novo portal de informações em desenvolvimento sobre florestas plantadas da Nova Zelândia.

Skevakis, Tsinaraki e demais autores (2014) descrevem o MoM-NOCS, uma aplicação desenvolvida pra auxiliar as comunidades de pesquisa na coleta e compartilhamento de observações multimídia de objetos ou eventos usando dispositivos. Segundo os autores, as observações são automaticamente associadas aos metadados contextuais que permitem a visualização sobreposta em mapas 2D ou 3D. Além disso, os autores informam que é possível gerenciar essas observações e permitir que outros usuários contribuam com notas. Essa forma de registro colaborativo de achados de campo permite ampliar a precisão dos dados e, conseqüentemente facilita a geração de conhecimento. Os autores destacam que o MoM-NOCS é complementar e apresenta interoperabilidade com sistemas gerenciados por museus de história natural como MMAT e sistemas de gerenciamento de metadados da biodiversidade como BIOCASE e GBIF.

Kandel, Huettmann e demais autores (2015) compilaram e disponibilizaram publicamente em um repositório institucional (usando dSPACE) conjunto de dados de presença do panda vermelho (*Ailurus fulgens*). Essa é uma espécie globalmente ameaçada que vive na região multinacional do Hindu-Kush no Himalaia. Segundo os autores há uma tendência de declínio da população devido a pressões antropogênicas, com a previsão de agravamento devido mudanças climáticas. Os autores relatam que faltam informações quantitativas e transparentes sobre o nicho ecológico dessa espécie nas vastas e complexas regiões em que há registro de presença. Integrar essas informações é crucial para identificar e monitorar populações, bem como para compreender a evolução biogeográfica para um

monitoramento eficiente e para priorizar ações de preservação. Os autores explicam que desenvolveram o primeiro modelo preditivo para a espécie na região e a plataforma de dados do *Centre for Integrated Mountain Development* (ICIMOD) integrada à rede GBIF. Para o desenvolvimento do modelo preditivo calcularam a média de todos os modelos disponíveis em soluções de Sistema de Informações Geográficas (SIG) e utilizaram algoritmos de mineração de dados e de *machine learning* (e.g. *high-performance classification*; TreeNet; Árvores de regressão; *Random Forest*; *multivariate adaptive regression splines* - MARS).

Pfaff, König-Ries e demais autores (2015) relatam que embora os periódicos científicos exijam cada vez com mais frequência os conjuntos de dados primários utilizados nas pesquisas como um pré-requisito para uma publicação, em geral, há perda do que o vínculo entre os artigos e os dados. Os autores consideram que o compartilhamento dos dados primários têm o potencial de aumentar significativamente a precisão e a reprodutibilidade das descobertas científicas e isso pode ser aperfeiçoado ao documentar as etapas de pós-processamento da vinculação desses dados. Nesse sentido, o autores apresentam o pacote rBEFdata R como complemento da plataforma colaborativa de gerenciamento de dados BEFdata. Esse pacote permite pesquisar dados e integrar a pesquisa a tesouros externos para melhorar a descoberta dos dados, também possibilita baixar e importar dados e metadados para o R, inclusive com recurso de download em lote disponibilizado pelo BEFdata. Esse recurso do BEFdata também permite agrupar dados e metadados primários. A funcionalidade de upload do pacote R em combinação com o mecanismo do portal BEFdata permite anexar produtos e scripts de dados derivados diretamente do R, abordando, assim, os principais aspectos da documentação do pós-processamento de dados. Por fim, os autores utilizam um exemplo de análise ecológica e discutem o potencial da documentação de pós-processamento de dados, que vincula os dados primários a produtos e conhecimentos derivados desses dados.

Minadakis, Marketakis e demais autores (2016) apontam que os dados da biodiversidade são caracterizados por seu caráter interdisciplinar, pela variedade extremamente ampla de tipos e estruturas de dados e pela variedade de conceitos semânticos que englobam. Além disso, relatam que há uma infinidade de fontes de dados diferentes, fornecendo recursos para a mesma informação de maneira heterogênea. De forma geral, os dados sobre biodiversidade permanecem desconectados e amplamente

distribuídos entre diferentes formatos. Nesse sentido, os autores apresentam ações realizadas para integrar dados da biodiversidade no portal *LifeWatch* Grécia (catalogação e publicação de todas as informações relevantes de metadados do domínio da biodiversidade grego; definição de modelos e integração de dados de fontes heterogêneas; desenvolvimento de ferramentas para descoberta de dados. Este trabalho foi explorado, avaliado e cientificamente confirmado pela comunidade da biodiversidade através dos serviços prestado).

Schindel, Miller e demais autores (2016) apresentam o registro global de repositórios de biodiversidade – Global Registry of Biodiversity Repositories (GRBio). Trata-se de um agregador internacional de dados sobre coleções de história natural que permite conectar registros das amostras físicas às suas citações na literatura, bem como aos registros digitalizados em bancos de dados. Essa ferramenta também permite aos pesquisadores localizar mais facilmente recursos em coleções de história natural e identificar as políticas de acesso e empréstimos das amostras. Os autores destacam que os periódicos científicos têm papel fundamental para estimular os pesquisadores a contribuir adequadamente e citar as coleções em seus trabalhos, apontando as políticas estabelecidas pelos periódicos *ZooKeys* e a *Pensoft Publishing* como precursores na colaboração para esse avanço. O GRBio fornece informações de contato e endereço, características das instituições e coleções usando vocabulários controlados e descrições de texto livre, bem como identificadores exclusivos para cada instituição e registro de coleção. Inclui também campos de texto para políticas de empréstimo e usa uma variedade de outros descritores. Por fim, os autores destacam que o avanço dos padrões de metadados, em especial do *Darwin Core* tem sido fundamental para aperfeiçoar o GRBio.

Leigh, Laporte e demais autores (2017) descrevem que a pesquisa e o gerenciamento ecológico contemporâneo demandam compreensão das interações entre biodiversidade, processos ecossistêmicos e os serviços ecossistêmicos diante das mudanças globais. Isso é particularmente importante para a biodiversidade de água doce no contexto das mudanças no regime de vazão dos rios. Os rios intermitentes (que fluir e secam por determinados períodos do ano) são ecossistemas dinâmicos e predominantes globalmente. Segundo os autores, as pesquisas sobre rios intermitentes estão se expandindo rapidamente com a era do *big data*, entretanto, os dados permanecem amplamente fragmentados, limitando

análises a estudos caso a caso. Em resposta a esses problemas, os autores apresentam as soluções desenvolvidas no projeto *Intermittent River Biodiversity Analysis and Synthesis* (IRBAS), que envolveu a coleta, análise e síntese dos dados de todo o mundo sobre a biodiversidade e as características ambientais dos rios intermitentes. O IRBAS abriga mais de 2000 amostras de dados coletadas de seis países em três continentes, descrevendo principalmente os taxa de invertebrados aquáticos que habitam rios intermitentes durante as fases hidrológicas. Dessa forma, essa solução integra e fornece acesso gratuito a esses dados globais, contribuindo para evolução do conhecimento teórico e aplicado sobre os sistemas fluviais. Trata-se de um portal voltado ao armazenamento, padronização e descoberta, com ferramentas que facilitam a coleta, síntese e análise de dados para elucidar fenômenos relacionados à biodiversidade de rios.

Em uma publicação desenvolvida com contribuições de pesquisadores de mais de 500 instituições distribuídas globalmente (HUDSON, NEWBOLD, *et al.*, 2017) são apresentados os resultados do projeto *Projecting Responses of Ecological Diversity In Changing Terrestrial Systems* (PREDICTS). Esse projeto reuniu, a partir de estudos publicados, um grande banco de dados razoavelmente representativo de amostras da biodiversidade de vários locais. A partir dessa base de dados foram coletadas evidências para desenvolver modelos estatísticos globais e regionais sobre como a biodiversidade responde considerando a natureza ou intensidade de impactos humanos relacionados ao uso da terra. Em sua versão de 2016, o banco de dados do PREDICTS disponibilizou mais de 3,2 milhões de registros amostrados em mais de 26.000 locais e representando mais de 47.000 espécies. O artigo descreve exemplos de como esse banco de dados pode ajudar a responder a perguntas em ecologia e biologia da conservação. A publicação aponta que o PREDICTS é o maior e mais representativo banco de dados geográfico e taxonomicamente representativo das comparações espaciais da biodiversidade que foram coletadas e, por isso, fundamental para potencializar os esforços internacionais em pesquisas voltadas a modelar e entender o *status* global da biodiversidade.

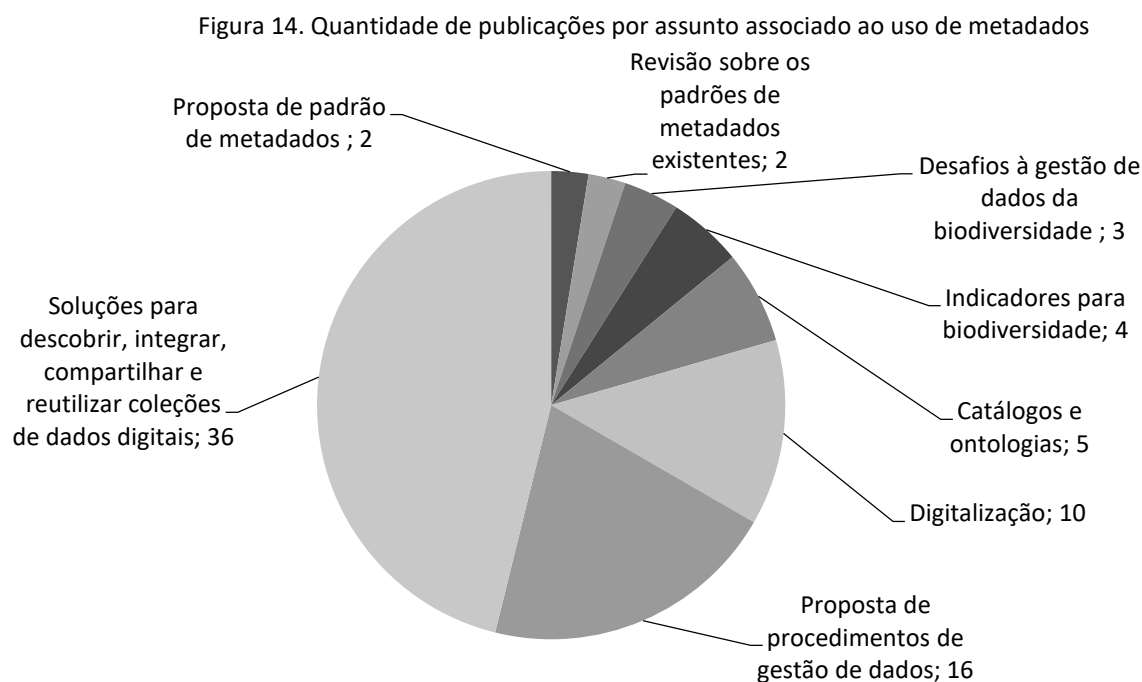
Singer, Love e Page (2018) apontam que mudanças recentes na infraestrutura de tecnologia da informação de instituições e nos métodos de armazenamento de dados de coleções melhoraram drasticamente a acessibilidade de dados por meio de bancos de dados digitais e agregadores de dados. Os autores analisam as coleções de peixes digitalizadas nos

EUA depositadas no iDigBio usando o pacote *ridigbio* e o aplicativo da web *fishfindR*. O iDigBio é um agregador de dados de coleções depositadas de forma voluntária, por isso, os autores apontam que nem todas as coleções têm os dados compartilhados no iDigBio. Entretanto, as 38 das 143 coleções de peixes conhecidas nos EUA que estão no iDigBio concentram a maioria das amostras de peixes. Segundo os autores, essa análise permitiu compreender como as informações dos agregadores de dados, neste caso iDigBio, podem ser extraídas e analisadas. Os autores concluem que apesar de fornecer dados para pesquisas, educação e várias atividades de extensão, os dados acessíveis por meio de agregadores de dados também podem ser usados para desenvolver relatórios precisos e atualizados de informações sobre coleções institucionais, que podem orientar os esforços direcionados a tornar os dados mais transparentes para os usuários. Por fim, os autores destacam que esses relatórios são fundamentais para compreender a disponibilidade digital das coleções para alavancar apoio financeiro e institucional.

Velásquez-Tibatá, Olaya-Rodríguez e demais autores (2019) apontam que as plataformas de acesso aberto e disseminação dos dados da biodiversidade enfrentam o desafio de incorporar parâmetros de modelos preditivos, que subsidiam decisões com base em ocorrência de espécies e dados ecológicos. Nesse sentido, os autores descrevem o BioModelos, um sistema com abordagem voltada à modelagem que recebe contribuições de rede de especialistas para o desenvolvimento de modelos de distribuição de espécies, avaliando a qualidade dos dados de ocorrência, identificando variáveis ambientais potencialmente limitantes, estabelecendo áreas de ocorrência de espécies e validando previsões de modelagem qualitativamente. Os modelos desenvolvidos por meio do BioModelos, depois de validados por especialistas, são disponibilizados gratuitamente e publicamente, aumentando o uso em aplicações de conservação. Os autores trabalham com essa abordagem na Colômbia desde 2013 e já contam com uma rede de quase 500 especialistas. Esses especialistas colaboram para melhorar o conhecimento sobre a distribuição de espécies e ajudam no desenvolvimento de vários produtos de apoio à decisão, como manuais nacional de avaliação de riscos e compensação de biodiversidade. Os autores concluem que o BioModelos é um exemplo de operacionalização de uma ferramenta essencial para a gestão da biodiversidade na Colômbia, que por meio da de uma infraestrutura de pesquisa aprimora a qualidade dos dados abertos.

4.2.2.4 Síntese integrativa dos dados coletados nos artigos

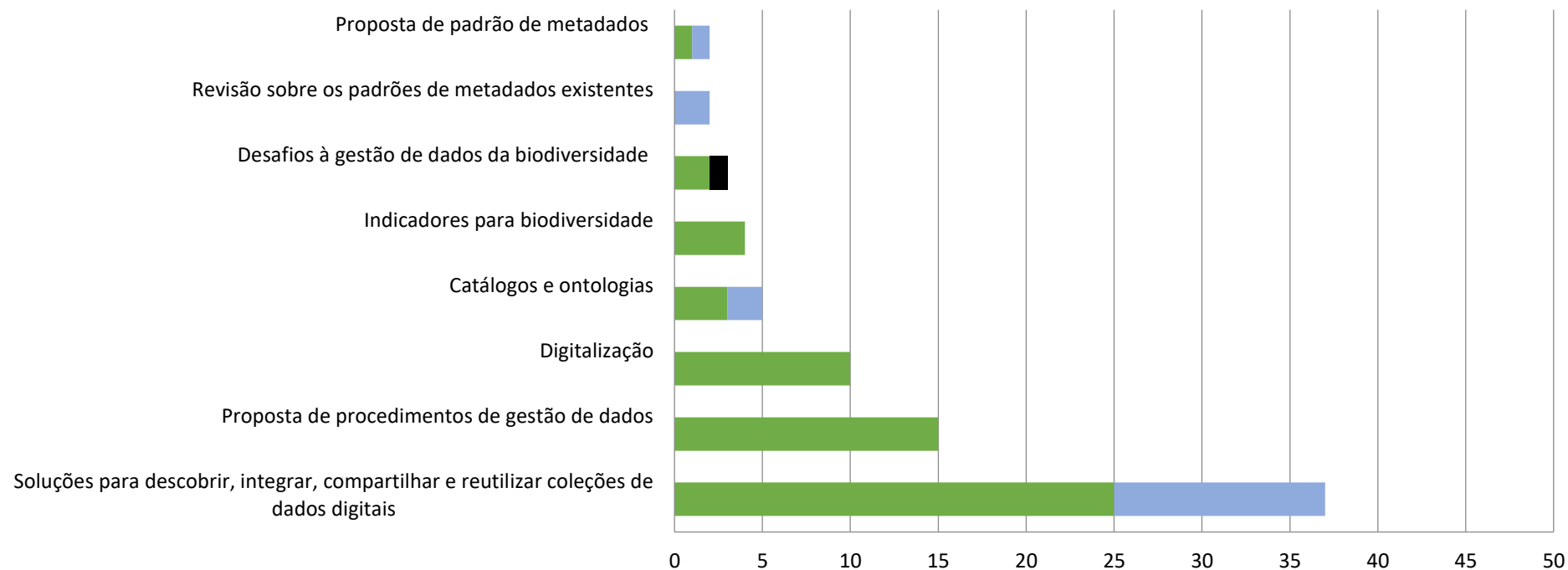
A análise dos 78 artigos relacionados ao uso de metadados na temática biodiversidade permitiu identificar 8 assuntos associados considerados centrais nessas pesquisas: (i) proposta de padrão de metadados ou extensão de padrão; (ii) desafios à gestão de dados da biodiversidade; (iii) revisão comparativa dos padrões de metadados existentes; (iv) indicadores para biodiversidade; (v) catálogos e ontologias; (vi) digitalização de coleções de história natural e anotações de campo; (vii) procedimentos de gestão de dados; e (viii) soluções computacionais (e.g. algoritmos, softwares, aplicações, portais, bases de dados) voltadas à descoberta, integração, compartilhamento e reutilização de conjuntos de dados digitais. A distribuição da quantidade de artigos por cada um dos assuntos identificados é apresentada na figura 14.



Fonte: Elaboração própria.

Já a figura 15 ilustra a distribuição da quantidade de publicações relacionadas a cada assunto em cada uma das três décadas analisadas. Isso permite visualizar a evolução na diversificação dos assuntos ao longo desses períodos, bem como analisar comparativamente com que frequência esses assuntos são registrados em cada década.

Figura 15. Quantidade de publicações por década e por assunto associado ao uso de metadados



	Soluções para descobrir, integrar, compartilhar e reutilizar coleções de dados digitais	Proposta de procedimentos de gestão de dados	Digitalização	Catálogos e ontologias	Indicadores para biodiversidade	Desafios à gestão de dados da biodiversidade	Revisão sobre os padrões de metadados existentes	Proposta de padrão de metadados
■ década 2010	25	15	10	3	4	2	0	1
■ década 2000	12	0	0	2	0	0	2	1
■ década 1990	0	0	0	0	0	1	0	0

Fonte: Elaboração própria.

A década de 1990 foi marcada por acontecimentos relevantes para a gestão dos dados de pesquisa de forma geral, bem como para às questões relacionadas aos dados da biodiversidade. Houve um movimento de institucionalização da gestão de dados de pesquisa, com formalização de normativas. Entretanto, de forma geral, esses acontecimentos começaram a refletir impactos positivos nas pesquisas apenas na década seguinte. Há apenas uma publicação identificada já no final da década de 1990 relacionada ao uso de metadados na temática biodiversidade que levanta os desafios para descoberta, compartilhamento e reuso dos conjuntos de dados da biodiversidade na CHINA (HAIGEN XU, ZHENNING GAO, *et al.*, 1999). Essa pesquisa já representa uma resposta à demanda por desenvolvimento de mecanismos de *clearinghouse* (CMH) em atendimento à CDB (1992).

Na década de 2000 há duas publicações voltadas identificar e analisar comparativamente os padrões de metadados para biodiversidade (HAASA, HENJUMB, *et al.*, 2003; CHULIN MENG, 2004). Isso se resulta do processo de discussão e desenvolvimento desses padrões que estava em andamento no período.

Há também três publicações referentes à criação de padrão de metadados. A primeira da década de 2000, apresenta o padrão de metadados para biodiversidade desenvolvido pela China (HAIGEN XU, YIMIN LI e HUI DING, 2004) utilizado para consolidar o CHM dos dados de pesquisa dessa área. As outras duas publicações, mais relacionadas à extensão de padrões existentes, tratam da necessidade de esquemas de metadados para atender áreas do conhecimento mais específicas e para aperfeiçoar a integração de diferentes parâmetros de dados (RASAIAH, BELLMAN, *et al.*, 2015; POOTER, APPELTANS, *et al.*, 2017).

Desde a década de 1990, grande parte dos artigos analisados apontam secundariamente os desafios relacionados à gestão, organização e comunicação dos dados de pesquisa da biodiversidade (BERENDSOHN e SELTMANN, 2010; BAIRD, 2010; VOLLMAR, MACKLIN e FORD, 2010; HAIGEN XU, ZHENNING GAO, *et al.*, 1999; KWANG-TSAO SHAO, CHING-I PENG, *et al.*, 2007) e abordam outros assuntos de forma central. Os problemas mais comuns relatados referem-se às discrepâncias entre as descrições dos diversos conjuntos de dados e a ausência de padrões na estruturação dos sistemas e portais, que dificultam a interoperabilidade e compartilhamento dos dados, bem como falta de incentivos aos pesquisadores para a publicação dos dados. Mas na década de 2010 há dois artigos que

tratam especificamente desses desafios como tema central, abordando questões relacionadas à descoberta, interoperabilidade, compartilhamento e reuso dos dados de pesquisa de biodiversidade (DAVIS, TENOPIR, *et al.*, 2014; BENDIX, NIESCHULZE e MICHENER, 2012).

Considerando as 78 publicações analisadas, o assunto mais frequente (35 artigos) refere-se ao desenvolvimento ou uso de soluções computacionais (aplicações, *softwares*, bases de dados, portais e algoritmos) para descobrir, integrar, compartilhar e reutilizar coleções de dados digitais. A partir de 2000, houve um aumento significativo das pesquisas voltadas ao desenvolvimento dessas soluções, que em geral possuíam esquemas próprios de estruturação dos metadados voltados a solucionar demandas de análises específicas. A partir da final da década de 2000 e durante toda a década de 2010 se intensificam as pesquisas voltadas a comparar muitas das soluções existentes e desenvolver recursos para promover a interoperabilidade de bases de metadados apresentadas em formatos diversos. Observa-se também que houve um movimento em busca de adequar a descrição dos conjuntos de dados aos padrões de metadados mais recomendados e utilizados pelas comunidades científicas (e.g. EML, DWC e ABCD). Com isso surgiram nessa última década diversas pesquisas voltadas a propor procedimentos para a descrição e gestão dos conjuntos de dados.

Durante a década de 2010 destacam-se a ampliação da diversidade de assuntos associados ao uso de metadados na temática biodiversidade e o aumento significativo da quantidade de publicações (60 nesse período), que representa mais de dois terços do total de 78 analisadas. Também considerando o total analisado neste estudo, essa década concentra a maioria dos artigos relacionados às soluções computacionais (1º assunto mais frequente) e registra o surgimento de publicações associadas à proposição de procedimentos de gestão de dados de pesquisa da biodiversidade (2º assunto mais frequente) e à digitalização de coleções de história natural (3º assunto mais frequente).

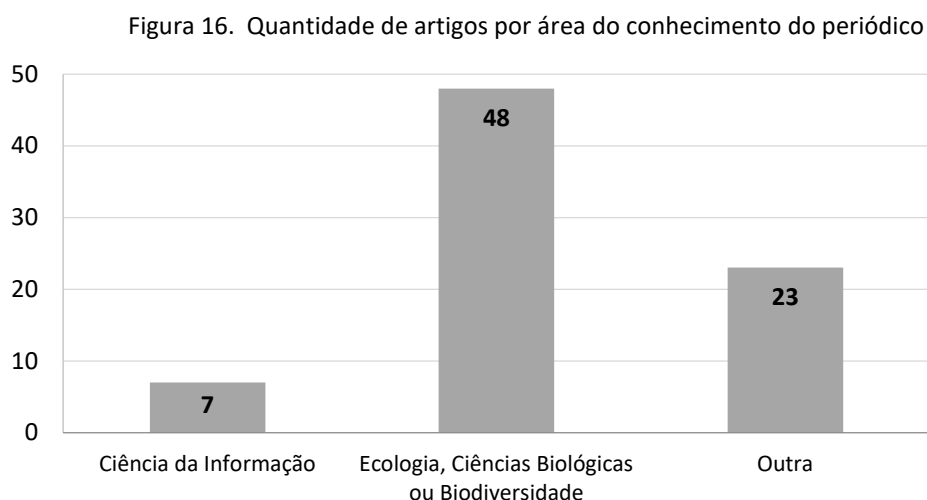
Entre as publicações sobre procedimentos de gestão de dados, destacam-se dois estudos. O primeiro referente à expansão das políticas de depósito de dados de pesquisa – *Public Data Archiving* (PDA) – e os desafios para qualificar a acessibilidade, descoberta e reuso dos dados de pesquisa vinculados aos artigos científicos (ROCHE, KRUIK, *et al.*, 2015). E o segundo estudo, apresenta os princípios FAIR (WILKINSON, DUMONTIER, *et al.*, 2016)

como uma forte resposta aos desafios apontados por diversos autores relacionados a dificuldades de descoberta e reuso de dados de pesquisa.

Os princípios FAIR representam uma resposta aos desafios na gestão de dados de pesquisa já conhecidos e discutidos por diversas comunidades científicas. Por isso, apesar de da formalização recente (WILKINSON, DUMONTIER, *et al.*, 2016), esses princípios já são referência em processo de absorção por muitas comunidades de pesquisa. Entre outras iniciativas, a rede GBIF aderiu ao movimento de aperfeiçoamento da gestão de dados de pesquisa com base na adequação aos princípios FAIR desde 2017 (GBIF, 2016) e passou a promover eventos relacionados ao tema.

Outro ponto relevante nas últimas duas décadas, que demonstra evolução na gestão e organização dos dados de pesquisa da biodiversidade, refere-se às pesquisas associadas à proposição e recomendação de uso de catálogos e ontologias. Isso demonstra o avanço de em busca de soluções semânticas que permitam análises mais complexas e melhor aproveitamento dos dados de pesquisa para subsidiar decisões relacionadas à gestão dos recursos ambientais e da biodiversidade (WALLS, GURALNICK, *et al.*, 2014; LUCARINI, GIGANTE, *et al.*, 2014; GURALNICK, CELLINESE, *et al.*, 2015; BOSCH, LENDVAI, *et al.*, 2009; MADIN, BOWERS, *et al.*, 2008). Além disso, foram identificadas três publicações que resultam de pesquisas desenvolvidas por brasileiros (BARROSA, LAENDERA, *et al.*, 2008; TORRES, MEDEIROS, *et al.*, 2006; PÔÇAS, GONÇALVES, *et al.*, 2014).

Os artigos analisados estão vinculados a periódicos científicos de diferentes áreas. A figura 16 apresenta a distribuição dessas publicações por área do conhecimento.

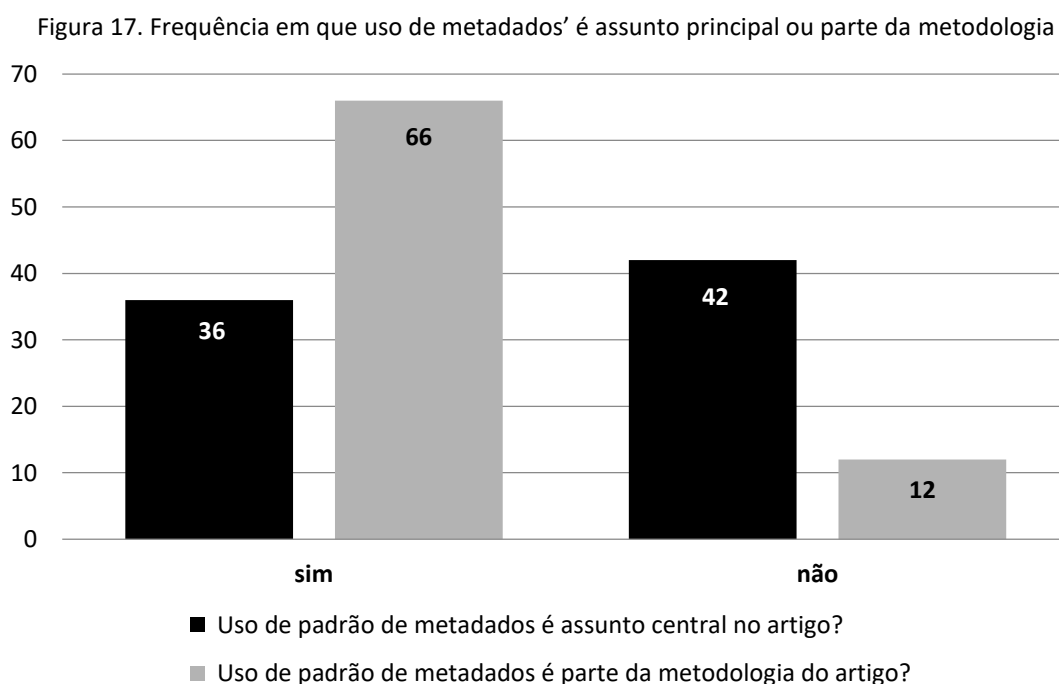


Fonte: Elaboração própria.

A maior parte das publicações analisadas foi publicada em periódicos ligados às Ciências Biológicas (48 artigos). Também há uma parcela representativa vinculada a periódicos de temáticas multidisciplinares (23 artigos - e.g. Ciências da Terra; Ciências de Dados; Gestão do Meio Ambiente; Soluções computacionais aplicadas). Por fim, foram identificadas sete publicações em periódicos ligados às Ciências da Informação, das quais três estão associadas à proposição de procedimentos de gestão de dados (NIU, 2016a; NIU, 2016b; SKEVAKIS, MAKRIS, *et al.*, 2014), duas às soluções computacionais (DENG, 2010; TORRES, MEDEIROS, *et al.*, 2006) e duas à revisão de padrões de metadados da biodiversidade (HAASA, HENJUMB, *et al.*, 2003; CHULIN MENG, 2004).

Essa distribuição em periódicos de diversas áreas do conhecimento confirma a importância das buscas por artigos abrangerem a maior quantidade possível de bases da literatura científica também vinculadas a diferentes áreas do conhecimento. Isso permitiu alcançar, em especial, os periódicos vinculados a áreas multidisciplinares, que estão vinculados a quase um terço das publicações analisadas neste estudo.

Considerando o “uso de metadados” como tema central desse estudo, foi verificado também quantas publicações apresentavam esse termo como parte do assunto principal da pesquisa e quantas apresentavam como parte da metodologia – figura 17 .

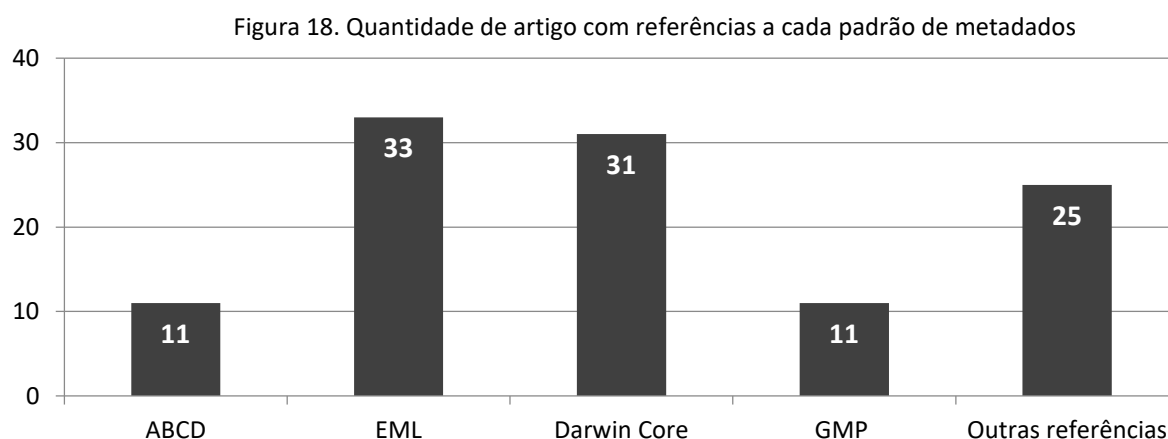


Fonte: Elaboração própria.

Destaca-se que o uso de metadados como tema central ou como parte da metodologia em um mesmo artigo não são aspectos excludentes, mas complementares. Um mesmo artigo, portanto, pode registrar sim ou não para os dois aspectos simultaneamente.

A maioria dos artigos analisados apresenta o uso de metadados como parte da metodologia estabelecida para a pesquisa desenvolvida, seja qual for o tema associado. Além disso, diante do fato de que a maioria das publicações são vinculadas a periódicos relacionados à temática biodiversidade (figura 16), é relevante notar que quase metade dos artigos analisados apresenta o uso de metadados como tema central do objeto da pesquisa (figura 17). Isso demonstra a relevância que a gestão de dados e o uso de metadados vêm alcançando nas comunidades científicas e instituições ligadas às pesquisas sobre biodiversidade.

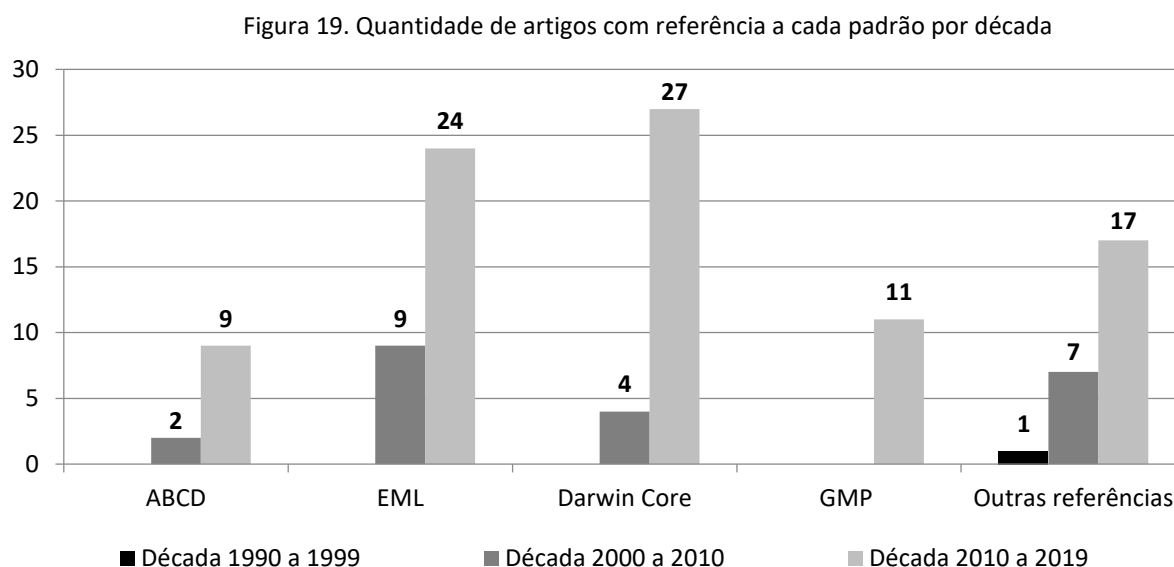
Na etapa I deste estudo, foi identificado que os padrões de metadados recomendados e mais utilizados para descrição de dados de pesquisa da biodiversidade são o EMLA, ABCD, DwC e o perfil de metadados GMP (*GBIF metada profile*). Durante a análise dos artigos foram identificadas evidências de uso dos padrões de metadados, conforme apresentado na figura 18.



Fonte: Elaboração própria.

O padrão EML surgiu em 1997, mas apenas a partir de 2002 passou a ser mantido de forma colaborativa com a comunidade, com a última versão atualizada em 2019. Este é o padrão de metadados mais mencionado nos artigos analisados. Seguido do padrão DWC, que começou a ser desenvolvido em 1999, mas foi ratificado pela TWGD em 2009, quando passou a contar com maior colaboração da comunidade. Já o ABCD e o GMP são mais

recentes. O padrão ABCD começou a ser desenvolvido em 2001 e foi ratificado pela TWGD em 2005. A primeira versão do GMP foi documentada em 2011 e atualizada em 2017. Considerando o tempo de existência para a consolidação desses padrões, é interessante analisar a quantidade de artigos relacionada a cada um deles por década – figura 19.



Fonte: Elaboração própria.

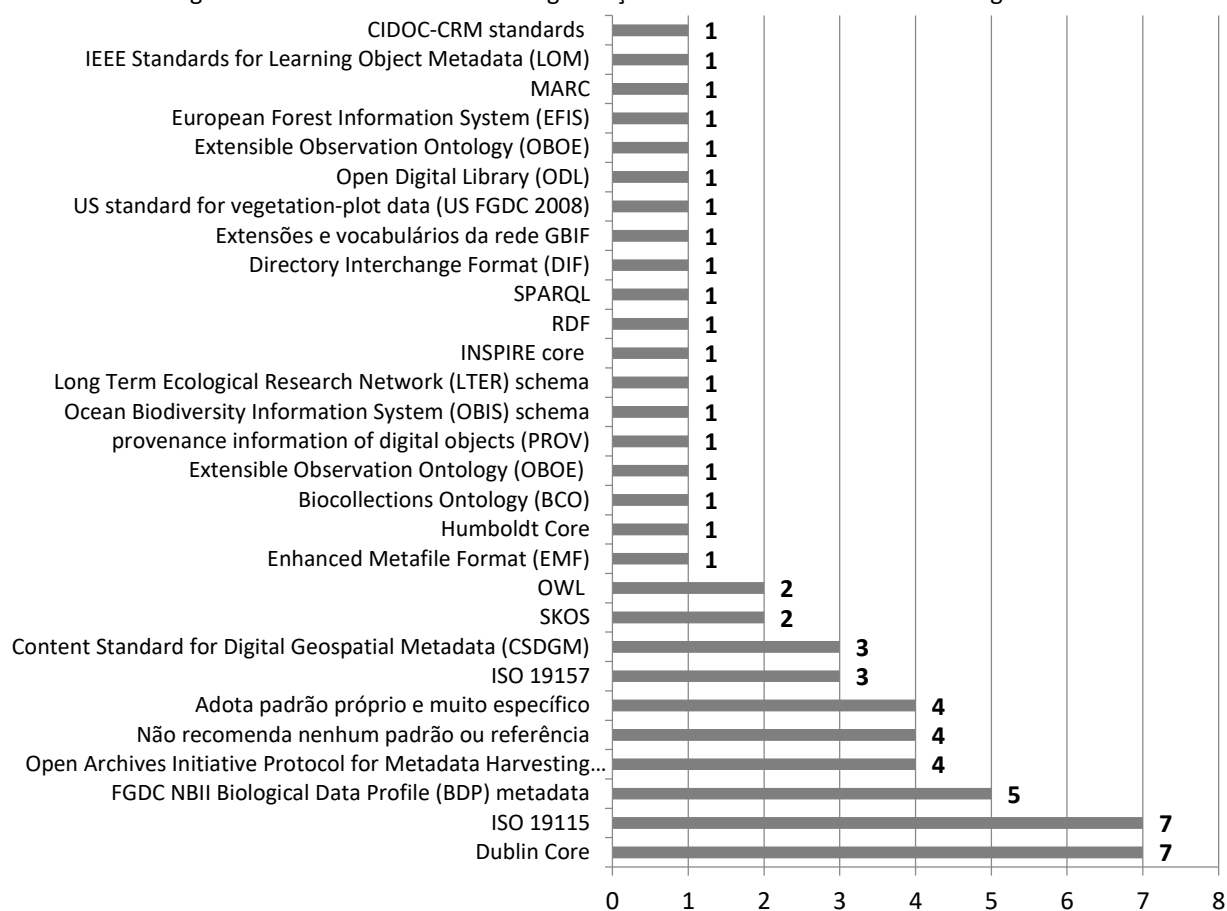
O único artigo identificado no final da década de 1990 refere-se à necessidade de criar padrão de metadados da biodiversidade para a China e não faz referência aos demais padrões – figura 19. Já na década de 2000, os principais padrões de metadados da biodiversidade passaram a ser discutidos e receber colaborações das comunidades de pesquisa, isso se refletiu na quantidade de artigos que apresentam referência a esses padrões. Por fim, a década de 2010 apresenta o maior número de artigos que tratam do uso de metadados e, conseqüentemente, também houve um crescimento na quantidade de artigos que abordam os padrões.

Destaca-se que na década de 2000 há predominância de artigos relacionados ao padrão EML, mas na década de 2010, o padrão DwC passou a ocupar lugar de destaque também. Isso possivelmente se deve a estruturação dos arquivos DwC (DwC-A) que permitem estabelecer diversas relações entre os dados que integram o conjunto de dados, mas também ao crescimento do uso desse padrão em redes e plataformas de referência, como ocorreu no caso da GBIF.

Além disso, por trata-se de um perfil de metadados que abrange recomendações, boas práticas, vocabulários e esquemas dos demais padrões de metadados, o GMP vem ganhando espaço na última década. As ferramentas da GBIF facilitam a publicação dos dados nos demais padrões e associam campos fundamentais estabelecido no GMP. Esses achados, corroboram com os resultados encontrados na etapa I deste estudo, demonstrando que os padrões identificados são de fato absorvidos pelas comunidades de pesquisa relacionadas à biodiversidade.

Destaca-se que um mesmo artigo pode abordar mais de um dos padrões de metadados. O item 'outras referências' apresenta a quantidade de artigos que não tratam de nenhum dos padrões de metadados para biodiversidade identificados, mas podem apresentar outras referências. A figura 20 lista e apresenta a quantidade de ocorrências em que apareceram outras referências relacionadas à organização dos dados de pesquisa nos artigos analisados.

Figura 20. Outras referências de organização de dados identificadas nos artigos



Fonte: Elaboração própria.

Apesar de não exaustiva, a revisão sistematizada da literatura científica associada ao método *berrypicking* permitiu reunir um conjunto representativo de artigos. Isso viabilizou a análise abrangente das pesquisas desenvolvidas e, conseqüentemente, da evolução do uso dos metadados nas últimas três décadas. De forma geral, os resultados evidenciam o aumento dos esforços em pesquisas (e.g. financiamentos, projetos de colaboração internacional, desenvolvimento de soluções computacionais considerando padrões de metadados e outros), bem como apontam o potencial de ações a serem desenvolvidas para a superação dos desafios relacionados à gestão dos dados da biodiversidade.

4.2.3 COLETA DOS METADADOS DOS CONJUNTOS DE DADOS DA BIODIVERSIDADE

A maior compilação aberta de dados sobre biodiversidade encontra-se no portal da rede GBIF (www.gbif.org), que agrega principalmente dados de ocorrência de espécies de diferentes coleções, com descrições variadas dos metadados (HUDSON, NEWBOLD, *et al.*, 2017). A portal da rede GBIF contém o total de 53.729 conjuntos de dados (*datasets*) depositados até a data 13 de julho de 2010. Esses conjuntos de dados são classificados em quatro categorias: (i) checklist; (ii) ocorrência; (iii) evento; e (iv) recursos de metadados. A descrição de cada uma dessas categorias e respectivos metadados foi apresentada na subseção 4.1.5.1.

Considerando os 53.729 conjuntos de dados depositados na GBIF, os registros de ocorrência somam o total de 1.576.468.473 registros de espécimes catalogados. Esses registros apresentam a seguinte distribuição por reinos: *Animalia* (1.223.177.396); *Plantae* (296.094.522); *Fungi* (19.584.251); *Bacteria* (14.551.267); *Chromista* (12.336.630); *Protozoa* (1.149.284); *Archaea* (266,653); *Viruses* (44,011); e *incertae sedis* ou desconhecidos (9.264.459).

Apesar da natureza não estruturada de parte dos dados depositados na GBIF, esses recursos têm sido cada vez mais usados na pesquisas (HUDSON, NEWBOLD, *et al.*, 2017), devido a diversas iniciativas implantadas pela rede (e.g. identificadores únicos para conjuntos de dados; estímulo à publicação de *data papers*; treinamento nas ferramentas desenvolvidas pela GBIF para facilitar e qualificar a descrição dos conjuntos de dados; eventos e capacitações visando à conscientização dos pesquisadores sobre a relevância da descrição dos conjuntos de dados; estímulo à citação de conjuntos de dados e *data papers*).

Neste estudo, primeiramente foram coletados os metadados que descrevem os 53.729 conjuntos de dados – chamados metadados do recurso – depositados na GBIF. Essa consulta foi realizada por meio da ferramenta de consulta de *dataset* disponível no portal GBIF. Em seguida foram coletados os metadados que descrevem os dados contidos em cada um dos conjuntos de dados. Esses dados apresentam acesso aberto no portal GBIF.

Essa coleta foi realizada utilizando um *script* de captura e extração automatizada de dados diretamente nas páginas do portal da GBIF. Os dados coletados foram estruturados na forma da lista de verificação – modelo apresentado no APÊNDICE B. Essa estruturação viabilizou a análise apresentada a seguir. O resultado dessa coleta consta no ANEXO 11.

4.2.4 ANÁLISE DOS METADADOS DE CONJUNTOS DE DADOS

Instituições e comunidades de pesquisa já têm adotado o uso de padrões de metadados e vêm desenvolvendo recomendações e esquemas temáticos para estimular e facilitar a descrição dos conjuntos de dados (GORDON e HABERMANN, 2018).

Os dados extraídos da rede GBIF são evidências relevantes para retratar o uso dos metadados em pesquisas sobre a biodiversidade em todo o mundo. Os resultados são apresentados e discutidos a seguir considerando as seguintes abordagens: (i) abrangência espacial; (ii) abrangência temporal; (iii) elementos de metadados dos conjuntos de dados; (iv) elementos de metadados dos dados de biodiversidade; (v) conjuntos de dados de referência; e (vi) análise das contribuições da América do Sul e Brasil.

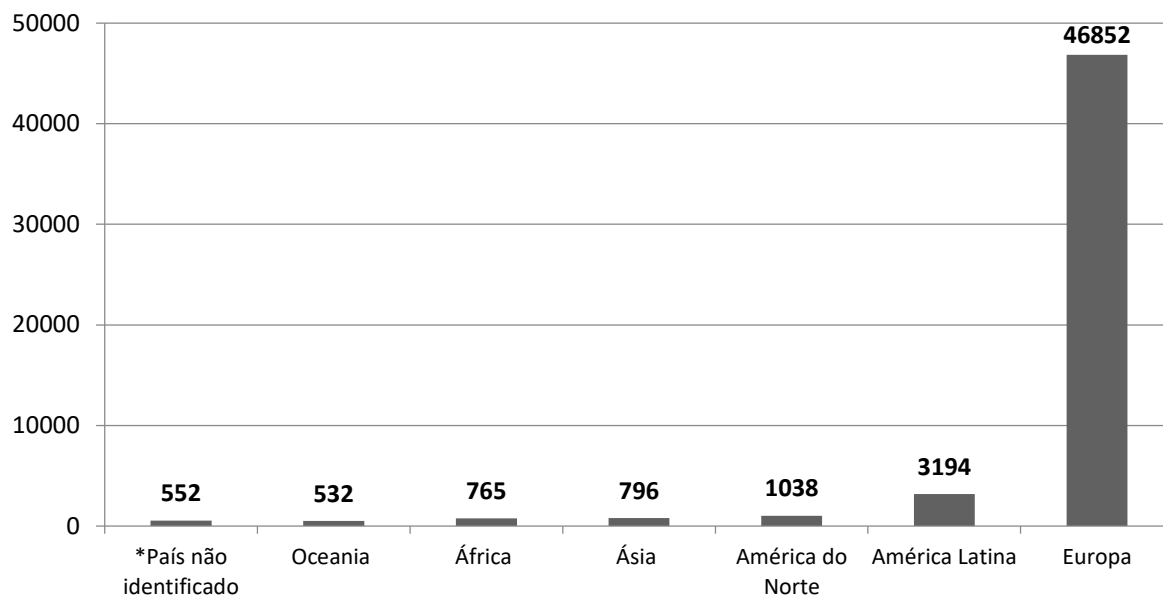
4.2.4.1 Abrangência espacial

A abrangência e representatividade espacial dos conjuntos de dados depositados na rede GBIF devem ser analisadas sob duas perspectivas complementares. A primeira considera a localidade em que está situada a instituição responsável pela gestão do conjunto de dados. A segunda considera a distribuição espacial dos dados da biodiversidade que integram o conjuntos de dados.

Considerando a primeira perspectiva, a figura 21 apresenta a quantidade de conjuntos de dados depositados na GBIF por continente. Nessa análise, cabe destacar que a identificação da localidade considera o país que sedia a instituição responsável pelo depósito

do conjunto de dados. Portanto, ainda não reflete a distribuição geográfica dos dados da biodiversidade que integram os conjuntos de dados.

Figura 21. Quantidade de conjuntos de dados por continente.



*Não identificados = ausência do registro de país.

Fonte: elaboração própria autora baseado nos dados extraídos dados da GBIF (2020e).

As instituições sediadas na Europa ocupam um papel central, responsáveis por 85,5% dos conjuntos de dados. O quadro 23 lista os 120 países responsáveis pelo depósito de dados na GBIF, indicando também a quantidade total e de cada classe (checklist; ocorrência; eventos; ou recursos de metadados). Considerando países com o maior número de conjuntos de dados, destaca-se que a Suíça é responsável por 56,9 %, isso possivelmente decorre das ações governamentais voltadas a estabelecer iniciativas de incentivo ao compartilhamento de dados como políticas nacionais. Os 10 primeiros países da lista são responsáveis por 89,6% dos depósitos registrados na rede GBIF.

Quadro 23. Quantidade de instituições depositárias e quantidade de conjuntos de dados (*datasets*) depositados por país na rede GBIF.

	PAÍS	TOTAL POR CLASSE DOS DATASETS				Total <i>dataset</i>	Total Instituições
		<i>checklist</i>	ocorrência	evento	metadado do recurso		
1	Suíça	30579	23	0	1	30603	15
2	Alemanha	36	8757	9	0	8802	35
3	França	1	2566	1	0	2568	50

	PAÍS	TOTAL POR CLASSE DOS DATASETS				Total dataset	Total Instituições
		checklist	ocorrência	evento	metadado do recurso		
4	Reino Unido	64	822	909	1	1796	145
5	Colômbia	217	976	52	0	1245	140
6	México	2	1025	0	0	1027	4
7	Estados Unidos da America	45	705	90	5	845	213
8	Japão	0	463	3	2	468	2
9	Brasil	15	367	11	16	409	99
10	Bulgária	314	59	0	3	376	3
11	Austrália	1	350	2	0	353	83
12	Rússia	29	278	41	5	353	67
13	Espanha	11	297	9	2	319	101
14	Itália	291	16	1	0	308	5
15	Países Baixos	2	143	3	154	302	34
16	Argentina	1	272	3	2	278	32
17	Noruega	3	212	51	7	273	30
18	Bélgica	25	144	85	4	258	20
19	Canadá	14	179	0	0	193	34
20	Dinamarca	91	78	4	1	174	23
21	Benin	15	146	0	4	165	17
22	Portugal	11	66	85	0	162	21
23	República da Coreia	1	142	0	0	143	31
24	Suécia	4	43	82	0	129	11
25	Polônia	0	102	0	0	102	30
26	Nova Zelândia	7	82	6	0	95	6
27	África do Sul	1	89	2	0	92	6
28	Irlanda	0	69	0	0	69	1
29	Togo	38	26	0	5	69	3
30	Finlândia	0	60	1	0	61	11
31	Quênia	10	42	6	0	58	4
32	Malawi	1	38	14	1	54	4
33	Taiwan	5	37	11	1	54	11
34	Grécia	0	2	50	1	53	1
35	Madagáscar	1	39	0	4	44	10
36	Venezuela	1	24	17	1	43	2
37	China	4	34	1	1	40	3
38	Nova Caledônia	1	36	0	0	37	2

	PAÍS	TOTAL POR CLASSE DOS DATASETS				Total <i>dataset</i>	Total Instituições
		<i>checklist</i>	ocorrência	evento	metadado do recurso		
39	Chile	9	25	1	1	36	17
40	Guiné	0	36	0	0	36	6
41	Trindade e Tobago	0	35	0	0	35	1
42	Ucrânia	2	28	0	0	30	10
43	Suriname	3	25	0	0	28	3
44	Tanzânia	7	19	0	1	27	4
45	Estônia	1	21	0	3	25	3
46	Gana	4	19	1	1	25	9
47	Israel	0	22	2	0	24	3
48	Nigéria	3	20	0	0	23	9
49	Uganda	0	19	3	0	22	4
50	Indonésia	0	17	4	0	21	4
51	Botsuana	0	14	6	0	20	3
52	Senegal	11	7	0	0	18	2
53	Costa Rica	0	17	0	0	17	5
54	Filipinas	4	11	2	0	17	3
55	Índia	2	14	0	0	16	4
56	Tonga	0	16	0	0	16	2
57	Camarões	4	11	0	0	15	7
58	Zimbábue	7	8	0	0	15	2
59	Áustria	0	14	0	0	14	13
60	Costa do Marfim	0	13	1	0	14	5
61	Jamaica	1	9	4	0	14	2
62	Barbados	0	12	0	0	12	2
63	Congo	0	12	0	0	12	4
64	Libéria	2	10	0	0	12	2
65	Belize	0	11	0	0	11	1
66	Fiji	1	10	0	0	11	3
67	Peru	2	5	2	1	10	6
68	Moçambique	0	9	0	0	9	5
69	Vanuatu	1	8	0	0	9	3
70	Equador	1	6	1	0	8	3
71	Islândia	2	6	0	0	8	1
72	Andorra	0	7	0	0	7	1
73	Guiana	4	3	0	0	7	1

	PAÍS	TOTAL POR CLASSE DOS DATASETS				Total <i>dataset</i>	Total Instituições
		<i>checklist</i>	ocorrência	evento	metadado do recurso		
74	Paquistão	0	7	0	0	7	6
75	Uruguai	1	6	0	0	7	2
76	Uzbequistão	0	7	0	0	7	4
77	Vietnã	0	7	0	0	7	1
78	Eslovênia	0	6	0	0	6	3
79	Irã	0	6	0	0	6	1
80	Maurícia	4	1	1	0	6	2
81	Níger	0	6	0	0	6	5
82	Tailândia	0	6	0	0	6	1
83	Tchéquia	0	6	0	0	6	2
84	Ilhas Salomão	0	5	0	0	5	2
85	Mauritânia	3	2	0	0	5	2
86	Angola	0	4	0	0	4	2
87	Bielorrússia	2	2	0	0	4	3
88	Macedônia do Norte	0	4	0	0	4	1
89	Nepal	0	4	0	0	4	2
90	Zâmbia	0	4	0	0	4	2
91	Luxemburgo	1	2	0	0	3	1
92	Burkina Faso	0	2	0	0	2	2
93	Camboja	1	1	0	0	2	1
94	Gabão	0	1	1	0	2	1
95	Guatemala	1	1	0	0	2	2
96	Guiana Francesa	0	2	0	0	2	2
97	Ruanda	0	2	0	0	2	1
98	Samoa	0	2	0	0	2	1
99	Burundi	0	1	0	0	1	1
100	Cingapura	0	1	0	0	1	1
101	Croácia	0	1	0	0	1	1
102	Eslováquia	0	1	0	0	1	1
103	Micronésia	0	1	0	0	1	1
104	Honduras	0	1	0	0	1	1
105	Hong Kong	0	1	0	0	1	1
106	Hungria	0	1	0	0	1	1
107	Letônia	0	1	0	0	1	1
108	Líbano	0	1	0	0	1	1

	PAÍS	TOTAL POR CLASSE DOS DATASETS				Total <i>dataset</i>	Total Instituições
		<i>checklist</i>	ocorrência	evento	metadado do recurso		
109	Malásia	0	1	0	0	1	1
110	Marrocos	0	1	0	0	1	1
111	Mongólia	0	1	0	0	1	1
112	Nicarágua	0	1	0	0	1	1
113	Palau	0	1	0	0	1	1
114	Papua Nova Guiné	0	1	0	0	1	1
115	Porto Rico	0	1	0	0	1	1
116	Sérvia	0	1	0	0	1	1
117	Somália	0	1	0	0	1	1
118	Sudão do Sul	0	1	0	0	1	1
119	Turquia	1	0	0	0	1	1
120	Wallis e Futuna	0	1	0	0	1	1
	*País não identificado	225	235	12	80	552	34
		32.151	19.680	1.590	308	** 53.729	*** 1.570

Fonte: elaboração própria autora - Fonte dos dados: GBIF (2020e). *instituições e *datasets* sem país vinculado; **total de *datasets*; ***total de instituições.

Existem 2025 instituições cadastradas para depositar dados na GBIF, entretanto, 1570 são responsáveis por pelo menos um ou mais *datasets*. Dessas há 34 que apresentam registro desvinculado de país, pois são entidades de abrangência global, com extensa rede de representantes, dirigidas ou geridas por mecanismos participativos guiados por comitês ou conselhos com representantes de diversas nacionalidades – listadas no quadro 24.

Quadro 24. Instituições de abrangência global e sem referência a um único país.

NOME DAS INSTITUIÇÕES	
1	Alaska Ocean Observing System
2	Albertine Rift Conservation Society (ARCOS)
3	Antarctic Biodiversity Information Facility (ANTABIF)
4	Australian Antarctic Data Centre
5	Bioinformatics Research Group, Biological Science Department, Faculty of Bioscience & Bioengineering
6	Bioversity International
7	CABI (Centre for Agriculture and Biosciences International)
8	Conservation of Arctic Flora and Fauna

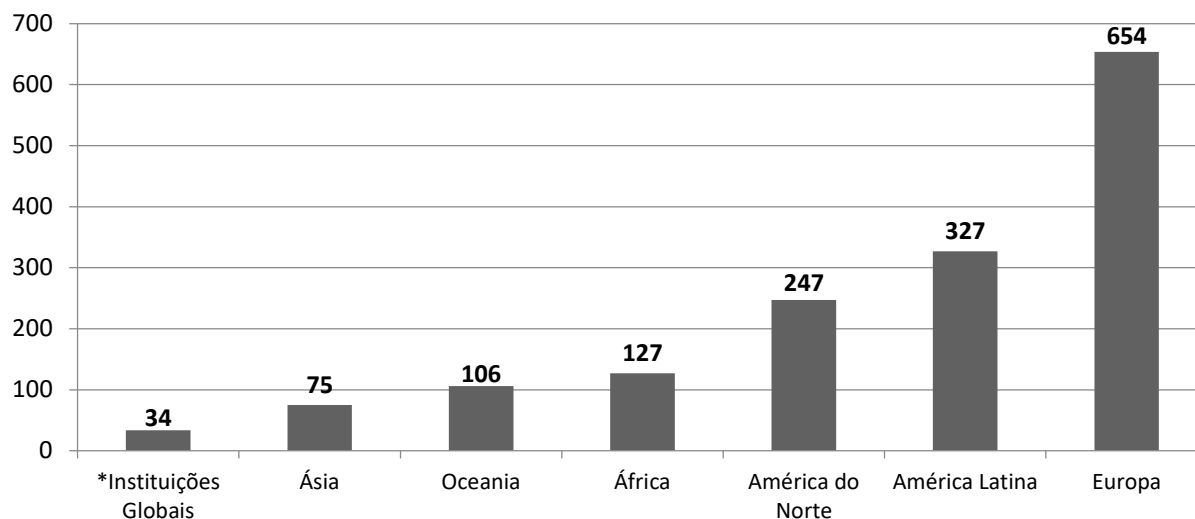
NOME DAS INSTITUIÇÕES	
9	Crop Research Institute (CRI)
10	Deep-sea OBIS node
11	Encyclopedia of Life (EOL)
12	European Environment Agency
13	Fauna Europaea Consortium
14	GBIF Secretariat
15	Index Fungorum Partnership
16	International Centre for Insect Physiology and Ecology
17	International Centre for Integrated Mountain Development
18	International Commission on Zoological Nomenclature
19	International Committee on Taxonomy of Viruses
20	International Compositae Alliance
21	International Ornithologists Union (IOU)
22	International Union for Conservation of Nature
23	Marine Biology Laboratory
24	Nordic Genetic Resource Center (NORDGEN)
25	Northern Territory Department of Environment and Natural Resources
26	Public Domain Datasets
27	Reef Life Survey
28	SCAR - Microbial Antarctic Resource System
29	Senckenberg - CeDAMar Provider
30	South Australia, Department for Environment and Water
31	The Catalogue of Life Partnership
32	The International Barcode of Life Consortium
33	The International Plant Names Index Collaborators
34	The Tree of Life Web Project

Fonte: elaboração própria autora - Fonte dos dados: GBIF (2020e).

O ANEXO 9 lista os 552 *datasets* vinculados às instituições globais desvinculadas de país. Em geral, são conjuntos de dados considerados referência, pois resultam de grandes censos, programas de monitoramento de ampla abrangência, integração de grandes bases

de dados ou revisões exaustivas sobre determinados grupos da biodiversidade. A distribuição das instituições por continente é apresentada na figura 22.

Figura 22. Quantidade de instituições por continente.



*Instituições sem registro vinculado a país e com abrangência global

Fonte: elaboração própria autora - Fonte dos dados: GBIF (2020e).

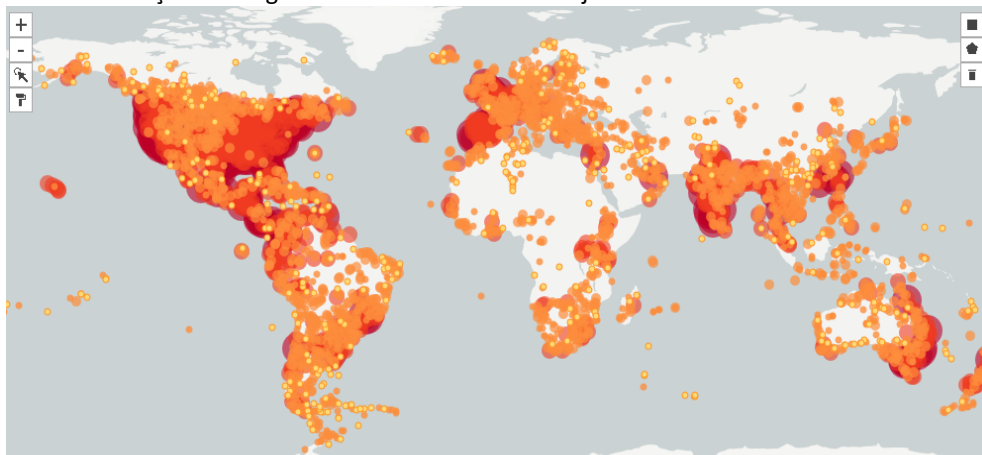
A Europa possui o maior número de instituições responsáveis pela gestão dos conjuntos de dados da biodiversidade depositados na rede GBIF, mais de 40% são sediadas nesse continente. Destaca-se a concentração de cerca de 20% das instituições na América Latina. As 34 instituições globais referem-se àquelas com registro desvinculado de qualquer país.

Em uma segunda perspectiva, a abrangência espacial também pode ser analisada considerando a distribuição dos dados da biodiversidade. Para melhor ilustrar a diferença entre as duas perspectivas de análise espacial dos conjuntos de dados, considera-se como exemplo um conjunto de dados do tipo 'ocorrência', o *eBird Observation Dataset* (EOD), que reúne grande quantidade de dados de observação ornitológica (aves). A instituição responsável pelo depósito do EOD é *Cornell Lab of Ornithology*, localizada nos Estados Unidos. O EOD contém o total de 705.008.469 ocorrências de avistamento de aves, coletadas por pesquisadores colaboradores em 250 países do mundo (e.g. Canadá: 70.337.655; Austrália: 21.643.122; Índia: 16.897.707; México: 8.871.197; Espanha: 8.562.570; Reino Unido: 8.314.877; Costa Rica: 8.022.073; Colômbia: 6.138.347; Brasil: 5.907.917 e diversos outros países). Nesse caso, no quadro 23 acima, o conjunto de dados

EOD é contabilizado nos Estados Unidos, pois é o país em que se encontra a instituição responsável por seu depósito.

Já em relação à distribuição espacial dos dados de observação de aves que integram o EOD, apenas em dezembro de 2019, foram adicionados mais de 8 milhões de registros de ocorrências. O mapa a seguir apresenta a distribuição desses dados – figura 23.

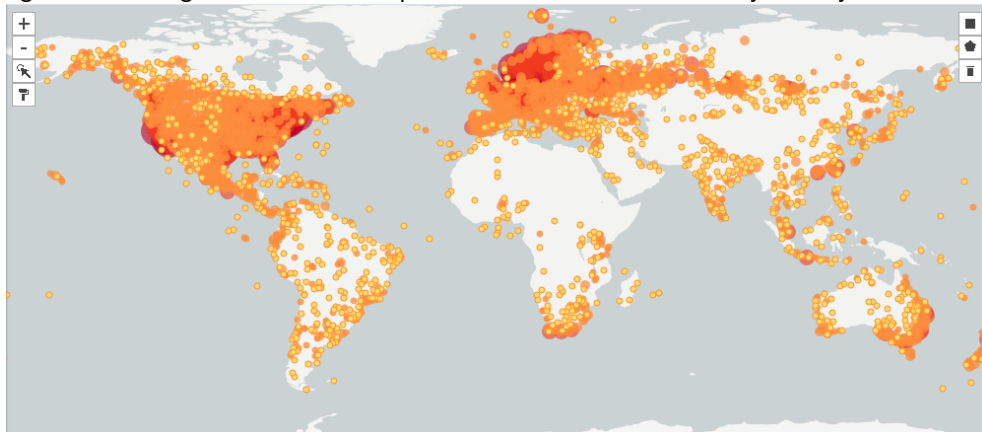
Figura 23. Distribuição dos registros de ocorrências do conjunto de dados EOD em dezembro de 2019



Fonte: GBIF (2020e).

A distribuição espacial dos dados da biodiversidade é, portanto, ainda mais significativa, pois abrange praticamente todo o planeta. Em outro exemplo, o mapa (figura 24) a seguir ilustra a distribuição de ocorrência de espécies de depósitos de dados realizados apenas nos meses de junho (1.674.839) e julho (768.503) de 2020 – considerando dados depositados até 2 de agosto de 2020.

Figura 24. Abrangência dos dados depositados na GBIF nos meses de junho e julho de 2020.



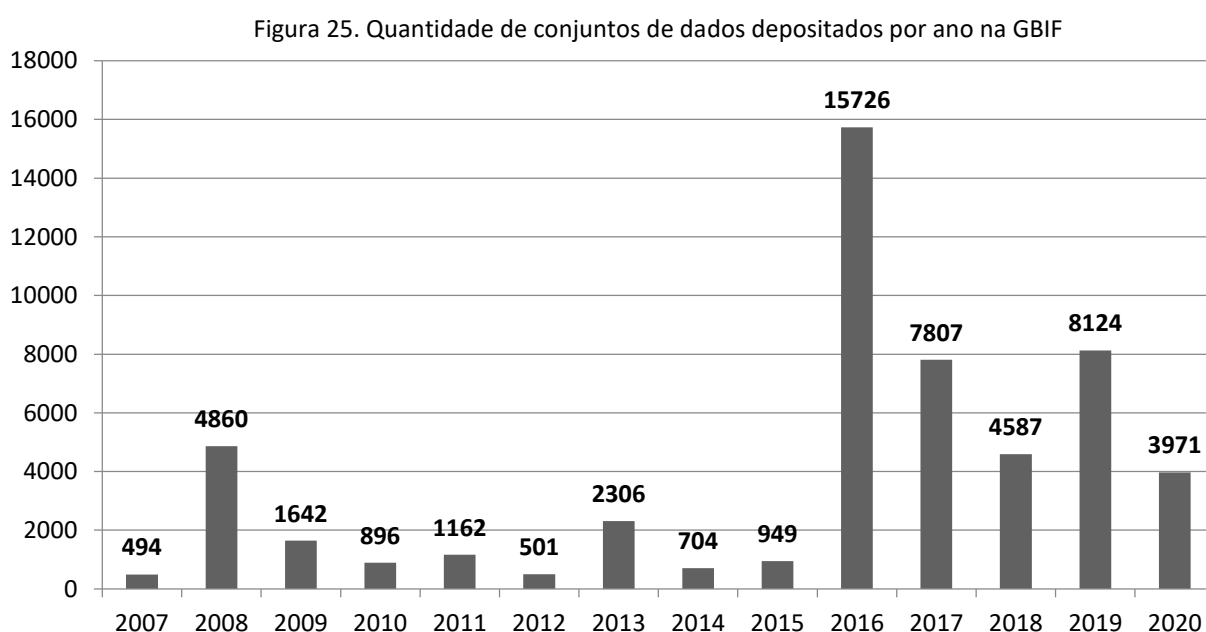
Fonte: GBIF (2020e).

Devido ao volume expressivo de registros de ocorrência de espécies (mais de 1,5 bilhão), não é possível representar de forma adequada a distribuição de todos os dados depositados na GBIF em um único mapa. Entretanto, é possível gerar mapas ou relatórios diversos a partir de referências mais específicas conforme a necessidade de pesquisa (e.g. geográficas; taxonômicas; instituição responsável; conjuntos de dados; e outros).

4.2.4.2 Abrangência temporal

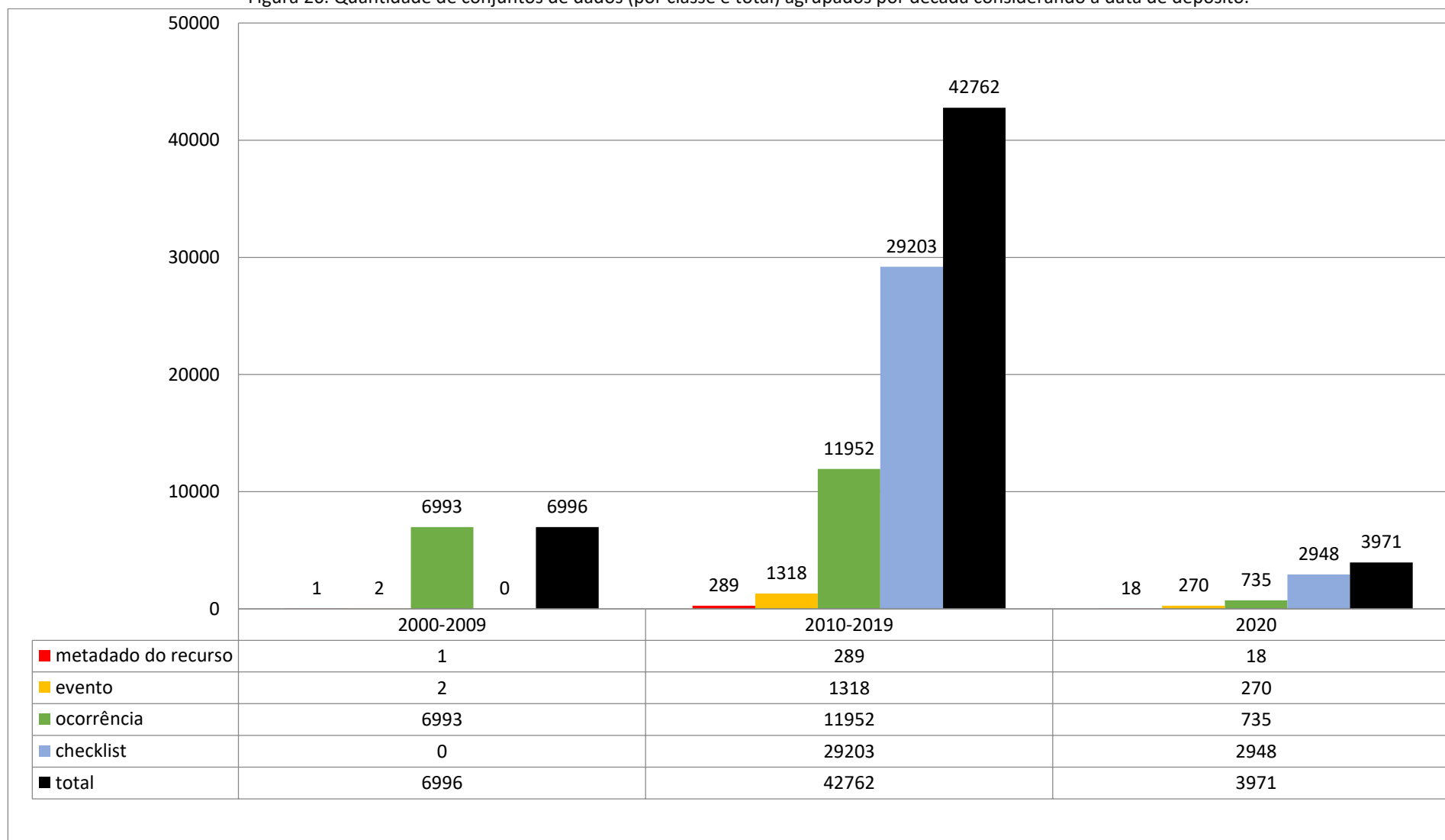
A abrangência temporal dos conjuntos de dados depositados na rede GBIF deve ser analisada sob duas perspectivas complementares. A primeira considera a data do depósito dos conjuntos de dados na GBIF. A segunda considera a cobertura temporal dos dados da biodiversidade, considerando desde a data mais antiga até a mais recente dos registros que integram cada conjunto de dados.

Nessa primeira perspectiva, a figura 25 apresenta a quantidade de conjuntos de dados depositados por ano na rede GBIF. O aumento expressivo de depósitos em 2016 é resultante de coleções digitalizadas de herbários, museus de história natural e museus de instituições de pesquisa (14.370 da classe *checklist*), bem como da integração com portais de temas específicos, em especial com soluções de coleta baseadas em observações humanas (1.285 da classe ocorrência). Na figura 26 é possível comparar a quantidade dos depósitos de conjuntos de dados (total e por classe) nas últimas décadas.



Fonte: elaboração própria autora - Fonte dos dados: GBIF (2020e).

Figura 26. Quantidade de conjuntos de dados (por classe e total) agrupados por década considerando a data de depósito.



Fonte: elaboração própria autora - Fonte dos dados: GBIF (2020e).

A década de 2010 concentra 64,5% dos depósitos, pois foi um período de grande mobilização para integração dos dados da biodiversidade. Esse foi o período com maior número de depósitos de *datasets* das classes *checklist* (e.g. coleções de herbários, museus institucionais e de história natural) e ocorrência (e.g. registros de avistamento; observações humanas). Isso corrobora com as evidências de uso dos metadados identificadas na revisão sistematizada dos artigos científicos. Essa tendência deve se manter, pois até 13 de julho (data de coleta dos dados), os conjuntos de dados depositados no ano de 2020 já representam 7,3% do total.

O aumento dos conjuntos de dados do tipo evento (que incluem informações sobre o evento de coleta) também tem ganhado espaço nas comunidades de pesquisa desde a possibilidade de inclusão do núcleo evento no arquivo modelo do padrão DwC (DwC-A), que facilitou a documentação e relacionamento com os demais tipos de dados. Também é relevante o registro dos conjuntos de dados da classe recurso, pois permite que a comunidade científica tome conhecimento da existência de dados ainda não disponibilizados em formato digital e estruturado, permitindo a descoberta, bem como o estabelecimento de políticas e planos para a digitalização e disponibilização desses recursos.

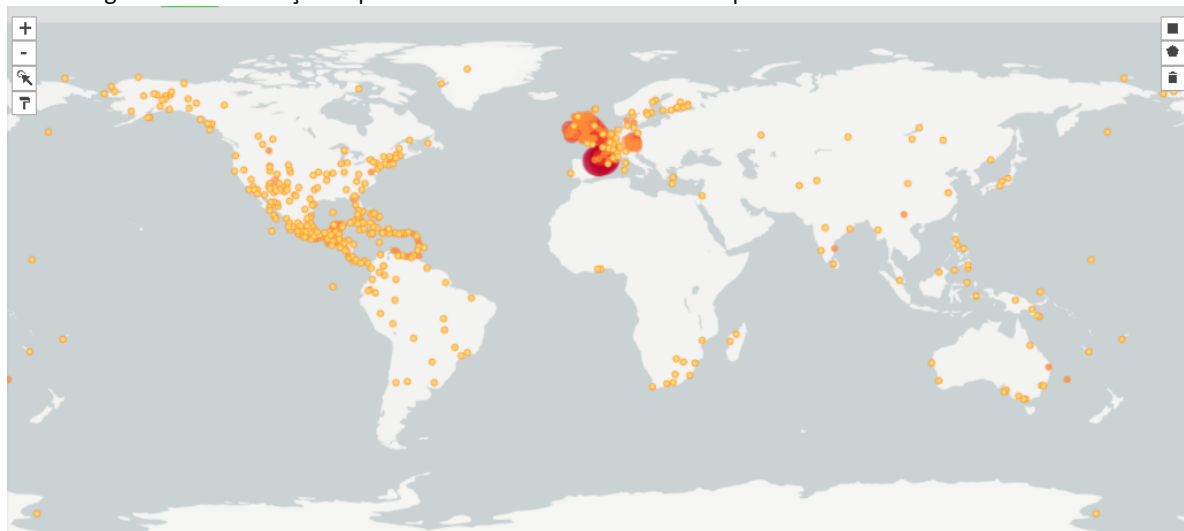
A segunda perspectiva de análise considera a cobertura temporal dos dados da biodiversidade. Considerando a data mais antiga, há na GBIF registros de ocorrências a partir do ano 1600. Os mapas a seguir apresentam a distribuição espacial de dados que datam de 1600 (figura 27), do período de 1600 a 1700 (figura 28) e do ano 2020 (figura 29).

Figura 27. Distribuição espacial dos dados de ocorrência do ano 1600.



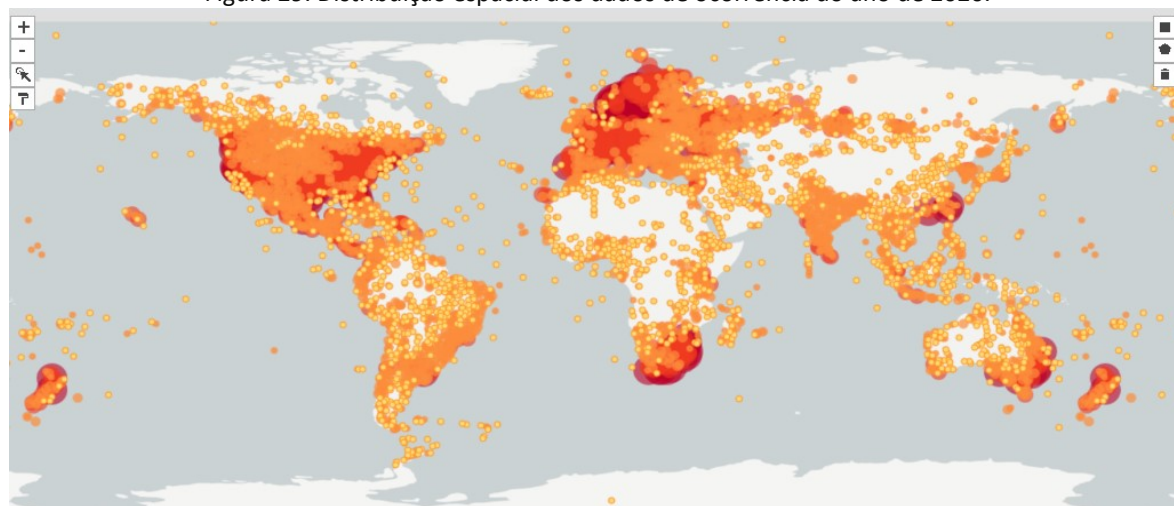
Fonte: GBIF (2020e).

Figura 28. Distribuição espacial dos dados de ocorrência do período entre os anos 1600 a 1700.



Fonte: GBIF (2020e).

Figura 29. Distribuição espacial dos dados de ocorrência do ano de 2020.



Fonte: GBIF (2020e).

As ferramentas disponibilizadas no portal GBIF permitem restringir as buscas escolhendo períodos do tempo. Conforme exemplificado, considerando a extensão da dimensão temporal, não é possível representar adequadamente todos os dados depositados na GBIF desde o ano 1600 em um único mapa ou relatório. Por isso, a GBIF disponibiliza ferramentas com diversos parâmetros que podem ser utilizados para explorar os dados e delimitar as pesquisas.

4.2.4.3 Campos de metadados usados para a descrição dos conjuntos de dados

Há dois níveis de metadados de descrição coletados. O primeiro nível refere-se aos campos de metadados utilizados para descrever o conjunto de dados (*dataset*). O segundo nível refere-se aos campos utilizados para descrever os dados que integram os *datasets*.

Considerando o primeiro nível, os dados analisados foram coletados diretamente das páginas que descrevem cada um dos *datasets* no portal GBIF. O quadro a seguir apresenta as evidências do uso de metadados para a descrição de cada classe de *dataset* – quadro 25. A lista completa dos 53.729 *datasets* analisados consta no ANEXO 11.

Quadro 25. Elemento de metadados utilizados na descrição dos conjuntos de dados.

	Elemento de Metadado	Tipo de conjunto de dados			
		<i>checklist</i>	ocorrência	evento	metadado do recurso
1	abbreviation	X	-	-	-
2	alias	X	-	-	-
3	deleted	X	-	-	-
4	duplicateofdatasetkey	-	X	-	-
5	parentdatasetkey	X	-	-	-
6	rights	-	X	-	-
7	subtype	X	X	-	-
8	additionalinfo	X	X	X	X
9	bibliographiccitations	X	X	X	X
10	citation	X	X	X	X
11	collections	X	X	X	X
12	comments	X	X	X	X
13	contacts	X	X	X	X
14	countrycoverage	X	X	X	X
15	created	X	X	X	X
16	createdby	X	X	X	X
17	curatorialunits	X	X	X	X
18	datadescriptions	X	X	X	X
19	datalanguage	X	X	X	X
20	description	X	X	X	X
21	doi	X	X	X	X
22	endpoints	X	X	X	X
23	external	X	X	X	X
24	geographiccoverages	X	X	X	X
25	homepage	X	X	X	X
26	identifiers	X	X	X	X

	Elemento de Metadado	Tipo de conjunto de dados			
		<i>checklist</i>	ocorrência	evento	metadado do recurso
27	installationkey	X	X	X	X
28	key	X	X	X	X
29	keywordcollections	X	X	X	X
30	language	X	X	X	X
31	license	X	X	X	X
32	lockedforautoupdate	X	X	X	X
33	logourl	X	X	X	X
34	machinetags	X	X	X	X
35	maintenancedescription	X	X	X	X
36	maintenanceupdatefrequency	X	X	X	X
37	modified	X	X	X	X
38	modifiedby	X	X	X	X
39	network	X	X	X	X
40	numconstituents	X	X	X	X
41	project	X	X	X	X
42	pubdate	X	X	X	X
43	publishingorganizationkey	X	X	X	X
44	purpose	X	X	X	X
45	samplingdescription	X	X	X	X
46	tags	X	X	X	X
47	taxonomiccoverages	X	X	X	X
48	temporalcoverages	X	X	X	X
49	title	X	X	X	X
50	type	X	X	X	X
51	version	X	X	X	X

(X) presente ou (-) ausente na classe de conjunto de dados. Fonte: elaboração própria autora - Fonte dos dados: GBIF (2020e).

As evidências do uso de metadados coletadas no portal GBIF apresentam diversas referências à descrição dos dados (e.g. temporais, espaciais, origem e classe). Conforme apresentado no quadro acima, o uso de metadados para a descrição dos datasets é similar em todas as classes de *datasets*, com exceção de alguns elementos (e.g. *abbreviation*; *alias*; *deleted*; *parentdatasetkey*; *subtype*; *duplicateofdatasetkey*; *rights*).

Os *datasets* depositados na GBIF apresentam uma descrição mais detalhada por encontrarem-se inseridos em um contexto dedicado a atender necessidades das pesquisas sobre a biodiversidade. Essa especialização apresentada no portal é relevante, pois disponibiliza parâmetros fundamentais para a descoberta e reuso dos dados.

4.2.4.4 Campos de metadados usados para a descrição dos dados da biodiversidade

O segundo nível de uso dos metadados refere-se aos elementos (também chamados campos ou atributos) que descrevem os dados registrados nos *datasets*. Essas evidências do uso de metadados foram coletadas dos registros existentes em cada um dos 53.729 *datasets*. A documentação dos dados também é disponibilizada no portal GBIF na página de cada dataset ao acessar as informações dos registros. As quatro classes de *datasets* foram analisadas em 3 partes: (i) metadados de recursos; (ii) ocorrência e eventos; e (iii) checklist.

(i) Classe metadados de recursos

Os *datasets* da classe metadados de recurso (*resource metadata*), em geral não apresentaram registros de dados. Entretanto, isso já era previsto devido sua função da classe. Esses *datasets* representam apontamentos da existência de objetos de informação ainda não disponíveis, por isso são relevantes para que pesquisadores de todo o mundo tomem conhecimento da existência desses recursos e também fundamentais para planejamento de políticas e projetos de gestão dos dados da biodiversidade.

Nessa classe, há 308 *datasets* registrados no portal da GBIF e apenas 9 deles apresentam registros de dados. A lista completa dos *datasets* da classe metadados de recurso encontra-se no ANEXO 11. Portanto, o uso de metadados para essa classe está ligada a descrição dos conjuntos, conforme apresentado na subseção anterior – 4.2.4.3.

(ii) Classes ocorrência e eventos

As classes ocorrência (*occurrence*) e evento (*sampling event*) apresentam maior similaridade, pois tratam respectivamente dos registros de ocorrências de espécies e dos eventos realizados para coletar essas ocorrências. Nessas classes, a análise do uso dos metadados na descrição dos dados iniciou pela identificação dos elementos/campos utilizados nos 21.270 *datasets* (19.680 ocorrência; 1.590 eventos). O quadro a seguir apresenta a lista completa dos elementos de metadados identificados e a indicação da quantidade de *datasets* em estão presentes – quadro 26.

Quadro 26. Frequência do uso do elementos/campos de metadados para descrição dos dados em *datasets* da classe ocorrência e da classe eventos no portal GBIF.

	Elemento/campo de metadados	Total dataset		Frequência (F)		
		Ocorrência	Evento	*Fa	**Fr	***Fp
1	basisOfRecord	19.398	1.515	20.913	0,983216	98,3216%
2	crawlId	19.398	1.515	20.913	0,983216	98,3216%
3	datasetKey	19.398	1.515	20.913	0,983216	98,3216%
4	extensions	19.398	1.515	20.913	0,983216	98,3216%
5	facts	19.398	1.515	20.913	0,983216	98,3216%
6	gbifID	19.398	1.515	20.913	0,983216	98,3216%
7	identifiedByIDs	19.398	1.515	20.913	0,983216	98,3216%
8	identifiers	19.398	1.515	20.913	0,983216	98,3216%
9	installationKey	19.398	1.515	20.913	0,983216	98,3216%
10	issues	19.398	1.515	20.913	0,983216	98,3216%
11	key	19.398	1.515	20.913	0,983216	98,3216%
12	kingdom	19.398	1.515	20.913	0,983216	98,3216%
13	kingdomKey	19.398	1.515	20.913	0,983216	98,3216%
14	lastInterpreted	19.398	1.515	20.913	0,983216	98,3216%
15	lastParsed	19.398	1.515	20.913	0,983216	98,3216%
16	license	19.398	1.515	20.913	0,983216	98,3216%
17	media	19.398	1.515	20.913	0,983216	98,3216%
18	protocol	19.398	1.515	20.913	0,983216	98,3216%
19	publishingOrgKey	19.398	1.515	20.913	0,983216	98,3216%
20	recordedByIDs	19.398	1.515	20.913	0,983216	98,3216%
21	relations	19.398	1.515	20.913	0,983216	98,3216%
22	scientificName	19.398	1.515	20.913	0,983216	98,3216%
23	taxonKey	19.398	1.515	20.913	0,983216	98,3216%
24	taxonRank	19.398	1.515	20.913	0,983216	98,3216%
25	lastCrawled	19.397	1.515	20.912	0,983169	98,3169%
26	publishingCountry	19.175	1.503	20.678	0,972167	97,2167%
27	acceptedScientificName	19.250	1.373	20.623	0,969582	96,9582%
28	acceptedTaxonKey	19.250	1.373	20.623	0,969582	96,9582%
29	taxonomicStatus	19.250	1.373	20.623	0,969582	96,9582%
30	genericName	19.235	1.373	20.608	0,968876	96,8876%
31	identifier	18.676	1.514	20.190	0,949224	94,9224%
32	phylum	18.688	1.274	19.962	0,938505	93,8505%
33	phylumKey	18.688	1.274	19.962	0,938505	93,8505%
34	eventDate	17.365	1.350	18.715	0,879878	87,9878%
35	year	17.365	1.350	18.715	0,879878	87,9878%
36	decimalLatitude	17.357	1.298	18.655	0,877057	87,7057%

	Elemento/campo de metadados	Total dataset		Frequência (F)		
		Ocorrência	Evento	*Fa	**Fr	***Fp
37	decimalLongitude	17.357	1.298	18.655	0,877057	87,7057%
38	geodeticDatum	17.357	1.298	18.655	0,877057	87,7057%
39	institutionCode	18.087	394	18.481	0,868876	86,8876%
40	month	16.942	1.303	18.245	0,857781	85,7781%
41	country	16.587	1.380	17.967	0,844711	84,4711%
42	countryCode	16.587	1.380	17.967	0,844711	84,4711%
43	day	16.684	1.270	17.954	0,844100	84,4100%
44	class	16.466	1.088	17.554	0,825294	82,5294%
45	classKey	16.466	1.088	17.554	0,825294	82,5294%
46	order	15.948	1.069	17.017	0,800047	80,0047%
47	orderKey	15.948	1.069	17.017	0,800047	80,0047%
48	family	15.888	1.016	16.904	0,794734	79,4734%
49	familyKey	15.888	1.016	16.904	0,794734	79,4734%
50	genus	15.535	863	16.398	0,770945	77,0945%
51	genusKey	15.535	863	16.398	0,770945	77,0945%
52	collectionCode	14.986	295	15.281	0,718430	71,8430%
53	recordedBy	14.757	355	15.112	0,710484	71,0484%
54	catalogNumber	14.800	207	15.007	0,705548	70,5548%
55	species	13.093	577	13.670	0,642689	64,2689%
56	speciesKey	13.093	577	13.670	0,642689	64,2689%
57	specificEpithet	13.083	577	13.660	0,642219	64,2219%
58	modified	12.843	226	13.069	0,614433	61,4433%
59	occurrenceID	10.267	1.514	11.781	0,553879	55,3879%
60	datasetID	11.046	168	11.214	0,527221	52,7221%
61	locality	7.943	852	8.795	0,413493	41,3493%
62	identifiedBy	6.269	294	6.563	0,308557	30,8557%
63	stateProvince	6.044	267	6.311	0,296709	29,6709%
64	geologicalContextID	6.186	1	6.187	0,290879	29,0879%
65	county	5.857	97	5.954	0,279925	27,9925%
66	coordinateUncertaintyInMeters	4.636	272	4.908	0,230748	23,0748%
67	eventID	3.230	1.514	4.744	0,223037	22,3037%
68	occurrenceStatus	4.126	495	4.621	0,217254	21,7254%
69	taxonID	3.892	56	3.948	0,185614	18,5614%
70	language	3.290	190	3.480	0,163611	16,3611%
71	http://rs.tdwg.org/dwc/terms/organismQuantityType	2.385	1.081	3.466	0,162953	16,2953%
72	organismQuantityType	2.385	1.081	3.466	0,162953	16,2953%
73	type	3.140	323	3.463	0,162811	16,2811%
74	http://rs.tdwg.org/dwc/terms/organismQuantity	2.360	1.069	3.429	0,161213	16,1213%

	Elemento/campo de metadados	Total dataset		Frequência (F)		
		Ocorrência	Evento	*Fa	**Fr	***Fp
75	nameAccordingTo	3.428	1	3.429	0,161213	16,1213%
76	organismQuantity	2.318	1.055	3.373	0,158580	15,8580%
77	individualCount	2.894	330	3.224	0,151575	15,1575%
78	datasetName	2.500	362	2.862	0,134556	13,4556%
79	elevation	2.766	93	2.859	0,134415	13,4415%
80	vernacularName	2.698	118	2.816	0,132393	13,2393%
81	institutionID	2.484	162	2.646	0,124401	12,4401%
82	continent	2.474	140	2.614	0,122896	12,2896%
83	dynamicProperties	1.651	872	2.523	0,118618	11,8618%
84	identificationVerificationStatus	2.507	6	2.513	0,118148	11,8148%
85	originalNameUsage	2.435	2	2.437	0,114575	11,4575%
86	elevationAccuracy	2.244	56	2.300	0,108134	10,8134%
87	rightsHolder	2.058	176	2.234	0,105031	10,5031%
88	samplingProtocol	1.662	536	2.198	0,103338	10,3338%
89	projectId	1.940	192	2.132	0,100235	10,0235%
90	municipality	1.945	147	2.092	0,098354	9,8354%
91	habitat	1.821	269	2.090	0,098260	9,8260%
92	depth	1.656	403	2.059	0,096803	9,6803%
93	higherClassification	2.030	3	2.033	0,095581	9,5581%
94	preparations	1.891	95	1.986	0,093371	9,3371%
95	collectionID	1.946	21	1.967	0,092478	9,2478%
96	ownerInstitutionCode	1.747	212	1.959	0,092102	9,2102%
97	depthAccuracy	1.497	398	1.895	0,089093	8,9093%
98	verbatimCoordinateSystem	1.700	33	1.733	0,081476	8,1476%
99	scientificNameID	1.503	227	1.730	0,081335	8,1335%
100	dateIdentified	1.580	97	1.677	0,078843	7,8843%
101	locationID	1.224	388	1.612	0,075787	7,5787%
102	occurrenceRemarks	1.511	87	1.598	0,075129	7,5129%
103	dataGeneralizations	1.578	17	1.595	0,074988	7,4988%
104	locationRemarks	1.471	78	1.549	0,072826	7,2826%
105	recordNumber	1.518	12	1.530	0,071932	7,1932%
106	bibliographicCitation	1.450	18	1.468	0,069017	6,9017%
107	eventRemarks	453	957	1.410	0,066291	6,6291%
108	fieldNumber	1.286	106	1.392	0,065444	6,5444%
109	footprintWKT	1.273	104	1.377	0,064739	6,4739%
110	startDayOfYear	1.309	19	1.328	0,062435	6,2435%
111	eventTime	1.180	133	1.313	0,061730	6,1730%
112	georeferenceSources	1.109	66	1.175	0,055242	5,5242%

	Elemento/campo de metadados	Total dataset		Frequência (F)		
		Ocorrência	Evento	*Fa	**Fr	***Fp
113	http://rs.tdwg.org/dwc/terms/sampleSizeUnit	41	1.104	1.145	0,053832	5,3832%
114	sampleSizeUnit	41	1.104	1.145	0,053832	5,3832%
115	http://rs.tdwg.org/dwc/terms/sampleSizeValue	40	1.099	1.139	0,053550	5,3550%
116	sampleSizeValue	30	1.082	1.112	0,052280	5,2280%
117	identificationRemarks	223	846	1.069	0,050259	5,0259%
118	higherGeography	1.047	13	1.060	0,049835	4,9835%
119	acceptedNameUsage	1.024	32	1.056	0,049647	4,9647%
120	identificationID	1.038	0	1.038	0,048801	4,8801%
121	verbatimEventDate	973	40	1.013	0,047626	4,7626%
122	parentNameUsage	992	0	992	0,046638	4,6638%
123	taxonConceptID	985	3	988	0,046450	4,6450%
124	namePublishedInYear	960	5	965	0,045369	4,5369%
125	materialSampleID	117	841	958	0,045040	4,5040%
126	verbatimLocality	853	93	946	0,044476	4,4476%
127	identificationReferences	64	862	926	0,043535	4,3535%
128	references	865	8	873	0,041044	4,1044%
129	relativeOrganismQuantity	0	840	840	0,039492	3,9492%
130	sex	744	26	770	0,036201	3,6201%
131	waterBody	620	128	748	0,035167	3,5167%
132	otherCatalogNumbers	689	4	693	0,032581	3,2581%
133	nomenclaturalCode	639	27	666	0,031312	3,1312%
134	networkKeys	520	137	657	0,030889	3,0889%
135	lifeStage	619	29	648	0,030465	3,0465%
136	accessRights	591	55	646	0,030371	3,0371%
137	samplingEffort	392	231	623	0,029290	2,9290%
138	georeferencedBy	538	23	561	0,026375	2,6375%
139	verbatimElevation	493	44	537	0,025247	2,5247%
140	programmeAcronym	465	20	485	0,022802	2,2802%
141	georeferenceVerificationStatus	427	3	430	0,020216	2,0216%
142	associatedReferences	405	23	428	0,020122	2,0122%
143	georeferenceProtocol	397	26	423	0,019887	1,9887%
144	disposition	403	7	410	0,019276	1,9276%
145	infraspecificEpithet	391	6	397	0,018665	1,8665%
146	endDayOfYear	377	5	382	0,017960	1,7960%
147	georeferencedDate	367	4	371	0,017442	1,7442%
148	previousIdentifications	360	6	366	0,017207	1,7207%
149	verbatimSRS	331	22	353	0,016596	1,6596%
150	coordinatePrecision	265	87	352	0,016549	1,6549%

	Elemento/campo de metadados	Total dataset		Frequência (F)		
		Ocorrência	Evento	*Fa	**Fr	***Fp
151	georeferenceRemarks	262	20	282	0,013258	1,3258%
152	informationWithheld	266	13	279	0,013117	1,3117%
153	establishmentMeans	226	13	239	0,011236	1,1236%
154	identificationQualifier	232	6	238	0,011189	1,1189%
155	rights	228	0	228	0,010719	1,0719%
156	locationAccordingTo	201	7	208	0,009779	0,9779%
157	island	167	27	194	0,009121	0,9121%
158	fieldNotes	184	6	190	0,008933	0,8933%
159	taxonRemarks	144	43	187	0,008792	0,8792%
160	organismID	181	1	182	0,008557	0,8557%
161	parentEventID	32	150	182	0,008557	0,8557%
162	http://unknown.org/recordId	155	0	155	0,007287	0,7287%
163	typeStatus	152	0	152	0,007146	0,7146%
164	reproductiveCondition	134	12	146	0,006864	0,6864%
165	namePublishedIn	122	2	124	0,005830	0,5830%
166	verbatimTaxonRank	106	8	114	0,005360	0,5360%
167	islandGroup	91	19	110	0,005172	0,5172%
168	associatedTaxa	92	2	94	0,004419	0,4419%
169	http://unknown.org/recordEnteredBy	83	0	83	0,003902	0,3902%
170	footprintSRS	20	61	81	0,003808	0,3808%
171	behavior	69	11	80	0,003761	0,3761%
172	nameAccordingToID	55	7	62	0,002915	0,2915%
173	verbatimDepth	47	14	61	0,002868	0,2868%
174	acceptedNameUsageID	25	27	52	0,002445	0,2445%
175	earliestPeriodOrLowestSystem	52	0	52	0,002445	0,2445%
176	higherGeographyID	49	2	51	0,002398	0,2398%
177	earliestEpochOrLowestSeries	45	1	46	0,002163	0,2163%
178	formation	44	0	44	0,002069	0,2069%
179	associatedSequences	41	1	42	0,001975	0,1975%
180	earliestEraOrLowestErathem	42	0	42	0,001975	0,1975%
181	organismRemarks	41	1	42	0,001975	0,1975%
182	associatedOccurrences	33	1	34	0,001598	0,1598%
183	organismName	31	2	33	0,001551	0,1551%
184	typifiedName	32	0	32	0,001504	0,1504%
185	latestPeriodOrHighestSystem	28	0	28	0,001316	0,1316%
186	earliestAgeOrLowestStage	25	0	25	0,001175	0,1175%
187	latestEpochOrHighestSeries	24	1	25	0,001175	0,1175%
188	latestEraOrHighestErathem	24	0	24	0,001128	0,1128%

	Elemento/campo de metadados	Total dataset		Frequência (F)		
		Ocorrência	Evento	*Fa	**Fr	***Fp
189	earliestEonOrLowestEonothem	17	0	17	0,000799	0,0799%
190	group	14	1	15	0,000705	0,0705%
191	parentNameUsageID	14	0	14	0,000658	0,0658%
192	source	14	0	14	0,000658	0,0658%
193	latestAgeOrHighestStage	13	0	13	0,000611	0,0611%
194	pointRadiusSpatialFit	12	1	13	0,000611	0,0611%
195	latestEonOrHighestEonothem	11	0	11	0,000517	0,0517%
196	http://unknown.org/georeferenceRemarks	10	0	10	0,000470	0,0470%
197	organismScope	8	2	10	0,000470	0,0470%
198	member	9	0	9	0,000423	0,0423%
199	minimumDistanceAboveSurfaceInMeters	6	3	9	0,000423	0,0423%
200	nomenclaturalStatus	9	0	9	0,000423	0,0423%
201	associatedOrganisms	6	0	6	0,000282	0,0282%
202	http://unknown.org/occurrenceDetails	6	0	6	0,000282	0,0282%
203	lithostratigraphicTerms	5	1	6	0,000282	0,0282%
204	maximumDistanceAboveSurfaceInMeters	3	3	6	0,000282	0,0282%
205	http://unknown.org/organismQuantityType	2	3	5	0,000235	0,0235%
206	lowestBiostratigraphicZone	5	0	5	0,000235	0,0235%
207	footprintSpatialFit	4	0	4	0,000188	0,0188%
208	namePublishedInID	4	0	4	0,000188	0,0188%
209	highestBiostratigraphicZone	3	0	3	0,000141	0,0141%
210	http://unknown.org/license	3	0	3	0,000141	0,0141%
211	http://unknown.org/organismQuantity	1	2	3	0,000141	0,0141%
212	http://unknown.org/rightsHolder	3	0	3	0,000141	0,0141%
213	http://unknown.org/sampleSizeUnit	1	2	3	0,000141	0,0141%
214	http://unknown.org/sampleSizeValue	1	2	3	0,000141	0,0141%
215	Bed	1	0	1	0,000047	0,0047%
216	created	1	0	1	0,000047	0,0047%
217	http://unknown.org/class	1	0	1	0,000047	0,0047%
218	http://unknown.org/habitatproperty	1	0	1	0,000047	0,0047%
219	http://unknown.org/nick	1	0	1	0,000047	0,0047%
220	http://unknown.org/parentEventID	0	1	1	0,000047	0,0047%
221	http://unknown.org/vegetationLayer	0	1	1	0,000047	0,0047%

Fonte: elaboração própria autora - Fonte dos dados: GBIF (2020e).

*Fa = frequência absoluta que o metadado aparece nos datasets. **Fr = frequência relativa ao total de datasets. ***Fp = frequência percentual.

Foi identificados o total de 221 elementos de metadados. Destaca-se que há 357 *datasets* (282 ocorrência; 75 evento) que não apresentam registros de dados. Por isso, no quadro acima não há nenhum campo de metadados que esteja em uso na totalidade dos *datasets* analisados para essas classes. Considerando o percentual de *datasets* em que cada elemento está presente (coluna frequência relativa percentual no Quadro 26), a análise de tendência do uso do metadados (elementos/campos) é apresentada utilizando *decis* e *quartis* como medidas de posição – figura 27.

Quadro 27. Intervalos estabelecidos para análise tendências do uso de metadados nos *datasets*.

DECIS		QUARTIS	
1% - 10%	Q1	0% - 25%	Q1
11% - 20%	Q2	26% - 50%	Q2
21% - 30%	Q3	51% - 75%	Q3
31% - 40%	Q4	76% - 100%	Q4
41% - 50%	Q5		
51% - 60%	Q6		
61% - 70%	Q7		
71% - 80%	Q8		
81% - 90%	Q9		
91% - 100%	Q10		

Aplicações :

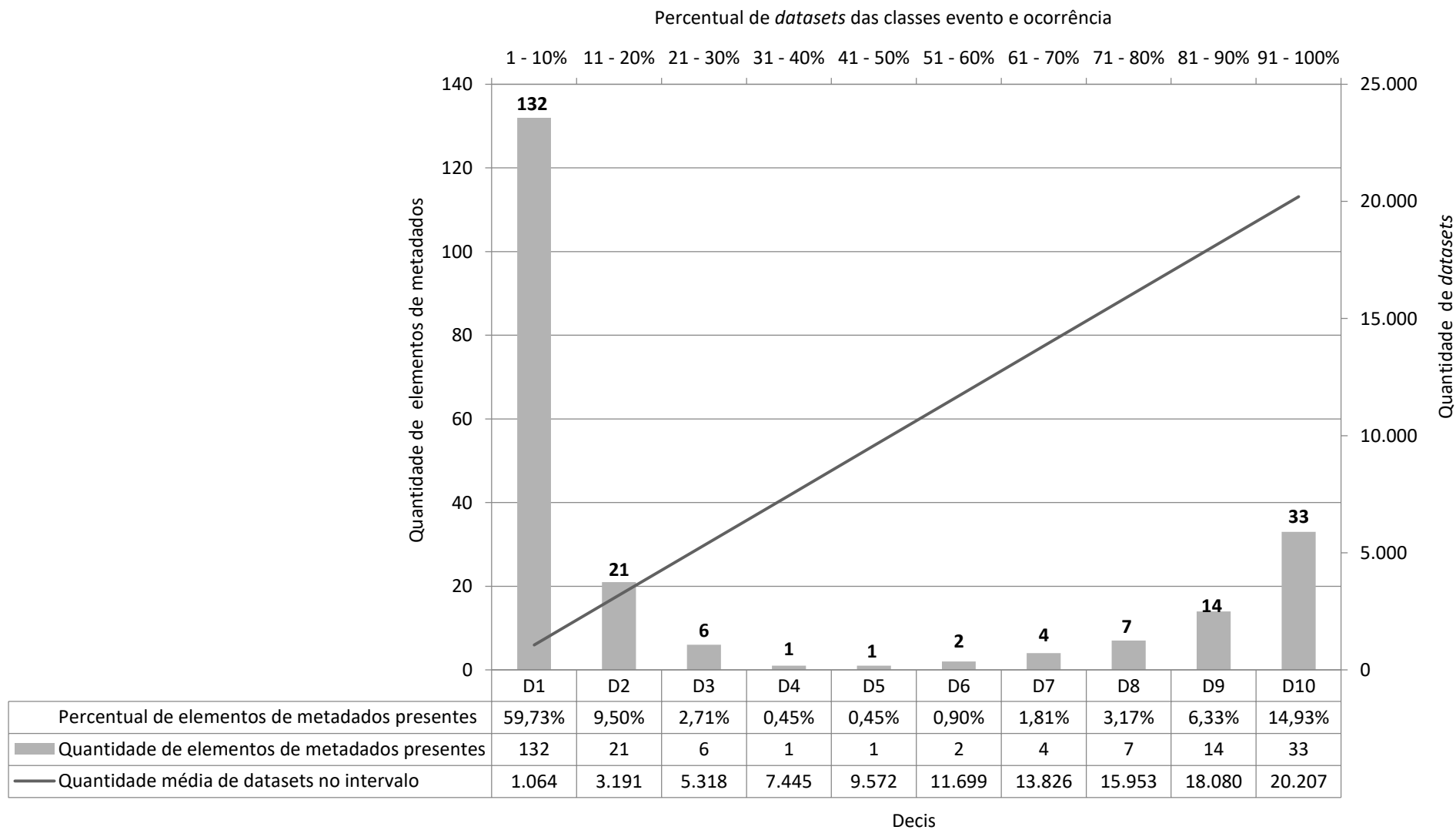
Perspectiva 1- Percentual de *datasets* que usam cada elemento.

Perspectiva 2 - Percentual de elementos em uso em cada *dataset*.

Fonte: elaboração própria autora.

Dessa forma, na primeira perspectiva de análise, os elementos de metadados usados no menor número *datasets* estão posicionados no Q_1 (intervalo que abrange 0 a 10% dos *datasets*) e aqueles usados no maior número *datasets* estão posicionados no Q_{10} (intervalo que abrange 90 a 100% dos *datasets*). Na segunda perspectiva de análise, também os *datasets* com menor número de elementos de metadados em uso estão posicionados em Q_1 (intervalo que abrange 0 a 10% dos elementos em uso) e aqueles com maior número estão posicionados no Q_{10} (intervalo que abrange 90 a 100% dos elementos em uso). A lista completa dos 21.270 *datasets* (ocorrência e evento) consta no ANEXO 11.

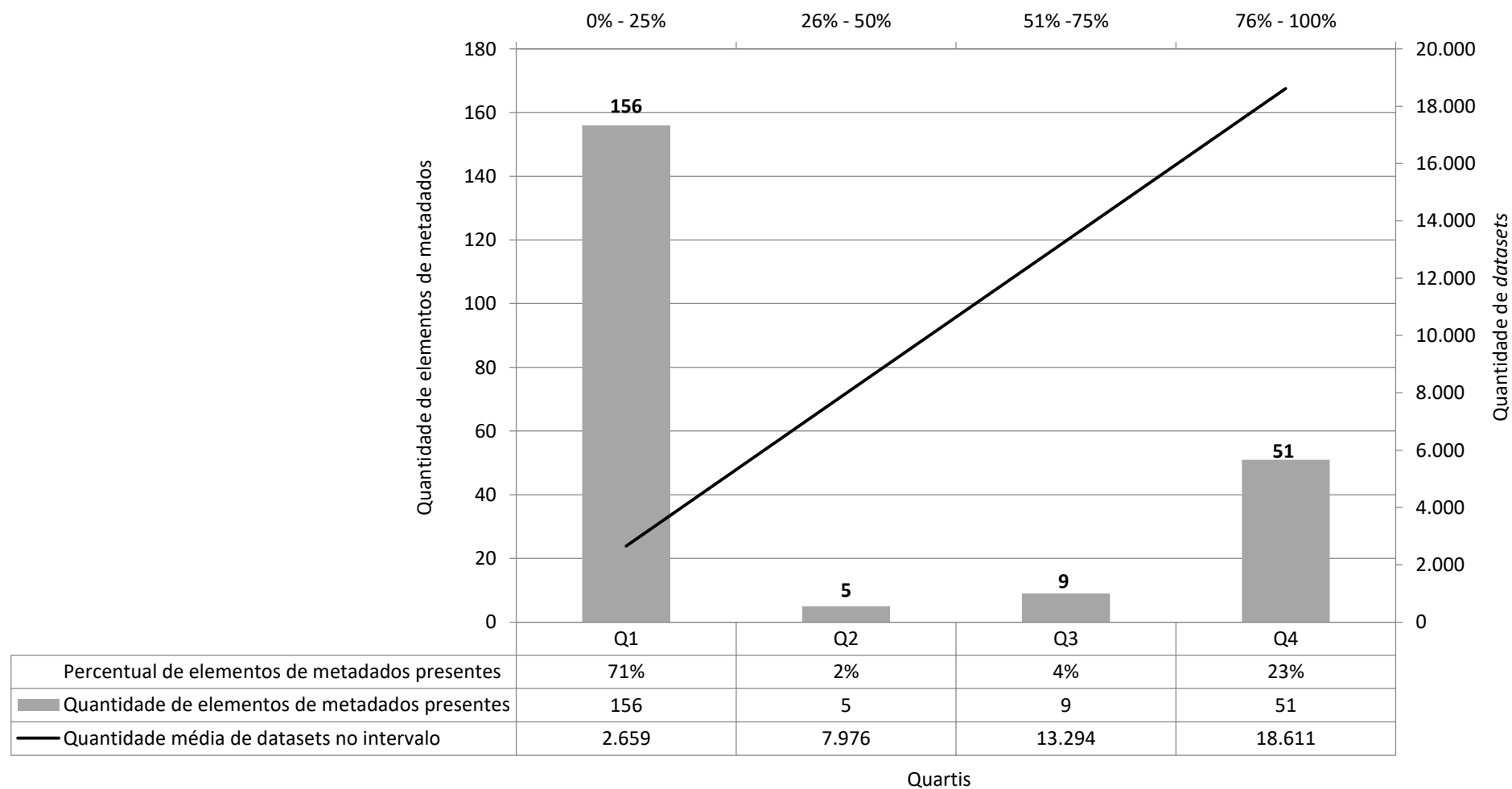
As figuras a seguir apresentam a relação entre quantidade de elementos de metadados presentes por percentual de *datasets* – figura 30 e figura 31. Nessa análise de tendência, com base na frequência dos elementos de metadados nos *datasets*, apresenta-se a quantidade de elementos em uso por percentual de *datasets* – *decis* e *quartis*.

Figura 30. Quantidade de elementos de metadados em uso por percentual de *datasets* (classes evento e ocorrência) – decis.

Fonte: elaboração própria autora - Fonte dos dados: GBIF (2020e).

Figura 31. Quantidade de elementos de metadados em uso por percentual de datasets (classes evento e ocorrência) – quartis.

Percentual de *datasets* das classes evento e ocorrência



Fonte: elaboração própria autora - Fonte dos dados: GBIF (2020e).

Considerando o total de 221 elementos de metadados identificados, verifica-se que a maior parte (132 elementos, quase 60%) está presente apenas em menos de 10% dos *datasets* – figura 30. Isso se deve possivelmente ao fato de que há elementos identificados relacionados a assuntos específicos dentro da temática biodiversidade e, portanto, não cabem a todos os *datasets*.

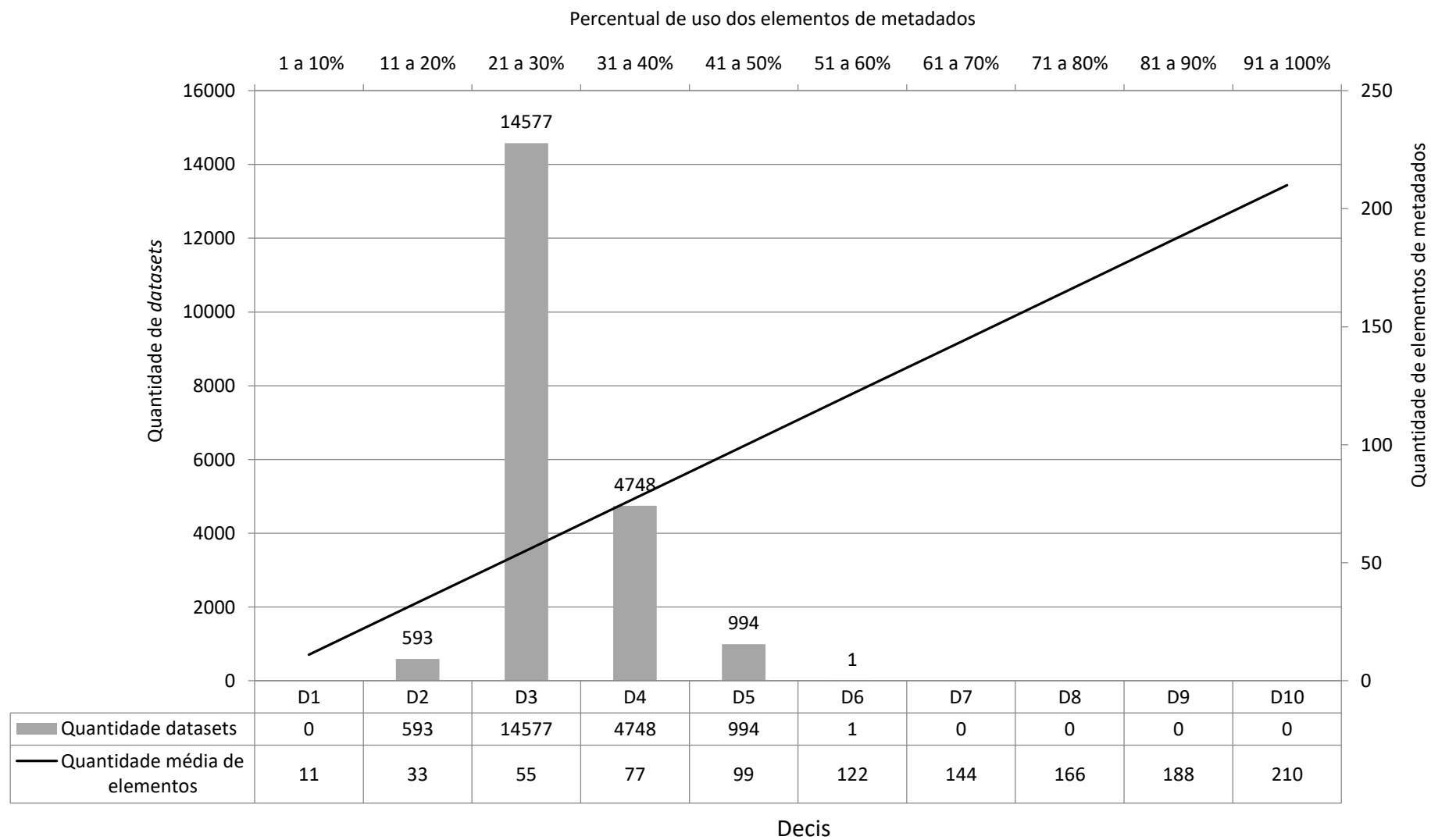
Além disso, o registro de *datasets* na GBIF apresenta campos obrigatórios e recomendados, mas permite também que existam elementos próprios dos projetos de pesquisas. Esses elementos de metadados estabelecidos especificamente para uso em determinados projetos também aparecem em poucos ou muitas vezes em apenas um único *dataset*.

Há também 33 elementos (15% dos 221 identificados) presentes em quase todos os conjuntos de dados analisados (mais de 90% dos *datasets*) – figura 30. Esses são os metadados com maior evidência de uso e referem-se aos elementos considerados obrigatórios, recomendados e fortemente recomendados no GMP e recomendações adicionais da GBIF.

De forma geral, o uso de cerca de 70% dos elementos de metadados se dá em menos de 25% dos *datasets* – figura 31. Por outro lado, cerca de 50 elementos de metadados estão em uso na descrição dos dados de 75 a 100% dos *datasets*. Ou seja, mais de 20% dos elementos de metadados identificados são utilizados para descrever a maior parte dos conjuntos de dados. Isto é bastante significativo e revela uma tendência à qualificação da descrição dos dados de pesquisa da biodiversidade.

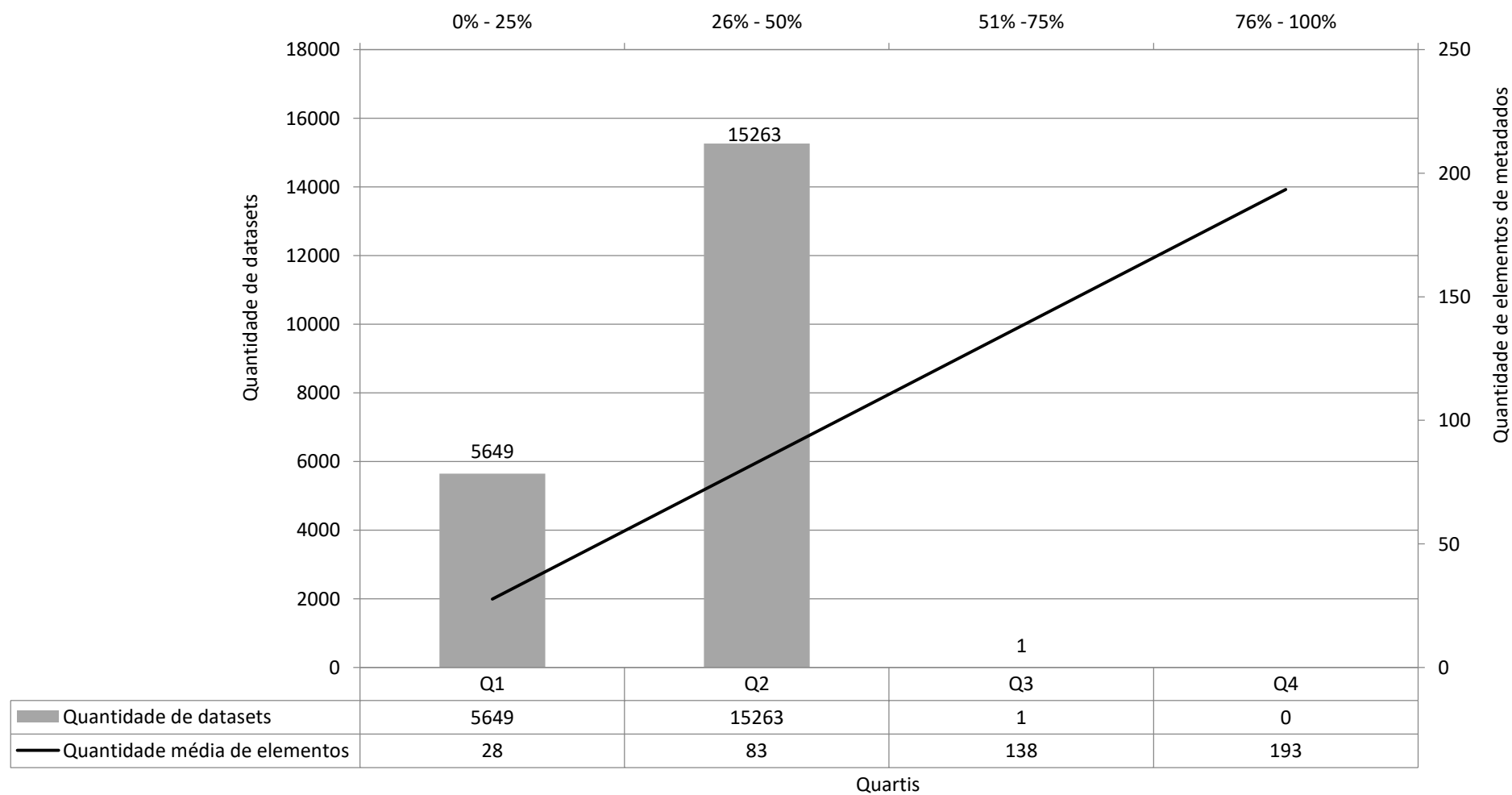
Essa análise permitiu identificar a frequência que os elementos de metadados aparecem nos *datasets* e, conseqüentemente, compreender quais apresentam maior uso na descrição dos dados. Outro aspecto relevante a ser observado refere-se à análise da quantidade de elementos de metadados presente em cada *dataset*, pois isso permite compreender qual a quantidade média de elementos em uso para descrever os dados.

As figuras a seguir apresentam a relação entre quantidade de *datasets* por percentual de elementos de metadados em uso – figura 32 e figura 33. Nessa análise de tendência, baseada na quantidade de elementos em uso por cada *dataset*, é apresentada a frequência relativa percentual que essa quantidade é encontrada nos 21.270 *datasets* das classes evento e ocorrência.

Figura 32. Quantidade de *datasets* por percentual de elementos de metadados em uso (classes evento e ocorrência) – decis.

Fonte: elaboração própria autora - Fonte dos dados: GBIF (2020e).

Figura 33. Quantidade de *datasets* por percentual de elementos de metadados em uso (classes evento e ocorrência) – quartis.
 Percentual de uso dos elementos de metadados



Fonte: elaboração própria autora - Fonte dos dados: GBIF (2020e).

Considerando o total de 21.270 *datasets*, quase 70% (14.577 *datasets*) utilizam para a descrição dos dados uma média de 55 elementos de metadados dos 221 identificados. Além disso, pouco mais de 20% dos *datasets* utilizam em média 77 elementos de metadados e menos de 5% utilizam em média 99 elementos. Apenas 1 *dataset* utiliza 120 elementos de metadados e nenhum utiliza todos – figura 32.

De forma geral, cerca de 70% dos *datasets* apresentam entre 55 a 110 elementos de metadados na descrição dos dados de pesquisa, uma média de 83 elementos por *dataset* – figura 33. Os outros 30% dos *datasets* apresentam entre 0 (existem 357 *datasets* vazios) e 55 elementos de metadados, uma média 28 elementos por *dataset*.

Por fim, os resultados encontrados demonstram que a maior parte dos *datasets* das classes evento e ocorrência apresentam descrição dos dados utilizando mais elementos de metadados do que apenas os obrigatórios ou recomendados. Também foi possível concluir que os elementos considerados obrigatórios e os recomendados estão presentes em mais de 90% dos *datasets*. A lista completa dos 21.270 *datasets* e a indicação de quantidade de elementos de metadados em cada um é apresentada no ANEXO 11.

(iii) Classe checklist

As classe *checklist* destaca-se das demais, pois está relacionada à descrição de espécies e normalização de nomenclaturas. A análise do uso dos metadados para a descrição dos dados iniciou pela identificação dos elementos utilizados nos 32.151 *datasets*. O quadro a seguir apresenta os elementos de metadados identificados e a indicação da quantidade de *datasets* em estão presentes – quadro 28.

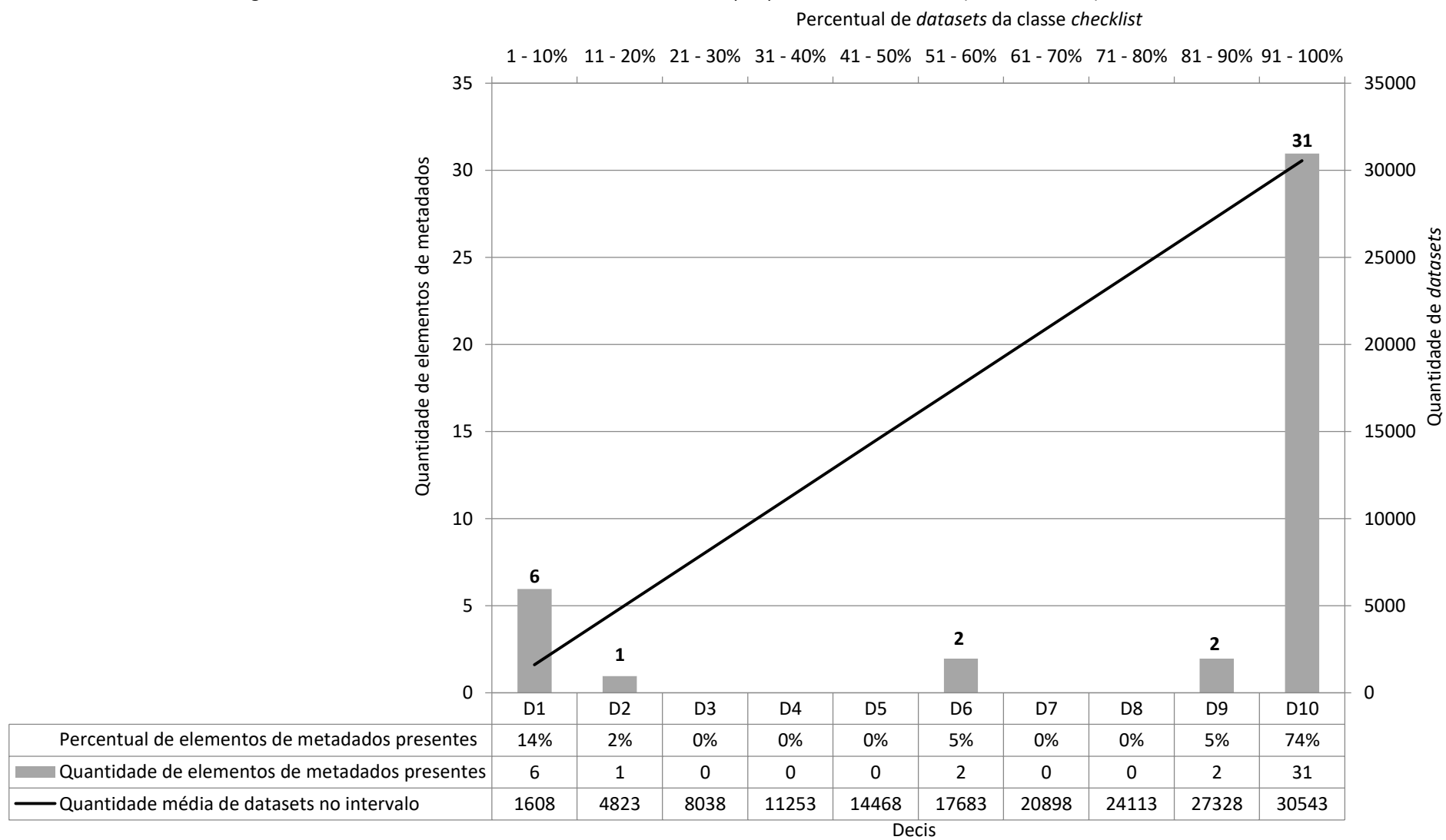
Quadro 28. Frequência do uso do elementos/campos de metadados para descrição dos dados em *datasets* da classe *checklist* no portal GBIF.

Classe Checklist					
	Elemento/campo de metadados	Frequência absoluta	Total de <i>Datasets</i>	Frequência relativa	Frequência percentual
1	key	31890	32151	0,991882	99,188%
2	nameKey	31890	32151	0,991882	99,188%
3	datasetKey	31890	32151	0,991882	99,188%
4	scientificName	31890	32151	0,991882	99,188%
5	authorship	31890	32151	0,991882	99,188%
6	nameType	31890	32151	0,991882	99,188%
7	origin	31890	32151	0,991882	99,188%
8	nomenclaturalStatus	31890	32151	0,991882	99,188%

Classe Checklist					
	Elemento/campo de metadados	Frequência absoluta	Total de Datasets	Frequência relativa	Frequência percentual
9	numDescendants	31890	32151	0,991882	99,188%
10	lastCrawled	31890	32151	0,991882	99,188%
11	lastInterpreted	31890	32151	0,991882	99,188%
12	issues	31890	32151	0,991882	99,188%
13	synonym	31890	32151	0,991882	99,188%
14	taxonomicStatus	31888	32151	0,991820	99,182%
15	canonicalName	31869	32151	0,991229	99,123%
16	rank	31723	32151	0,986688	98,669%
17	kingdom	31639	32151	0,984075	98,408%
18	kingdomKey	31639	32151	0,984075	98,408%
19	taxonID	31120	32151	0,967933	96,793%
20	parentKey	30911	32151	0,961432	96,143%
21	parent	30911	32151	0,961432	96,143%
22	references	30182	32151	0,938758	93,876%
23	phylum	29930	32151	0,930920	93,092%
24	phylumKey	29930	32151	0,930920	93,092%
25	classKey	29853	32151	0,928525	92,852%
26	class	29853	32151	0,928525	92,852%
27	order	29828	32151	0,927747	92,775%
28	orderKey	29828	32151	0,927747	92,775%
29	nubKey	29680	32151	0,923144	92,314%
30	family	29641	32151	0,921931	92,193%
31	familyKey	29641	32151	0,921931	92,193%
32	genus	28562	32151	0,888371	88,837%
33	genusKey	28562	32151	0,888371	88,837%
34	species	16128	32151	0,501633	50,163%
35	speciesKey	16128	32151	0,501633	50,163%
36	publishedIn	5869	32151	0,182545	18,254%
37	accordingTo	159	32151	0,004945	0,495%
38	remarks	68	32151	0,002115	0,212%
39	acceptedKey	52	32151	0,001617	0,162%
40	accepted	52	32151	0,001617	0,162%
41	constituentKey	2	32151	0,000062	0,006%
42	sourceTaxonKey	1	32151	0,000031	0,003%

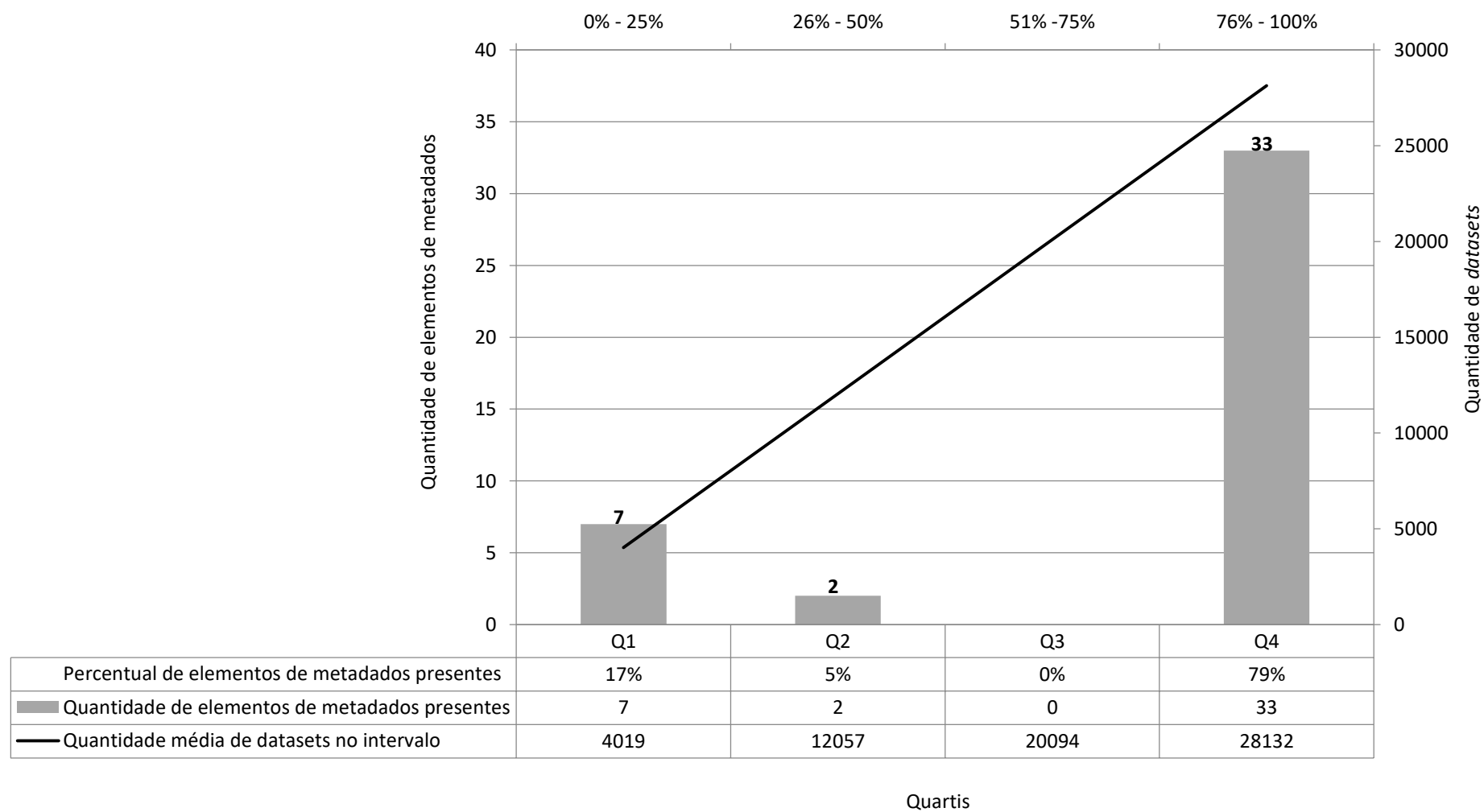
Fonte: elaboração própria autora - Fonte dos dados: GBIF (2020e).

A lista completa de *datasets* consta no ANEXO 11. Há 261 *datasets* da classe *checklist* sem registros de dados. Por isso, nenhum elemento identificado está em uso na totalidade dos *datasets* analisados para essa classe. A análise de tendência apresentada a seguir também utilizou decis e quartis como medidas de posição.

Figura 34. Quantidade de elementos de metadados em uso por percentual de datasets (classe *checklist*) – decis.

Fonte: elaboração própria autora - Fonte dos dados: GBIF (2020e).

Figura 35. Quantidade de elementos de metadados em uso por percentual de datasets (classe checklist) – decis.

Percentual de *datasets* da classe *checklist*

Fonte: elaboração própria autora - Fonte dos dados: GBIF (2020e).

Considerando o total de 42 elementos de metadados identificados, verifica-se que a maior parte (31 elementos, 74% dos 42 identificados) está presente em mais de 90% dos *datasets* – figura 34. Portanto, esses apresentam maior evidência de uso e referem-se aos elementos considerados obrigatórios, recomendados e fortemente recomendados no GMP e recomendações adicionais da GBIF.

Isso se deve principalmente ao fato de que essa classe apresenta como função principal apresentar listas de espécies e normalizar nomenclaturas, por isso, independente do grupo da biodiversidade apresentado, a maior parte dos elementos identificados é considerada item recomendado. Consequentemente, também há poucos elementos estabelecidos para uso específico em determinado *dataset*.

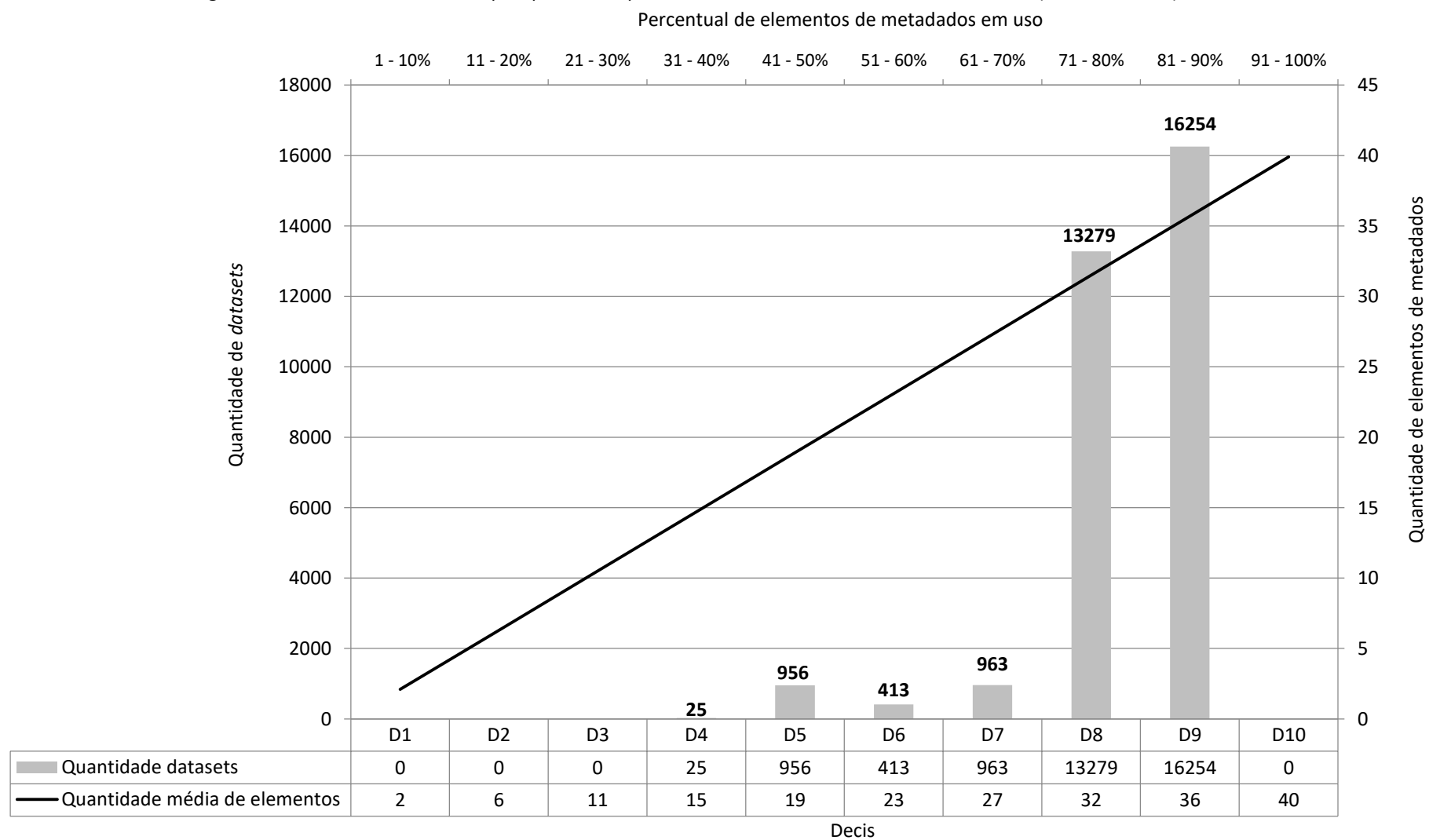
De forma geral, o uso de cerca de 80% dos elementos de metadados se dá em mais de 75% dos *datasets* – figura 35. Ou seja, 33 elementos de metadados identificados são utilizados para descrever a maior parte dos *datasets* da classe *checklist*. Isto é bastante significativo e revela maior uniformidade na descrição dos dados dessa classe.

Analisar a frequência que os elementos de metadados aparecem nos *datasets* permitiu compreender quais apresentam maior uso na descrição dos dados. Outro aspecto relevante a ser observado refere-se à análise da quantidade de elementos de metadados presente em cada *dataset*, pois isso permite compreender qual a quantidade média de elementos em uso para descrever os dados.

As figuras a seguir apresentam a relação entre quantidade de *datasets* por percentual de elementos de metadados em uso – figura 36 e figura 37. Nessa análise de tendência, baseada na quantidade de elementos em uso por cada *dataset*, é apresentada a frequência relativa percentual que essa quantidade é encontrada nos 32.151 *datasets* da classe *checklist*.

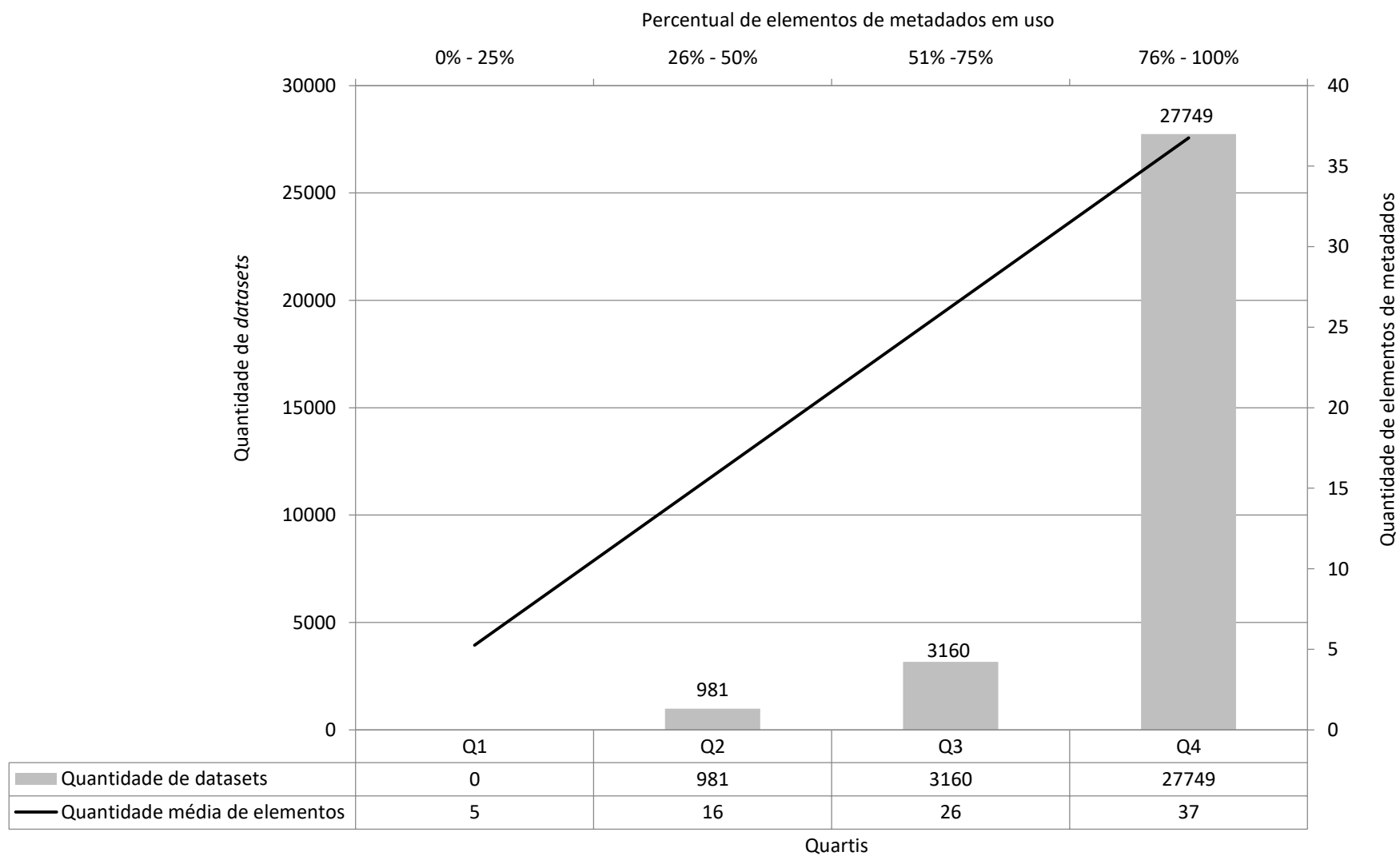
Considerando o total de *datasets* da classe *checklist*, cerca de 50% (16.254 *datasets*) utilizam para a descrição dos dados uma média de 36 elementos de metadados dos 42 identificados – figura 36. Além disso, todos os *datasets* usam no mínimo 15 elementos. De forma geral, 84% dos *datasets* utilizam mais de 75% dos elementos de metadados – figura 37.

Figura 36. Quantidade de *datasets* por quantidade percentual de elementos de metadados em uso (classe *checklist*) - decis.



Fonte: elaboração própria autora - Fonte dos dados: GBIF (2020e).

Figura 37. Quantidade de datasets por quantidade percentual de elementos de metadados em uso (classe checklist) – quartis.



Fonte: elaboração própria autora - Fonte dos dados: GBIF (2020e).

Por fim, os resultados encontrados demonstram que a maior parte dos *datasets* da classe *checklist* apresenta descrição dos dados utilizando elementos de metadados que vão além dos considerados obrigatórios ou recomendados. Também foi possível concluir que os elementos considerados obrigatórios e os recomendados estão presentes em mais de 90% dos *datasets*. A lista completa dos 32.151 *datasets* e a indicação de quantidade de elementos de metadados apresentada em cada um é apresentada no ANEXO 11.

4.2.4.5 Contribuições do Brasil

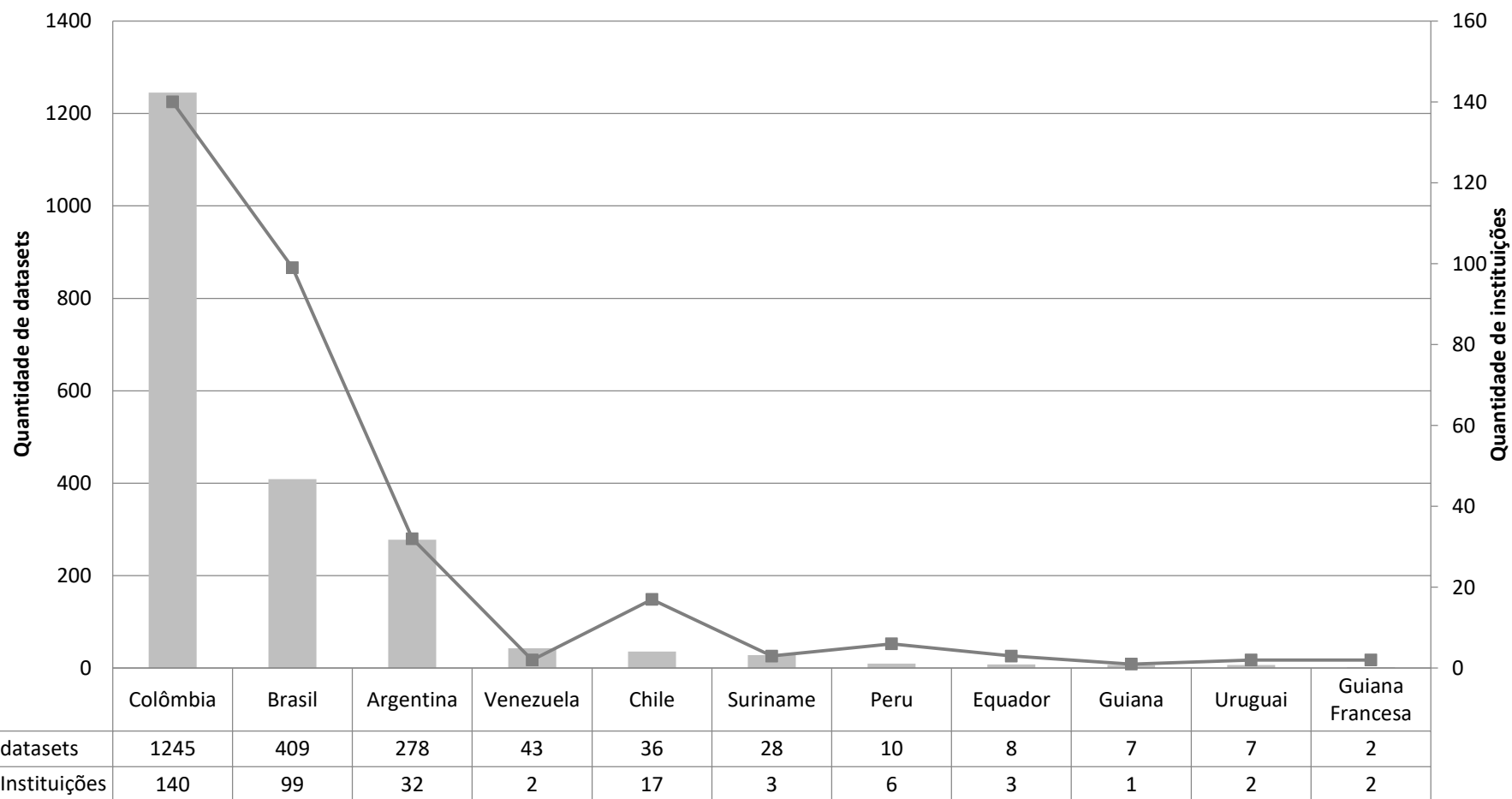
Cerca de 6% dos *datasets* está vinculado a da América do Sul. Há uma acentuada concentração de *datasets* e instituições no continente europeu (responsável por 87% dos *datasets*). Apesar da disparidade em relação a Europa, o continente Sul Americano ocupa a segunda posição na classificação geral e também se destacada em relação aos demais continentes.

A Colômbia (em 5ª posição) e o Brasil (em 9ª posição) estão entre os 10 países do mundo com maior número de *datasets* e também com mais instituições publicadoras (responsáveis pelos *datasets*). Além disso, representam a liderança em pesquisa sobre biodiversidade na América do Sul – figura 38.

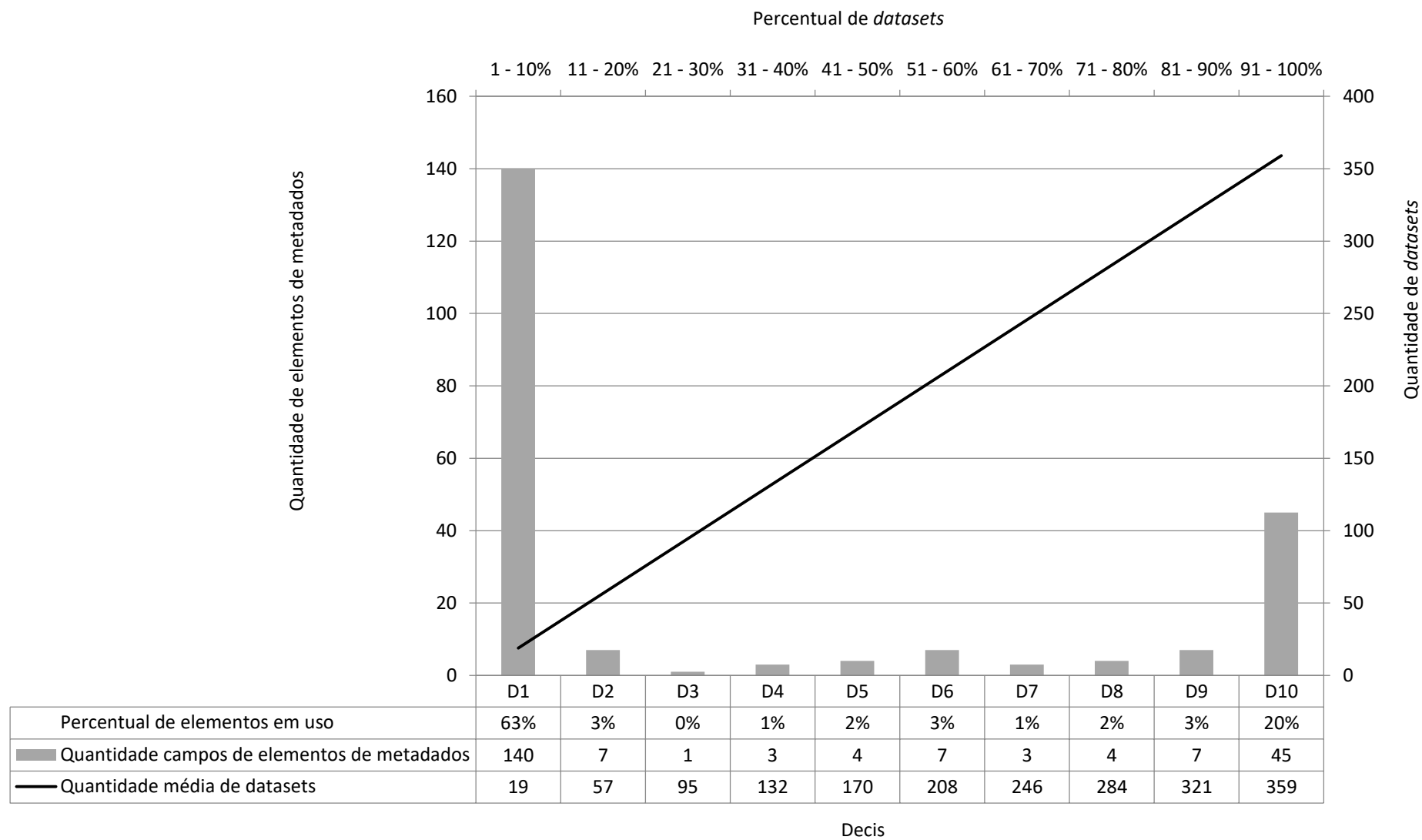
O Brasil tem 99 instituições de pesquisa cadastradas na GBIF, responsáveis por 409 *datasets* (16 da classe metadado de recurso; 367 da classe ocorrência; 15 da classe *checklist*; e 11 da classe eventos). Há 18 *datasets* que não apresentam registros de dados (15 da classe metadado de recurso; 3 da classe ocorrência). A lista completa de *datasets* e instituições vinculados ao Brasil consta no ANEXO 10. Considerando os elementos identificados (total de 221) para a descrição dos dados de *datasets* da classe ocorrência e evento, a seguir é apresentada a análise do uso dos metadados da amostra vinculada ao Brasil.

Analisando a perspectiva da frequência do uso dos metadados, utilizou-se decis como medida de posição para indicar o percentual dos *datasets* do Brasil que utiliza cada um dos 221 elementos – figura 39. Outra perspectiva analisada refere-se à quantidade de elementos utilizados em cada um dos 409 *datasets* – figura 40.

Figura 38. Panorama na América do Sul: quantidade de datasets e instituições por país.

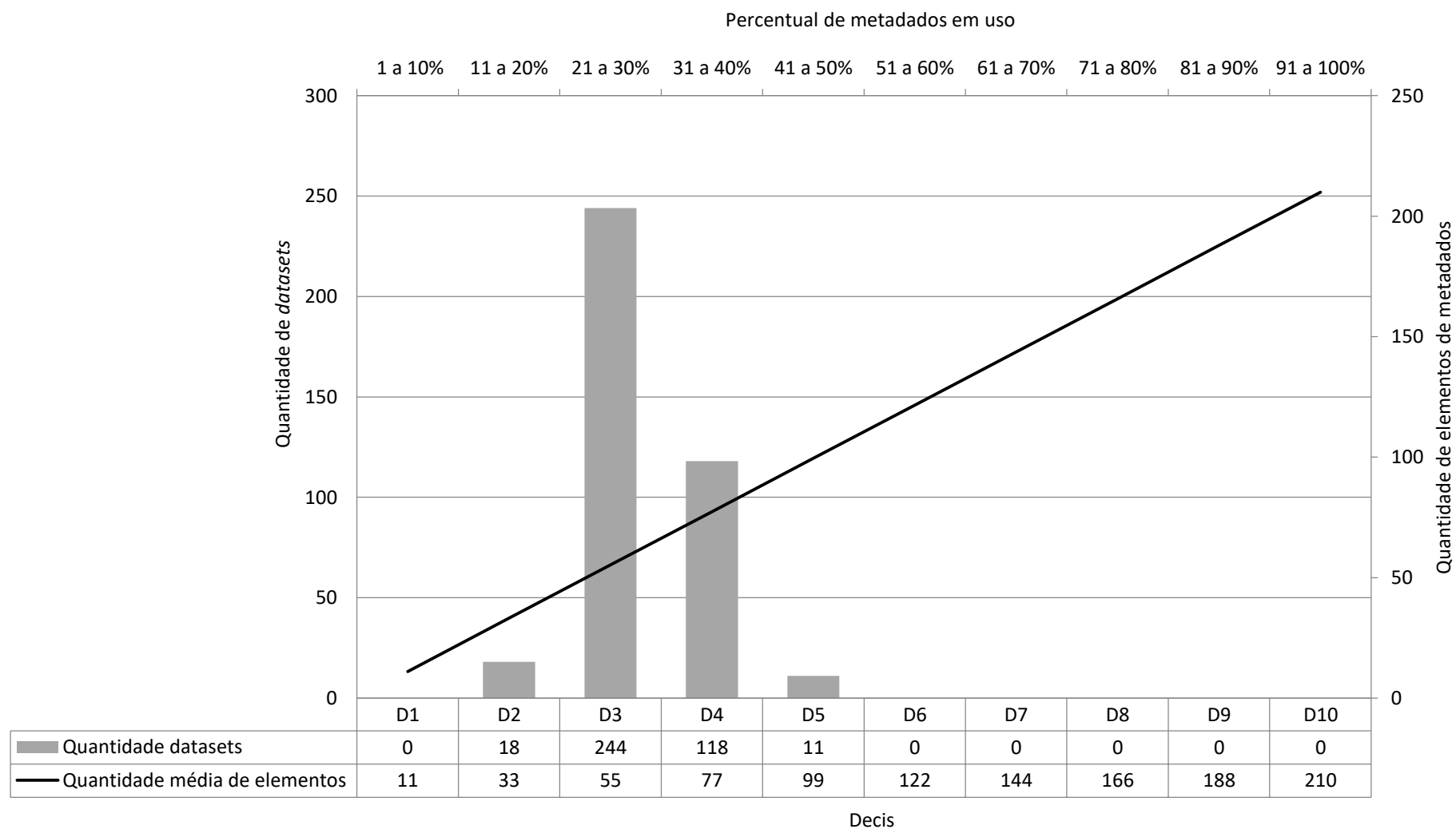


Fonte: elaboração própria autora - Fonte dos dados: GBIF (2020e).

Figura 39. Brasil: quantidade de elementos de metadados em uso por percentual de *datasets* (classes evento e ocorrência) – decis.

Fonte: elaboração própria autora - Fonte dos dados: GBIF (2020e).

Figura 40. Brasil: quantidade de *datasets* por percentual de elementos de metadados em uso (classes evento e ocorrência) – decis



. Fonte: elaboração própria autora - Fonte dos dados: GBIF (2020e).

A maior parte dos elementos de metadados (140 de 221 identificados) está presente em cerca de 10% dos *datasets* brasileiros. Além disso, há cerca de 20% dos elementos (45 de 221 identificados) que está presente em mais de 90% dos *datasets* – figura 39. Já na perspectiva de quantidade de elementos por *dataset*, mais de 50% dos *datasets* apresentam em média 55 elementos de metadados para descrição dos dados. Além disso, 118 *datasets* (28%) utilizam em média 77 elementos para a descrição dos dados. Esses resultados sugerem que o Brasil mantém as tendências observadas na análise global referentes às classes ocorrência e eventos – figura 30 e Figura 32. Há 15 *datasets* da classe *checklist* vinculados a instituições brasileiras, 13 apresentam 33 elementos para descrição, 1 apresenta 30 elementos e 1 apresenta 28 elementos. Dessa forma, dos 42 elementos de metadados identificados na análise global para a classe *checklist* a maior parte está presente em todos os *datasets* brasileiros.

4.3 ETAPA III – IDENTIFICAÇÃO DE TENDÊNCIAS

Na etapa I deste estudo – subseção 4.1 – foram identificados os padrões de metadados relevantes para descrição dos dados de pesquisa na temática biodiversidade, que são o EML, DwC, ABCD e o perfil de metadados da GBIF (GMP). Esses achados constituem a linha de base deste estudo, utilizada como referência para análise comparativa com as evidências de uso de metadados apresentadas na etapa II – subseção 4.2. A partir da análise dos 53.729 conjuntos de dados depositados na rede GBIF, foram identificados os elementos (também chamados atributos ou campos) de metadados utilizados para a descrição dos *datasets* e respectivos dados de pesquisa que os integram. A seguir, é apresentada a análise comparativa entre os resultados encontrados na etapa I e II.

Iniciando pela descrição dos *datasets*, identificou-se o padrão de metadados relacionado a cada um dos 51 elementos de metadados coletados, conforme apresentado no quadro 29 – GMP (GBIF, 2020a); EML (KNB, 2019); ABCD (TDWG, 2019b); DwC (TDWG, 2015a). Em seguida foram identificados elementos de metadados utilizados para a descrição dos dados de pesquisa. O quadro 30 (classe evento e ocorrência) e o quadro 31 (classe *checklist*) apresentam a relação desses elementos com a documentação dos padrões de metadados – GMP (GBIF, 2020a); EML (KNB, 2019); ABCD (TDWG, 2019b); DwC (TDWG, 2015a).

Quadro 29. Identificação dos padrões relacionados aos elementos de metadados em uso nas descrição dos *datasets*.

	Elemento	EML	ABCD	DwC	GBIF Metadata Profile
1	abbreviation	X	Element ResourceGroup / alternateIdentifier		
2	additionalinfo	X	Element ResourceGroup / additionalInfo		
3	alias	X	Element ResourceGroup / shortName		
4	bibliographiccitations	X	Element DatasetType / usageCitation		
5	citation	X	Element DatasetType / usageCitation		
6	collections		X Element Unit / NamedCollectionsOrSurveys		
7	comments		X Element TaxonIdentified / NameComments		
8	contacts	X	Element ResourceGroup / creator		
9	countrycoverage	X	Element ResourceGroup / coverage		
10	created	X	Element ResourceGroup / creator		
11	createdby	X	Element ResourceGroup / creator		
12	curatorialunits			X DwC curatorial extension	
13	datadescriptions	X	Element ResourceGroup / creator		
14	datalanguage				X Classe Dataset
15	deleted	X	Element ResourceGroup / creator		
16	description	X	Element ResourceGroup / creator		
17	doi		X Element Reference / ResourceURIs		
18	duplicateofdatasetkey				X Classe Dataset
19	endpoints				X Classe Dataset

	Elemento	EML	ABCD	DwC	GBIF Metadata Profile
20	external				X Classe Dataset
21	geographiccoverages	X	Element ResourceGroup / coverage		
22	homepage	X	Complex Type UriType		
23	identifiers				X Classe Dataset
24	installationkey				X Classe Dataset
25	key				X Classe Dataset
26	keywordcollections	X	Element ResourceGroup / keywordSet		
27	language	X	Element ResourceGroup / language		
28	license	X	Element ResourceGroup / licensed		
29	lockedforautoupdate				X Classe Dataset
30	logourl		X Element ContentMetadata / LogoURL		
31	machinetags				X Classe Dataset
32	maintenancedescription	X	Element MaintenanceType / changeHistory		
33	maintenanceupdatefrequency	X	Element MaintenanceType / changeHistory		
34	modified	X	Element MaintenanceType / changeHistory		
35	modifiedby	X	Element MaintenanceType / changeHistory		
36	network				X Classe Dataset
37	numconstituents				X Classe Dataset
38	parentdatasetkey				X Classe Dataset

	Elemento	EML	ABCD	DwC	GBIF Metadata Profile	
39	project	X Element DatasetType / project				
40	pubdate	X Element ResourceGroup / pubDate				
41	publishingorganizationkey	X Element ResourceGroup / metadataProvider				
42	purpose	X Element DatasetType / purpose				
43	rights	X Element ResourceGroup / intellectualRights				
44	samplingdescription	X Element MethodsType / sampling / samplingDescription				
45	subtype (DatasetSubtype)			X DwC extensions		
46	tags				X Classe Dataset	
47	taxonomiccoverages	X Element ResourceGroup / coverage				
48	temporalcoverages	X Element ResourceGroup / coverage				
49	title	X Element ResourceGroup / title				
50	type (datasetType)			X DwC-A core type		
51	version		X Element ContentMetadata / Version			
* 30			** 5		*** 3	
**** 13						

Fonte: própria autora. Fonte de dados: (GBIF, 2020c; GBIF, 2020d; TDWG, 2019b; TDWG, 2015a; GBIF, 2020f)

*Total de elementos de metadados herdados da EML.

**Total de elementos de metadados herdados do ABCD.

***Total de elementos de metadados herdados do DwC.

****Total de elementos de metadados exclusivos do GMP.

Quadro 30. Identificação dos padrões relacionados aos elementos de metadados em uso para descrição dos dados de pesquisa das classes eventos e ocorrência

CLASSES OCORRÊNCIA E EVENTOS		PADRÃO DE METADADOS					Quantidade Datasets*
Elementos de metadados identificados		DwC	GBIF		Outro		
1	basisOfRecord	X	Classe Record-level				20913
2	crawlId			X	Classe VerbatimOccurrence		20913
3	datasetKey			X	Classe Dataset		20913
4	Extensions			X	Classe VerbatimOccurrence		20913
5	Facts			X	Classe Occurrence		20913
6	gbifID			X	Classe Enum Constant		20913
7	identifiedByIDs			X	Classe Enum Constant		20913
8	Identifiers			X	Classe Enum Constant		20913
9	installationKey			X	Classe Dataset		20913
10	Issues			X	Classe Enum OccurrenceIssue		20913
11	Key			X	Classe VerbatimOccurrence		20913
12	Kingdom	X	Classe Taxon				20913
13	kingdomKey			X	Classe Occurrence		20913
14	lastInterpreted			X	Classe Enum Constant		20913
15	lastParsed			X	Classe Enum Constant		20913
16	License	X	Classe Record-level				20913
17	Media			X	Classe Occurrence		20913
18	Protocol			X	Classe Dataset		20913
19	publishingOrgKey			X	Classe Dataset		20913
20	recordedByIDs			X	Classe Enum Constant		20913
21	Relations			X	Classe OccurrenceRelation		20913
22	scientificName	X	Classe Taxon				20913
23	taxonKey			X	Classe Occurrence		20913
24	taxonRank	X	Classe Taxon				20913

CLASSES OCORRÊNCIA E EVENTOS		PADRÃO DE METADADOS				Quantidade Datasets*
Elementos de metadados identificados		DwC	GBIF	Outro		
25	lastCrawled		X Classe Enum Constant		20912	
26	publishingCountry		X Classe VerbatimOccurrence		20678	
27	acceptedScientificName		X Classe Enum Constant		20623	
28	acceptedTaxonKey		X Classe Occurrence		20623	
29	taxonomicStatus	X Classe Taxon			20623	
30	genericName		X Classe Enum Constant		20608	
31	Identifier		X Classe Enum Constant		20190	
32	Phylum	X Classe Taxon			19962	
33	phylumKey		X Classe Occurrence		19962	
34	eventDate	X Classe Event			18715	
35	Year	X Classe Event			18715	
36	decimalLatitude	X Classe Location			18655	
37	decimalLongitude	X Classe Location			18655	
38	geodeticDatum	X Classe UseWithIRI			18655	
39	institutionCode	X Classe Record-level			18481	
40	Month	X Classe Event			18245	
41	Country	X Classe Location			17967	
42	countryCode	X Classe Location			17967	
43	Day	X Classe Event			17954	
44	Class	X Classe Taxon			17554	
45	classKey		X Classe Occurrence		17554	
46	Order	X Classe Taxon			17017	
47	orderKey		X Classe Occurrence		17017	
48	Family	X Classe Taxon			16904	
49	familyKey		X Classe Occurrence		16904	
50	Genus	X Classe Taxon			16398	

CLASSES OCORRÊNCIA E EVENTOS		PADRÃO DE METADADOS				Quantidade Datasets*
Elementos de metadados identificados		DwC	GBIF	Outro		
51	genusKey		X Classe Occurrence			16398
52	collectionCode	X	Classe Record-level			15281
53	recordedBy	X	Classe Occurrence			15112
54	catalogNumber	X	Classe Occurrence			15007
55	species		X Classe Occurrence			13670
56	speciesKey		X Classe Occurrence			13670
57	specificEpithet	X	Classe Taxon			13660
58	modified	X	Classe Record-level			13069
59	occurrenceID	X	Classe Occurrence			11781
60	datasetID	X	Classe Record-level			11214
61	locality	X	Classe Location			8795
62	identifiedBy	X	Classe Identification			6563
63	stateProvince	X	Classe Location			6311
64	geologicalContextID	X	Classe GeologicalContext			6187
65	county	X	Classe Location			5954
66	coordinateUncertaintyInMeters	X	Classe Location			4908
67	eventID	X	Classe Event			4744
68	occurrenceStatus	X	Classe Occurrence			4621
69	taxonID	X	Classe Taxon			3948
70	language	X	Classe Record-level			3480
71	http://rs.tdwg.org/dwc/terms/organismQuantityType	X	Classe Occurrence			3466
72	organismQuantityType	X	Classe Occurrence			3466
73	type	X	Classe Record-level			3463
74	http://rs.tdwg.org/dwc/terms/organismQuantity	X	Classe Occurrence			3429
75	nameAccordingTo	X	Classe Taxon			3429
76	organismQuantity	X	Classe Occurrence			3373

CLASSES OCORRÊNCIA E EVENTOS		PADRÃO DE METADADOS				Quantidade Datasets*
Elementos de metadados identificados		DwC	GBIF	Outro		
77	individualCount	X	Classe Occurrence			3224
78	datasetName	X	Classe Record-level			2862
79	elevation			X	Classe Enum Constant	2859
80	vernacularName	X	Classe Taxon			2816
81	institutionID	X	Classe Record-level			2646
82	continent	X	Classe Location			2614
83	dynamicProperties	X	Classe Record-level			2523
84	identificationVerificationStatus	X	Classe Identification			2513
85	originalNameUsage	X	Classe Taxon			2437
86	elevationAccuracy			X	Classe Enum Constant	2300
87	rightsHolder	X	Classe Record-level			2234
88	samplingProtocol	X	Classe Event			2198
89	projectId			X	Classe VerbatimOccurrence	2132
90	municipality	X	Classe Location			2092
91	habitat	X	Classe UseWithIRI			2090
92	depth			X	Classe Enum Constant	2059
93	higherClassification	X	Classe Taxon			2033
94	preparations	X	Classe Occurrence			1986
95	collectionID	X	Classe Record-level			1967
96	ownerInstitutionCode	X	Classe Record-level			1959
97	depthAccuracy			X	Classe Enum Constant	1895
98	verbatimCoordinateSystem	X	Classe Location			1733
99	scientificNameID	X	Classe Taxon			1730
100	dateIdentified	X	Classe Identification			1677
101	locationID	X	Classe Location			1612
102	occurrenceRemarks	X	Classe Occurrence			1598

CLASSES OCORRÊNCIA E EVENTOS		PADRÃO DE METADADOS				Quantidade Datasets*
Elementos de metadados identificados		DwC	GBIF	Outro		
103	dataGeneralizations	X	Classe Record-level			1595
104	locationRemarks	X	Classe Location			1549
105	recordNumber	X	Classe Occurrence			1530
106	bibliographicCitation	X	Classe Record-level			1468
107	eventRemarks	X	Classe Event			1410
108	fieldNumber	X	Classe UseWithIRI			1392
109	footprintWKT	X	Classe Location			1377
110	startDayOfYear	X	Classe Event			1328
111	eventTime	X	Classe Event			1313
112	georeferenceSources	X	Classe UseWithIRI			1175
113	http://rs.tdwg.org/dwc/terms/sampleSizeUnit	X	Classe Event			1145
114	sampleSizeUnit	X	Classe Event			1145
115	http://rs.tdwg.org/dwc/terms/sampleSizeValue	X	Classe Event			1139
116	sampleSizeValue	X	Classe Event			1112
117	identificationRemarks	X	Classe Identification			1069
118	higherGeography	X	Classe Location			1060
119	acceptedNameUsage	X	Classe Taxon			1056
120	identificationID	X	Classe Identification			1038
121	verbatimEventDate	X	Classe Event			1013
122	parentNameUsage	X	Classe Taxon			992
123	taxonConceptID	X	Classe Taxon			988
124	namePublishedInYear	X	Classe Taxon			965
125	materialSampleID	X	Classe MaterialSample			958
126	verbatimLocality	X	Classe Location			946
127	identificationReferences	X	Classe Identification			926
128	references	X	Classe Record-level			873

CLASSES OCORRÊNCIA E EVENTOS		PADRÃO DE METADADOS				Quantidade Datasets*
Elementos de metadados identificados		DwC	GBIF	Outro		
129	relativeOrganismQuantity		X Classe Enum Constant		840	
130	sex	X	Classe UseWithIRI		770	
131	waterBody	X	Classe Location		748	
132	otherCatalogNumbers	X	Classe Occurrence		693	
133	nomenclaturalCode	X	Classe Taxon		666	
134	networkKeys		X Classe Dataset		657	
135	lifeStage	X	Classe Occurrence		648	
136	accessRights	X	Classe Record-level		646	
137	samplingEffort	X	Classe Event		623	
138	georeferencedBy	X	Classe UseWithIRI		561	
139	verbatimElevation	X	Classe Location		537	
140	programmeAcronym		X Classe VerbatimOccurrence		485	
141	georeferenceVerificationStatus	X	Classe UseWithIRI		430	
142	associatedReferences	X	Classe Occurrence		428	
143	georeferenceProtocol	X	Classe Location		423	
144	disposition	X	Classe Occurrence		410	
145	infraspecificEpithet	X	Classe Taxon		397	
146	endDayOfYear	X	Classe Event		382	
147	georeferencedDate	X	Classe Location		371	
148	previousIdentifications	X	Classe Organism		366	
149	verbatimSRS	X	Classe Location		353	
150	coordinatePrecision	X	Classe Location		352	
151	georeferenceRemarks	X	Classe Location		282	
152	informationWithheld	X	Classe UseWithIRI		279	
153	establishmentMeans	X	Classe UseWithIRI		239	
154	identificationQualifier	X	Classe Identification		238	

CLASSES OCORRÊNCIA E EVENTOS		PADRÃO DE METADADOS				Quantidade Datasets*
Elementos de metadados identificados		DwC	GBIF	Outro		
155	rights		X Classe Enum XmpRightsTerm			228
156	locationAccordingTo	X	Classe Location			208
157	island	X	Classe Location			194
158	fieldNotes	X	Classe UseWithIRI			190
159	taxonRemarks	X	Classe Taxon			187
160	organismID	X	Classe Organism			182
161	parentEventID	X	Classe Event			182
162	http://unknown.org/recordId			X	desconhecido	155
163	typeStatus	X	Classe Identification			152
164	reproductiveCondition	X	Classe Occurrence			146
165	namePublishedIn	X	Classe Taxon			124
166	verbatimTaxonRank	X	Classe Taxon			114
167	islandGroup	X	Classe Location			110
168	associatedTaxa	X	Classe Occurrence			94
169	http://unknown.org/recordEnteredBy			X	desconhecido	83
170	footprintSRS	X	Classe Location			81
171	behavior	X	Classe UseWithIRI			80
172	nameAccordingToID	X	Classe Taxon			62
173	verbatimDepth	X	Classe Location			61
174	acceptedNameUsageID	X	Classe Taxon			52
175	earliestPeriodOrLowestSystem	X	Classe GeologicalContext			52
176	higherGeographyID	X	Classe Location			51
177	earliestEpochOrLowestSeries	X	Classe GeologicalContext			46
178	formation	X	Classe GeologicalContext			44
179	associatedSequences	X	Classe Occurrence			42
180	earliestEraOrLowestErathem	X	Classe GeologicalContext			42

CLASSES OCORRÊNCIA E EVENTOS		PADRÃO DE METADADOS				Quantidade Datasets*
Elementos de metadados identificados		DwC	GBIF	Outro		
181	organismRemarks	X	Classe Organism			42
182	associatedOccurrences	X	Classe Organism			34
183	organismName	X	Classe Organism			33
184	typifiedName			X	Classe Enum Constant	32
185	latestPeriodOrHighestSystem	X	Classe GeologicalContext			28
186	earliestAgeOrLowestStage	X	Classe GeologicalContext			25
187	latestEpochOrHighestSeries	X	Classe GeologicalContext			25
188	latestEraOrHighestErathem	X	Classe GeologicalContext			24
189	earliestEonOrLowestEonothem	X	Classe GeologicalContext			17
190	group	X	Classe GeologicalContext			15
191	parentNameUsageID	X	Classe Taxon			14
192	source			X	Classe VernacularName	14
193	latestAgeOrHighestStage	X	Classe GeologicalContext			13
194	pointRadiusSpatialFit	X	Classe Location			13
195	latestEonOrHighestEonothem	X	Classe GeologicalContext			11
196	http://unknown.org/georeferenceRemarks				X desconhecido	10
197	organismScope	X	Classe Organism			10
198	member	X	Classe GeologicalContext			9
199	minimumDistanceAboveSurfaceInMeters	X	Classe Location			9
200	nomenclaturalStatus	X	Classe Taxon			9
201	associatedOrganisms	X	Classe Organism			6
202	http://unknown.org/occurrenceDetails				X desconhecido	6
203	lithostratigraphicTerms	X	Classe GeologicalContext			6
204	maximumDistanceAboveSurfaceInMeters	X	Classe Location			6
205	http://unknown.org/organismQuantityType				X desconhecido	5
206	lowestBiostratigraphicZone	X	Classe GeologicalContext			5

CLASSES OCORRÊNCIA E EVENTOS		PADRÃO DE METADADOS				Quantidade Datasets*
Elementos de metadados identificados		DwC	GBIF	Outro		
207	footprintSpatialFit	X	Classe Location			4
208	namePublishedInID	X	Classe Taxon			4
209	highestBiostratigraphicZone	X	Classe GeologicalContext			3
210	http://unknown.org/license				X desconhecido	3
211	http://unknown.org/organismQuantity				X desconhecido	3
212	http://unknown.org/rightsHolder				X desconhecido	3
213	http://unknown.org/sampleSizeUnit				X desconhecido	3
214	http://unknown.org/sampleSizeValue				X desconhecido	3
215	bed	X	Classe GeologicalContext			1
216	created				X desconhecido	1
217	http://unknown.org/class			X	Classe Dataset	1
218	http://unknown.org/habitatproperty				X desconhecido	1
219	http://unknown.org/nick				X desconhecido	1
220	http://unknown.org/parentEventID				X desconhecido	1
221	http://unknown.org/vegetationLayer				X desconhecido	1
		162		*44		****15

Fonte: própria autora. Fonte de dados: (GBIF, 2020c; GBIF, 2020d; TDWG, 2019b; TDWG, 2015a; GBIF, 2020f)

*Total de datasets das classes ocorrência e eventos que utilizam o elemento de metadados identificado.

**Total de elementos de metadados nativos do padrão Darwin Core.

***Total de elementos de metadados exclusivos do GMP.

****Total de elementos de metadados com padrão desconhecido.

Quadro 31. Identificação dos padrões relacionados aos elementos de metadados em uso para descrição dos dados de pesquisa da classe *checklist*.

Classe <i>checklist</i>		PADRÃO DE METADADOS			Quantidade
Elementos de metadados		DwC	GBIF		Datasets*
1	authorship		X	Classe NameUsage	31890
2	datasetKey		X	Classe Dataset	31890
3	issues		X	Classe NameUsage	31890
4	key		X	Classe VerbatimNameUsage	31890
5	lastCrawled		X	Classe Enum Constant	31890
6	lastInterpreted		X	Classe Enum Constant	31890
7	nameKey		X	Classe NameUsage	31890
8	nameType		X	Classe NameUsage	31890
9	nomenclaturalStatus	X	Classe Taxon		31890
10	numDescendants		X	Classe NameUsage	31890
11	origin		X	Classe NameUsage	31890
12	scientificName	X	Classe Taxon		31890
13	synonym		X	Classe NameUsage	31890
14	taxonomicStatus	X	Classe Taxon		31888
15	canonicalName		X	Classe NameUsage	31869
16	rank		X	Classe NameUsage	31723
17	kingdom	X	Classe Taxon		31639
18	kingdomKey		X	Classe Enum Constant	31639
19	taxonID	X	Classe Taxon		31120
20	parent		X	Classe NameUsage	30911
21	parentKey		X	Classe NameUsage	30911
22	references	X	Classe Record-level		30182
23	phylum	X	Classe Taxon		29930
24	phylumKey		X	Classe NameUsage	29930
25	class	X	Classe Taxon		29853
26	classKey		X	Classe NameUsage	29853
27	order	X	Classe Taxon		29828
28	orderKey		X	Classe Enum Constant	29828
29	nubKey		X	Classe NameUsage	29680
30	family	X	Classe Taxon		29641
31	familyKey		X	Classe Enum Constant	29641
32	genus	X	Classe Taxon		28562
33	genusKey		X	Classe Enum Constant	28562
34	species		X	Classe NameUsage	16128
35	speciesKey		X	Classe NameUsage	16128
36	publishedIn		X	Classe NameUsage	5869
37	accordingTo		X	Classe NameUsage	159

Classe <i>checklist</i>		PADRÃO DE METADADOS		Quantidade <i>Datasets</i> *
Elementos de metadados		DwC	GBIF	
38	remarks		X Classe NameUsage	68
39	accepted		X Classe NameUsage	52
40	acceptedKey		X Classe NameUsage	52
41	constituentKey		X Classe NameUsage	2
42	sourceTaxonKey		X Classe NameUsage	1
		** 11	*** 31	

Fonte: própria autora. Fonte de dados: (GBIF, 2020c; GBIF, 2020d; TDWG, 2019b; TDWG, 2015a; GBIF, 2020f)

*Total de datasets das classes *checklist* que utilizam o elemento de metadados identificado.

**Total de elementos de metadados nativos do padrão Darwin Core.

***Total de elementos de metadados exclusivos do GMP.

Na descrição dos *datasets*, os resultados da análise comparativa demonstram que há predominância de elementos nativos da EML, mas também são encontradas algumas referências herdadas do padrão ABCD, bem como referências próprias do perfil da GBIF (GMP). Destaca-se o elemento “type”, que define os tipos de *datasets*, é herdado das classes estabelecidas como núcleo dos arquivos do padrão DwC (DwC-A).

Referente à descrição dos dados, para as classes eventos e ocorrência há predominância dos termos que integram o padrão DwC e alguns elementos nativos dos vocabulários da GBIF. Já para a classe *checklist* há elementos herdados do DwC, mas predominam os elementos exclusivos do GMP, em geral, apresentam a função de normalização semântica. Destaca-se que o GMP apresenta chaves (Key) específicas para os registros das classes.

Também foram encontrados termos identificados pela API da GBIF como padrão desconhecido, entretanto alguns constam no vocabulário do GMP. Referente aos termos em redundância, identifica-se que alguns *datasets* não apontam corretamente o protocolo de origem (padrão e vocabulários de referência), resultando em inconsistências. Isso pode ser originário da descrição do dataset, bem como um problema resultante da integração dos esquemas de diferentes padrões ou do apontamento pela API da GBIF. Ainda assim, a soma dos apontamentos identificados como desconhecidos e as redundâncias de alguns termos não alcançam 10% dos elementos de metadados identificados e também ocorrem em poucos *datasets*.

Outra perspectiva relevante para identificar tendências refere-se à evolução das soluções adotadas pelas comunidades de pesquisa em resposta (ou relacionadas) aos

assuntos associados ao uso de metadados identificados na revisão dos artigos científicos – 4.2.2.4. Considerando a abrangência global das instituições que integram a GBIF, as ações adotadas por essa rede refletem as principais iniciativas relacionada ao uso de metadados adotadas pelas comunidades de pesquisa na última década, bem como aquelas já planejadas para os próximos anos.

Nessa última análise de tendência, portanto, são comparados os assuntos identificados nos artigos científicos com as iniciativas adotadas pelas comunidades de pesquisa identificadas com base nas ações da GBIF. Essa comparação permite compreender como ocorreram os desdobramentos dos assuntos associados ao uso de metadados diante das ações adotadas pelas comunidades de pesquisa relacionadas à temática biodiversidade – Quadro 32.

Quadro 32. Comparação dos assuntos associados ao uso de metadados identificados nos artigos científicos e as iniciativas globais das comunidades de pesquisa

Assunto associado ao uso de metadados Artigos científicos	Evidências identificadas Rede GBIF
(i) Proposta de padrão de metadados ou extensão de padrão	<ul style="list-style-type: none"> • Recomendação de uso dos padrões EML, ABDC, Darwin Core • Disponibilização e constante atualização da documentação de extensões ao Darwin Core • Perfil de metadado embasado predominantemente em EML e DwC, mas com referências herdadas do ABCD
(ii) Desafios à gestão dos dados da biodiversidade	<ul style="list-style-type: none"> • Realização de fóruns e eventos voltados à mobilização e capacitação dos diversos seguimentos envolvidos com as pesquisas na temática biodiversidade. • Plano de adoção dos princípios FAIR (iniciado em 2017) • Incentivo à publicação de <i>data papers</i> • Possibilidade de registrar DOI para os <i>datasets</i> depositados • Possibilidade de relacionar estudos, <i>data papers</i> e <i>datasets</i> de uma pesquisa utilizando campos para registro do DOI • Disponibilização de métricas relacionadas à citação de <i>datasets</i> e <i>data papers</i> • Validação de <i>datasets</i> de referência com foco na normalização semântica • Estabelecimentos de nós (núcleos) representativos da GBIF em diversas regiões para estimular a união de esforços e o

Assunto associado ao uso de metadados Artigos científicos	Evidências identificadas Rede GBIF
	<p>estabelecimento de projetos voltados ao aperfeiçoamento da gestão de dados de pesquisa</p> <ul style="list-style-type: none"> • Acesso aos estudos que citam o uso dos <i>datasets</i> • Debate amplo e participativo das iniciativa e soluções • Definição de frente de trabalho estratégicas com comitês compostos a partir da candidatura de pesquisadores • Premiação anual de trabalhos científicos
(iii) revisão comparativa dos padrões de metadados existentes	<ul style="list-style-type: none"> • Constante atualização do Perfil de metadado (GMP), dos vocabulários e extensões aos metadados • Integração de esquemas com outras redes de dados, normalização de vocabulários
(iv) indicadores para biodiversidade	<ul style="list-style-type: none"> • Integração com o <i>Group on Earth Observations - Biodiversity Observation Network</i> (GEO BON) voltado à descoberta de <i>datasets</i> de interesse para desenvolvimento das <i>Essential Biodiversity Variables</i> (EBVs) e outros indicadores. • Disponibilização de ferramentas de análise de cenários e modelagem de dados
(v) catálogos e ontologias	<ul style="list-style-type: none"> • Utilização de diversos elementos de metadados voltados à qualificação semântica dos modelos, em especial da classe UseWithIRI do padrão DwC, que estabelece <i>Internationalized Resource Identifiers – IRIs</i> • Validação de <i>datasets</i> de nomenclatura e vocabulário taxonômico, em geral, da classe <i>checklist</i>, indicando os termos a serem adotados, definidos a partir da comparação de vários catálogos
(vi) digitalização de coleções de história natural e anotações de campo	<ul style="list-style-type: none"> • Suporte aos estudos voltados a definição de parâmetros para o planejamento da digitalização de coleções (e.g. evento, treinamento, recomendações, financiamento) • Uso da classe de <i>datasets</i> chamada recurso de metadados, que permite dar publicidade a coleções ou outros recursos informacionais não disponibilizados na rede • Suporte às grandes inciativas voltadas à digitalização de coleções • Incentivo ao cadastro de recursos informacionais, mesmo que com categorias básicas de informações • Identificação da existência de diversas coleções a serem digitalizadas

Assunto associado ao uso de metadados Artigos científicos	Evidências identificadas Rede GBIF
	<ul style="list-style-type: none"> • Suporte ao cadastro de coleções científicas • Incentivo a sugerir <i>datasets</i>
(vii) procedimentos de gestão de dados	<ul style="list-style-type: none"> • Elaboração e disponibilização de guias • Disponibilização de soluções para validação de <i>datasets</i> • Documentação do GMP e vocabulários em constante atualização e disponibilizada no portal da GBIF • Ampla divulgação de recomendações • Definição de atributos (elementos ou campos) de metadados obrigatórios para efetivar o depósito na rede • Disponibilização de modelos de documentação dos dados e <i>datasets</i> de referência para cada classe • O processamento dos <i>datasets</i> inclui procedimentos de normalização, interpretação e persistência dos dados • Programas e projetos em todos os continentes voltados ao aperfeiçoamento da gestão dos dados de pesquisa
(viii) soluções computacionais	<ul style="list-style-type: none"> • Desenvolvimento e documentação do portal e APIs baseados em EML • Disponibilização de API temáticas (e.g. <i>datasets</i> taxonômicos, de ocorrência, constantes enumerativas e outros) • Ferramenta de apoio à descrição, validação e registro dos <i>datasets</i> – <i>Integrated Publishing Toolkit</i> (IPT) • Integração a outras redes (<i>Endpoints</i>) de dados • Disponibilização de catálogo de ferramentas

Fonte: própria autora.

As tendências apontam para a ampliação do uso da EML na descrição dos *datasets* e do uso de DwC na descrição dos dados de pesquisa registrados nos *datasets*. O GMP apresenta um relevante papel na integração dos dados, com elementos voltados a redução de inconsistências e identificação precisa de registros.

Também há um crescimento dos esforços para padronização dos vocabulários e integração dos dados, com um aumento da cultura de descrição dos dados nas comunidades de pesquisa. Por fim, há um amadurecimento das soluções computacionais em busca da interoperabilidade.

5 CONCLUSÕES

A crescente demanda global por informações sobre a biodiversidade se reflete no direcionamento de esforços para o aperfeiçoamento da gestão dos dados de pesquisa nessa temática. “Os dados da biodiversidade são diferenciados por seu caráter interdisciplinar, pela variedade extremamente ampla de tipos e estruturas de dados e pela variedade de conceitos semânticos que eles englobam” (MINADAKIS, MARKETAKIS, *et al.*, 2016).

Há claramente um consenso global sobre a necessidade comum de uso integrado dos dados de biodiversidade como insumo fundamental ao avanço das pesquisas, à qualificação das decisões e à formulação de políticas relacionadas ao monitoramento e manejo da biodiversidade, bem como a questões ainda mais complexas, como mudanças climáticas e sustentabilidade ambiental. Esse consenso se reflete no surgimento de diversas iniciativas estabelecidas globalmente para aperfeiçoar o acesso aos dados de pesquisa da biodiversidade. Tais iniciativas têm impulsionado o uso de padrões de metadados para a descrição dos dados desde a coleta até a disseminação.

O uso de metadados apresenta-se nas pesquisas associado principalmente ao desenvolvimento de soluções computacionais interoperáveis e voltadas à integração dos *datasets* provenientes de diferentes fontes em diversas partes do mundo. Nesse sentido, a GBIF destaca-se por ser uma rede de abrangência global tanto pelas entidades e países membros, como pela distribuição espacial dos dados de pesquisa depositados e da localização das 2.025 instituições de pesquisa registradas e habilitadas para o depósito de *datasets*. Dessas instituições, há 1.570 vinculadas à publicação dos 53.729 *datasets* analisados neste estudo.

Considerando os *datasets* disponíveis na GBIF, a Europa destaca-se como continente responsável pelo maior número de depósitos de *datasets* (87%) e também maior quantidade de instituições de pesquisa registradas (41%). A América do Sul ocupa a segunda posição com 20% das instituições de pesquisa registradas e 6% dos *datasets* depositados na GBIF. A Colômbia (em 5ª posição) e o Brasil (em 9ª posição) estão entre os 10 países do mundo com maior número de *datasets* e também com mais instituições publicadoras (responsáveis pelos *datasets*). Dessa forma, também representam a liderança em pesquisa sobre biodiversidade na América do Sul.

A análise dos *datasets* coletados na GBIF permitiu identificar 51 elementos de metadados para a descrição dos *datasets* depositados na GBIF. Há quatro classes de *datasets* registrados na rede GBIF: metadados de recurso (em geral apenas apontam a existência de coleções); ocorrência; eventos e *checklist*. Em um segundo nível, analisando a descrição dos dados de pesquisa registrados nesses *datasets*, foram identificados 221 elementos (classe evento e classe ocorrência) e 42 elementos (classe *checklist*).

Nas classes evento e ocorrência, 33 elementos (15% dos total de 221) estão presentes em mais de 90% dos *datasets* e 70% dos *datasets* apresentam em média 55 elementos para a descrição dos dados. Na classe *checklist*, 31 elementos (do total de 42) estão presentes em mais de 90% dos *datasets* e 50% dos *datasets* apresentam em média 36 elementos para a descrição dos dados.

Verificou-se que a descrição dos *datasets* coletados apresenta atributos que vão além dos obrigatórios e recomendados pela GBIF. Nas classes evento e ocorrência há maior uso do Dwc. Já para a classe *checklist* predominam elementos nativos do perfil de metadados da GBIF (GMP), principalmente voltados à normalização semântica. Além disso, os resultados apontam a prevalência do uso da EML na descrição dos *datasets*. Também há elementos nativos do GMP e alguns herdados do padrão ABCD, com destaque aos elementos relacionados ao registro do *Digital Object Identifier* (DOI) para *datasets*.

Há diversos guias, vocabulários controlados e arquivos modelos disponibilizados pela GBIF para orientar o preenchimento dos atributos de metadados. Além dessa documentação, a disponibilização da ferramenta *Integrated Publishing Toolkit* (IPT) para apoiar a descrição e publicação dos *datasets* impactou positivamente a comunidade de pesquisa em biodiversidade. Embasado no GMP, o IPT aprimorou a publicação de *datasets* permitindo alcançar periódicos tradicionais, como já acontece com artigos e *data papers*. (ROBERTSON, DÖRING, *et al.*, 2014). Isso ampliou o reconhecimento dos profissionais e instituições responsáveis pelos *datasets* e gerou forte incentivo ao aumento e qualificação da publicação desses recursos.

Dessa forma, verifica-se que o incentivo ao compartilhamento de dados de pesquisa não se sustenta apenas com políticas de obrigatoriedade de depósito para a publicação de artigos em periódicos científicos. O expressivo volume de *datasets* registrados na GBIF resulta de uma conjugação de ações e esforços direcionados a melhoria da gestão dos dados

de biodiversidade. Isso demonstra a necessidade de disponibilização de infraestrutura adequada para o depósito dos *datasets* associada a ferramentas, guias e treinamentos voltados a orientar a descrição dos dados.

Cabe destacar também que na última década houve um esforço expressivo empregado na identificação e digitalização de coleções de história natural. Isso segue como tendência nos próximos anos, pois ainda há um grande volume de recursos identificados a serem digitalizados. Tanto a publicação de dados de pesquisa, como a digitalização de coleções demandam investimentos financeiros para a desenvolvimento e manutenção da infraestrutura de depósito e de ferramentas adequadas que auxiliem o processamento dos dados. Também é relevante a disponibilização de guias e documentação de orientação, treinamento das comunidades de pesquisa e mecanismos de reconhecimento como incentivo ao depósito.

As mudanças nas atitudes e união de esforços da comunidade de pesquisa associadas ao aumento da facilidade no compartilhamento de dados contribuíram para o rápido crescimento de compilações abertas de dados estruturados de biodiversidade (HUDSON, NEWBOLD, *et al.*, 2017). A maior disponibilidade de *datasets* registrados na GBIF oferece oportunidades significativas para estudos em larga escala sobre a biodiversidade. As informações sobre distribuição de espécies são reconhecidas como um insumo crucial para a conservação e manejo da biodiversidade. Por isso, recursos consideráveis foram dedicados ao aumento da quantidade e disponibilidade de dados de ocorrência de espécies, aumentando o potencial de uso na modelagem de distribuição de espécies.

De forma geral, a tendência para os próximos anos aponta para o aumento do espaço dedicado à publicação de *datasets* em periódicos científicos, atribuindo o mesmo valor de reconhecimento que aos artigos científicos tradicionais e *data papers*. Entretanto, há também o desafio de ampliar a infraestrutura e o suporte necessário aos depósitos de *datasets*. Nesse sentido, em estudo futuros, é relevante aprofundar a compreensão do uso de metadados e incorporação dos padrões de metadados na temática biodiversidade também sob a ótica aspectos sociais, do movimento de acesso livre e da adequação aos princípios FAIR.

Por fim, outro desafio relevante para a próxima década refere-se a incorporar atributos de descrição de modelos de predição e indicadores, que permitam aperfeiçoar

análises e decisões (VELÁSQUEZ-TIBATÁ, OLAYA-RODRÍGUEZ, *et al.*, 2019). Nesse sentido, as tendências apontam para o crescimento das pesquisas voltadas à qualificação semântica associadas ao desenvolvimento soluções que permitam análises com modelagem dos dados.

6 REFERÊNCIAS BIBLIOGRÁFICAS

AGUILAR, R. et al. A flexible online metadata editing and management system. **Ecological Informatics**, v. 5, p. 26-31, 2010. Disponível em: <DOI:10.1016/j.ecoinf.2009.08.012>.

ALMEIDA, M. B. Níveis de representação para interoperabilidade entre dados médicos. In: CAMPOS, M. L. D. A., et al. **Produção, tratamento, disseminação e uso recursos informacionais heterogêneos [recurso eletrônico]: diálogos interdisciplinares**. Niterói: IACS/UFF, v. 5, 2018. Cap. 7, p. 101-109. ISBN 978-85-86918-08-7.

ALVES, J. C.; CAFÉ, L. M. A. Análise focada em metadados sob a luz do padrão MTD-BR, Porto Alegre, v. 16, p. 179-202, jul./dez. 2010. Disponível em: <<https://seer.ufrgs.br/EmQuestao/article/download/12930/10431>>. Acesso em: 14 ago. 2019.

ALVES, M. D. D. R.; SOUZA, M. I. F. Estudo de correspondência de elementos metadados: Dublin Core e MARC 21. **Revista Digital de Biblioteconomia e Ciência da Informação**, Campinas, v. 4, n. 2, p. 20-38, jan./jun. 2007. ISSN 1678-765X. Disponível em: <<http://www.brapci.inf.br/index.php/article/download/8610>>. Acesso em: 04 mai 2019.

ALVES, R. C. V. **Metadados como elementos do processo de catalogação - Tese (Doutorado em Ciência da Informação)**. Universidade Estadual Paulista. Marília, p. 132. 2010.

APPOLINÁRIO, F. **Metodologia da ciência: filosofia e prática da pesquisa**. São Paulo: Cengage Learning, v. xii, 2006. 209 p.

ARAÚJO, C. A. Á. A ciência da informação como ciência social. **Ci. Inf.**, Brasília, v. 32, n. 3, p. 21-27, set./dez. 2003.

ARAÚJO, C. A. Á. **Ciência da Informação: origem e evolução**. In: Arquivologia, biblioteconomia, museologia e ciência da informação: o diálogo possível. Brasília: Brinquet de Lemos, 2014. 99-151 p.

AUER, S.; BRYL, V.; TRAMP, S. (Eds.). **Linked Open Data – Creating Knowledge Out of Interlinked Data - Results of the LOD2 Project**. [S.l.]: Springer Open, 2014. ISBN 978-3-319-09846-3/1611-3349.

BACHMAN, S. et al. Supporting Red List threat assessments with GeoCAT: geospatial conservation assessment tool. **ZooKeys**, v. 150, p. 117-126, 2011. Disponível em: <DOI: 10.3897/zookeys.150.2109>.

BAIRD, R. Leveraging the fullest potential of scientific collections through digitization. **Biodiversity Informatics**, v. 7, p. 130-136, 2010.

BAKER, E.; JOHNSON, K. G.; YOUNG, J. R. The future of the past in the present: biodiversity informatics and geological time, p. 397-405, 2011. Disponível em: <DOI: 10.3897/zookeys.150.2350>.

- BARBOSA, E. C. Produção e Recepção de Informações: Teorias e Métodos. **e-Disciplinas: Sistema de Apoio às Disciplinas USP**, 2018. Disponível em: <<https://edisciplinas.usp.br/course/view.php?id=6209>>. Acesso em: 06 março 2019.
- BARRETO, A. D. A. Mudança estrutural no fluxo do conhecimento: a comunicação eletrônica. **Ciência da Informação**, Brasília, v. 27, n. 2, p. 122-127, maio/ago 1998.
- BARROSA, E. G. et al. A digital library environment for integrating, disseminating and exploring ecological data. **Ecological Informatics**, v. 3, p. 295-308, 2008. Disponível em: <DOI: 10.1016/j.ecoinf.2008.09.006>.
- BATES, M. J. The design of browsing and berrypicking techniques for the online search interface". **Online Review**, v. 13, n. 5, p. 407-424, 1989. ISSN <https://doi.org/10.1108/eb024320>.
- BCH. The Biosafety Clearing-House. **Biosafety Clearing-House (BCH)**, 2016. Disponível em: <<https://bch.cbd.int/about/>>. Acesso em: 15 jun 2020.
- BELL, G. Prefácio. In: HEY, T., et al. **O quarto paradigma: descobertas científicas na era da e-science**. São Paulo: Oficina de Textos, 2011. p. 11-15.
- BENDIX, J.; NIESCHULZE, J.; MICHENER, W. K. Data platforms in integrative biodiversity research. **Ecological Informatics**, v. 11, p. 1-4, 2012. Disponível em: <DOI:10.1016/j.ecoinf.2012.04.001>.
- BERENDSOHN, W. G.; CHAVAN, V.; MACKLIN, J. A. Recommendations of the GBIF task group on the global strategy and action plan for the mobilization of natural history collections data. **Biodiversity Informatics**, 7, 2010, pp.67 – 71, v. 67-71, n. 7, 2010.
- BERENDSOHN, W. G.; SELTMANN, P. Using geographical and taxonomic metadata to set priorities in specimen digitization. **Biodiversity Informatics**, v. 7, p. 120-129, 2010.
- BERNERS-LEE, T. Axioms of Web Architecture: Metadata. **World Wide Web Consortium - W3C**, 06 jan. 1997. Disponível em: <<https://www.w3.org/DesignIssues/Metadata.html>>. Acesso em: 05 mai. 2019.
- BERTZKY, M.; STOLL-KLEEMANN, S. Multi-level discrepancies with sharing data on protected areas: What we have and what we need for the global village. **Journal of Environmental Management**, v. 90, p. 8-24, 2009. Disponível em: <DOI: 10.1016/j.jenvman.2007.11.001>.
- BIOCASE. Home - News - New Data Usage Statistics Online. **Biological Collection Access Service (BioCASE)**, 2016. Disponível em: <<https://www.biocase.org/index.shtml>>. Acesso em: 18 dez 2019.
- BIOCASE. The BioCASe Protocol. **Biological Collection Access Service (BioCASE)**, 2019. Disponível em: <<https://www.biocase.org/products/protocols>>. Acesso em: 18 dez. 2019.
- BLAGODEROV, V. et al. No specimen left behind: industrial scale digitization of natural history collections. **ZooKeys**, v. 209, p. 133-146, 2012. Disponível em: <DOI: 10.3897/zookeys.209.3178>.
- BONET, F. J. et al. Documenting, storing, and executing models in Ecology: A conceptual framework and real implementation in a global change monitoring program. **International Journal of Geographical Information Science**, v. 28, n. 11, p. 2356-2371, 2014. Disponível em: <DOI: 10.1080/13658816.2014.924627>.

- BORKO, H. Information science: what is it? **American Documentation**, 19, n. 1, 1968.
- BOSCH, A. V. D. et al. Weaving a New Fabric of Natural History. **INTERDISCIPLINARY SCIENCE REVIEWS**, v. 34, n. 2-3, p. 206-223, 2009. Disponível em: <DOI: 10.1179/174327909X441117>.
- BOUGHIDA, K. B. CDWA Lite for Cataloging Cultural Objects (CCO): A New XML Schema for the Cultural Heritage Community. In: BOONSTRA, O.; BEURNE, L.; DOORN, P. **Humanities, Computers and Cultural Heritage: Proceedings of the XVI International Conference of the Association for**. Amsterdam: Royal Netherlands Academy of Arts and Sciences, 2005. Cap. 9, p. 49-54. ISBN 90-6984-456-7. Disponível em: <<https://repository.ubn.ru.nl/bitstream/handle/2066/56711/56711.pdf?sequence.>>. Acesso em: 27 abr. 2019.
- BRASCHER, M.; CAFÉ, L. **Organização da Informação ou Organização do Conhecimento?** ANCIB - Encontro Nacional de Pesquisa em Ciência da Informação. IX ENCONTRO ENANCIB. GT 2 - Organização e Representação do Conhecimento. São Paulo: [s.n.]. 2008. p. 1-14.
- BRASIL. MCTIC. PPBIO. **Conhecendo a Biodiversidade**. Ministério da Ciência, Tecnologia, Inovações e Comunicações (MCTIC); Programa de Pesquisa em Biodiversidade (PPBio). Brasília, p. 196. 2016. (ISBN 978-85-63100-08-5).
- BRASIL. MCTIC.CAPES. Portal de Periódicos da Capes - Institucional. **Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES)**, 2019. Disponível em: <<https://www.periodicos.capes.gov.br>>. Acesso em: 21 set. 2019.
- BURKHARDT, U. et al. The Edaphobase project of GBIF-Germany—A new onlinesoil-zoological data warehouse. **Applied Soil Ecology**, v. 83, p. 3-2, 2014. Disponível em: <DOI: 10.1016/j.apsoil.2014.03.021>.
- BUSH, V. As we may may think. **Atlantic Monthly**, v. 176, n.1, p. 101-108, 1995. Disponível em: <<http://web.mit.edu/STS.035/www/PDFs/think.pdf>>. Acesso em: 30 maio 2018.
- CAFÉ, L. M. A.; SALES, R. D. Organização da informação: Conceitos básicos e breve fundamentação teórica. In: ROBREDO, J.; BRÄSCHER, M. **Passeios pelo bosque da informação: estudos sobre representação e organização da informação e do conhecimento - EROIC**. Brasília: IBICT, 2010. Cap. 6. 329 p, p. 115-129. ISBN 978-85-7013-072-3.
- CALAME, P.; TALAMANT, A. **A Questão do Estado no Coração do Futuro**. 1.ed. ed. Rio de Janeiro: Vozes, 2001. 168 p.
- CALARESU, M.; SHIRI, A. Understanding semantic web: A conceptual model. **Library Review**, v. 64, p. 82-100, 2015.
- CALDERA-SERRANO, J.; FREIRE-ANDINO, R. O. Los metadatos asociados a la información audiovisual televisiva por “agentes externos” al servicio de documentación: validez, uso y posibilidades. **Biblios**, n. 62-75, 2016. ISSN 1562-4730. Disponível em: <<http://biblios.pitt.edu/ojs/index.php/biblios/article/view/285>>. Acesso em: 04 abr 2019.
- CAMPOS, L. F. B. Metadados digitais: revisão bibliográfica da evolução e tendências por meio de categorias funcionais. **Encontros Bibli: Revista Eletrônica de Biblioteconomia e Ciência da Informação**, 12, n. 23, 2007. 16-46. Disponível em: <<http://www.brapci.inf.br/v/a/4431>>. Acesso em: 11 jun. 2018.

CAMPOS, M. L. M.; CAMPOS, M. L. D. A.; CAMPOS, L. M. Web semântica e a gestão de conteúdos informacionais. In: SAYÃO, L. **Bibliotecas digitais: saberes e práticas**. Salvador, Brasília: EDUFBA, IBICT, 2005. p. 55-75. ISBN 8523203508. Disponível em: <<http://livroaberto.ibict.br/handle/1/1013>>.

CAPURRO, R. Epistemologia e Ciência da Informação. **Apresentação realizada no V Encontro Nacional de Pesquisa em Ciência da Informação**, Belo Horizonte, 2003. Disponível em: <http://www.capurro.de/enancib_p.htm>. Acesso em: 28 maio 2018.

CAVALCANTE, D. G. S. M.; KOBASHI, N. Y. Integração de acervos multimídia no contexto da web semântica: análise de instrumentos e metodologias para representação e recuperação. In: CAMPOS, M. L. D. A., et al. **Produção, tratamento, disseminação e uso recursos informacionais heterogêneos [recurso eletrônico]: diálogos interdisciplinares**. Niterói: IACS/UFF, v. 5, 2018. Cap. 16, p. 142-145. ISBN 978-85-86918-08-7.

CDB. Text of the Convention. **Convention on Biological Diversity (CBD)**, 1992. Disponível em: <<https://www.cbd.int/convention/text/>>. Acesso em: 27 abr. 2019.

CEOS. About CEOS. **Committee on Earth Observation Satellites (CEOS)**, 2020. Disponível em: <<http://ceos.org/about-ceos>>. Acesso em: 05 fev. 2020.

CHANDLER, M. et al. Contribution of citizen science towards international biodiversity monitoring. **Biological Conservation**, v. 213, p. 280-294, 2017. Disponível em: <DOI: 10.1016/j.biocon.2016.09.004>.

CHAVAN, V.; PENEV, L.; HOBERN, D. Cultural Change in Data Publishing Is Essential. **BioScience**, v. 63, n. 6, p. 419-420, 2013. ISSN <https://doi.org/10.1525/bio.2013.63.6.3>.

CHIGNARD, S. A brief history of Open Data. **ParisTech Review**, Paris, 2013. Disponível em: <<http://www.paristechreview.com/2013/03/29/brief-history-open-data/>>.

CHULIN MENG. Biological Information Standards. **Bulletin of the American Society for Information Science and Technology**, p. 15-17, 2004.

CODATA; PFEIFFENBERGER, H.; UHLIR, P. **Twenty-Year Review of GBIF**. Committee on Data of the International Science Council (CODATA). Paris, p. 15. 2020. (DOI: 10.35035/ctzm-hz97).

COSTA, M. M.; CUNHA, M. B. D. O bibliotecário no tratamento de dados oriundos da e-science: considerações iniciais. **Perspectivas em Ciência da Informação**, v. 19, n. 3, p. 189-206, set. 2014. ISSN 19815344. Disponível em: <<http://portaldeperiodicos.eci.ufmg.br/index.php/pci/article/view/1900/1445>>. Acesso em: 13 nov 2018.

COSTA, S. M. D. S.; LEITE, F. C. L.; TAVARES, R. B. **Comunicação da informação, gestão da informação e gestão do conhecimento**. 1ª. ed. Brasília: Ibict, 2018. 415 p. ISBN 978-85-7013-147-8. Disponível em: <<http://livroaberto.ibict.br/handle/123456789/1071>>. Acesso em: 2019 jan. 11.

COSTA, S. M. S. O novo papel das tecnologias digitais na comunicação científica. In: SAYÃO, L. **Bibliotecas Digitais**. Salvador. Brasília: UFBA, IBICT, v. 1, 2005. p. 165-183. Disponível em: <<http://repositorio.unb.br/handle/10482/1437>>. Acesso em: 11 março 2019.

COSTELLO, M. J.; WIECZOREK, J. Best practice for biodiversity data management and publication. **Biological Conservation**, v. 173, p. 68-73, 2014. ISSN <https://doi.org/10.1016/j.biocon.2013.10.018>.

CRAIN, I. K. **The Biodiversity Information Clearing House: Concept and Challenges**. [S.l.]: World Conservation Monitoring Centre, World Conservation Press, 1994. 34 p.

CRESWELL, J. W. **PROJETO DE PESQUISA: MÉTODOS QUALITATIVO, QUANTITATIVO E MISTO; TRADUÇÃO**. Tradução de Magda Lopes. 3ª. ed. Porto Alegre: ARTMED, 2010. 296 p.

DAVIS, M. et al. Facilitating Access to Biodiversity Information: A Survey of Users' Needs and Practices. **Environmental Management**, v. 53, p. 690-701, 2014. Disponível em: <DOI 10.1007/s00267-014-0229-7>.

DCC. Data discovery and reuse - Data catalogues. **Digital Curation Center - DCC**, 2019. Disponível em: <<http://www.dcc.ac.uk/tailored-support/data-discovery>>. Acesso em: 03 maio 2019.

DCC. ABCD - Access to Biological Collection Data. **Digital Curation Center (DCC)**, 2019a. Disponível em: <<http://www.dcc.ac.uk/resources/metadata-standards/abcd-access-biological-collection-data>>. Acesso em: 16 out 2019.

DCC. Darwin Core. **Digital Curation Center (DCC)**, 2019b. Disponível em: <<http://www.dcc.ac.uk/resources/metadata-standards/darwin-core>>. Acesso em: 5 nov 2019.

DCMI. Dublin Core Metadata Element Set - Version 1.1: Reference Description. **Dublin Core Metadata Initiative (DCMI)**, 2012. Disponível em: <<https://dublincore.org/specifications/dublin-core/dces/>>. Acesso em: 06 jul 2020.

DEMPSEY, L.; HEERY, R. **Specification for resource description methods. Part 1. A review of metadata: a survey of current resource description formats**. Telematics - Fourth Framework Programme - European Union. Bath, p. 92. 1997.

DENG, S. Optimizing Workflow through Metadata Repurposing and Batch Processing. **Journal of Library Metadata**, v. 10, n. 4, p. 219-237, 2010. ISSN ISSN: 1938-6389 DOI: 10.1080/19386389.2010.524862.

DENGLER, J. et al. The Global Index of Vegetation-Plot Databases (GIVD): a new resource for vegetation science. **Journal of Vegetation Science**, v. 22, p. 582-597, 2011. Disponível em: <DOI: 10.1111/j.1654-1103.2011.01265.x>.

DEVARAKONDA, R. et al. Mercury: reusable metadata management, data discovery and access system. **Earth Sci Inform**, v. 3, p. 87-94, 2010. Disponível em: <DOI: 10.1007/s12145-010-0050-7>.

DEVARAKONDA, R. et al. Data sharing and retrieval using OAI-PMH. **Earth Sci Inform**, v. 4, p. 1-5, 2011. Disponível em: <DOI: 10.1007/s12145-010-0073-0>.

DZIEKANIAK, G. V. Mapeamento do uso de metadados por comunidades científicas. **BI-BLOS - Revista do Instituto de Ciências Humanas e da Informação**, Rio Grande, v. 20, n. 1, p. 229-243, 2007. Disponível em: <<https://periodicos.furg.br/biblos/article/view/732/225>>. Acesso em: 04 mai 2019.

ECOINFORMATICS. Tools - Metadata Specification. **Ecoinformatics**, 2019. Disponível em: <<http://ecoinformatics.org/tools.html>>. Acesso em: 10 dez. 2019.

ELLISON, A. M. et al. Analytic Webs Support the Synthesis of Ecological Data Sets. **Ecology**, v. 87, n. 6, p. 1345-1358, 2006.

EUA. FGDC. Federal Geographic Data Committee - HISTORY. **Federal Geographic Data Committee (FGDC)**, 2020. Disponível em: <<https://www.fgdc.gov/who-we-are/history#7>>. Acesso em: 05 jun. 2020.

EUA. FGDC. USGS. **Content Standard for Digital Geospatial Metadata - Part 1: Biological Data Profile**. Federal Geographic Data Committee and (FGDC). United States Geological Survey (USGS). Washington, D.C., p. 54. 1999.

EUA. NATIONAL RESEARCH COUNCIL. **A Biological Survey for the Nation**. 2ª. ed. Washington, DC: The National Academies Press, 1994. 224 p. ISBN 978-0-309-04984-9. Disponível em: <<https://doi.org/10.17226/2243>>.

EUA. OMB. FGDC. **M-11-03 Issuance of OMB Circular A-16 Supplemental Guidance**. Office of Management and Budget (OMB). [S.l.], p. 34. 2010.

EUA.NSF. Cyberinfrastructure Vision for 21st Century Discovery. **National Science Foundation (NSF)**, Arlington, 2007. Disponível em: <<https://www.nsf.gov/pubs/2007/nsf0728/>>. Acesso em: 11 abr. 2018.

EUROPEAN COMMISSION. G20. G20 Leaders' Communique Hangzhou Summit - STATEMENT/16/2967. **European Commission website**, 2016. Disponível em: <https://ec.europa.eu/commission/presscorner/detail/en/STATEMENT_16_2967>. Acesso em: 16 jun. 2020.

FILHO, D. B. F. et al. O que é, para que serve e como se faz uma meta-análise? **Revista Teoria & Pesquisa**, v. 23, n. 2, p. 205-228, 2014. Disponível em: <<http://dx.doi.org/10.4322/tp.2014.018>>.

FINK, E. E. American Art Collaborative (AAC) Linked Open Data (LOD) Initiative -Overview and Recommendations for Good Practices. **The American Art Collaborative (AAC)**, p. 80, 2018. Disponível em: <http://americanartcollaborative.org/wp-content/uploads/2018/03/AAC_LOD_Overview_Recommendations.pdf>. Acesso em: 06 dez 2018.

GARVEY, W. D. **Communication: the essence of science - facilitating information among Librarians, Scientists, Engineers and Students**. 1ª. ed. Oxford: Pegamon, 1979. 344 p. ISBN 9781483182070.

GBIF. Darwin Core Extension - Darwin Core Taxon. **Global Biodiversity Information Facility (GBIF)**, 2015a. Disponível em: <http://rs.gbif.org/core/dwc_taxon_2015-04-24.xml>. Acesso em: 20 maio 2020.

GBIF. Darwin Core Extension - Darwin Core Occurrence. **Global Biodiversity Information Facility (GBIF)**, 2015b. Disponível em: <http://rs.gbif.org/core/dwc_occurrence_2015-07-02.xml>. Acesso em: 10 maio 2020.

GBIF. Darwin Core Extension - Darwin Core Event. **Global Biodiversity Information Facility (GBIF)**, 2015c. Disponível em: <http://rs.gbif.org/core/dwc_event_2016_06_21.xml>. Acesso em: 15 maio 2020.

GBIF. GBIF.org implements licensing changes, updates terms of service. **Global Biodiversity Information Facility (GBIF)**, 2016. Disponível em: <<https://www.gbif.org/pt/news/82868/gbiforg-implements-licensing-changes-updates-terms-of-service>>. Acesso em: 04 mar 2019.

GBIF. GBIF Metadata Profile – How-to Guide. **Global Biodiversity Information Facility (GBIF)**, 2017a. Disponível em: <<https://github.com/gbif/ipt/wiki/GMPHowToGuide>>. Acesso em: 19 dez. 2019.

GBIF. Darwin Core Archives – How-to Guide. **Global Biodiversity Information Facility (GBIF)**, 2017b. Disponível em: <<https://github.com/gbif/ipt/wiki/DwCAHowToGuide>>. Acesso em: 10 mar. 2020.

GBIF. Darwin Core Archives – How-to Guide. **Global Biodiversity Information Facility (GBIF)**, 2018a. Disponível em: <<https://github.com/gbif/ipt/wiki/DwCAHowToGuide>>. Acesso em: 18 maio 2020.

GBIF. How to publish biodiversity data through GBIF.org. **Global Biodiversity Information Facility (GBIF)**, 2018b. Disponível em: <<https://github.com/gbif/ipt/wiki/howToPublish#instructions>>. Acesso em: 2019 dez 03.

GBIF. Best Practices in Publishing Species Checklists. **Global Biodiversity Information Facility (GBIF)**, 2018c. Disponível em: <<https://github.com/gbif/ipt/wiki/BestPracticesChecklists>>. Acesso em: 07 fev. 2020.

GBIF. The IPT User Manual. **Global Biodiversity Information Facility (GBIF)**, 22 jul 2019a. Disponível em: <<https://github.com/gbif/ipt/wiki/IPT2ManualNotes.wiki>>. Acesso em: 19 dez. 2019.

GBIF. The GBIF Network. **Global Biodiversity Information Facility (GBIF)**, 2019b. Disponível em: <<https://www.gbif.org/the-gbif-network>>. Acesso em: 02 dez 2019.

GBIF. Data standards. **Global Biodiversity Information Facility (GBIF)**, 2019c. Disponível em: <<https://www.gbif.org/pt/standards>>. Acesso em: 09 dez. 2019.

GBIF. Dataset classes. **Global Biodiversity Information Facility (GBIF)**, 2019d. Disponível em: <<https://www.gbif.org/en/dataset-classes>>. Acesso em: 10 dez. 2019.

GBIF. What is Darwin Core, and why does it matter? **Global Biodiversity Information Facility (GBIF)**, 2019e. Disponível em: <<https://www.gbif.org/en/darwin-core>>. Acesso em: 17 dez. 2019.

GBIF. Data papers - Getting scholarly recognition for your datasets. **Global Biodiversity Information Facility (GBIF)**, 2019f. Disponível em: <<https://www.gbif.org/en/data-papers>>. Acesso em: 18 dez. 2019.

GBIF. What is GBIF? **Global Biodiversity Information Facility (GBIF)**, 2019g. Disponível em: <<https://www.gbif.org/what-is-gbif>>. Acesso em: 30 nov. 2019.

GBIF. GBIF API model registry. **Global Biodiversity Information Facility (GBIF)**, 2020a. Disponível em: <<http://gbif.github.io/parsers/apidocs/>>. Acesso em: 10 jun 2020.

GBIF. Data quality requirements: Occurrence datasets. **Global Biodiversity Information Facility (GBIF)**, 14 mai. 2020b. Disponível em: <<https://www.gbif.org/data-quality-requirements-occurrences>>. Acesso em: 20 jun. 2020.

GBIF. GBIF Common : API 0.116-SNAPSHOT API. **Global Biodiversity Information Facility (GBIF)**, 2020c. Disponível em: <<https://gbif.github.io/gbif-api/apidocs/index-all.html>>. Acesso em: 14 jul 2020.

GBIF. Darwin Core API 1.28-SNAPSHOT API. **Global Biodiversity Information Facility (GBIF)**, 2020d. Disponível em: <<https://gbif.github.io/dwc-api/apidocs/>>. Acesso em: 14 jul 2020.

GBIF. Get Data - Datasets - Search Datasets. **Global Biodiversity Information Facility (GBIF)**, 2020e. Disponível em: <<https://www.gbif.org/dataset/search>>. Acesso em: 13 jul 2020.

GBIF. Repository of Schemas. **Global Biodiversity Information Facility (GBIF)**, 2020f. Disponível em: <<https://rs.gbif.org/>>. Acesso em: 14 jun 2020.

GILLILAND, A. J. Setting the stage. In: BACA, M. **Introduction to Metadata**. 3ª. ed. Los Angeles: Getty Research Institute, 2016. ISBN 9781606064795 (pbk.); ISBN 9781606064801 (epub); ISBN 9781606065006 (online). Disponível em: <<http://www.getty.edu/publications/intrometadata/>>. Acesso em: 03 maio 2019.

GLASS, G. V. Primary, secondary and meta-analysis of research. **Educational Researcher**, v. 5, n. 10, p. 3-8, 1976. Disponível em: <<https://doi.org/10.3102/0013189X005010003>>.

GOMES, H. E. ARQUIVOLOGIA E CIÊNCIA DA INFORMAÇÃO – PONTOS DE INTERSEÇÃO. In: CAMPOS, M. L. D. A., et al. **Produção, tratamento, disseminação e uso recursos informacionais heterogêneos [recurso eletrônico]: diálogos interdisciplinares**. Niterói: IACS/UFF, v. 5; 241p, 2018. Cap. 1, p. 12-28. ISBN 978-85-86918-08-7..

GONZALES, R. et al. SFMN GeoSearch: An interactive approach to the visualization and exchange of point-based ecological data. **Ecological Informatics**, v. 4, p. 196-205, 2009. Disponível em: <DOI: 10.1016/j.ecoinf.2009.07.007>.

GORDON, S.; HABERMANN, T. The influence of community recommendations on metadata completeness. **Ecological Informatics**, v. 43, p. 38-51, 2018. Disponível em: <DOI: 10.1016/j.ecoinf.2017.09.005>.

GRAHAM, M.; KENNEDY, J. Vesper: Visualising species archives. **Ecological Informatics**, v. 24, p. 132-147, 2014. Disponível em: <DOI: 10.1016/j.ecoinf.2014.08.004>.

GÜNTSCH, A. et al. Efficient rescue of threatened biodiversity data using reBiND workflows. **Plant Biosystems**, v. 146, n. 4, p. 752-755, 2012. Disponível em: <DOI: 10.1080/11263504.2012.740086>.

GURALNICK, R. P. et al. Community Next Steps for Making Globally Unique Identifiers Work for Biocollections Data. **ZooKeys**, v. 494, p. 133-154, 2015. Disponível em: <DOI: 10.3897/zookeys.494.9352>.

HAASA, S. C. et al. Darwin and MARC: A voyage of metadata discovery. **Library Collections, Acquisitions & Technical Services**, v. 27, p. 291-304, 2003. Disponível em: <DOI: 10.1016/S1464-9055(03)00071-X>.

HAIGEN XU et al. China National Biodiversity Information Query System. **Journal of Environmental Management**, 56, mai 1999. 55-59.

HAIGEN XU; DEHUI WANG; XUEFENG SUN. Biodiversity Clearing-House Mechanism in China: present status and future needs. **Biodiversity and Conservation**, 9, 2000. 361–378.

HAIGEN XU; YIMIN LI; HUI DING. The establishment of China's core biodiversity metadata standard. **Biodiversity and Conservation**, v. 13, p. 1009-1022, 2004.

HALE, S. S.; HOLLISTER, J. W. Beyond data management: how ecoinformatics can benefit environmental monitoring programs, p. 227-235, 2009. Disponível em: <DOI: 10.1007/s10661-008-0675-x>.

HARPRING, P. **Introdução aos vocabulários controlados**: terminologia para arte, arquitetura e outras obras culturais. São Paulo: Se. Tradução de Christina Maria Müller. São Paulo: Secretaria da Cultura do Estado: Pinacoteca de São Paulo: ACAM Portinari, v. 4, 2016. 288 p. ISBN 978-85-63566-19-5. Disponível em: <<https://www.sisemsp.org.br/wp-content/uploads/2013/12/Vocabularios%20Controlados%20-%20Digital.pdf>>. Acesso em: 12 março 2019.

HILBERT, M.; LOPEZ, P. How to measure the worlds technological capacity to communicate, store and compute information part I: Results and scope. **International Journal of Communication**, Los Angeles, v. 6, p. 956-979, 2012.

HOCK, B.; PAYN, T.; HEAPHY, M. Issues in the re-use of non-forestry specific spatial data sources for national environmental reporting on planted forests in New Zealand. **Journal of Spatial Science**, v. 59, n. 2, p. 237-252, 2014. Disponível em: <DOI: 10.1080/14498596.2014.913508>.

HOLETSCHEK, J. et al. The ABCD of primary biodiversity data access. **Plant Biosystems**, v. 146, n. 4, p. 771-779, 2012. Disponível em: <DOI: 10.1080/11263504.2012.740085>.

HONDROS, C. Standardizing legal content with OWL and RDF. In: WOOD, D. **Linking Enterprise Data**. New York: Springer, 2010. p. 221-240. Disponível em: <<http://linkeddatadeveloper.com/Projects/Linking-Enterprise-Data/Manuscript/led-hondros.html>>. Acesso em: 02 dez. 2018.

HUANG, Y.-L. et al. Using collections data to infer biogeographic, environmental, and host structure in communities of endophytic fungi. **Mycologia**, v. 110, n. 1, p. 47-62, 2018. Disponível em: <DOI: 10.1080/00275514.2018.1442078>.

HUDSON, L. N. et al. The database of the PREDICTS (Projecting Responses of Ecological Diversity In Changing Terrestrial Systems) project. **Ecology and Evolution**, v. 7, p. 145-188, 2017. Disponível em: <DOI: 10.1002/ece3.2579>.

HUETTMANN, F. et al. Predictions of 27 Arctic pelagic seabird distributions using public environmental variables, assessed with colony data: a first digital IPY and GBIF open access synthesis platform. **Marine Biodiversity**, v. 41, p. 141-179, 2011. Disponível em: <DOI: 10.1007/s12526-011-0083-2>.

IDC. The Digital Universe in 2020: Big Data, Bigger Digital Shadows, Biggest Growth in the Far East. **International Data Corporation (IDC)**, 2012. Disponível em: <<https://www.emc.com/leadership/digital-universe/2012iview/index.htm>>. Acesso em: 31 maio 2018.

INTERNET WORLD STATS. World Internet User Statistics and 2018 World Population Stats. **Internet World Stats - Usage and Population Statistics**, 25 set. 2018. Disponível em: <<https://www.internetworldstats.com/stats.htm>>. Acesso em: 28 nov. 2018.

ISOTANI, S.; BITTENCOURT, I. I. **Dados abertos conectados**. [S.l.]: Novatec, 2015. Disponível em: <<http://ceweb.br/livros/dados-abertos-conectados/>>. Acesso em: 04 jun. 2018. versão html.

JACKENDOFF, R. If computers are so smart, why can't they use simple English? **Linguistic Society of America**, 2012. Disponível em: <<https://www.linguisticsociety.org/resource/faq-why-cant-computers-use-english>>. Acesso em: 28 nov. 2018.

JONES, M. B. et al. The New Bioinformatics: Integrating Ecological Data from the Gene to the Biosphere. **Annual Review of Ecology, Evolution, and Systematics**, v. 37, p. 519-544, 2006.

KANDEL, K. et al. Rapid multi-nation distribution assessment of a charismatic conservation species using open access ensemble model GIS predictions: Red panda (*Ailurus fulgens*) in the Hindu-Kush Himalaya region. **Biological Conservation**, v. 181, p. 150-161, 2015. Disponível em: <DOI: 10.1016/j.biocon.2014.10.007>.

KAUARK, F.; MANHÃES, F. C.; MEDEIROS, C. H. **Metodologia da pesquisa: guia prático**. Itauna: VIA LITTERARUM, 2010. 86 p.

KENNEY, A. R.; RIEGER, O. Y.; ENTLICH, R. **Levando la teoria a la práctica: tutorial de digitalización de imágenes**. Biblioteca de la Universidad de Cornell. Mountain View, p. 188. 2003.

KING, N. et al. Improving access to biodiversity data for, and from, EIAs – a data publishing framework built to global standards. **Impact Assessment and Project Appraisal**, 30:3, 148-156, v. 30, n. 3, p. 148-156, 2012. ISSN ISSN: 1461-5517; DOI: 10.1080/14615517.2012.705068.

KISSLING, W. D. et al. Building essential biodiversity variables (EBVs) of species distribution and abundance at a global scale. **Biological Reviews**, v. 93, p. 600-625, 2018. Disponível em: <DOI: 10.1111/brv.12359>.

KNB. About KNB. **The Knowledge Network for Biocomplexity (KNB)**, 2019. Disponível em: <<https://knb.ecoinformatics.org/about>>. Acesso em: 17 dez. 2019.

KNB. EML Specification. **Knowledge Network for Biocomplexity (KNB)**, 2019. Disponível em: <<https://eml.ecoinformatics.org/>>. Acesso em: 18 dez. 2019.

KOBASHI, N. Y. Linguística Textual e Elaboração de Informações Documentárias. In: GASPAR, N. R.; ROMÃO, L. M. S.; (ORGS) **Discurso e texto: multiplicidade de sentidos na ciência da informação**. São Carlos: EdUFSCar, 2008. p. 359. ISBN 9788576001379; eISBN:9788576003700; DOI:<https://doi.org/10.7476/9788576003700>.

KOBASHI, N. Y. Organização e recuperação de informação de documentos de anistia política. In: CAMPOS, M. L. D. A., et al. **Produção, tratamento, disseminação e uso recursos informacionais heterogêneos [recurso eletrônico]: diálogos interdisciplinares**. Niterói: IACS/UFF, v. 5, 2018. Cap. 8, p. 89-100. ISBN 978-85-86918-08-7.

KOTHARI, C. R. **Research Methodology: Methods and Techniques**. 2ª. ed. Nova Deli: NEW AGE INTERNATIONAL (P) LIMITED, PUBLISHERS, 2004. 401 p.

KWANG-TSAO SHAO et al. INTEGRATION OF BIODIVERSITY DATABASES IN TAIWAN AND LINKAGE TO GLOBAL DATABASES. **Data Science Journal**, v. 6, 2007.

- LAIHONEN, P. et al. Geospatially structured biodiversity information as a component of a regional biodiversity clearing house. **Biodiversity and Conservation**, v. 12, p. 103-120, 2003. Disponível em: <DOI: 10.1023/A:1021238212962>.
- LANCASTER, F. W.; SMITH, L. C. Science, scholarship and the communication of knowledge. **Library Trends**, v. 27, n. 3, p. 367-387, 1978.
- LAUFER, C. Guia de Web Semântica. **CEWEB.br**, São Paulo, p. 133, 2015. Disponível em: <http://ceweb.br/media/docs/publicacoes/13/Guia_Web_Semantica.pdf>. Acesso em: 28 nov. 2018.
- LE COADIC, Y. F. **A ciência da informação**. 2ª. ed. Brasília: Briquet de Lemos, 2004.
- LEIGH, C. et al. IRBAS: An online database to collate, analyze, and synthesize data on the biodiversity and ecology of intermittent rivers worldwide. **Ecology and Evolution**, v. 7, p. 815-823, 2017. Disponível em: <DOI: 10.1002/ece3.2679>.
- LEITE, F. C. L. Gestão do conhecimento científico no contexto acadêmico: proposta de um modelo conceitual - Dissertação (Mestrado em Ciência da Informação). **Universidade de Brasília - Repositório Institucional da Universidade de Brasília**, Brasília, p. 240, 2006. Disponível em: <<http://repositorio.unb.br/handle/10482/3975>>. Acesso em: 30 abr. 2019.
- LEITE, F. C. L. Modelo genérico de gestão da informação científica para instituições de pesquisa na perspectiva da comunicação científica e do acesso aberto - Tese (Doutorado em Ciência da Informação). **Repositório Institucional da Universidade de Brasília**, Brasília, p. 262, 2011. Disponível em: <<http://repositorio.unb.br/handle/10482/9753>>. Acesso em: 10 agosto 2018.
- LENGYEL, S. et al. Evaluating and benchmarking biodiversity monitoring: Metadata-based indicators for sampling design, sampling effort and data analysis. **Ecological Indicators**, v. 85, p. 624-633, 2018. Disponível em: <DOI: 10.1016/j.ecolind.2017.11.012>.
- LI, K.; GREENBERG, J.; DUNIC, J. Data objects and documenting scientific processes: An analysis of data events in biodiversity data papers. **Journal of the Association for Information Science & Technology**, v. 71, n. 2, p. 172-182, 2020. Disponível em: <DOI: 10.1002/asi.24226>.
- LIMA, J. A. D. O. Modelo Genérico de Relacionamentos e a Organização da Informação Legislativa e Jurídica - Tese (Doutorado em Ciência da Informação). **Repositório Institucional da Universidade de Brasília**, Brasília, p. 289, 2008. Disponível em: <<https://repositorio.unb.br/handle/10482/3398>>. Acesso em: 10 set. 2018.
- LIMA, J. A. D. O.; CUNHA, M. B. D. Modelo genérico de relacionamento e a organização da informação. **CONFERÊNCIA IBEROAMERICANA DE PUBLICAÇÕES ELETRÔNICAS NO CONTEXTO DA COMUNICAÇÃO CIENTÍFICA. Anais.**, Campo Grande, v. 1, p. 67-74, 2006.
- LOURENÇO, C. A. Metadados: O Grande Desafio na Organização da WEB. **Informação & Sociedade**, João Pessoa, 17, 2017..
- LUCARINI, D. et al. The anArchive taxonomic Checklist for Italian botanical data banking and vegetation analysis: Theoretical basis and advantages. **Plant Biosystems**, v. 146, n. 6, p. 958-965, 2014. Disponível em: <DOI: 10.1080/11263504.2014.984010>.
- MADIN, J. S. et al. Advancing ecological research with ontologies. **Trends in Ecology and Evolution**, v. 23, n. 3, p. 159-168, 2008. Disponível em: <DOI: 10.1016/j.tree.2007.11.007>.

MARCHIONINI, G. Information Science Roles in the Emerging. **Journal of Data and Information Science**, v. 1, n. 2, p. 1-6, 2016. DOI: 10.20309/jdis.201609.

MEADOWS, A. J. **A Comunicação Científica. Communicating Research**. Tradução de Antonio Agenor Briquet de Lemos. Brasília: Briquet de Lemos Livros, 1999. 268 p. ISBN: 85-85637-15-3.

MENDEL, T. **Liberdade de informação: um estudo de direito comparado**. 2ª. ed. Brasília: UNESCO, 2009. 162 p.

MENDES, K. D. S.; SILVEIRA, R. C. D. C. P.; GALVAO, C. M. Revisão integrativa: método de pesquisa para a incorporação de evidências na saúde e na enfermagem, Florianópolis, 2008. Disponível em: <<http://dx.doi.org/10.1590/S0104-07072008000400018>>. Acesso em: 12 ago. 2019.

MEY, E. S. A.; SILVEIRA, N. C. **Catálogo no plural**. Brasília: Briquet de Lemos, 2009. 217 p. ISBN 978.85.85637.39.2.

MINADAKIS, N. et al. LifeWatch Greece data-services: Discovering Biodiversity Data using Semantic Web Technologies. **Biodiversity Data Journal**, v. 4, n. e8443, 2016. Disponível em: <DOI: 10.3897/BDJ.4.e8443>.

MINICANTE, S. A. et al. Building a Natural and Cultural Heritage Repository for the Storage and Dissemination of Knowledge: The Algarium Veneticum and the Archivio di Studi Adriatici Case Study. **Journal of Library Metadata**, v. 17, n. 2, p. 111-125, 2017. ISSN ISSN: 1938-6389 ; DOI: 10.1080/19386389.2017.1355165.

MINISTÉRIO DO MEIO AMBIENTE - MMA. Convenção da Diversidade Biológica. **Ministério do Meio Ambiente**, 2019. Disponível em: <<http://www.mma.gov.br/biodiversidade/convenção-da-diversidade-biológica.html>>. Acesso em: 27 abr. 2019.

MORENO, F. P.; LEITE, F. C. L.; ARELLANO, M. Á. M. Acesso livre a publicações e repositórios digitais em ciência da informação no Brasil. , Belo Horizonte , v. 11, n. 1, p. 82-94, Apr. 2006. A. **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 11, n. 1, p. 82-94, jan./abr. 2006. ISSN 1981-5344. Disponível em: <<http://dx.doi.org/10.1590/S1413-99362006000100007>>.

MORITZ, T. et al. Towards mainstreaming of biodiversity data publishing: recommendations of the GBIF Data Publishing Framework Task Group. **BMC Bioinformatics**, v. 12, n. 15:S1, 2011. ISSN <https://doi.org/10.1186/1471-2105-12-S15-S1>.

MUELLER, S. P. M. A ciência, o sistema de comunicação científica e a literatura científica. In: CAMPELLO, B. S.; CENDÓN, B. V.; KREMER, J. M. **Fontes de informação para pesquisadores e profissionais**. Belo Horizonte: UFMG, 2000. Cap. 1, p. 21 a 34 de 319.

MUELLER, S. P. M. A CIÊNCIA, O SISTEMA DE COMUNICAÇÃO CIENTÍFICA E A LITERATURA CIENTÍFICA. In: CAMPELLO, B. S., et al. **Fontes de informação para pesquisadores e profissionais**. Belo Horizonte: UFMG, 2000. Cap. 1, p. 319.

MUELLER, S. P. M.; PASSOS, E. J. L. As questões da comunicação científica e a ciência da informação. In: MUELLER, S. P. M.; PASSOS, E. J. L.; (ORGS) **Comunicação científica**. Brasília: Ciência da Informação, 2000. p. 13-22. Disponível em: <<http://repositorio.unb.br/handle/10482/1444>>.

- NISO. **Understanding metadata what is metadata, and what is it for?** National Information Standards Organization (NISO). Baltimore, p. 45. 2017.
- NIU, J. Aggregate control of scientific data. **Archives and Records**, v. 37, n. 1, p. 53-64, 2016a. Disponível em: <DOI: 10.1080/23257962.2016.1145578>.
- NIU, J. Organisation and description of datasets. **Archives and Manuscripts**, v. 44, n. 2, p. 73-85, 2016b. ISSN ISSN: 0157-6895; DOI: 10.1080/01576895.2016.1179585.
- OBIS, O. B. I. S. Darwin Core Terms. **gCube**, 2013. Disponível em: <https://gcube.wiki.gcube-system.org/gcube/Darwin_Core_Terms>. Acesso em: 17 dez. 2019.
- OECD. **Final Report of the OECD Megascience Forum**. Organisation for Economic Co-operation and Development (OECD). Paris, p. 74. 1999.
- OECD. **Memorandum of Understanding for the Global Biodiversity Information Facility**. Organisation for Economic Co-operation and Development (OECD). Copenhagen, p. 17. 2010. (ISBN 87-92020-17-8).
- ONU. 17 OBJETIVOS PARA TRANSFORMAR NOSSO MUNDO. **Organização das Nações Unidas (ONU)**, 2015. Disponível em: <<https://nacoesunidas.org/pos2015>>. Acesso em: 05 fev. 2020.
- ONU. Convention on Biodiversity. **Organização das Nações Unidas (United Nations) - Biodiversity**, 2019. Disponível em: <<https://www.un.org/en/events/biodiversityday/convention.shtml>>. Acesso em: 27 abr. 2019.
- ONU. A ONU e o meio ambiente. **Organização das Nações Unidas (ONU)**, 2020. Disponível em: <<https://nacoesunidas.org/acao/meio-ambiente/>>. Acesso em: 11 jun 2020.
- ONU. CDB. United Nations Decade on Biodiversity - Strategic Plan for Biodiversity 2011-2020. **Organização das Nações Unidas (ONU). Convention on Biological Diversity (CBD)**, 2010. Disponível em: <<https://www.cbd.int/2011-2020/>>. Acesso em: 28 jan. 2020.
- OROZCO GARCÍA-MAYORCA, A. La construcción de metadatos en el proceso de organización análisis documental y recuperación de la información em los archivos de imágenes em movimiento, 2001. Disponível em: <<http://www.patrimoniofilmico.org.co/anterior/docs/metadatos.pdf>>. Acesso em: 02 abr 2018.
- OSTROM, E. **Governing the Commons - The Evolution of Institutions for Collective Action**. 2003. ed. [S.l.]: Cambridge University Press, 1990. ISBN 0 521405998.
- OSTROM, E.; HESS, C. (Eds.). **Understanding Knowledge as a Commons - From Theory to Practice**. Cambridge; London: MIT Press, 2007. 367 p. ISBN 978-0-262-08357-7.
- OTEGUI, J. et al. Assessing the Primary Data Hosted by the Spanish Node of the Global Biodiversity Information Facility (GBIF). **PLoS ONE**, v. 8, n. 1, 2013. Disponível em: <DOI: 10.1371/journal.pone.0055144>.
- PETERS, D. P. C. Accessible ecology: synthesis of the long, deep, and broad. **Trends in Ecology & Evolution**, v. 25, n. 10 - Special Issue: Long-term ecological research, 2010. ISSN <https://doi.org/10.1016/j.tree.2010.07.005>.

PFAFF, C.-T. et al. rBEFdata: documenting data exchange and analysis for a collaborative data management platform. **Ecology and Evolution**, v. 5, n. 14, p. 2890-2897, 2015. Disponível em: <DOI: 10.1002/ece3.1547>.

PÔÇAS, I. et al. Evaluating the fitness for use of spatial data sets to promote quality in ecological assessment and monitoring4. **International Journal of Geographical Information Science**, v. 28, n. 11, p. 2356-2371, 2014. Disponível em: <DOI: 10.1080/13658816.2014.924627>.

POLYDORATOU, P.; NICHOLAS, D. Familiarity With and Use of Metadata Formats and Metadata Registries amongst Those Working in Diverse Professional Communities Within the Information Sector. **ASLIB Proceedings**, v. 53, n. 8, p. 309-324, 2001. Disponível em: <<http://dx.doi.org/10.1108/EUM0000000007064D>>. Acesso em: 05 mai. 2019.

POOTER, D. D. et al. Toward a new data standard for combined marine biological and environmental datasets - expanding OBIS beyond species occurrences. **Biodiversity Data Journal**, v. 5, n. e10989, 2017. Disponível em: <DOI: 10.3897/BDJ.5.e10989>.

RASAI AH, B. A. et al. Towards an Interoperable Field Spectroscopy Metadata Standard with Extended Support for Marine Specific Applica. **Remote Sensing**, v. 7, n. 1, p. 15668-1570, 2015. Disponível em: <DOI: 10.3390/rs71115668>.

REMSEN, D. et al. From text to structured data: Converting a wordprocessed floristic checklist into Darwin Core Archive format. **PhytoKeys**, v. 9, n. 1, 2012. Disponível em: <DOI: 10.3897/phytokeys.9.2770>.

ROBERTSON, T. et al. The GBIF Integrated Publishing Toolkit: Facilitating the Efficient Publishing of Biodiversity Data on the Internet. **PLoS ONE**, v. 9, n. 8, p. e102623, 2014. Disponível em: <DOI: 10.1371/journal.pone.0102623>.

ROCHE, D. G. et al. Public Data Archiving in Ecology and Evolution: How Well Are We Doing? **PLoS Biol**, v. 13, n. 11, 2015. Disponível em: <DOI: 10.1371/journal.pbio.1002295>.

RODRÍGUEZ, E. M. M. **Metadatos y recuperación de la información: estándares, problemas y aplicabilidad en bibliotecas digitales - Tese (Programa de Doctorado en Documentación)**. Universidad Carlos III de Madrid. Madrid, p. 845. 2001.

ROSCOE, D. D.; JENKINS, S. A meta-analysis of campaign contributions impact on roll call voting. **Social Science Quarterly**, v. 86, n. 1, p. 52-68, 2005. Disponível em: <<https://doi.org/10.1111/j.0038-4941.2005.00290.x>>.

RUBEN, B. D. Integrating concepts for the Information age: communication, information. In: RUBEN, B. D.; SCHEMENT, J. R. **Between communication and information. Information and Behavior**. New Jersey: Transaction Publishers, v. 4, 1993. Cap. 10, p. 219-236. ISSN: 0740-5502 e ISBN 13: 978-1-56000-037-2 (hbk).

SANTOS, P. L. V. A. D. C.; SIMIONATO, A. C.; ARAKAKI, F. A. Definição de metadados para recursos informacionais: apresentação da metodologia BEAM. **Informação & Informação**, v. 19, p. 146-163, fev. 2014. ISSN 1981-8920. Disponível em: <<http://www.uel.br/revistas/uel/index.php/informacao/article/view/15251>>. Acesso em: 02 mai. 2019.

SARACEVIC, T. Ciência da Informação: origem, evolução e relações. **Perspectivas em Ciência da Informação**, v. 1, n. 1, p. 4-62, 1996.

SAYÃO, L. **Bibliotecas digitais: saberes e práticas**. Salvador; Brasília: EDUFBA; IBICT, 2005. 342 p.

SAYÃO, L. F. Padrões para bibliotecas digitais abertas e interoperáveis. **Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação**, Florianópolis, dez. 2007. 18-47. Disponível em: <<https://periodicos.ufsc.br/index.php/eb/article/view/1518-2924.2007v12nesp1p18/436>>. Acesso em: 14 ago. 2019.

SAYÃO, L. F. Uma outra face dos metadados: informações para a gestão da preservação digital. **Encontros Bibli: Revista Eletrônica de Biblioteconomia e Ciência da Informação**, Florianópolis, v. 15, n. 30, p. 1-31, 2010. Disponível em: <<https://periodicos.ufsc.br/index.php/eb/article/view/1518-2924.2010v15n30p1>>. Acesso em: 04 maio 2019.

SAYÃO, L. F.; SALES, L. F. **Guia de Gestão de Dados de Pesquisa para Bibliotecários e Pesquisadores**. CNEN/IEN. Rio de Janeiro, p. 90. 2015. (978-85-61905-03-3).

SCDB. Thematic Programmes and Cross-Cutting Issues. **Secretariat of the Convention on Biological Diversity - SCDB**, 2019. Disponível em: <<https://www.cbd.int/programmes/>>. Acesso em: 27 abr. 2019.

SCHINDEL, D. E. et al. The Global Registry of Biodiversity Repositories: A Call for Community Curation. **Biodiversity Data Journal**, v. 4, n. e10293, 2016. Disponível em: <DOI: 10.3897/BDJ.4.e10293>.

SCHMIDT-KLOIBER, A. et al. The WISER metadatabase: the key to more than 100 ecological datasets from European rivers, lakes and coastal waters, p. 29-38, 2013. Disponível em: <DOI: 10.1007/s10750-012-1295-6>.

SCHUCKA, A. et al. The European Forest Information System - an Internet based interface between information providers and the user community. **Computers and Electronics in Agriculture**, v. 47, n. 3, p. 185-206, 2005. ISSN <https://doi.org/10.1016/j.compag.2004.12.004>.

SEREGIN, A. P. Digital Herbarium of Moscow State University: The Largest Russian Biodiversity Database. **Biology Bulletin**, v. 44, n. 6, p. 584-590, 2017. ISSN 1062-3590; DOI: 10.1134/S1062359017060103.

SILVA, D. L. et al. **Diretrizes para a Integração de Dados de Biodiversidade**. Brasília: Ministério do Meio Ambiente, 2015. 100 p. ISBN 978-85-7738-195-1.

SILVA, T. E. D.; TOMAÉL, M. I. Política de Informação: tendências internacionais. **Anais - GT5 - X Encontro Nacional de Pesquisa em Ciência da Informação**, João Pessoa, 2009. Disponível em: <<http://enancib.ibict.br/index.php/enancib/xenancib/paper/viewFile/3257/2383>>. Acesso em: 27 nov. 2018.

SINGER, R. A.; LOVE, K. J.; PAGE, L. M. A survey of digitized data from U.S. fish collections in the iDigBio data aggregator. **PLoS ONE 13(12)**; v. 13, n. 12 e0207636, 2018. Disponível em: <DOI: 10.1371/journal.pone.0207636>.

SIQUEIRA, I. C. P.; SILVA, J. F. M. D. Metadados: o fio de ariadne ou a coragem de teseu? **Bibliotecas Universitárias: pesquisas, experiências e perspectivas**, 1, 2011. Disponível em: <<http://hdl.handle.net/20.500.11959/brapci/92353>>. Acesso em: 6 mar. 2019.

SKEVAKIS, G. et al. A crowdsourcing framework for the management of mobile multimedia nature observations. **International Journal of Pervasive Computing and Communications**, v. 10, n. 3, p. 216-238, 2014. Disponível em: <DOI 10.1108/IJPC-06-2014-0038>.

SKEVAKIS, G. et al. Metadata management, interoperability and Linked Data publishing support for Natural History Museums. **International Journal on Digital Libraries**, v. 14, p. 127-140, 2014. Disponível em: <DOI 10.1007/s00799-014-0114-2>.

SMIRAGLIA, R. P. **Metadata: a cataloger's primer**. Nova York: Haworth Information Press, 2005. 312 p. ISBN 13: 978-0789028013; ISBN-10: 0789028018.

SMIT, J. W.; KOBASHI, N. Y. **Como elaborar vocabulário controlado para aplicação em arquivos**. Como Fazer. ed. São Paulo: [s.n.], v. 10, 2002. 56 p. ISBN 85-86726-52-4 e 85-7060-216-2.

SOUZA, J. F. D. Web Semântica. **Universidade Federal de Juiz de Fora (UFJF)**, 2013. Disponível em: <http://www.ufjf.br/jairo_souza/files/2013/05/Web-Sema%CC%82ntica.pdf>. Acesso em: 21 fev. 2019.

SPANG-HANSEN, H. How to teach about information as related to documentation? **Human IT**, 2001. Disponível em: <<https://humanit.hb.se/article/download/168/186>>. Acesso em: 20 nov. 2018.

TAKAHASHI, T. (Ed.). **Sociedade da informação no Brasil: livro verde**. Brasília: Ministério da Ciência e Tecnologia, 2000. 192 p. ISBN 85-88063-01-8. Disponível em: <<https://www.governodigital.gov.br/documentos-e-arquivos/livroverde.pdf>>. Acesso em: 04 maio 2019.

TDWG. Darwin Core. **Biodiversity Information Standards**, 2015a. Disponível em: <<https://dwc.tdwg.org/>>. Acesso em: 25 nov 2019.

TDWG. Simple Darwin Core. **Biodiversity Information Standards (TDWG)**, 2015b. Disponível em: <<http://rs.tdwg.org/dwc/simple/>>. Acesso em: 17 dez. 2019.

TDWG. Darwin Core Community. **Biodiversity Information Standards**, 2015c. Disponível em: <<https://www.tdwg.org/community/dwc/>>. Acesso em: 25 nov 2019.

TDWG. About TDWG. **Biodiversity Information Standards**, 2019a. Disponível em: <<https://www.tdwg.org/about>>. Acesso em: 09 dez. 2019.

TDWG, B. I. S. ABCD 3.0. **Access to Biological Collection Data**, 2019b. Disponível em: <<https://abcd.tdwg.org/3.0/>>. Acesso em: 16 out 2019.

TESSAROLO, G. et al. Temporal degradation of data limits biodiversity research. **Ecology and Evolution**, v. 7, p. 6863-6870, 2017. Disponível em: <DOI: 10.1002/ece3.3259>.

THOMER, A. et al. From documents to datasets: A MediaWiki-based method of annotating and extracting species observations in century-old field notebooks. **ZooKeys**, v. 209, p. 235-253, 2012. Disponível em: <DOI: 10.3897/zookeys.209.3247>.

TORRES, R. D. S. et al. A digital library framework for biodiversity information systems. **International Journal on Digital Libraries**, v. 6, n. 1, p. 3-17, 2006. Disponível em: <DOI: 10.1007/s00799-005-0124-1>.

- UN. Act for nature forum. **United Nations (UN) - United Nations Environment Assembly**, 2020. Disponível em: <<https://environmentassembly.unenvironment.org/act-for-nature-forum>>. Acesso em: 23 jan 2020.
- VANDERBILT, K. L. et al. Fostering ecological data sharing: collaborations in the International Long Term Ecological Research Network. **Ecosphere**, v. 6, n. 10 (204), 2015. Disponível em: <DOI: 10.1890/ES14-00281.1>.
- VEEN, L. E. et al. A semantically integrated, user-friendly data model for species observation data. **Ecological Informatics**, v. 8, p. 1-9, 2012. Disponível em: <DOI:10.1016/j.ecoinf.2011.11.002>.
- VELÁSQUEZ-TIBATÁ, J. et al. BioModelos: A collaborative online system to map species distributions. **PLoS ONE**, v. 14, n. 3 e0214522, 2019. Disponível em: <DOI: 10.1371/journal.pone.0214522>.
- VOLLMAR, A.; MACKLIN, J. A.; FORD, L. S. Natural history specimen digitization: challenges and concerns. **Biodiversity Informatics**, v. 7, p. 93-112, 2010.
- W3C. W3C Semantic Web Activity. **W3C**, 2013. Disponível em: <<https://www.w3.org/2001/sw/>>. Acesso em: 28 nov. 2018.
- WALLS, R. L. et al. Meeting report: advancing practical applications of biodiversity ontologies. **Walls et al. Standards in Genomic Sciences**, v. 9, n. 17, 2014. Disponível em: <DOI: 10.1186/1944-3277-9-17>.
- WEATHERDON, L. V. et al. Blueprints of Effective Biodiversity and Conservation Knowledge. **Frontiers in Marine Science**, 2017. Disponível em: <DOI: 10.3389/fmars.2017.00096>.
- WERSIG, G.; NEVELING, U. The phenomena of interest to information science. **The Information Scientist**, v. 9, n.4, 1975.
- WILKINSON, M. D. et al. The FAIR Guiding Principles for scientific data management and stewardship. **Scientific Data**, v. 3, n. 160018, mar 2016. ISSN 2052-4463; DOI: 10.1038/sdata.2016.18.
- WILKINSON, M. D. et al. Interoperability and FAIRness through a novel combination of Web technologies. **PeerJ Computer Science**, v. 3, n. 110, 2017. Disponível em: <DOI: 10.7717/peerj-cs.110>.
- WILSON, T. D. Models in information behaviour research. **Journal of Documentation**, n. 3, p. 249-279, 1999.
- WONG, I. W. et al. Species at risk: Data and knowledge management within the WILDSPACE Decision Support System. **Environmental Modelling & Software**, v. 22, p. 423-430, 2007.
- ZILIOLI, M. et al. Feeding Essential Biodiversity Variables (EBVs): actual and potential contributions from LTER-Italy. **Nature Conservation**, n. 34, p. 477-503, 2019. Disponível em: <DOI: 10.3897/natureconservation.34.30735>.
- ZLMAN, J. **Public knowledge**. London: Cambridge University Press, 1968.

APÊNDICES

APÊNDICE A.	Roteiro de estudo para coleta de dados nos artigos científicos	257
APÊNDICE B.	Lista de verificação do uso de metadados nos <i>datasets</i>	258
APÊNDICE C.	Dados coletados nos artigos científicos – roteiro parte 01	259
APÊNDICE D.	Dados coletados nos artigos científicos – roteiro parte 02	262
APÊNDICE E.	Dados coletados nos artigos científicos – roteiro parte 03	265
APÊNDICE F.	Dados coletados nos artigos científicos – roteiro parte 04	271
APÊNDICE G.	Lista de artigos elegíveis para análise	278
APÊNDICE H.	Artigos científicos excluídos por não tratar do uso de metadados para biodiversidade	282
APÊNDICE I.	Artigos científicos selecionados para análise sobre uso de metadados	284
APÊNDICE J.	Artigos excluídos após análise refinada	291

APÊNDICE A. Roteiro de estudo para coleta de dados nos artigos científicos

A. Área do conhecimento do periódico da publicação
1. Ciência da Informação
2. Ecologia, Ciências Biológicas ou Biodiversidade
3. Outra (sempre multidisciplinar/interdisciplinar)
B. Uso de padrão de metadados é assunto central?
1. sim
2. não
C. Uso de padrão de metadados é parte do método
1. sim
2. não
D. Abordagem relacionada ao uso de metadados
E. Outros assuntos associados:
F. Padrões de metadados da biodiversidade mencionados no artigo
1. ABCD
2. EML
3. Darwin Core
4. GMP (GBIF metadata profile)
G. Outras preferências mencionadas no artigo
H. Termos-chave utilizados nas buscas que estão presentes no artigo.
a) Biodiversity
b) Metadata
c) ABCD
d) Darwin Core
e) EML
f) GBIF profile
I. Palavras-chave
J. Assunto do artigo

APÊNDICE B. Lista de verificação do uso de metadados nos *datasets*

LISTA DE VERIFICAÇÃO
1 - Identificação do conjunto de dados (nome/título e DOI)
2 - Identificação da instituição ou pesquisador responsável pelo depósito
3 - Classe do conjunto de dados (<i>checklist</i> ; ocorrência; evento de amostragem; metadados de recurso)
4 - Ano de publicação/depósito
5 - Uso de metadados para descrição do <i>dataset</i> : Campos de metadados presente (lista completa de metadados descritos para o conjunto de dados)
6 - Uso de metadados para descrição dos dados de pesquisa registrados no <i>dataset</i> : Campos de metadados presente (lista completa de metadados descritos para o conjunto de dados)

APÊNDICE C. Dados coletados nos artigos científicos – roteiro parte 01

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	
A. Área do conhecimento do periódico da publicação																															
1. Ciência da Informação																X		X												X	
2. Ecologia, Ciências Biológicas ou Biodiversidade		X		X	X	X	X	X		X	X	X	X	X	X		X		X	X	X	X	X					X			
3. Outra (sempre multidisciplinar/interdisciplinar)	X		X						X																X	X	X		X		X
B. Uso de padrão de metadados é assunto central?																															
1. sim					X	X		X							X	X			X	X			X						X	X	
2. não	X	X	X	X			X		X	X	X	X	X	X			X	X			X	X		X	X	X	X				X
C. Uso de padrão de metadados é parte do método																															
1. sim		X	X	X	X	X			X	X	X	X		X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
2. não	X						X	X					X																X		
D. Abordagem relacionada ao uso de metadados																															
1. Proposta de padrão de metadado																															
2. Adoção de padrão de metadado			X						X						X			X	X											X	
3. Uso de metadados na gestão de dados	X								X								X		X					X	X						X
4. Uso de metadados na análise de dados				X	X					X													X			X	X				
5. Perfil de metadados																							X								
6. Gestão e/ou Análise de metadados		X		X	X	X										X	X							X			X				X
7. Solução para coleta e registro de metadados																											X				

	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	
A. Área do conhecimento do periódico da publicação																															
1. Ciência da Informação																						X									
2. Ecologia, Ciências Biológicas ou Biodiversidade		X		X	X	X	X	X	X		X	X		X	X	X	X	X	X		X		X	X			X		X	X	
3. Outra (sempre multidisciplinar/interdisciplinar)	X		X							X			X							X					X	X		X			
B. Uso de padrão de metadados é assunto central?																															
1. sim	X	X	X	X	X		X			X		X	X					X		X	X	X		X	X			X	X		
2. não						X		X	X		X			X	X	X	X		X				X			X	X			X	
C. Uso de padrão de metadados é parte do método																															
1. sim	X	X	X	X	X		X	X	X	X	X	X	X	X	X	X	X	X		X	X	X	X	X		X		X	X	X	
2. não						X													X						X		X				
D. Abordagem relacionada ao uso de metadados																															
1. Proposta de padrão de metadado					X						X																				
2. Adoção de padrão de metadado			X	X			X	X	X		X	X	X			X						X									
3. Uso de metadados na gestão de dados	X						X			X			X		X	X		X						X	X	X		X			
4. Uso de metadados na análise de dados		X													X						X				X					X	
5. Perfil de metadados	X		X	X							X																				
6. Gestão e/ou Análise de metadados		X		X													X	X		X		X		X				X		X	
7. Solução para coleta e registro de metadados								X					X					X				X			X			X		X	

	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78
A. Área do conhecimento do periódico da publicação																		
1. Ciência da Informação			X					X										X
2. Ecologia, Ciências Biológicas ou Biodiversidade				X	X		X		X	X			X				X	X
3. Outra (sempre multidisciplinar/interdisciplinar)	X	X				X				X	X	X		X	X			
B. Uso de padrão de metadados é assunto central?																		
1. sim	X					X	X	X				X	X	X			X	X
2. não		X	X	X	X				X	X	X					X	X	
C. Uso de padrão de metadados é parte do método																		
1. sim	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X			
2. não																X	X	X
D. Abordagem relacionada ao uso de metadados																		
1. Proposta de padrão de metadado							X											
2. Adoção de padrão de metadado	X			X														X
3. Uso de metadados na gestão de dados		X	X	X	X					X	X	X	X			X	X	X
4. Uso de metadados na análise de dados				X		X				X			X		X			
5. Perfil de metadados																		
6. Gestão e/ou Análise de metadados						X		X		X	X	X	X					
7. Solução para coleta e registro de metadados	X		X	X		X					X							X

APÊNDICE D. Dados coletados nos artigos científicos – roteiro parte 02

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	
E. Outros assuntos associados:																															
1. Base de dados ou integração de dados	X		X			X	X		X	X	X	X		X	X	X	X	X	X	X		X	X								X
2. Modelos de análise da biodiversidade	X																									X					
3. Definição de variáveis essenciais (EBV)				X									X																		
4. Subsidiar decisões e definição políticas								X																			X				
5. Modelo semântico para biodiversidade																X													X		
6. Identificadores únicos para conjuntos de dados																				X											
7. Repositórios e portais																			X			X									
8. Compartilhamento e reuso de dados																					X	X									
9. Análise de conjunto de dados			X	X	X		X			X				X			X					X								X	
10. Procedimentos de gestão de dados							X	X	X		X	X				X	X	X				X	X	X	X				X	X	
11. Vocabulários e ontologia																												X			
12. Análise da procedência dos dados (qualidade)																									X						
13. Digitalização e estruturação de coleções											X																				
F. Padrões de metadados mencionados no artigo																															
1. ABCD								X								X					X								X		
2. EML	X	X		X		X		X							X	X		X			X		X			X			X		
3. Darwin Core	X		X	X				X	X						X	X	X	X	X		X							X		X	
4. GMP (GBIF metadata profile)	X							X							X													X			

	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78
E. Outros assuntos associados:																		
1. Base de dados ou agregador de dados	X	X	X	X	X	X				X		X	X				X	
2. Modelos de análise da biodiversidade																		
3. Definição de variáveis essenciais (EBV)																		
4. Subsidiar decisões e definição políticas									X						X		X	
5. Modelo semântico para biodiversidade																		
6. Identificadores únicos para conjuntos de dados																		
7. Repositórios, bancos de dados e portais					X							X			X		X	
8. Compartilhamento e reuso de dados	X			X	X	X			X	X		X	X					
9. Análise de conjunto de dados															X			
10. Procedimentos de gestão de dados				X	X							X					X	X
11. Vocabulários e ontologia																X		
12. Análise da procedência dos dados (qualidade)	X		X			X			X	X								
13. Digitalização e estruturação de coleções																		X
F. Padrões de metadados mencionados no artigo																		
1. ABCD																X		X
2. EML	X			X	X			X									X	X
3. Darwin Core	X				X			X										X
4. GMP (GBIF metadata profile)															X	X		

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30		
25. European Forest Information System (EFIS)																																
26. MARC																																
27. IEEE Standards for Learning Object Metadata (LOM)																																
28. Não recomenda nenhum padrão ou referência																																
29. Adota padrão próprio e muito específico	X						X																									
H. Termos presentes no artigo																																
a) Biodiversity	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
b) Metadata	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
c) ABCD								X												X										X		
d) Darwin Core			X	X			X	X							X	X	X	X	X	X									X			
e) EML		X		X		X	X								X	X		X			X		X						X			
f) GBIF profile															X	X													X			

	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78
26. MARC								X										
27. IEEE Standards for Learning Object Metadata (LOM)								X										
28. Não recomenda nenhum padrão ou referência		X																
29. Adota padrão próprio e muito específico							X			X								
H. Termos presentes no artigo																		
a) Biodiversity	X	X	X	X	X	X	X	X	X	X	X	X	X		X	X	X	X
b) Metadata	X	X	X	X	X	X	X	X	X	X	X	X	X		X	X	X	X
c) ABCD																		X
d) Darwin Core	X				X													X
e) EML	X			X	X												X	X
f) GBIF profile					X								X			X	X	X

APÊNDICE F. Dados coletados nos artigos científicos – roteiro parte 04

	I. Palavras-chave		J. Assunto do artigo
1	2019	Biodiversity; Internet; Statistical Models	Collaborative species mapping; Information on species distribution; modeling approach supported by an online system
2	2019	Ecology, Essential Biodiversity Variables, LTER-Italy, DEIMS-SDR, metadata analysis, research infrastructure assessment, EML, EMF	Metadata analysis to access Essential Biodiversity Variables (EBVs)
3	2018	Data Collection – Analysis; Fish Communities – Research; Fish Communities – Information Management	Aggregated fish collections
4	2018	big data, biodiversity monitoring, data interoperability, ecological sustainability, environmental policy, global change research, indicators, informatics, metadata, research infrastructures.	'Big Data' approach to building global EBV data products across taxa and spatiotemporal scales, focusing on species distribution and abundance.
5	2018	2020 target; Assessment; Biodiversity; observation network; Biodiversity strategy; Citizen science; Conservation funding; Environmental policy; Evidence-based conservation; Statistical power; Surveillance	Based on metadata on species and habitat monitoring schemes, developed indicators for sampling design, sampling effort, and data analysis to evaluate monitoring practices and improve current practices in sampling design, sampling effort and data analysis
6	2018	LTER network; Metadata completeness; Ecological metadata language (EML); Content Standard for Digital Geographic; Metadata (CSDGM); Information management; DataONE; Collection analysis; Community recommendations; Metadata dialects; Data analysis; Concept occurrence	Compared the number of elements/record for all levels of the recommendation using the z-test for a difference between two means
7	2018	Ascomycota; biodiversity; biogeography; fungaria; herbaria; metadata; symbiosis; Botany	Use herbarium collection to examine endophyte community structure in an ecological and evolutionary context
8	2017	science-policy interface, marine policy, knowledge exchange, biodiversity informatics, knowledge products, evidence-based conservation	Use of biodiversity data to develop and implement relevant management and conservation measures and actions Transforming these marine biodiversity and conservation datasets into knowledge products that convey the information required by policy- and decision-makers
9	2017	algal collection; Algarium Veneticum; Archivio di Studi Adriatici; repository; Darwin core; scientific collections	Digitized specimens and metadata, compiled using Dublin Core and Simple Darwin Core formats.
10	2017	Citizen science Community-based monitoring Databases; Essential biodiversity variables (EBV); Global Biodiversity Information Facility (GBIF) Group on Earth Observations Biodiversity Observation Network (GEO BON)	Essential Biodiversity Variable framework to describe the range of biodiversity data needed to track progress towards global biodiversity targets, and weassess strengths and gaps in geographical and taxonomic coverage in CS and CMB programs.
11	2017	Universities and Colleges ; Biodiversity ; Web Portals ; Metadata ; Databases	Results of digitizing three-fourths of the collection and the procedure of the obtaining new collections in the Herbarium of Moscow

	I. Palavras-chave		J. Assunto do artigo
12	2017	aquatic and terrestrial biodiversity, data sharing, global change, meta-analysis, temporary river, water management	Intermittent River Biodiversity Analysis and Synthesis (IRBAS; http://irbas.cesab.org) project has collated, analyzed, and synthesized data from across the world on the biodiversity and environmental characteristics of intermittent rivers.
13	2017	biodiversity data, data degradation, data quality, maps of ignorance, species distribution models, temporal decay	Conceptual framework for understanding the influence of time on biodiversity data quality
14	2017	data sharing, global biodiversity modeling, global change, habitat destruction, land use	PREDICTS project—Projecting Responses of Ecological Diversity In Changing Terrestrial Systems
15	2017	Darwin Core Archive, sample event, species occurrence, environmental data, ecosystem data, telemetry data, data standardisation, oceanographic data	Darwin Core Archive to sample event in OBIS database
16	2016	metadata; scientific data; data management	Investigation of metadata standards for scientific data-sets and the catalogues and websites of some scientific data repositories and portals
17	2016	Data Integration, Biodiversity Metadata, Data Discovery, Semantic Web	Model metadata with respect to a set of semantic models, specifically designed a new model for the purposes of the LifeWatch Greece project
18	2016	Data curation; dataset; digital archiving; metadata	Improve data description and support better data discovery and use based on similarities and differences between the knowledge organisation methods of national archives and of dataset repositories
19	2016	Collections, Repositories, Registry, institutionCode, collectionCode, Darwin Core Triplet, Authority File	Global Registry of Biodiversity Repositories is an online metadata resource for biodiversity collections
20	2015	Biocollections, identifiers, Globally Unique Identifiers, GUIDs, field collections, legacy collections, linked open data, semantic publishing	Unique identifiers for data associated with biocollections
21	2015	collaboration; culture; data sharing; ecological networks; ILTER; information management; LTER; long term ecological research; metadata; multilingual; Special Feature: International LTER	International Long Term Ecological Research (ILTER) Network and lessons learned to collaborate internationally on projects that depend on data sharing
22	2015	Red panda (<i>Ailurus fulgens</i>); Hindu-Kush Himalaya (HKH); Open access GBIF data; Machine learning; Ensemble model GIS prediction	Predictive model is the first of its kind and allows to assess the red panda distribution based on empirical open access data
23	2015	Biodiversity ecosystem functioning, data postprocessing, data sharing, ecological metadata language, metadata, open science, open source, paper proposals, R, reproducible science, semantic integration, workflow	EFdata (Biodiversity and Ecosystem Functioning Data) download and import data and metadata into R for analysis B

	I. Palavras-chave		J. Assunto do artigo
24	2014	Multimedia capturing, Observation, Annotation, Crowdsourcing, Ontology, Interactive map, Visualization, Event, Path, Mobile multimedia	describe MoM-NOCS, a Framework and a System that support communities with common interests in nature to capture and share multimedia observations of nature objects or events using mobile devices
25	2014	Algorithms; Metadata; Model Documenting; Workflow	storing, documenting, and executing ecological models and analytical procedures
26	2014	application context; critical factors; external data quality; geospatial metadata; quality indicator; spatial data quality	use of spatial data sets for environmental and ecological applications, focusing on user requirements for specified application contexts
27	2014	Biodiversity; Science information; Information needs; Information access; Information providers	survey about biodiversity information needs of environmental researchers, managers, and decision makers
28	2014	Ontology, Biodiversity, Population, Community, Darwin Core, OWL, RDF, Microbial ecology, Sequencing	Biological Collections Ontology (BCO) and the Population and Community Ontology (PCO)
29	2014	Digital Curation; Preservation Metadata; SKOS; Linked Data; Europeana; BioCASE; GBIF	Architecture, deployment and evaluation of the Natural Europe infrastructure allowing the curators to publish, semantically describe and manage the museums' cultural heritage objects, as well as disseminate them to Europeana.eu and BioCASE/GBIF
30	2014	Soil invertebrates; Multi-source datasets; Habitat parameters; Data standardization; Metadata analyses; Distribution patterns	tEdaphobase is a non-commercial data warehouse on soil organisms (up to now including Chilopoda, Collembola, Diplopoda, Enchytraeidae, Gamasina, Lumbricidae, Nematoda, Oribatida) integrated in the Global Biodiversity Information Facility (GBIF) network
31	2014	Computer Programs ; Integration ; Software ; Data Processing ; Check Lists ; Biodiversity ; Internet ; Historical Account ; Crises ; Biological Diversity ; Internet ; Issues in Sustainable Development ; Bioinformatics & Computer Applications	Integrated Publishing Toolkit (IPT) is a software package developed to support biodiversity dataset publication in a common format.
32	2014	Information visualisation; Data quality; Darwin Core Archive; Open source; Biodiversity	DwC-A quality checking tool
33	2013	Biodiversity; Databases, Factual	Assessing the Primary Data Hosted by the Spanish Node of the Global Biodiversity Information Facility (GBIF) and found patterns in the availability and accrual of data that seem to arise naturally from the digitization processes and may provide an indication of the overall fitness-for-use in these data.
34	2013	Aquatic metadata; Ecological databases; Online query tool; Data accessibility; Intellectual property rights; Water Framework Directive	WISER project (Water bodies in Europe: Integrative Systems to assess Ecological status and Recovery) and a designed metadatabase to allow scientists to find the optimal data for their analyses

		I. Palavras-chave	J. Assunto do artigo
35	2012	Data integration; Semantic integration; Data model; Species observations	data model for species observation data
36	2012	Biodiversity; Multidisciplinary Data; Metadata; Ontology; Repository	Data repositories in environmental sciences — concepts, definitions, technical solutions and user requirements with a special focus on Data platforms in integrative biodiversity research
37	2012	Biodiversity, research data, data rescue workflow, reBiND, BioCASE, ABCD	reBiND project workflow for rescuing biodiversity data sets
38	2012	Field notes, notebooks, crowd sourcing, digitization, biodiversity, transcription, text-mining, Darwin Core, Junius Henderson, annotation, taxonomic referencing, natural history, Wikisource, Colorado, species occurrence records	Discover or re-use of field notebooks. Workflow to generate structured outputs while also maintaining links to the original texts.
39	2012	Data mining, taxonomic checklists, Darwin Core Archive	Convert a conventional floristic checklist, written in a standard word processing program, into structured data in the Darwin Core Archive format
40	2012	Biodiversity data; EIA; Biodiversity Data Publishing Framework; GBIF; South Africa; India	EIA Biodiversity Data Publishing Framework', based on the Global Biodiversity Information Facility (GBIF) global standards
41	2012	Digitization, imaging, specimen metadata, natural history collections, biodiversity informatics	Workflow for the mass-digitization of natural history museum collections
42	2012	ABCD, BioCASE, XML, primary biodiversity data networks, GBIF, open access	Biological Collections Access Service (BioCASE) and the access to rich natural history data using complex schemas like ABCD (Access to Biological Collection Data)
43	2011	Mercury search system; Scientific data search; OAI-PMH; jOAI; Data sharing; Metadata; Ecological informatics; Climate change; Environmental informatics; Spatiotemporal data	Mercury, is a open-source metadata harvesting, data discovery, and access system, built for researchers to search for, share and obtain spatiotemporal data used across a range of climate and ecological sciences.
44	2011	Pelagic circumpolar seabird distribution; Open access online databases; GIS (Geographic Information System); Circumpolar seabird colonies; International Polar Year (IPY); Arctic biodiversity; Global Biodiversity Information Facility (GBIF); Data mining synthesis	Model for compilation, quantification and summary of 27 seabird species presence data.
45	2011	Red List, Conservation, Open Source, Biodiversity, Mapping, IUCN, GBIF, Flickr, Geospatial, Google maps, HTML5, JSON, AJAX	Geospatial analysis to ease process of Red Listing taxa
46	2011	Palaeontology, Biodiversity Informatics, Scratchpads, web services, GBIF	System to biological and palaeontological communities by providing a webservice of geological time data via the GBIF Vocabularies site and ensured that the system can adapt to changes in the

I. Palavras-chave		J. Assunto do artigo
		definition of geological time intervals and is capable of querying datasets independently of the format of geological age data used
47	2011 Biodiversity; Data sharing; Ecoinformatics; GBIF; Global change; Macroecology; Metadata; Phytosociology; Releve´ ; Scientific reward	Design and initial content of the Global Index of Vegetation-Plot Databases (GIVD), structure and data quality of databases registered in Global Index of Vegetation-Plot Databases (GIVD)
48	2010 XML schema parser; Metadata editor; EML; Xforms; Xquery; Orbeon	Metadata editing and management system
49	2010 biodiversity; metadata; digitization; science infrastructure	Access to digitized specimen data. A globally coordinated approach to the digitization of a critical mass of scientific specimens and specimen-related data is highly desirable. An action plan of this scale is required to maximize the value of these collections to civil society and to support the advancement of our scientific knowledge globally.
50	2010 Metadata . Metadata management . Metadata discovery . Scientific data search . Reusable search engine, ORNL DAAC	Mercury - metadata harvesting, search and retrieval tool
51	2010 Natural history collections, survey, collections, specimens, specimen data, metadata, digitization, GBIF, biodiversity research.	Survey on the challenges and concerns involved with digitizing natural history specimens
52	2010 metadata repurposing, metadata processing, metadata workflow, metadata management, Excel, OCLC, SOAR, DSpace batch import, Google Summer of Code, MarcEdit, MARC, Dublin Core, Darwin Core	Metadata repurposing and batch processing
53	2010 Natural history collections; collections; specimens; specimen data; metadata; digitization; GBIF; biodiversity research	Strategy and Action Plan for the Mobilization of Natural History Collections Data established by the Global Biodiversity Information Facility (GBIF)
54	2010 Natural history collections; collections; specimens; specimen data; metadata; digitization; GBIF; biodiversity research.	Creating a “Geotaxonomic Index” providing metadata on the number of specimens from a specific geographic region belonging to a specific higher taxonomic category may provide a means to attract the attention of researchers and governments towards relevant non-digitized holdings of the collections and set priorities for their digitization according to the needs of information users outside the taxonomic community
55	2009 Ecoinformatics; Ecological databases; Environmental monitoring; Taxonomic databases; U.S. National Coastal Assessment	Review ways in which the new discipline of ecoinformatics is changing how environmental monitoring data are managed, synthesized, and analyzed
56	2009 Protected areas; Conservation achievements; Management effectiveness; Monitoring and evaluation; Open-access data; Databases; Metadata; World Wide Web; Data policies; Data sharing; Scientific competitiveness; Standardization	Discrepancies with sharing data. Extensive online search and review of existing monitoring data from freely accessible online databases for its use in an assessment of conservation achievements in a larger sample of protected areas

		I. Palavras-chave	J. Assunto do artigo
57	2009	Data sharing; Geographic web portal; Virtual globes; Geobrowser; Spatiotemporal visualization	Web-based application that facilitates the sharing of scientific data within a research network using the now-common "virtual globe" in combination with advanced visualization methods designed for geographically distributed scientific data
58	2009	Natural history, Taxonomy, Metadata discovery, Ontologies, Natural language processing	Automated discovery of metadata from textual object databases: first, the automatic detection of new metadata in existing free-text database columns, and second, the discovery of new ontological relations between metadata elements
59	2008	Digital libraries; Ecological data management; Data integration; User evaluation	Digital-library-based environment developed to support the integration, dissemination and exploration of ecological data
60	2008	Environmental Sciences; Biology; Ecology	Ontologies to enhance ecological information management
61	2007	Biodiversity, Information Integration, TaiBIF, TaiBNET, GBIF	The information of Catalog of Life, specimens, and alien species were integrated by the Darwin core standard and allowed the biodiversity information of Taiwan to connect with global databases/biodiversity databases in Taiwan
62	2007	Species at risk; Decision support system; Database management system; Data mining; Ecoinformatics; Knowledge management	WILDSPACE Decision Support System to support information requirements for the Species at Risk Act (SARA)
63	2006	Biodiversity information system; Content-based image retrieval; OAI	Digital library framework for biodiversity information systems. Digital library (DL) architecture for managing heterogeneous data about living beings and their ecosystems
64	2006	analytic web; eddy covariance; EML; metadata; process; synthesis; SciWalker; XML.	Prototype software tool SciWalker. Support system for both the synthesis and the validation of ecological data sets. Formal representation, an "analytic web," that aids both producers and consumers of data sets by providing complete and precise definitions of scientific processes used to process raw and derived data set
65	2006	ecoinformatics, data integration, data sharing, metadata, ontology, scientific workflows, semantics Abstract Bioinformatics	Describe informatics frameworks for ecology, from subject-specific data warehouses, to generic data collections that use detailed metadata descriptions and formal ontologies to catalog and cross-reference information.
66	2005	EFIS; Forestry; Information systems; Information services; Resource Discovery; Exploratory data analysis; Visualisation Toolkit	European Forest Information and Communication System (EFICS)
67	2004	Biodiversity, China, Information sharing, Metadata, Standard	China's core biodiversity metadata standard
68	2003	Metadata; GIS; Geographic headings; MARC; Museum specimens; Darwin Core	Overview of major schemes in vogue related to natural history, and discusses a project that enhanced MARC metadata to provide taxonomic and geospatial fields

I. Palavras-chave		J. Assunto do artigo
69	2003 Archipelago Sea, Clearing House Mechanism, Convention on Biological Diversity, Georeferencing, Geospatiality, Information, Information technology	(1) process and provide existing biodiversity metadata, data and information, especially for environmental officials and other decision-makers, (2) test the possibilities of using a geospatially oriented information network to serve as an intermediary of information to target audiences, and (3) assess the methods needed to transform local biodiversity metadata and data into information usable in an information network.
70	2000 biodiversity, China, Clearing-House Mechanism, database, information networks	Problems in construction of the Biodiversity Clearing-House Mechanism in China, such as insufficient funding, unreasonable organizational mechanism, and low-level information sharing. Future needs for the CHM construction were discussed
71	1999 biodiversity, metadata, database, Clearing-House Mechanism	China National Biodiversity Information Query System (CNBIQS)
72	2017 FAIR data, Interoperability, Data integration, Semantic web, Linked data, REST	set of resource-oriented Web design patterns for data discovery, accessibility, transformation, and integration that can be implemented by any general- or specialpurpose repository as a means to assist users in finding and reusing their data holdings.
73	2015 Archives; Evolutionary biology; Public policy; Data processing; Reproducibility; Science policy; Scientific publishing; Evolutionary genetics	Policies that mandate public data archiving (PDA) successfully increase accessibility to data underlying scientific publications. However, is the data quality sufficient to allow reuse and reanalysis?
74	2015 metadata; databases; remote sensing; databases; in situ observations; field spectroscopy; marine	Spectroscopy metadata standards. Approach to developing robust metadata standards for specific applications that serves to ensure a high level of reliability and interoperability for a spectroscopy dataset.
75	2014 national reporting; data accuracy; environmental indicators; sustainability; forestry	Evaluating the applicability and usefulness of existing national environmental data sources for reporting, including the compatibility of attribute information, spatial data resolution and coverage, and extending existing data through analysis.
76	2014 Biodiversity informatics, database, harmonization, nomenclature, plant names, synonymity	anArchive taxonomic Checklist, an online synonymized list of botanical species names, developed to support the botanical data banking and vegetation analysis
77	2011 Biology	GBIF data publishing framework - a systematic, standard approach to the formal definition and public disclosure of data - in the context of biodiversity data
78	2004 Biological Information Standards	Biological informatics already has established some standards for sharing data, from nomenclature to exchange protocols. This paper presents a brief introduction of those data standards.

APÊNDICE G. Lista de artigos elegíveis para análise

	ANO	Título
1	2019	A fundamental difference between macrobiota and microbial eukaryotes: protistan plankton has a species maximum in the freshwater-marine transition zone of the Baltic Sea
2	2019	BioModelos: A collaborative online system to map species distributions
3	2019	Checklist of terrestrial Parasitengona mites in Fennoscandia with new species- and distribution records (Acariformes: Prostigmata)
4	2019	Evidence of the environmental impact of noise pollution on biodiversity: a systematic map protocol
5	2019	Feeding essential biodiversity variables (EBVs): Actual and potential contributions from LTER-Italy
6	2019	The relative effectiveness of seed addition methods for restoring or re-creating species rich grasslands: a systematic review protocol.(Report)
7	2018	A New Daily Observational Record from Grytviken, South Georgia: Exploring Twentieth-Century Extremes in the South Atlantic
8	2018	A survey of digitized data from U.S. fish collections in the iDigBio data aggregator
9	2018	AgroPortal: A vocabulary and ontology repository for agronomy
10	2018	An updated checklist of the Tenebrionidae sec. Bousquet et al. 2018 of the Algodones Dunes of California, with comments on checklist data practices
11	2018	Benthic macroinvertebrate diversity in the middle doce River Basin, Brazil
12	2018	BioTIME: A database of biodiversity time series for the Anthropocene.(Report)
13	2018	Building essential biodiversity variables (EBVs) of species distribution and abundance at a global scale
14	2018	EML: An I/O scheduling algorithm in large-scale-application environments
15	2018	Evaluating and benchmarking biodiversity monitoring: Metadata-based indicators for sampling design, sampling effort and data analysis.(Report)
16	2018	Generic names and mislabeling conceal high species diversity in global fisheries markets.(Report)
17	2018	GLOSSary: the GLObal Ocean 16S subunit web accessible resource
18	2018	The global latitudinal diversity gradient pattern in spiders.(Report)
19	2018	The influence of community recommendations on metadata completeness
20	2018	Towards robust and repeatable sampling methods in eDNA-based studies
21	2018	Using collections data to infer biogeographic, environmental, and host structure in communities of endophytic fungi
22	2017	A world of sequences: can we use georeferenced nucleotide databases for a robust automated phylogeography?
23	2017	An ontology-based search engine for digital reconstructions of neuronal morphology

24	2017	Blueprints of Effective Biodiversity and Conservation Knowledge Products That Support Marine Policy
25	2017	Building a Natural and Cultural Heritage Repository for the Storage and Dissemination of Knowledge: The Algae Veneticum and the Archivio di Studi Adriatici Case Study
26	2017	Contribution of citizen science towards international biodiversity monitoring
27	2017	Digital Herbarium of Moscow State University: The Largest Russian Biodiversity Database.(Report)
28	2017	Ecological and Functional Traits in 99 Bird Species over a Large-Scale Gradient in Germany
29	2017	Interoperability and FAIRness through a novel combination of Web technologies
30	2017	IRBAS : An online database to collate, analyze, and synthesize data on the biodiversity and ecology of intermittent rivers worldwide
31	2017	Mobilization and integration of bacterial phenotypic data--Enabling next generation biodiversity analysis through the BacDive metadatabase.(Report)
32	2017	Responses of fish and invertebrates to floods and droughts in Europe.(Report)
33	2017	Temporal degradation of data limits biodiversity research
34	2017	The database of the PREDICTS (Projecting Responses of Ecological Diversity In Changing Terrestrial Systems) project
35	2017	The Genomic Observatories Metadatabase (GeOMe): A new repository for field and sampling event metadata associated with genetic samples
36	2017	Toward a new data standard for combined marine biological and environmental datasets - expanding OBIS beyond species occurrences
37	2016	A comparison of the gut microbiome between long-term users and non-users of proton pump inhibitors.(Report)
38	2016	Aerobiology Over Antarctica - A New Initiative for Atmospheric Ecology
39	2016	Aggregate control of scientific data
40	2016	Eliciting and Representing High-Level Knowledge Requirements to Discover Ecological Knowledge in Flower-Visiting Data
41	2016	LifeWatch Greece data-services: Discovering Biodiversity Data using Semantic Web Technologies
42	2016	Mitogenome metadata: current trends and proposed standards
43	2016	Organisation and description of datasets
44	2016	The Global Registry of Biodiversity Repositories: A Call for Community Curation
45	2015	Community Next Steps for Making Globally Unique Identifiers Work for Biocollections Data
46	2015	Computer-assisted melanoma diagnosis: a new integrated system
47	2015	Fostering ecological data sharing: collaborations in the International Long Term Ecological Research Network
48	2015	Measuring energy consumption using EML (energy measurement library)
49	2015	Rapid multi-nation distribution assessment of a charismatic conservation species using open access ensemble model GIS predictions: Red panda (<i>Ailurus fulgens</i>) in the Hindu-Kush Himalaya region
50	2015	rBEFdata: documenting data exchange and analysis for a collaborative data management platform

51	2015	Standardizing metadata and taxonomic identification in metabarcoding studies
52	2015	What Drives Academic Data Sharing?
53	2014	A crowdsourcing framework for the management of mobile multimedia nature observations
54	2014	Documenting, storing, and executing models in Ecology: A conceptual framework and real implementation in a global change monitoring program
55	2014	Evaluating the fitness for use of spatial data sets to promote quality in ecological assessment and monitoring
56	2014	Facilitating Access to Biodiversity Information: A Survey of Users' Needs and Practices
57	2014	Meeting report: advancing practical applications of biodiversity ontologies
58	2014	Metadata management, interoperability and Linked Data publishing support for Natural History Museums
59	2014	The Edaphobase project of GBIF-Germany-A new online soil-zoological data warehouse
60	2014	The GBIF Integrated Publishing Toolkit: Facilitating the Efficient Publishing of Biodiversity Data on the Internet
61	2014	Vesper: Visualising species archives
62	2013	Assessing the Primary Data Hosted by the Spanish Node of the Global Biodiversity Information Facility (GBIF)
63	2013	Incorporating clinical metadata with digital image features for automated identification of cutaneous melanoma
64	2013	Publishing and sharing library resources with the ABCD Site
65	2013	The WISER metadatabase: the key to more than 100 ecological datasets from European rivers, lakes and coastal waters
66	2012	A dataset from bottom trawl survey around Taiwan
67	2012	A semantically integrated, user-friendly data model for species observation data
68	2012	Conserving and promoting evenness: organic farming and fire-based wildland management as case studies
69	2012	Data platforms in integrative biodiversity research
70	2012	Efficient rescue of threatened biodiversity data using reBiND workflows
71	2012	From documents to datasets: A MediaWiki-based method of annotating and extracting species observations in century-old field notebooks
72	2012	From text to structured data: Converting a word-processed floristic checklist into Darwin Core Archive format
73	2012	Improving access to biodiversity data for, and from, EIAs - a data publishing framework built to global standards
74	2012	No specimen left behind: Industrial scale digitization of natural history collections
75	2012	The ABCD of primary biodiversity data access
76	2012	The Darwin Core extension for genebanks opens up new opportunities for sharing genebank datasets
77	2012	Web Approach for Ontology-Based Classification, Integration, and Interdisciplinary Usage of Geoscience Metadata
78	2011	A Survey of BiodiversityMetadata Standards

79	2011	Data sharing and retrieval using OAI-PMH
80	2011	Predictions of 27 Arctic pelagic seabird distributions using public environmental variables, assessed with colony data: a first digital IPY and GBIF open access synthesis platform
81	2011	Supporting Red List threat assessments with GeoCAT: geospatial conservation assessment tool
82	2011	The future of the past in the present: biodiversity informatics and geological time
83	2011	The Global Index of Vegetation-Plot Databases (GIVD): a new resource for vegetation science
84	2010	A flexible online metadata editing and management system
85	2010	Leveraging the fullest potential of scientific collections through digitisation
86	2010	Mercury: reusable metadata management, data discovery and access system
87	2010	Natural History Specimen Digitization: Challenges and Concerns
88	2010	Optimizing Workflow through Metadata Repurposing and Batch Processing
89	2010	Summary of Recommendations of the GBIF Task Group on the Global Strategy and Action Plan for the Digitisation of Natural History Collections
90	2010	Using geographical and taxonomic metadata to set priorities in specimen digitization
91	2009	Beyond data management: how ecoinformatics can benefit environmental monitoring programs
92	2009	Improving the assessment and monitoring of forest biodiversity
93	2009	Multi-level discrepancies with sharing data on protected areas: What we have and what we need for the global village
94	2009	SFMN GeoSearch: An interactive approach to the visualization and exchange of point-based ecological data
95	2009	Towards a Generalized Architecture for the Integration of Tools in LMSs
96	2009	Weaving a New Fabric of Natural History
97	2008	A digital library environment for integrating, disseminating and exploring ecological data
98	2008	Advancing ecological research with ontologies
99	2007	Integration of Biodiversity Databases in Taiwan and Linkage to Global Databases
100	2007	Species at risk: Data and knowledge management within the WILDSPACEa"cents Decision Support System.(Report)
101	2006	A digital library framework for biodiversity information systems
102	2006	Analytic Webs Support the Synthesis of Ecological Data Sets
103	2006	The New Bioinformatics: Integrating Ecological Data from the Gene to the Biosphere
104	2005	The European Forest Information System -- an Internet based interface between information providers and the user community.(Report)
105	2004	The establishment of China's core biodiversity metadata standard
106	2003	Analysis of molecular data of <i>Arabidopsis thaliana</i> (L.) Heynh. (Brassicaceae) with Geographical Information Systems (GIS)

107	2003	Darwin and MARC: a voyage of metadata discovery
108	2003	Geospatially structured biodiversity information as a component of a regional biodiversity clearing house
109	2001	Biodiversity informatics, Biota/FAPESP and the future a personal view
110	2000	Biodiversity Clearing-House Mechanism in China: present status and future needs
111	1999	China National Biodiversity Information Query System

APÊNDICE H. Artigos científicos excluídos por não tratar do uso de metadados para biodiversidade

	ano	Título	Assunto
1	2019	A fundamental difference between macrobiota and microbial eukaryotes: protistan plankton has a species maximum in the freshwater-marine transition zone of the Baltic Sea	Confiabilidade dos dados considerando a disponibilidade de metadados. Caso muito específico sobre uma espécie e sem discussão real de metadados
2	2019	Evidence of the environmental impact of noise pollution on biodiversity: a systematic map protocol	avaliação de impactos e poluição sonora
3	2019	The relative effectiveness of seed addition methods for restoring or re-creating species rich grasslands: a systematic review protocol.(Report)	agricultura - revisão da literatura para obter dados sobre plantio de sementes
4	2018	A New Daily Observational Record from Grytviken, South Georgia: Exploring Twentieth-Century Extremes in the South Atlantic	Mudanças climáticas e impactos sobre a biodiversidade. Análise da variação de temperatura
5	2018	AgroPortal: A vocabulary and ontology repository for agronomy	agricultura. vocabulário controlado para agricultura - ontologia
6	2018	EML: An I/O scheduling algorithm in large-scale-application environments	EML (equal-length multi-level algorithm)
7	2018	Generic names and mislabeling conceal high species diversity in global fisheries markets.(Report)	restreabilidade de pescado
8	2018	GLOSSary: the GLocal Ocean 16S subunit web accessible resource	Genoma, RNA
9	2018	The global latitudinal diversity gradient pattern in spiders.(Report)	estudo muito específico sobre gradientes de ocorrência de aracnídeos. Não trata de metadados
10	2018	Towards robust and repeatable sampling methods in eDNA-based studies	DNA genoma
11	2017	A world of sequences: can we use georeferenced nucleotide databases for a robust automated phylogeography?	genoma, sequencias de DNA e RNA . espacialização de dados de genoma para análise filogeografica.

12	2017	An ontology-based search engine for digital reconstructions of neuronal morphology	neuociência e biomedicina
13	2017	Interoperability and FAIRness through a novel combination of Web technologies	interoperabilidade com princípios FAIR, utilizando várias tecnologias, não trata do uso de metadados para biodiversidade
14	2017	Mobilization and integration of bacterial phenotypic data--Enabling next generation biodiversity analysis through the BacDive metadatabase.(Report)	digitalização de base dados base de dados sobre bactérias - sequenciamento de material genético
15	2017	Responses of fish and invertebrates to floods and droughts in Europe.(Report)	Revisão estatística sobre modelagem hidrológica e enchentes
16	2017	The Genomic Observatories Metadatabase (GeOMe): A new repository for field and sampling event metadata associated with genetic samples	genoma
17	2016	A comparison of the gut microbiome between long-term users and non-users of proton pump inhibitors.(Report)	genoma
18	2016	Aerobiology Over Antarctica - A New Initiative for Atmospheric Ecology	Sugestão de método para registro de metadados referentes a sequenciamento genético
19	2016	Eliciting and Representing High-Level Knowledge Requirements to Discover Ecological Knowledge in Flower-Visiting Data	Método de análise de dados
20	2016	Mitogenome metadata: current trends and proposed standards	genoma
21	2015	Computer-assisted melanoma diagnosis: a new integrated system	melanoma
22	2015	Measuring energy consumption using EML (energy measurement library)	EML (energy measurement library)
23	2015	Standardizing metadata and taxonomic identification in metabarcoding studies	sequenciamento genético
24	2015	What Drives Academic Data Sharing?	comportamento dos pesquisadores e instituições no compartilhamento de dados
25	2013	Incorporating clinical metadata with digital image features for automated identification of cutaneous melanoma	genética e biomedicina
26	2013	Publishing and sharing library resources with the ABCD Site	OMS medicina e saúde
27	2012	Conserving and promoting evenness: organic farming and fire-based wildland management as case studies	agricultura
28	2012	The Darwin Core extension for genebanks opens up new opportunities for sharing genebank datasets	genética e bancos de dados de genoma
29	2012	Web Approach for Ontology-Based Classification, Integration, and Interdisciplinary Usage of Geoscience Metadata	Padrão DIF, sem relação com biodiversidade. Sobre organização de dados espaciais em busca de prover compressão semântica

30	2009	Improving the assessment and monitoring of forest biodiversity	Indicadores ecológicos
31	2009	Towards a Generalized Architecture for the Integration of Tools in LMSs	Educational modeling languages - EML e Learning Management Systems(LMSs)
32	2003	Analysis of molecular data of <i>Arabidopsis thaliana</i> (L.) Heynh. (Brassicaceae) with Geographical Information Systems (GIS)	sequencia de DNA

APÊNDICE I. Artigos científicos selecionados para análise sobre uso de metadados

	ano	Título	Periódico	Identificador
1	2019	BioModelos: A collaborative online system to map species distributions	Publicado em: PLoS One, Mar 2019, Vol.14(3), p.e0214522	E-ISSN: 19326203 , DOI: 10.1371/journal.pone.0214522
2	2019	Checklist of terrestrial Parasitengona mites in Fennoscandia with new species- and distribution records (Acariformes: Prostigmata)	Publicado em: Biodiversity Data Journal, 2019, Vol.7	E-ISSN: 13142828 , DOI: 10.3897/BDJ.7.e36094
3	2019	Feeding essential biodiversity variables (EBVs): Actual and potential contributions from LTER-Italy	Publicado em: Nature Conservation, 3 May 2019, Vol.34, pp.477-503	ISSN: 13146947 , E-ISSN: 13143301 , DOI: 10.3897/natureconservation.34.30735
4	2018	A survey of digitized data from U.S. fish collections in the iDigBio data aggregator	Publicado em: PLoS One, Dec 2018, Vol.13(12), p.e0207636	E-ISSN: 19326203 , DOI: 10.1371/journal.pone.0207636
5	2018	An updated checklist of the Tenebrionidae sec. Bousquet et al. 2018 of the Algodones Dunes of California, with comments on checklist data practices	Publicado em: Biodiversity Data Journal, 2018(6)	E-ISSN: 1314-2828 , DOI: 10.3897/BDJ.6.e24927 , PMID: 29942173
6	2018	Benthic macroinvertebrate diversity in the middle doce River Basin, Brazil	Publicado em: Data, 2018, Vol.3(2)	E-ISSN: 23065729 , DOI: 10.3390/data3020017
7	2018	BioTIME: A database of biodiversity time series for the Anthropocene.(Report)	Publicado em: Global Ecology and Biogeography, 2018, Vol.27(7), p.760(27)	ISSN: 1466-822X , DOI: 10.1111/geb.12729
8	2018	Building essential biodiversity variables (EBVs) of species distribution and abundance at a global scale	Publicado em: Biological Reviews, February 2018, Vol.93(1), pp.600-625	ISSN: 1464-7931 , E-ISSN: 1469-185X , DOI: 10.1111/brv.12359
9	2018	Evaluating and benchmarking biodiversity monitoring: Metadata-based indicators for sampling design, sampling effort and data analysis.(Report)	Publicado em: Ecological Indicators, 2018, Vol.85, p.624	ISSN: 1470-160X , DOI: 10.1016/j.ecolind.2017.11.012

10	2018	The influence of community recommendations on metadata completeness	Publicado em: Ecological Informatics, January 2018, Vol.43, pp.38-51	ISSN: 1574-9541 , DOI: 10.1016/j.ecoinf.2017.09.005
11	2018	Using collections data to infer biogeographic, environmental, and host structure in communities of endophytic fungi	Publicado em: Mycologia, 2018, Vol.110(1), pp.47-62	E-ISSN: 1557-2536 , PMID: 29863996 Version:1 , DOI: 10.1080/00275514.2018.1442078
12	2017	Blueprints of Effective Biodiversity and Conservation Knowledge Products That Support Marine Policy	Frontiers in Marine Science, Volume 4, pp.96-112	ISSN=2296-7745 DOI: 10.3389/fmars.2017.00096
13	2017	Building a Natural and Cultural Heritage Repository for the Storage and Dissemination of Knowledge: The Algarium Veneticum and the Archivio di Studi Adriatici Case Study	Publicado em: Journal of Library Metadata, 03 April 2017, Vol.17(2), pp.111-125	ISSN: 1938-6389 , E-ISSN: 1937-5034 , DOI: 10.1080/19386389.2017.1355165
14	2017	Contribution of citizen science towards international biodiversity monitoring	Publicado em: Biological Conservation, September 2017, Vol.213, pp.280-294	ISSN: 0006-3207 , DOI: 10.1016/j.biocon.2016.09.004
15	2017	Digital Herbarium of Moscow State University: The Largest Russian Biodiversity Database.(Report)	Publicado em: Biology Bulletin, 2017, Vol.44(6), p.584(7)	ISSN: 1062-3590 , DOI: 10.1134/S1062359017060103
16	2017	Ecological and Functional Traits in 99 Bird Species over a Large-Scale Gradient in Germany	Publicado em: Data, Jun 2017, Vol.2(2)	E-ISSN: 23065729 , DOI: 10.3390/data2020012
17	2017	IRBAS : An online database to collate, analyze, and synthesize data on the biodiversity and ecology of intermittent rivers worldwide	Publicado em: Ecology and Evolution, Feb 2017, pp.815-823	E-ISSN: 20457758 , DOI: 10.1002/ece3.2679
18	2017	Temporal degradation of data limits biodiversity research	Publicado em: Ecology and Evolution, Sep 2017, pp.6863-6870	E-ISSN: 20457758 , DOI: 10.1002/ece3.3259
19	2017	The database of the PREDICTS (Projecting Responses of Ecological Diversity In Changing Terrestrial Systems) project	Publicado em: Ecology and Evolution, Jan 2017, pp.145-188	E-ISSN: 20457758 , DOI: 10.1002/ece3.2579
20	2017	Toward a new data standard for combined marine biological and environmental datasets - expanding OBIS beyond species occurrences	Publicado em: Biodiversity Data Journal, 2017(5)	E-ISSN: 1314-2828 , DOI: 10.3897/BDJ.5.e10989 , PMCID: 5345125 , PMID: 28325978
21	2016	Aggregate control of scientific data	Publicado em: Archives and Records, 02 January 2016, Vol.37(1), pp.53-64	ISSN: 2325-7962 , E-ISSN: 2325-7989 , DOI: 10.1080/23257962.2016.1145578
22	2016	LifeWatch Greece data-services: Discovering Biodiversity Data using Semantic Web Technologies	Publicado em: Biodiversity Data Journal, 2016(4)	E-ISSN: 1314-2828 , DOI: 10.3897/BDJ.4.e8443 , PMCID: 5136645 , PMID: 27932908
23	2016	Organisation and description of datasets	Publicado em: Archives and Manuscripts, 03 May 2016, Vol.44(2), pp.73-85	ISSN: 0157-6895 , E-ISSN: 2164-6058 , DOI: 10.1080/01576895.2016.1179585

24	2016	The Global Registry of Biodiversity Repositories: A Call for Community Curation	Publicado em: Biodiversity data journal, 2016(4), pp.e10293	ISSN: 1314-2828 , PMID: 27660523 Version:1 , DOI: 10.3897/BDJ.4.e10293
25	2015	Community Next Steps for Making Globally Unique Identifiers Work for Biocollections Data	ZooKeys 494: 133–154 (2015)	doi: 10.3897/zookeys.494.9352
26	2015	Fostering ecological data sharing: collaborations in the International Long Term Ecological Research Network	Publicado em: Ecosphere, 2015 Oct, Vol.6(10)	ISSN: 2150-8925 , DOI: 10.1890/ES14-00281.1
27	2015	Rapid multi-nation distribution assessment of a charismatic conservation species using open access ensemble model GIS predictions: Red panda (<i>Ailurus fulgens</i>) in the Hindu-Kush Himalaya region	Publicado em: Biological Conservation, January 2015, Vol.181, pp.150-161	ISSN: 0006-3207 , DOI: 10.1016/j.biocon.2014.10.007
28	2015	rBEFdata: documenting data exchange and analysis for a collaborative data management platform	Publicado em: Ecology And Evolution, 2015 Jul(14), pp.2890-2897	ISSN: 2045-7758 , DOI: 10.1002/ece3.1547
29	2014	A crowdsourcing framework for the management of mobile multimedia nature observations	Publicado em: International Journal of Pervasive Computing and Communications, 2014, Vol.10(3), pp.216-238	ISSN: 1742-7371
30	2014	Documenting, storing, and executing models in Ecology: A conceptual framework and real implementation in a global change monitoring program	Publicado em: Environmental Modelling & Software, 2014 Feb, Vol.52, pp.192-199	ISSN: 1364-8152 , DOI: 10.1016/j.envsoft.2013.10.027
31	2014	Evaluating the fitness for use of spatial data sets to promote quality in ecological assessment and monitoring	Publicado em: International Journal Of Geographical Information Science, 2014 Nov 2, Vol.28(11), pp.2356-2371	ISSN: 1365-8816 , DOI: 10.1080/13658816.2014.924627
32	2014	Facilitating Access to Biodiversity Information: A Survey of Users' Needs and Practices	Publicado em: Environmental Management, March 2014, Vol.53(3), pp.690-701	ISSN: 0364-152X , E-ISSN: 1432-1009 , DOI: 10.1007/s00267-014-0229-7
33	2014	Meeting report: advancing practical applications of biodiversity ontologies	Publicado em: Standards in Genomic Sciences, 2014, Vol.9, p.17-17	E-ISSN: 1944-3277 , DOI: 10.1186/1944-3277-9-17 , PMCID: 4334987
34	2014	Metadata management, interoperability and Linked Data publishing support for Natural History Museums	Publicado em: International Journal on Digital Libraries, 2014, Vol.14(3), pp.127-140	ISSN: 1432-5012 , E-ISSN: 1432-1300 , DOI: 10.1007/s00799-014-0114-2

35	2014	The Edaphobase project of GBIF-Germany-A new online soil-zoological data warehouse	Publicado em: Applied Soil Ecology, 2014 Nov, Vol.83, pp.3-12	ISSN: 0929-1393 , DOI: 10.1016/j.apsoil.2014.03.021 , E-ISSN: 1873-0272
36	2014	The GBIF Integrated Publishing Toolkit: Facilitating the Efficient Publishing of Biodiversity Data on the Internet	Publicado em: PLoS One, Aug 2014, Vol.9(8), p.e102623	E-ISSN: 19326203 , DOI: 10.1371/journal.pone.0102623
37	2014	Vesper: Visualising species archives	Publicado em: Ecological Informatics, November 2014, Vol.24, pp.132-147	ISSN: 1574-9541 , DOI: 10.1016/j.ecoinf.2014.08.004
38	2013	Assessing the Primary Data Hosted by the Spanish Node of the Global Biodiversity Information Facility (GBIF)	Publicado em: PLoS One, Jan 2013, Vol.8(1), p.e55144	E-ISSN: 19326203 , DOI: 10.1371/journal.pone.0055144
39	2013	The WISER metadatabase: the key to more than 100 ecological datasets from European rivers, lakes and coastal waters	Publicado em: Hydrobiologia, Mar 2013, Vol.704(1), pp.29-38	ISSN: 0018-8158 , E-ISSN: 1573-5117 , DOI: 10.1007/s10750-012-1295-6
40	2012	A dataset from bottom trawl survey around Taiwan	Publicado em: ZooKeys, 2012(198), p.103-109	ISSN: 1313-2989 , E-ISSN: 1313-2970 , DOI: 10.3897/zookeys.198.3032 , PMCID: 3368258 , PMID: 22707908
41	2012	A semantically integrated, user-friendly data model for species observation data	Publicado em: Ecological Informatics, 2012 Mar, Vol.8, pp.1-9	ISSN: 1574-9541 , DOI: 10.1016/j.ecoinf.2011.11.002
42	2012	Data platforms in integrative biodiversity research	Publicado em: Ecological Informatics, 2012 Sep, Vol.11, pp.1-4	ISSN: 1574-9541 , DOI: 10.1016/j.ecoinf.2012.04.001
43	2012	Efficient rescue of threatened biodiversity data using reBiND workflows	Publicado em: Plant Biosystems - An International Journal Dealing with all Aspects of Plant Biology, 01 December 2012, Vol.146(4), pp.752-755	ISSN: 1126-3504 , E-ISSN: 1724-5575 , DOI: 10.1080/11263504.2012.740086
44	2012	From documents to datasets: A MediaWiki-based method of annotating and extracting species observations in century-old field notebooks	Publicado em: ZooKeys, 2012(209), p.235-253	ISSN: 1313-2989 , E-ISSN: 1313-2970 , DOI: 10.3897/zookeys.209.3247 , PMCID: 3406479 , PMID: 22859891
45	2012	From text to structured data: Converting a word-processed floristic checklist into Darwin Core Archive format	Publicado em: PhytoKeys, 2012(9), pp.1-13	E-ISSN: 1314-2003 , PMID: 22371687 Version:1 , DOI: 10.3897/phytokeys.9.2770
46	2012	Improving access to biodiversity data for, and from, EIAs - a data publishing framework built to global standards	Publicado em: Impact Assessment and Project Appraisal, 01 September 2012, Vol.30(3), pp.148-156	ISSN: 1461-5517 , E-ISSN: 1471-5465 , DOI: 10.1080/14615517.2012.705068

47	2012	No specimen left behind: Industrial scale digitization of natural history collections	Publicado em: ZooKeys, 2012, Vol.209, pp.133-146	ISSN: 13132989 , E-ISSN: 13132970 , DOI: 10.3897/zookeys.209.3178
48	2012	The ABCD of primary biodiversity data access	Publicado em: Plant Biosystems - An International Journal Dealing with all Aspects of Plant Biology, 01 December 2012, Vol.146(4), pp.771-779	ISSN: 1126-3504 , E-ISSN: 1724-5575 , DOI: 10.1080/11263504.2012.740085
49	2011	A Survey of Biodiversity Metadata Standards	PNLA Quarterly 75:2 (Winter 2011)	ISSN: 0030-8188
50	2011	Data sharing and retrieval using OAI-PMH	Publicado em: Earth Science Informatics, Mar 2011, pp.1-5	ISSN: 18650473 , E-ISSN: 18650481 , DOI: 10.1007/s12145-010-0073-0
51	2011	Predictions of 27 Arctic pelagic seabird distributions using public environmental variables, assessed with colony data: a first digital IPY and GBIF open access synthesis platform	Publicado em: Marine Biodiversity, 2011, Vol.41(1), pp.141-179	ISSN: 1867-1616 , E-ISSN: 1867-1624 , DOI: 10.1007/s12526-011-0083-2
52	2011	Supporting Red List threat assessments with GeoCAT: geospatial conservation assessment tool	Publicado em: ZooKeys, 2011(150), p.117-126	ISSN: 1313-2989 , E-ISSN: 1313-2970 , DOI: 10.3897/zookeys.150.2109 , PMCID: 3234434 , PMID: 22207809
53	2011	The future of the past in the present: biodiversity informatics and geological time	Publicado em: ZooKeys, 2011(150), p.397-405	ISSN: 1313-2989 , E-ISSN: 1313-2970 , DOI: 10.3897/zookeys.150.2350 , PMCID: 3234446 , PMID: 22207819
54	2011	The Global Index of Vegetation-Plot Databases (GIVD): a new resource for vegetation science	Publicado em: Journal Of Vegetation Science, 2011 Aug, Vol.22(4), pp.582-597	ISSN: 1100-9233 , DOI: 10.1111/j.1654-1103.2011.01265.x
55	2010	A flexible online metadata editing and management system	Publicado em: Ecological Informatics, 2010, Vol.5(1), pp.26-31	ISSN: 1574-9541 , DOI: 10.1016/j.ecoinf.2009.08.012
56	2010	Leveraging the fullest potential of scientific collections through digitization	Publicado em: Biodiversity Informatics, 01 October 2010, Vol.7(2)	E-ISSN: 1546-9735 , DOI: 10.17161/bi.v7i2.3987
57	2010	Mercury: reusable metadata management, data discovery and access system	Publicado em: Earth Science Informatics, Jun 2010, Vol.3(1-2), pp.87-94	ISSN: 18650473 , E-ISSN: 18650481 , DOI: 10.1007/s12145-010-0050-7
58	2010	Natural History Specimen Digitization: Challenges and Concerns	Publicado em: Biodiversity Informatics, 01 October 2010, Vol.7(2)	E-ISSN: 1546-9735 , DOI: 10.17161/bi.v7i2.3992

59	2010	Optimizing Workflow through Metadata Repurposing and Batch Processing	Publicado em: Journal of Library Metadata, 17 November 2010, Vol.10(4), pp.219-237	ISSN: 1938-6389 , E-ISSN: 1937-5034 , DOI: 10.1080/19386389.2010.524862
60	2010	Summary of Recommendations of the GBIF Task Group on the Global Strategy and Action Plan for the Digitisation of Natural History Collections	Publicado em: Biodiversity Informatics, 01 October 2010, Vol.7(2)	E-ISSN: 1546-9735 , DOI: 10.17161/bi.v7i2.3989
61	2010	Using geographical and taxonomic metadata to set priorities in specimen digitization	Publicado em: Biodiversity Informatics, 01 October 2010, Vol.7(2)	E-ISSN: 1546-9735 , DOI: 10.17161/bi.v7i2.3988
62	2009	Beyond data management: how ecoinformatics can benefit environmental monitoring programs	Publicado em: Environmental Monitoring and Assessment, Mar 2009, Vol.150(1-4), pp.227-35	ISSN: 0167-6369 , E-ISSN: 1573-2959 , DOI: 10.1007/s10661-008-0675-x
63	2009	Multi-level discrepancies with sharing data on protected areas: What we have and what we need for the global village	Publicado em: Journal Of Environmental Management, 2009 Jan, Vol.90(1), pp.8-24	ISSN: 0301-4797 , DOI: 10.1016/j.jenvman.2007.11.001
64	2009	SFMN GeoSearch: An interactive approach to the visualization and exchange of point-based ecological data	Publicado em: Ecological Informatics, 2009 Sep, Vol.4(4), pp.196-205	ISSN: 1574-9541 , DOI: 10.1016/j.ecoinf.2009.07.007
65	2009	Weaving a New Fabric of Natural History	Publicado em: Interdisciplinary Science Reviews, Sept. 2009, Vol.34(2-3), pp.206-223	ISSN: 0308-0188 , DOI: 10.1179/174327909X441117
66	2008	A digital library environment for integrating, disseminating and exploring ecological data	Ecological Informatics, Volume 3, Issues 4–5, October 2008, Pages 295-308	https://doi.org/10.1016/j.ecoinf.2008.09.006
67	2008	Advancing ecological research with ontologies	Publicado em: Trends In Ecology & Evolution, 2008 Mar(3), pp.159-168	ISSN: 0169-5347 , DOI: 10.1016/j.tree.2007.11.007
68	2007	Integration of Biodiversity Databases in Taiwan and Linkage to Global Databases	Publicado em: Data Science Journal, 2007, Vol.6, pp.S2-S10	ISSN: 1683-1470 , DOI: 10.2481/dsj.6.S2
69	2007	Species at risk: Data and knowledge management within the WILDSPACEa"cents Decision Support System.(Report)	Publicado em: Environmental Modelling and Software, April, 2007, Vol.22(4), p.423(8)	ISSN: 1364-8152
70	2006	A digital library framework for biodiversity information systems	International Journal on Digital Libraries (2006) 6(1): 3–17	DOI 10.1007/s00799-005-0124-1
71	2006	Analytic Webs Support the Synthesis of Ecological Data Sets	Analytic Webs Support the Synthesis of Ecological Data Sets. Ecology 87: 1345-1358.	ISSN: 0012-9658

72	2006	The New Bioinformatics: Integrating Ecological Data from the Gene to the Biosphere	Publicado em: 2006, Vol.37, p.519-544	ISSN: 1543-592X , DOI: 10.1146/annurev.ecolsys.37.091305.110031
73	2005	The European Forest Information System -- an Internet based interface between information providers and the user community.(Report)	Publicado em: Computers and Electronics in Agriculture, June, 2005, Vol.47(3), p.185(22)	ISSN: 0168-1699
74	2004	The establishment of China's core biodiversity metadata standard	Publicado em: Biodiversity and Conservation, May 2004, Vol.13(5), pp.1009-1022	ISSN: 09603115 , DOI: 10.1023/B:BIOC.0000014466.91355.0b
75	2003	Darwin and MARC: a voyage of metadata discovery	Publicado em: Library Collections, Acquisitions and Technical Services, 2003, Vol.27(3), pp.291-304	ISSN: 1464-9055 , E-ISSN: 1873-1821 , DOI: 10.1016/S1464-9055(03)00071-X
76	2003	Geospatially structured biodiversity information as a component of a regional biodiversity clearing house	Biodiversity and Conservation 12: 103–120, 2003.	DOI: 10.1023/A:1021238212962
77	2001	Biodiversity informatics, Biota/FAPESP and the future a personal view	Publicado em: Biota Neotropica, January 2001, Vol.1(1-2), pp.1-9	ISSN: 1676-0611 , E-ISSN: 1676-0611 , DOI: 10.1590/S1676-06032001000100004
78	2000	Biodiversity Clearing-House Mechanism in China: present status and future needs	Biodiversity and Conservation 9: 361–378, 2000.	https://doi.org/10.1023/A:1008958902306
79	1999	China National Biodiversity Information Query System	Publicado em: Journal of Environmental Management, May 1999, Vol.56(1), pp.45-59	ISSN: 03014797 , DOI: 10.1006/jema.1999.0266

APÊNDICE J. Artigos excluídos após análise refinada

Nº	ANO	TÍTULO	MOTIVO
2	2019	Checklist of terrestrial Parasitengona mites in Fennoscandia with new species- and distribution records (Acariformes: Prostigmata)	data paper
5	2018	An updated checklist of the Tenebrionidae sec. Bousquet et al. 2018 of the Algodones Dunes of California, with comments on checklist data practices	data paper
6	2018	Benthic macroinvertebrate diversity in the middle doce River Basin, Brazil	<i>data paper</i>
7	2018	BioTIME: A database of biodiversity time series for the Anthropocene.(Report)	<i>data paper</i>
16	2017	Ecological and Functional Traits in 99 Bird Species over a Large-Scale Gradient in Germany	<i>data paper</i>
40	2012	A dataset from bottom trawl survey around Taiwan	<i>data paper</i>
49	2011	A Survey of Biodiversity Metadata Standards	não é artigo científico
77	2001	Biodiversity informatics, Biota/FAPESP and the future a personal view	não trata do uso de metadados

ANEXOS

ANEXO 1.	Classes do padrão ABCD	293
ANEXO 2.	Relação entre Classes e Elementos do padrão ABCD	299
ANEXO 3.	Esquemas do padrão EML	324
ANEXO 4.	Esquemas e elementos do padrão EML.....	328
ANEXO 5.	Lista de TERMOS e CONCEITOS das propriedades apresentadas nos núcleos de um DwC-A.	343
ANEXO 6.	Lista de categorias e elementos que integram o GMP.	362
ANEXO 7.	Artigos resultantes das pesquisas por “metadata use” e “metadata standard use”	371
ANEXO 8.	Lista final de artigos utilizados na coleta de dados sobre o uso de metadados	372
ANEXO 9.	Lista de conjuntos de dados vinculados a instituições globais	378
ANEXO 10.	Lista de conjuntos de dados e instituições vinculados ao Brasil	403
ANEXO 11.	Dados coletados na rede GBIF: metadados dos conjuntos de dados.....	429

ANEXO 1. Classes do padrão ABCD

CLASSES	DESCRIÇÃO
<u>anyUriP</u>	xs:anyURI extended with Preferred attribute
<u>BiologicalStatus</u>	Code for the (propagation-)provenance of the unit.
<u>BotanicalGardenUnit</u>	Elements specific to units from botanical gardens
<u>CodeOfNomenclatureEnum</u>	An enumeration of the Codes of Nomenclature. However, note that this is work in progress, so there may be considerable differences between types with the same names in UBIF/SDD and ABCD.
<u>CollectingAcquisitionSource</u>	Code for the (location-)provenance of the unit. The coding scheme proposed can be used at 2 different levels of detail: either by using the general codes such as 10, 20, 30, 40 or by using the more specific codes such as 11, 12 etc
<u>Contact</u>	Contact details.
<u>ContactP</u>	Contact with preferred flag
<u>ContentMetadata</u>	Metadata referring to the principal source of the entire data collection (thus the metadata scope may be wider than the objects actually contained in the data set). If a history of the data collection (revised or expanded in various projects or at different institutions) exist, this must be reflected in the IPR statements and possibly in the list of Owners.
<u>Country</u>	Details of the country
<u>CultureCollectionUnit</u>	Elements only used for units in culture collections
<u>DateTime</u>	Temporal aspects of an event or activity.
<u>DateTimeISO</u>	The date and time expressed in a way conforming to a subset of ISO 8601
<u>decimalLatitudeDataType</u>	from http://rs.tdwg.org/dwc/basetypes.xsd to support compatibility with DwC Spatial Extensions
<u>decimalLongitudeDataType</u>	from http://rs.tdwg.org/dwc/basetypes.xsd to support compatibility with DwC Spatial Extensions
<u>Epithet</u>	A name element dependent on a genus or a species name under the bacteriological, botanical, and zoological code.

CLASSES	DESCRIÇÃO
<u>FAOInstituteCode</u>	FAO Institute Code. The codes consist of the 3-letter ISO 3166 country code of the country where the institute is located plus a three-digit number.
<u>Gathering</u>	Elements describing the event and site of collecting or observing a unit.
<u>GrowthConditionAtomized</u>	Atomized elements describing the growth conditions of the culture unit.
<u>HerbariumUnit</u>	Elements only used for herbarium units
<u>HigherTaxon</u>	Taxon name used to further classify the identification result.
<u>HigherTaxonRankEnum</u>	A taxonomic rank designation above the rank of genus.
<u>Identification</u>	The application of a name (or concept), i.e. the data pertaining to an identification event and the result of the identification itself.
<u>imageSize</u>	picture width and height in pixel
<u>InstitutionCode</u>	An abbreviated designation of an institution.
<u>Label</u>	Collection of language-specific label representations
<u>LabelRepr</u>	Language-specific simple label, using simple formatted text
<u>LegalStatement</u>	Text, optional Details (both free-form text) and optional website URL and resource URI.
<u>LegalStatements</u>	A sequence of statements related to Intellectual Property Rights, credit and acknowledgement.
<u>MeasurementOrFact</u>	General purpose measurement and fact (descriptor) recording.

CLASSES	DESCRIÇÃO
<u>MetadataDescriptionRepr</u>	Language-specific content metadata (title, description, etc.) with *required* language attribute added.
<u>Monomial</u>	A generic or higher taxon name (monomial) under the bacteriological, botanical, viral, and zoological code.
<u>MultimediaObject</u>	Elements to describe digital images, sound or video recordings, etc.
<u>MycologicalUnit</u>	Elements only used for mycological units
<u>NameBacterial</u>	An atomized name under the International Code of Nomenclature of Bacteria
<u>NameBotanical</u>	An atomized scientific name under the International Code of Botanical Nomenclature or the International Code of Nomenclature for Cultivated Plants
<u>NamedArea</u>	An atomized place name (an administrative or geoecological area) with possibility for hierarchical structuring.
<u>NameViral</u>	An atomized scientific name under the International Code of Virus Classification and Nomenclature
<u>NameZoological</u>	An atomized scientific name under the International Code of Zoological Nomenclature
<u>Organization</u>	identifier for the Organization or Organization Unit
<u>PaleontologicalUnit</u>	Elements only used for paleontological units
<u>Permit</u>	Container element for information on permits related to the Gathering event
<u>PersonName</u>	Name of an individual person

CLASSES	DESCRIÇÃO
<u>PGRUnit</u>	Extension for elements only used in plant genetic resource collections. Currently, all elements represent European Search Catalogue for Plant Genetic Resources (EURISCO) descriptors.
<u>positiveDouble</u>	from http://rs.tdwg.org/dwc/basetypes.xsd to support compatibility with DwC Spatial Extensions
<u>RankAbbreviation</u>	The abbreviation of a taxonomic rank as used in (botanical) name strings.
<u>RecordBasisEnum</u>	A standard designator for the nature of the object of the record. "MaterialSample" is added as "A resource describing the physical results of a sampling (or sub-sampling) event. In biological collections, the material sample is typically collected, and either preserved or destructively processed".
<u>Reference</u>	Published reference.
<u>RevisionData</u>	Revision data (creators, dates, revision) for the entire project/data set or individual objects.
<u>Role</u>	The role of an agent or contact.
<u>ScientificName</u>	Name of the taxon identified, formed according to the different Codes of Nomenclature which apply to scientific names.
<u>ScientificNameIdentified</u>	Name of the taxon identified, formed according to the different Codes of Nomenclature which apply to scientific names - with additional elements for suffixes or other expressions commonly used in taxonomic identifications.

CLASSES	DESCRIÇÃO
<u>Sequence</u>	A first proposal for this type of data. Note that this is used in the context of a collection unit and thus automatically refers to that unit. If your unit is a DNA Sample than please use the GGBN extension. This container here is meant to be used e.g. to provide Sequence information related to a specimen only without any further details.
<u>SexCodeEnum</u>	Sex of the organism (Male, Female, Unknown, Not applicable (e.g. for a fungus in a herbarium), MiXed).
<u>spatialFitDataType</u>	from http://rs.tdwg.org/dwc/basetypes.xsd to support compatibility with DwC Spatial Extensions
<u>StratigraphicTerm</u>	Descriptive term. May be taken from a controlled vocabulary maintained by a third party.
<u>StratigraphicTermL</u>	StratigraphicTerm type extended with language attribute.
<u>Stratigraphy</u>	Elements expressing stratigraphy at the gathering site.
<u>String</u>	normalized string required to contain at least 1 character (this removes the xml string anomaly, i. e. either element/attribute may be optional, but if they are required the content may not be an empty string)
<u>String255</u>	normalized string restricted to 1..255 character length (i. e. required, may not be empty string)
<u>String50</u>	normalized string restricted to 1..50 character length to be used for abbreviations (the recommended length of abbreviations is usually much shorter, but 50 characters should be a good margin).

CLASSES	DESCRIÇÃO
<u>StringL</u>	String (i. e. xs:string with minimum length=1) extended with language attribute
<u>StringL255</u>	String255 (i.e. xs:string with length 1-255), extended with language attribute
<u>StringLP</u>	String (i. e. xs:string with minimum length=1) extended with language and preferred attribute
<u>StringLP255</u>	String255 (i.e. xs:string with length 1-255), extended with language and preferred attribute
<u>StringP</u>	String (i. e. xs:string with minimum length=1) extended with preferred attribute
<u>StringP255</u>	String255 (i.e. xs:string with length 1-255), extended with preferred attribute
<u>TaxonIdentified</u>	Scientific or informal name as a result of an identification.
<u>Temperature</u>	Preferably a number, e.g. "-80", "room temperature". Use the attribute "scale" to denote the used temperature scale.
<u>TypeGermplasmStorage</u>	Code for the type of storage of the unit in the collection. If germplasm is maintained under different types of storage, multiple choices are allowed (separated by a semicolon). (Refer to FAO/IPGRI Genebank Standards 1994 for details on storage type.)
<u>Unit</u>	Elements for all unit-level data
<u>ZoologicalUnit</u>	Elements only used for zoological units

ANEXO 2. Relação entre Classes e Elementos do padrão ABCD

CLASSES	DESCRIÇÃO DA CLASSE	ELEMENTOS (*SUBCLASSE/ELEMENTOS - quando houver subclasse)
<u>BotanicalGardenUnit</u>	Elements specific to units from botanical gardens	<u>BotanicalGardenUnit/AccessionLineage</u> <u>BotanicalGardenUnit/AccessionMaterialType</u> <u>BotanicalGardenUnit/AccessionSpecimenNumbers</u> <u>BotanicalGardenUnit/AccessionStatus</u> <u>BotanicalGardenUnit/BreedingSystem</u> <u>BotanicalGardenUnit/Cultivation</u> <u>BotanicalGardenUnit/DonorCategory</u> <u>BotanicalGardenUnit/Hardiness</u> <u>BotanicalGardenUnit/IPEN</u> <u>BotanicalGardenUnit/LocationInGarden</u> <u>BotanicalGardenUnit/Perennation</u> <u>BotanicalGardenUnit/PlantingDate</u> <u>BotanicalGardenUnit/Propagation</u> <u>BotanicalGardenUnit/PropagationHistoryCode</u> <u>BotanicalGardenUnit/ProvenanceCategory</u>
<u>Contact</u>	Contact details. Subclass: - Person - Organization	<u>Contact/Addresses</u> <u>Contact/Addresses/Address</u> <u>Contact/EmailAddresses</u> <u>Contact/EmailAddresses/EmailAddress</u> <u>Contact/Organization</u> <u>Contact/Person</u> <u>Contact/ResourceURIs</u> <u>Contact/ResourceURIs/ResourceURI</u> <u>Contact/ResourceURIs/ResourceURI</u> <u>Contact/Roles/Role</u>

CLASSES	DESCRIÇÃO DA CLASSE	ELEMENTOS (*SUBCLASSE/ELEMENTOS - quando houver subclasse)
		<u>Contact/TelephoneNumbers</u> <u>Contact/TelephoneNumbers/TelephoneNumber</u> <u>Contact/UnformattedValue</u> <u>Contact/WebsiteURL</u>
<u>ContentMetadata</u>	<p>Metadata referring to the principal source of the entire data collection (thus the metadata scope may be wider than the objects actually contained in the data set). If a history of the data collection (revised or expanded in various projects or at different institutions) exist, this must be reflected in the IPR statements and possibly in the list of Owners.</p>	<u>ContentMetadata/Description</u> <u>ContentMetadata/Description/Representation</u> <u>ContentMetadata/DirectAccessURL</u> <u>ContentMetadata/InformationWithheld</u> <u>ContentMetadata/LegalStatements</u> <u>ContentMetadata/LogoURL</u> <u>ContentMetadata/Owners</u> <u>ContentMetadata/Owners/Owner</u> <u>ContentMetadata/RevisionData</u> <u>ContentMetadata/Scope</u> <u>ContentMetadata/Scope/GeoecologicalTerms</u> <u>ContentMetadata/Scope/GeoecologicalTerms/GeoecologicalTerm</u> <u>ContentMetadata/Scope/TaxonomicTerms</u> <u>ContentMetadata/Scope/TaxonomicTerms/TaxonomicTerm</u> <u>ContentMetadata/Version</u> <u>ContentMetadata/Version/DateCreated</u> <u>ContentMetadata/Version/Major</u> <u>ContentMetadata/Version/Minor</u> <u>ContentMetadata/Version/Modifier</u> <u>ContentMetadata/Version/VersionText</u>
<u>Country</u>	Details of the country	<u>Country/DerivedFlag</u> <u>Country/ISO3166Code</u> <u>Country/Name</u>

CLASSES	DESCRIÇÃO DA CLASSE	ELEMENTOS (*SUBCLASSE/ELEMENTOS - quando houver subclasse)
<u>CultureCollectionUnit</u>	Elements only used for units in culture collections	<u>CultureCollectionUnit/Applications</u> <u>CultureCollectionUnit/CultureNames</u> <u>CultureCollectionUnit/CultureNames/CultureName</u> <u>CultureCollectionUnit/FormOfSupply</u> <u>CultureCollectionUnit/Genotype</u> <u>CultureCollectionUnit/GrowthConditions</u> <u>CultureCollectionUnit/GrowthConditionsAtomized</u> <u>CultureCollectionUnit/GrowthConditionsAtomized/GrowthConditionAtomized</u> <u>CultureCollectionUnit/Hazard</u> <u>CultureCollectionUnit/InfrasubspecificName</u> <u>CultureCollectionUnit/Mutant</u> <u>CultureCollectionUnit/OrganismType</u> <u>CultureCollectionUnit/Pathovar</u> <u>CultureCollectionUnit/References</u> <u>CultureCollectionUnit/References/Reference</u> <u>CultureCollectionUnit/SerovarOrSerotype</u> <u>CultureCollectionUnit/Strain</u>
	TDWG root element: container for one or more DataSets.	<u>DataSets</u> <u>DataSets/DataSet</u> <u>DataSets/DataSet/ContentContacts</u> <u>DataSets/DataSet/ContentContacts/ContentContact</u> <u>DataSets/DataSet/DataCenter</u> <u>DataSets/DataSet/DataSetExtension</u> <u>DataSets/DataSet/GUID</u> <u>DataSets/DataSet/ID</u> <u>DataSets/DataSet/Metadata</u> <u>DataSets/DataSet/OtherProviders</u>

CLASSES	DESCRIÇÃO DA CLASSE	ELEMENTOS (*SUBCLASSE/ELEMENTOS - quando houver subclasse)
		<u>DataSets/DataSet/OtherProviders/OtherProvider</u> <u>DataSets/DataSet/ResourceURIs</u> <u>DataSets/DataSet/ResourceURIs/ResourceURI</u> <u>DataSets/DataSet/TechnicalContacts</u> <u>DataSets/DataSet/TechnicalContacts/TechnicalContact</u> <u>DataSets/DataSet/Units</u> <u>DataSets/DataSet/Units/Unit</u>
<u>DateTime</u>	Temporal aspects of an event or activity.	<u>DateTime/DateText</u> <u>DateTime/ISODateTimeBegin</u> <u>DateTime/ISODateTimeEnd</u> <u>DateTime/PeriodExplicit</u> <u>DateTime/TimeOfDayBegin</u> <u>DateTime/TimeOfDayEnd</u> <u>DateTime/TimeZone</u>
<u>Gathering</u>	Elements describing the event and site of collecting or observing a unit.	<u>Gathering/Altitude</u> <u>Gathering/Aspect</u> <u>Gathering/Biotope</u> <u>Gathering/Biotope/ClassificationScheme</u> <u>Gathering/Biotope/MeasurementsOrFacts</u> <u>Gathering/Biotope/MeasurementsOrFacts/MeasurementOrFact</u> <u>Gathering/Biotope/Name</u> <u>Gathering/Biotope/Text</u> <u>Gathering/CoordinateSets</u> <u>Gathering/CoordinateSets/CoordinateSet</u> <u>Gathering/CoordinateSets/CoordinateSet/CoordinatesGrid</u>

CLASSES	DESCRIÇÃO DA CLASSE	ELEMENTOS (*SUBCLASSE/ELEMENTOS - quando houver subclasse)
		<u>Gathering/CoordinateSets/CoordinateSet/CoordinatesGrid/GridCellCode</u> <u>Gathering/CoordinateSets/CoordinateSet/CoordinatesGrid/GridCellSystem</u> <u>Gathering/CoordinateSets/CoordinateSet/CoordinatesGrid/GridQualifier</u> <u>Gathering/CoordinateSets/CoordinateSet/CoordinatesLatLong</u> <u>Gathering/CoordinateSets/CoordinateSet/CoordinatesLatLong/Accuracy</u> <u>Gathering/CoordinateSets/CoordinateSet/CoordinatesLatLong/CoordinatesText</u> <u>Gathering/CoordinateSets/CoordinateSet/CoordinatesLatLong/ErrorDistanceInMeters</u> <u>Gathering/CoordinateSets/CoordinateSet/CoordinatesLatLong/ErrorMethod</u> <u>Gathering/CoordinateSets/CoordinateSet/CoordinatesLatLong/LatitudeDecimal</u> <u>Gathering/CoordinateSets/CoordinateSet/CoordinatesLatLong/LongitudeDecimal</u> <u>Gathering/CoordinateSets/CoordinateSet/CoordinatesLatLong/PointRadiusSpatialFit</u> <u>Gathering/CoordinateSets/CoordinateSet/CoordinatesLatLong/SpatialDatum</u> <u>Gathering/CoordinateSets/CoordinateSet/CoordinatesLatLong/VerbatimLatitude</u> <u>Gathering/CoordinateSets/CoordinateSet/CoordinatesLatLong/VerbatimLongitude</u> <u>Gathering/CoordinateSets/CoordinateSet/CoordinatesUTM</u> <u>Gathering/CoordinateSets/CoordinateSet/CoordinatesUTM/UTMDatum</u> <u>Gathering/CoordinateSets/CoordinateSet/CoordinatesUTM/UTMEasting</u> <u>Gathering/CoordinateSets/CoordinateSet/CoordinatesUTM/UTMNorthing</u> <u>Gathering/CoordinateSets/CoordinateSet/CoordinatesUTM/UTMNS</u> <u>Gathering/CoordinateSets/CoordinateSet/CoordinatesUTM/UTMSubzone</u> <u>Gathering/CoordinateSets/CoordinateSet/CoordinatesUTM/UTMText</u> <u>Gathering/CoordinateSets/CoordinateSet/CoordinatesUTM/UTMZone</u> <u>Gathering/CoordinateSets/CoordinateSet/CoordinatesUTM/UTMZoneFull</u> <u>Gathering/CoordinateSets/CoordinateSet/EPSGID</u> <u>Gathering/CoordinateSets/CoordinateSet/Method</u> <u>Gathering/CoordinateSets/CoordinateSet/Notes</u> <u>Gathering/CoordinateSets/CoordinateSet/References</u>

CLASSES	DESCRIÇÃO DA CLASSE	ELEMENTOS (*SUBCLASSE/ELEMENTOS - quando houver subclasse)
		<u>Gathering/CoordinateSets/CoordinateSet/References/Reference</u> <u>Gathering/CoordinateSets/CoordinateSet/VerificationStatus</u> <u>Gathering/Country</u> <u>Gathering/Date</u> <u>Gathering/Depth</u> <u>Gathering/Details</u> <u>Gathering/EstablishmentMeans</u> <u>Gathering/FootprintSpatialFit</u> <u>Gathering/FootprintWKT</u> <u>Gathering/GatheringAgents</u> <u>Gathering/GatheringAgents/GatheringAgent</u> <u>Gathering/GatheringAgentsText</u> <u>Gathering/GML</u> <u>Gathering/Height</u> <u>Gathering/ID</u> <u>Gathering/LocalityText</u> <u>Gathering/Method</u> <u>Gathering/MultimediaObjects</u> <u>Gathering/MultimediaObjects/MultimediaObject</u> <u>Gathering/NamedAreas</u> <u>Gathering/NamedAreas/NamedArea</u> <u>Gathering/NamedPlaceRelations</u> <u>Gathering/NamedPlaceRelations/NamedPlaceRelation</u> <u>Gathering/NamedPlaceRelations/NamedPlaceRelation/NearbyPlaceName</u> <u>Gathering/NamedPlaceRelations/NamedPlaceRelation/NearbyPlaceRelation</u> <u>Gathering/Notes</u> <u>Gathering/OtherMeasurementsOrFacts</u>

CLASSES	DESCRIÇÃO DA CLASSE	ELEMENTOS (*SUBCLASSE/ELEMENTOS - quando houver subclasse)
		<u>Gathering/OtherMeasurementsOrFacts/OtherMeasurementOrFact</u> <u>Gathering/Permits</u> <u>Gathering/Permits/Permit</u> <u>Gathering/PlatformName</u> <u>Gathering/Project</u> <u>Gathering/Project/Contact</u> <u>Gathering/Project/ResourceURIs</u> <u>Gathering/Project/ResourceURIs/ResourceURI</u> <u>Gathering/Project/Title</u> <u>Gathering/Project/WebsiteURL</u> <u>Gathering/Stratigraphy</u> <u>Gathering/Synecology</u> <u>Gathering/Synecology/AssociatedTaxa</u> <u>Gathering/Synecology/AssociatedTaxa/AssociatedTaxon</u> <u>Gathering/Synecology/Notes</u> <u>Gathering/Synecology/Syntaxon</u> <u>Gathering/ValidDistributionFlag</u> <u>Gathering/WFS</u> <u>Gathering/WMSURL</u>
<u>GrowthConditionAtomized</u>	Atomized elements describing the growth conditions of the culture unit.	<u>GrowthConditionAtomized/Aerobicity</u> <u>GrowthConditionAtomized/CultureMedium</u> <u>GrowthConditionAtomized/GrowthConditionsMeasurementsOrFacts</u> <u>GrowthConditionAtomized/GrowthConditionsMeasurementsOrFacts/GrowthConditionsMeasurementOrFact</u> <u>GrowthConditionAtomized/References</u> <u>GrowthConditionAtomized/References/Reference</u> <u>GrowthConditionAtomized/Temperature</u>

CLASSES	DESCRIÇÃO DA CLASSE	ELEMENTOS (*SUBCLASSE/ELEMENTOS - quando houver subclasse)
<u>HerbariumUnit</u>	Elements only used for herbarium units	<u>HerbariumUnit/CultivatedOccurrence</u> <u>HerbariumUnit/Exsiccatum</u> <u>HerbariumUnit/LoanDate</u> <u>HerbariumUnit/LoanDestination</u> <u>HerbariumUnit/LoanDispatchMethod</u> <u>HerbariumUnit/LoanForBotanist</u> <u>HerbariumUnit/LoanID</u> <u>HerbariumUnit/LoanReturnDate</u> <u>HerbariumUnit/LoanSequenceNumber</u> <u>HerbariumUnit/NaturalOccurrence</u>
<u>HigherTaxon</u>	Taxon name used to further classify the identification result.	<u>HigherTaxon/HigherTaxonName</u> <u>HigherTaxon/HigherTaxonRank</u>
<u>Identification</u>	The application of a name (or concept), i.e. the data pertaining to an identification event and the result of the identification itself.	<u>Identification/Date</u> <u>Identification/IdentificationSource</u> <u>Identification/IdentifierRole</u> <u>Identification/Identifiers</u> <u>Identification/Identifiers/Identifier</u> <u>Identification/Method</u> <u>Identification/NegativeIdentificationFlag</u> <u>Identification/Notes</u> <u>Identification/PreferredFlag</u> <u>Identification/References</u> <u>Identification/References/Reference</u> <u>Identification/Result</u> <u>Identification/Result/Extension</u> <u>Identification/Result/MaterialIdentified</u>

CLASSES	DESCRIÇÃO DA CLASSE	ELEMENTOS (*SUBCLASSE/ELEMENTOS - quando houver subclasse)
		<u>Identification/Result/TaxonIdentified</u> <u>Identification/ResultRole</u> <u>Identification/StoredUnderFlag</u> <u>Identification/VerificationLevel</u>
<u>imageSize</u>	picture width and height in pixel	<u>imageSize/Height</u> <u>imageSize/Width</u>
<u>LegalStatement</u>	Text, optional Details (both free-form text) and optional website URL and resource URI.	<u>Label/Representation</u> <u>LabelRepr/Abbreviation</u> <u>LabelRepr/Text</u> <u>LegalStatement/Details</u> <u>LegalStatement/ResourceURI</u> <u>LegalStatement/Text</u> <u>LegalStatement/WebsiteURL</u>
<u>LegalStatements</u>	A sequence of statements related to Intellectual Property Rights, credit and acknowledgement.	<u>LegalStatements/Acknowledgements</u> <u>LegalStatements/Acknowledgements/Acknowledgement</u> <u>LegalStatements/Licenses</u> <u>LegalStatements/Licenses/License</u> <u>LegalStatements/OtherLegalStatements</u> <u>LegalStatements/OtherLegalStatements/OtherLegalStatement</u> <u>LegalStatements/SuggestedCitations</u> <u>LegalStatements/SuggestedCitations/SuggestedCitation</u>
<u>MeasurementOrFact</u>	General purpose measurement and fact (descriptor) recording.	<u>MeasurementOrFact/MeasurementOrFactAtomized</u> <u>MeasurementOrFact/MeasurementOrFactAtomized/Accuracy</u> <u>MeasurementOrFact/MeasurementOrFactAtomized/AppliesTo</u> <u>MeasurementOrFact/MeasurementOrFactAtomized/Duration</u> <u>MeasurementOrFact/MeasurementOrFactAtomized/IsQuantitative</u> <u>MeasurementOrFact/MeasurementOrFactAtomized/LowerValue</u>

CLASSES	DESCRIÇÃO DA CLASSE	ELEMENTOS (*SUBCLASSE/ELEMENTOS - quando houver subclasse)
		MeasurementOrFact/MeasurementOrFactAtomized/MeasuredBy MeasurementOrFact/MeasurementOrFactAtomized/MeasurementDateTime MeasurementOrFact/MeasurementOrFactAtomized/MeasurementOrFactReference MeasurementOrFact/MeasurementOrFactAtomized/Method MeasurementOrFact/MeasurementOrFactAtomized/Parameter MeasurementOrFact/MeasurementOrFactAtomized/ReferenceSystem MeasurementOrFact/MeasurementOrFactAtomized/UnitOfMeasurement MeasurementOrFact/MeasurementOrFactAtomized/UpperValue MeasurementOrFact/MeasurementOrFactText
<u>MetadataDescriptionRepr</u>	Language-specific content metadata (title, description, etc.) with *required* language attribute added.	MetadataDescriptionRepr/Coverage MetadataDescriptionRepr/Details MetadataDescriptionRepr/Title MetadataDescriptionRepr/WebsiteURL
<u>MultimediaObject</u>	Elements to describe digital images, sound or video recordings, etc. Subclasses: - Video Object - Audio Object - Image Object	MultimediaObject/Audio MultimediaObject/Audio/AudioBitrate MultimediaObject/Audio/AudioChannel MultimediaObject/Audio/AudioEncoding MultimediaObject/Audio/CallType MultimediaObject/Audio/CarrierFrequency MultimediaObject/Audio/Chapters MultimediaObject/Audio/CodecName MultimediaObject/Audio/CodecQuality MultimediaObject/Audio/DurationTimeCode MultimediaObject/Audio/Microphone MultimediaObject/Audio/NumberOfNotes MultimediaObject/Audio/Parabola MultimediaObject/Audio/Pitch

CLASSES	DESCRIÇÃO DA CLASSE	ELEMENTOS (*SUBCLASSE/ELEMENTOS - quando houver subclasse)
		<u>MultimediaObject/Audio/PulseDistanceRegularity</u> <u>MultimediaObject/Audio/PulseEnergyContour</u> <u>MultimediaObject/Audio/PulseFrequencyContour</u> <u>MultimediaObject/Audio/PulseFrequencyModulations</u> <u>MultimediaObject/Audio/PulseGrouping</u> <u>MultimediaObject/Audio/PulseLength</u> <u>MultimediaObject/Audio/PulseMaximumAmplitude</u> <u>MultimediaObject/Audio/PulseRepetitionRate</u> <u>MultimediaObject/Audio/PulseTimeEncodedSignal</u> <u>MultimediaObject/Audio/SamplingRate</u> <u>MultimediaObject/Audio/Speed</u> <u>MultimediaObject/Audio/Volume</u> <u>MultimediaObject/CaptureEquipment</u> <u>MultimediaObject/Context</u> <u>MultimediaObject/Counter</u> <u>MultimediaObject/Creators</u> <u>MultimediaObject/Creators/Creator</u> <u>MultimediaObject/DateCreated</u> <u>MultimediaObject/FileFormat</u> <u>MultimediaObject/FileFormatVersion</u> <u>MultimediaObject/FileSize</u> <u>MultimediaObject/FileURL</u> <u>MultimediaObject/FilterUsedForRecording</u> <u>MultimediaObject/ID</u> <u>MultimediaObject/IDofContainingCollection</u> <u>MultimediaObject/Image</u> <u>MultimediaObject/Image/Brightness</u>

CLASSES	DESCRIÇÃO DA CLASSE	ELEMENTOS (*SUBCLASSE/ELEMENTOS - quando houver subclasse)
		<u>MultimediaObject/Image/Color</u> <u>MultimediaObject/Image/ColorSpace</u> <u>MultimediaObject/Image/Contrast</u> <u>MultimediaObject/Image/DigitalZoomRatio</u> <u>MultimediaObject/Image/ExposureMode</u> <u>MultimediaObject/Image/ExposureTime</u> <u>MultimediaObject/Image/Flash</u> <u>MultimediaObject/Image/FlashEnergy</u> <u>MultimediaObject/Image/FNumber</u> <u>MultimediaObject/Image/FocalLength</u> <u>MultimediaObject/Image/FocalLengthIn35mmFilm</u> <u>MultimediaObject/Image/Gamma</u> <u>MultimediaObject/Image/ImageResolution</u> <u>MultimediaObject/Image/ImageSize</u> <u>MultimediaObject/Image/ISOSpeed</u> <u>MultimediaObject/Image/LensModel</u> <u>MultimediaObject/Image/LightSource</u> <u>MultimediaObject/Image/PhotographicSensitivity</u> <u>MultimediaObject/Image/Saturation</u> <u>MultimediaObject/Image/Sharpness</u> <u>MultimediaObject/Image/ShutterSpeed</u> <u>MultimediaObject/Image/SpectralSensitivity</u> <u>MultimediaObject/Image/ThumbnailURL</u> <u>MultimediaObject/Image/WhiteBalance</u> <u>MultimediaObject/LegalStatements</u> <u>MultimediaObject/LifeStage</u> <u>MultimediaObject/LightCondition</u>

CLASSES	DESCRIÇÃO DA CLASSE	ELEMENTOS (*SUBCLASSE/ELEMENTOS - quando houver subclasse)
		<u>MultimediaObject/MakerNote</u> <u>MultimediaObject/Modified</u> <u>MultimediaObject/Notes</u> <u>MultimediaObject/PhysicalFormat</u> <u>MultimediaObject/PhysicalObjectID</u> <u>MultimediaObject/PlaybackUsed</u> <u>MultimediaObject/ProductURL</u> <u>MultimediaObject/Provider</u> <u>MultimediaObject/Ratings</u> <u>MultimediaObject/Ratings/Rating</u> <u>MultimediaObject/RecordingEnvironment</u> <u>MultimediaObject/Sex</u> <u>MultimediaObject/Source</u> <u>MultimediaObject/SubjectDistance</u> <u>MultimediaObject/SubjectOrientation</u> <u>MultimediaObject/SubjectPart</u> <u>MultimediaObject/Tags</u> <u>MultimediaObject/Tags/Tag</u> <u>MultimediaObject/TaxaInBackground</u> <u>MultimediaObject/TaxaInBackground/TaxonInBackground</u> <u>MultimediaObject/Temperature</u> <u>MultimediaObject/Title</u> <u>MultimediaObject/Type</u> <u>MultimediaObject/Video</u> <u>MultimediaObject/Video/AspectRatio</u> <u>MultimediaObject/Video/AudioBitrate</u> <u>MultimediaObject/Video/AudioChannel</u>

CLASSES	DESCRIÇÃO DA CLASSE	ELEMENTOS (*SUBCLASSE/ELEMENTOS - quando houver subclasse)
		<u>MultimediaObject/Video/AudioEncoding</u> <u>MultimediaObject/Video/CallType</u> <u>MultimediaObject/Video/CaptureFramerate</u> <u>MultimediaObject/Video/Chapters</u> <u>MultimediaObject/Video/Color</u> <u>MultimediaObject/Video/DurationTimeCode</u> <u>MultimediaObject/Video/Framerate</u> <u>MultimediaObject/Video/ImageSize</u> <u>MultimediaObject/Video/LensModel</u> <u>MultimediaObject/Video/LightSource</u> <u>MultimediaObject/Video/Microphone</u> <u>MultimediaObject/Video/NumberOfNotes</u> <u>MultimediaObject/Video/Subtitles</u> <u>MultimediaObject/Video/SubtitlesFormat</u> <u>MultimediaObject/Video/ThumbnailURL</u> <u>MultimediaObject/Video/VideoBitrate</u> <u>MultimediaObject/Video/VideoEncoding</u>
<u>MycologicalUnit</u>	Elements only used for mycological units	<u>MycologicalUnit/LichenMorphotype</u> <u>MycologicalUnit/MycologicalLiveStages</u> <u>MycologicalUnit/MycologicalLiveStages/MycologicalLiveStage</u> <u>MycologicalUnit/MycologicalSexualStage</u>
<u>NameBacterial</u>	An atomized name under the International Code of Nomenclature of Bacteria	<u>NameBacterial/AuthorTeamAndYear</u> <u>NameBacterial/GenusOrMonomial</u> <u>NameBacterial/NameApprobation</u> <u>NameBacterial/ParantheticalAuthorTeamAndYear</u> <u>NameBacterial/SpeciesEpithet</u> <u>NameBacterial/Subgenus</u>

CLASSES	DESCRIÇÃO DA CLASSE	ELEMENTOS (*SUBCLASSE/ELEMENTOS - quando houver subclasse)
		<u>NameBacterial/SubgenusAuthorAndYear</u> <u>NameBacterial/SubspeciesEpithet</u>
<u>NameBotanical</u>	An atomized scientific name under the International Code of Botanical Nomenclature or the International Code of Nomenclature for Cultivated Plants	<u>NameBotanical/AuthorTeam</u> <u>NameBotanical/AuthorTeamParenthesis</u> <u>NameBotanical/CultivarGroupName</u> <u>NameBotanical/CultivarName</u> <u>NameBotanical/FirstEpithet</u> <u>NameBotanical/GenusOrMonomial</u> <u>NameBotanical/HybridFlag</u> <u>NameBotanical/InfraspecificEpithet</u> <u>NameBotanical/Rank</u> <u>NameBotanical/TradeDesignationNames</u> <u>NameBotanical/TradeDesignationNames/TradeDesignationName</u>
<u>NamedArea</u>	An atomized place name (an administrative or geocological area) with possibility for hierarchical structuring.	<u>NamedArea/AreaClass</u> <u>NamedArea/Code</u> <u>NamedArea/CodeStandard</u> <u>NamedArea/Name</u> <u>NamedArea/Reference</u>
<u>NameViral</u>	An atomized scientific name under the International Code of Virus Classification and Nomenclature	<u>NameViral/Acronym</u> <u>NameViral/GenusOrMonomial</u> <u>NameViral/ViralSpeciesDesignation</u>
<u>NameZoological</u>	An atomized scientific name under the International Code of Zoological Nomenclature	<u>NameZoological/AuthorTeamOriginalAndYear</u> <u>NameZoological/AuthorTeamParenthesisAndYear</u> <u>NameZoological/Breed</u> <u>NameZoological/CombinationAuthorTeamAndYear</u> <u>NameZoological/GenusOrMonomial</u>

CLASSES	DESCRIÇÃO DA CLASSE	ELEMENTOS (*SUBCLASSE/ELEMENTOS - quando houver subclasse)
		<u>NameZoological/NamedIndividual</u> <u>NameZoological/SpeciesEpithet</u> <u>NameZoological/Subgenus</u> <u>NameZoological/SubspeciesEpithet</u>
<u>Organization</u>	identifier for the Organization or Organization Unit	<u>Organization/Divisions</u> <u>Organization/Divisions/Division</u> <u>Organization/GUID</u> <u>Organization/LogoURL</u> <u>Organization/Name</u>
<u>PaleontologicalUnit</u>	Elements only used for paleontological units	<u>PaleontologicalUnit/Preservation</u> <u>PaleontologicalUnit/Preservation/Completeness</u> <u>PaleontologicalUnit/Preservation/Form</u> <u>PaleontologicalUnit/Preservation/Matrix</u> <u>PaleontologicalUnit/Preservation/Mineralization</u> <u>PaleontologicalUnit/Preservation/Taphonomy</u> <u>PaleontologicalUnit/TimeRange</u>
<u>Permit</u>	Container element for information on permits related to the Gathering event	<u>Permit/Details</u> <u>Permit/ResourceURI</u> <u>Permit/Status</u> <u>Permit/Text</u> <u>Permit/Type</u> <u>Permit/WebsiteURL</u>
<u>PersonName</u>	Name of an individual person	<u>PersonName/AtomizedName</u> <u>PersonName/AtomizedName/GivenNames</u> <u>PersonName/AtomizedName/InheritedName</u> <u>PersonName/AtomizedName/PreferredName</u> <u>PersonName/AtomizedName/Prefix</u>

CLASSES	DESCRIÇÃO DA CLASSE	ELEMENTOS (*SUBCLASSE/ELEMENTOS - quando houver subclasse)
		<u>PersonName/AtomizedName/Suffix</u> <u>PersonName/FullName</u> <u>PersonName/SortingName</u>
<u>PGRUnit</u>	Extension for elements only used in plant genetic resource collections. Currently, all elements represent European Search Catalogue for Plant Genetic Resources (EURISCO) descriptors.	<u>PGRUnit/AccessionNames</u> <u>PGRUnit/AccessionNames/AccessionName</u> <u>PGRUnit/AccessionNameText</u> <u>PGRUnit/AncestralData</u> <u>PGRUnit/BiologicalStatus</u> <u>PGRUnit/BreedingCountryCode</u> <u>PGRUnit/BreedingInstitutionCode</u> <u>PGRUnit/CollectingAcquisitionSource</u> <u>PGRUnit/DecodedBreedingInstitute</u> <u>PGRUnit/DecodedGatheringInstitute</u> <u>PGRUnit/DecodedLocationSafetyDuplicates</u> <u>PGRUnit/GatheringInstitutionCode</u> <u>PGRUnit/LocationSafetyDuplicates</u> <u>PGRUnit/NationalInventoryCode</u> <u>PGRUnit/OtherIdentification</u> <u>PGRUnit/TypeGermplasmStorage</u>
<u>Reference</u>	Published reference.	<u>Reference/Details</u> <u>Reference/GUID</u> <u>Reference/ResourceURIs</u> <u>Reference/ResourceURIs/ResourceURI</u> <u>Reference/Text</u> <u>Reference/WebsiteURL</u>
<u>RevisionData</u>	Revision data (creators, dates, revision) for the	<u>RevisionData/Contributors</u> <u>RevisionData/Creators</u>

CLASSES	DESCRIÇÃO DA CLASSE	ELEMENTOS (*SUBCLASSE/ELEMENTOS - quando houver subclasse)
	entire project/data set or individual objects.	<u>RevisionData/DateCreated</u> <u>RevisionData/DateModified</u>
<u>ScientificName</u>	Name of the taxon identified, formed according to the different Codes of Nomenclature which apply to scientific names. Subclasses: - Botanical Scientific Name - Bacterial Scientific Name - Viral Scientific Name - Zoological Scientific Name	<u>ScientificName/FullScientificName</u> <u>ScientificName/NameAtomized</u> <u>ScientificName/NameAtomized/Bacterial</u> <u>ScientificName/NameAtomized/Botanical</u> <u>ScientificName/NameAtomized/Viral</u> <u>ScientificName/NameAtomized/Zoological</u>
<u>ScientificNameIdentified</u>	Name of the taxon identified, formed according to the different Codes of Nomenclature which apply to scientific names - with additional elements for suffixes or other expressions commonly used in taxonomic identifications.	<u>ScientificNameIdentified/IdentificationQualifier</u> <u>ScientificNameIdentified/NameAddendum</u>
<u>Sequence</u>	A first proposal for this type of data. Note that this is used in the context of a collection unit and thus automatically refers to that	<u>Sequence/DirectAccessURL</u> <u>Sequence/ID</u> <u>Sequence/Method</u> <u>Sequence/Reference</u> <u>Sequence/SequenceDatabase</u>

CLASSES	DESCRIÇÃO DA CLASSE	ELEMENTOS (*SUBCLASSE/ELEMENTOS - quando houver subclasse)
	unit. If your unit is a DNA Sample than please use the GGBN extension. This container here is meant to be used e.g. to provide Sequence information related to a specimen only without any further details.	<u>Sequence/SequencedPart</u> <u>Sequence/SequenceLength</u> <u>Sequence/SequencingAgent</u>
<u>StratigraphicTerm</u>	Descriptive term. May be taken from a controlled vocabulary maintained by a third party.	<u>StratigraphicTerm/Comment</u> <u>StratigraphicTerm/Domain</u> <u>StratigraphicTerm/Source</u> <u>StratigraphicTerm/Term</u>
<u>Stratigraphy</u>	Elements expressing stratigraphy at the gathering site.	<u>Stratigraphy/BiostratigraphicTerms</u> <u>Stratigraphy/BiostratigraphicTerms/BiostratigraphicTerm</u> <u>Stratigraphy/ChronostratigraphicTerms</u> <u>Stratigraphy/ChronostratigraphicTerms/ChronostratigraphicTerm</u> <u>Stratigraphy/LithostratigraphicTerms</u> <u>Stratigraphy/LithostratigraphicTerms/LithostratigraphicTerm</u> <u>Stratigraphy/StratigraphyText</u>
<u>TaxonIdentified</u>	Scientific or informal name as a result of an identification.	<u>TaxonIdentified/Code</u> <u>TaxonIdentified/HigherTaxa</u> <u>TaxonIdentified/HigherTaxa/HigherTaxon</u> <u>TaxonIdentified/InformalName</u> <u>TaxonIdentified/NameComments</u> <u>TaxonIdentified/ScientificName</u>
<u>Unit</u>	Elements for all unit-level data	<u>Unit/Age</u> <u>Unit/Annotations</u>

CLASSES	DESCRIÇÃO DA CLASSE	ELEMENTOS (*SUBCLASSE/ELEMENTOS - quando houver subclasse)
		<u>Unit/Annotations/Annotation</u> <u>Unit/Annotations/Annotation/Annotator</u> <u>Unit/Annotations/Annotation/Date</u> <u>Unit/Annotations/Annotation/Text</u> <u>Unit/Assemblages</u> <u>Unit/Assemblages/Assemblage</u> <u>Unit/Assemblages/Assemblage/AssemblageName</u> <u>Unit/Assemblages/Assemblage/ID</u> <u>Unit/Assemblages/Assemblage/ResourceURIs</u> <u>Unit/Assemblages/Assemblage/ResourceURIs/ResourceURI</u> <u>Unit/Associations</u> <u>Unit/Associations/Association</u> <u>Unit/Associations/Association/AssociationType</u> <u>Unit/Associations/Association/DirectAccessURL</u> <u>Unit/Associations/Association/GUID</u> <u>Unit/Associations/Association/ID</u> <u>Unit/Associations/Association/KindOfUnit</u> <u>Unit/Associations/Association/Notes</u> <u>Unit/Associations/Association/ResourceURIs</u> <u>Unit/Associations/Association/ResourceURIs/ResourceURI</u> <u>Unit/Associations/Association/SourceID</u> <u>Unit/Associations/Association/SourceInstitutionID</u> <u>Unit/BotanicalGardenUnit</u> <u>Unit/ContentContacts</u> <u>Unit/ContentContacts/ContentContact</u> <u>Unit/CultureCollectionUnit</u> <u>Unit/DateModified</u>

CLASSES	DESCRIÇÃO DA CLASSE	ELEMENTOS (*SUBCLASSE/ELEMENTOS - quando houver subclasse)
		<u>Unit/FieldNotes</u> <u>Unit/FieldNotesReferences</u> <u>Unit/FieldNotesReferences/FieldNotesReference</u> <u>Unit/FieldNumbers</u> <u>Unit/FieldNumbers/FieldNumber</u> <u>Unit/FieldNumbers/FieldNumber/CollectorName</u> <u>Unit/FieldNumbers/FieldNumber/ID</u> <u>Unit/Gathering</u> <u>Unit/GUID</u> <u>Unit/HerbariumUnit</u> <u>Unit/ID</u> <u>Unit/IdentificationHistory</u> <u>Unit/Identifications</u> <u>Unit/Identifications/Identification</u> <u>Unit/InformationWithheld</u> <u>Unit/KindOfUnit</u> <u>Unit/LastEditor</u> <u>Unit/LegalStatements</u> <u>Unit/MeasurementsOrFacts</u> <u>Unit/MeasurementsOrFacts/MeasurementOrFact</u> <u>Unit/MultimediaObjects</u> <u>Unit/MultimediaObjects/MultimediaObject</u> <u>Unit/MycologicalUnit</u> <u>Unit/NamedCollectionsOrSurveys</u> <u>Unit/NamedCollectionsOrSurveys/NamedCollectionOrSurvey</u> <u>Unit/Notes</u> <u>Unit/NumericID</u>

CLASSES	DESCRIÇÃO DA CLASSE	ELEMENTOS (*SUBCLASSE/ELEMENTOS - quando houver subclasse)
		<u>Unit/ObservationUnit</u> <u>Unit/ObservationUnit/ObservationUnitIDs</u> <u>Unit/ObservationUnit/ObservationUnitIDs/ObservationUnitID</u> <u>Unit/Owner</u> <u>Unit/PaleontologicalUnit</u> <u>Unit/PlantGeneticResourcesUnit</u> <u>Unit/RecordBasis</u> <u>Unit/ResourceURIs</u> <u>Unit/ResourceURIs/ResourceURI</u> <u>Unit/Sequences</u> <u>Unit/Sequences/Sequence</u> <u>Unit/Sex</u> <u>Unit/SourceID</u> <u>Unit/SourceInstitutionID</u> <u>Unit/SourceReference</u> <u>Unit/SpecimenUnit</u> <u>Unit/SpecimenUnit/Accessions</u> <u>Unit/SpecimenUnit/Accessions/Accession</u> <u>Unit/SpecimenUnit/Accessions/Accession/AccessionCatalogue</u> <u>Unit/SpecimenUnit/Accessions/Accession/AccessionDate</u> <u>Unit/SpecimenUnit/Accessions/Accession/AccessionNumber</u> <u>Unit/SpecimenUnit/Accessions/Accession/AccessionURI</u> <u>Unit/SpecimenUnit/Acquisition</u> <u>Unit/SpecimenUnit/Acquisition/AcquiredFrom</u> <u>Unit/SpecimenUnit/Acquisition/AcquisitionSourceText</u> <u>Unit/SpecimenUnit/Acquisition/AcquisitionType</u> <u>Unit/SpecimenUnit/Acquisition/Date</u>

CLASSES	DESCRIÇÃO DA CLASSE	ELEMENTOS (*SUBCLASSE/ELEMENTOS - quando houver subclasse)
		<u>Unit/SpecimenUnit/Disposition</u> <u>Unit/SpecimenUnit/DuplicatesDistributedTo</u> <u>Unit/SpecimenUnit/LoanRestrictions</u> <u>Unit/SpecimenUnit/LoanRestrictions/Blocked</u> <u>Unit/SpecimenUnit/LoanRestrictions/BlockedUntil</u> <u>Unit/SpecimenUnit/LoanRestrictions/LoanConditions</u> <u>Unit/SpecimenUnit/Marks</u> <u>Unit/SpecimenUnit/Marks/Mark</u> <u>Unit/SpecimenUnit/Marks/Mark/Images</u> <u>Unit/SpecimenUnit/Marks/Mark/Images/ImageID</u> <u>Unit/SpecimenUnit/Marks/Mark/MarkAuthor</u> <u>Unit/SpecimenUnit/Marks/Mark/MarkComment</u> <u>Unit/SpecimenUnit/Marks/Mark/MarkText</u> <u>Unit/SpecimenUnit/Marks/Mark/MarkType</u> <u>Unit/SpecimenUnit/Marks/Mark/PositionOnObject</u> <u>Unit/SpecimenUnit/NomenclaturalTypeDesignations</u> <u>Unit/SpecimenUnit/NomenclaturalTypeDesignations/NomenclaturalTypeDesignation</u> <u>Unit/SpecimenUnit/NomenclaturalTypeDesignations/NomenclaturalTypeDesignation/CodeAssessment</u> <u>Unit/SpecimenUnit/NomenclaturalTypeDesignations/NomenclaturalTypeDesignation/DoubtfulFlag</u> <u>Unit/SpecimenUnit/NomenclaturalTypeDesignations/NomenclaturalTypeDesignation/NomenclaturalReference</u> <u>Unit/SpecimenUnit/NomenclaturalTypeDesignations/NomenclaturalTypeDesignation/TypeStatus</u> <u>Unit/SpecimenUnit/NomenclaturalTypeDesignations/NomenclaturalTypeDesignation/TypifiedName</u> <u>Unit/SpecimenUnit/NomenclaturalTypeDesignations/NomenclaturalTypeDesignation/Verification</u>

CLASSES	DESCRIÇÃO DA CLASSE	ELEMENTOS (*SUBCLASSE/ELEMENTOS - quando houver subclasse)
		<p><u>onDate</u></p> <p><u>Unit/SpecimenUnit/NomenclaturalTypeDesignations/NomenclaturalTypeDesignation/VerificationNotes</u></p> <p><u>Unit/SpecimenUnit/NomenclaturalTypeDesignations/NomenclaturalTypeDesignation/Verifier</u></p> <p><u>Unit/SpecimenUnit/NomenclaturalTypeText</u></p> <p><u>Unit/SpecimenUnit/Owner</u></p> <p><u>Unit/SpecimenUnit/Preparations</u></p> <p><u>Unit/SpecimenUnit/Preparations/Preparation</u></p> <p><u>Unit/SpecimenUnit/Preparations/Preparation/PreparationAgent</u></p> <p><u>Unit/SpecimenUnit/Preparations/Preparation/PreparationDate</u></p> <p><u>Unit/SpecimenUnit/Preparations/Preparation/PreparationMaterials</u></p> <p><u>Unit/SpecimenUnit/Preparations/Preparation/PreparationMethod</u></p> <p><u>Unit/SpecimenUnit/Preparations/Preparation/PreparationType</u></p> <p><u>Unit/SpecimenUnit/Preparations/Preparation/References</u></p> <p><u>Unit/SpecimenUnit/Preparations/Preparation/References/Reference</u></p> <p><u>Unit/SpecimenUnit/Preparations/Preparation/SampleDesignations</u></p> <p><u>Unit/SpecimenUnit/Preparations/Preparation/SampleDesignations/SampleDesignation</u></p> <p><u>Unit/SpecimenUnit/Preparations/Preparation/SampleSize</u></p> <p><u>Unit/SpecimenUnit/Preparations/Preparation/Sieving</u></p> <p><u>Unit/SpecimenUnit/PreparationsText</u></p> <p><u>Unit/SpecimenUnit/Preservations</u></p> <p><u>Unit/SpecimenUnit/Preservations/Preservation</u></p> <p><u>Unit/SpecimenUnit/Preservations/Preservation/PreservationDateBegin</u></p> <p><u>Unit/SpecimenUnit/Preservations/Preservation/PreservationTemperature</u></p> <p><u>Unit/SpecimenUnit/Preservations/Preservation/PreservationType</u></p> <p><u>Unit/SpecimenUnit/PreviousUnits</u></p> <p><u>Unit/SpecimenUnit/PreviousUnits/PreviousUnit</u></p>

CLASSES	DESCRIÇÃO DA CLASSE	ELEMENTOS (*SUBCLASSE/ELEMENTOS - quando houver subclasse)
		<u>Unit/SpecimenUnit/PreviousUnits/PreviousUnit/Date</u> <u>Unit/SpecimenUnit/PreviousUnits/PreviousUnit/Notes</u> <u>Unit/SpecimenUnit/PreviousUnits/PreviousUnit/PreviousSourceID</u> <u>Unit/SpecimenUnit/PreviousUnits/PreviousUnit/PreviousSourceInstitutionID</u> <u>Unit/SpecimenUnit/PreviousUnits/PreviousUnit/PreviousUnitGUID</u> <u>Unit/SpecimenUnit/PreviousUnits/PreviousUnit/PreviousUnitID</u> <u>Unit/SpecimenUnit/PreviousUnits/PreviousUnit/PreviousUnitText</u> <u>Unit/SpecimenUnit/PreviousUnitsText</u> <u>Unit/SpecimenUnit/SpecimenMeasurementsOrFacts</u> <u>Unit/SpecimenUnit/SpecimenMeasurementsOrFacts/SpecimenMeasurementOrFact</u> <u>Unit/UnitExtensions</u> <u>Unit/UnitExtensions/UnitExtension</u> <u>Unit/UnitReferences</u> <u>Unit/UnitReferences/UnitReference</u> <u>Unit/WebsiteURL</u> <u>Unit/ZoologicalUnit</u>
<u>ZoologicalUnit</u>	Elements only used for zoological units	<u>ZoologicalUnit/PhasesOrStages</u> <u>ZoologicalUnit/PhasesOrStages/PhaseOrStage</u>

ANEXO 3. Esquemas do padrão EML

Esquemas	Descrição	Tipo de Esquema
eml	The eml module is a wrapper container that allows the inclusion of any metadata content in a single EML document. The eml module is used as a container to hold structured descriptions of ecological resources. In EML, the definition of a resource comes from the The Dublin Core Metadata Initiative , which describes a general element set used to describe "networked digital resources". The top-level structure of EML has been designed to be compatible with the Dublin Core syntax. In general, dataset resources, literature resources, software resources, and protocol resources comprise the list of information that may be described in EML. EML is largely designed to describe digital resources, however, it may also be used to describe non-digital resources such as paper maps and other non-digital media. In EML, the definition of a "Data Package" is the combination of both the data and metadata for a resource. So, data packages are built by using the <eml> wrapper, which will include all of the metadata, and optionally the data (or references to them). All EML packages must begin with the <eml> tag and end with the </eml> tag. The eml module may be extended to describe other resources by means of its optional sub-field, <additionalMetadata>. This field is largely reserved for the inclusion of metadata that may be highly discipline specific and not covered in this version of EML, or it may be used to internally extend fields within the EML standard.	Main Schema
eml - attribute	The eml-attribute module - Attribute level information within dataset entities. recommendedUsage: any dataset that uses dataTable, spatialRaster, spatialVector, storedProcedure, view or otherEntity or in a custom module where one wants to document an attribute (variable).	Data Structure Module
eml - constraint	The eml-constraint module - Relationships among and within dataset entities. recommendedUsage: All datasets where there are logical constraints between entities.	Data Structure Module
eml - dataTable	The eml-dataTable module - Logical information about data table entities. recommendedUsage: The EML dataTable Module is used to document datasets with one or more data tables.	Data Structure Module
eml - entity	The eml-entity module - Entity level information within datasets. recommendedUsage: This module is used to describe data entities.	Data Structure Module
eml - spatialRaster	The eml-spatialRaster module - Logical information about regularly gridded geospatial image data recommendedUsage: all spatial datasets that use spatial gridded data.	Data Structure Module
eml - spatialVector	The eml-spatialVector module - Logical information about non-gridded geospatial image data recommendedUsage: all spatial datasets that contain spatial data entities represented as vector features.	Data Structure Module
eml - access	The eml-access module - Access control rules for resources. recommendedUsage: all data where controlling user access to the dataset is an issue	Utility Modules

eml - coverage	The eml-coverage module - Geographic, temporal, and taxonomic extents of resources. RecommendedUsage: all datasets where spatial, temporal or taxonomic coverage is important.	Discovery and Interpretation Modules
eml - dataset	<p>The eml-dataset module - Dataset specific information. recommendedUsage: all datasets.</p> <p>The eml-dataset module contains general information that describes dataset resources. It is intended to provide overview information about the dataset: broad information such as the title, abstract, keywords, contacts, maintenance history, purpose, and distribution of the data themselves. The eml-dataset module also imports many other modules that are used to describe the dataset in fine detail. Specifically, it uses the eml-methods module to describe methodology used in collecting or processing the dataset, the eml-project module to describe the overarching research context and experimental design, the eml-access module to define access control rules for the data and metadata, and the eml-entity module to provide detailed information about the logical structure of the dataset. A dataset can be (and often is) composed of a series of data entities (tables) that are linked together by particular integrity constraints.</p> <p>The eml-dataset module, like other modules, may be "referenced" via the <references> tag. This allows a dataset to be described once, and then used as a reference in other locations within the EML document via its ID.</p>	Core
eml - literature	<p>The eml-literature module - Citation specific information. recommendedUsage: All datasets with literary citations</p> <p>The eml-literature module contains information that describes literature resources. It is intended to provide overview information about the literature citation, including title, abstract, keywords, and contacts. Citation types follow the conventions laid out by EndNote, and there is an attempt to represent a compatible subset of the EndNote citation types. These citation types include: article, book, chapter, edited book, manuscript, report, thesis, conference proceedings, personal communication, map, generic, audio visual, and presentation. The generic citation type would be used when one of the other types will not work.</p> <p>There are three unique CitationType elements that may be employed within a eml-dataset module, including the <literatureCited>, <usageCitation>, and <referencePublication> elements. The purpose and examples of each CitationType element type are detailed below.</p> <p>Similar to other eml modules, each of the CitationType elements may be referenced via the <references> tag. The <references> tag allows a citation to be described once then used as a reference in other locations within the EML document via its reference ID.</p> <p>As of EML 2.2.0, each CitationType element can use the <bibtex> element as an alternative to encoding citations in the EML XML structures. BibTeX entries generally play well inside of XML structures, but XML escaping is still needed for special characters so consider embedding BibTeX entries in CDATA blocks if XML escaping is cumbersome.</p>	Core
eml - methods	The eml-methods module - Methodological information for resources. recommendedUsage: All datasets.	Discovery and Interpretation Modules

eml - project	The eml-project module - Research context information for resources. recommendedUsage: Use eml-project to document the research context of any dataset or project.	Discovery and Interpretation Modules
eml - protocol	<p>The eml-protocol module - Research protocol specific information. recommendedUsage: Use eml-protocol to describe prescriptive procedures that can be associated with other descriptive or prescriptive procedures. The EML Protocol Module is used to define abstract, prescriptive procedures for generating or processing data. Conceptually, a protocol is a standardized method.</p> <p>Eml-protocol resembles eml-methods; however, eml-methods is descriptive (often written in the declarative mood: "I took five subsamples...") whereas eml-protocol is prescriptive (often written in the imperative mood: "Take five subsamples..."). A protocol may have versions, whereas methods (as used in eml-methods) should not.</p>	Core
eml - resource	<p>The eml-resource module - Base information for all resources. recommendedUsage: all dataset.</p> <p>The eml-resource module contains general information that describes dataset resources, literature resources, protocol resources, and software resources. Each of the above four types of resources share a common set of information, but also have information that is unique to that particular resource type. Each resource type uses the eml-resource module to document the information common to all resources, but then extend eml-resource with modules that are specific to that particular resource type. For instance, all resources have creators, titles, and perhaps keywords, but only the dataset resource would have a "data table" within it. Likewise, a literature resource may have an "ISBN" number associated with it, whereas the other resource types would not.</p> <p>The eml-resource module is exclusively used by other modules, and is therefore not a stand-alone module. The following four modules are used to describe separate resources: datasets, literature, software, and protocols. However, note that the dataset module makes use of the other top-level modules by importing them at different levels. For instance, a dataset may have been produced using a particular protocol, and that protocol may come from a protocol document in a library of protocols. Likewise, citations are used throughout the top-level resource modules by importing the literature module.</p>	Core
eml - semantics	The eml-semantics module - Semantic annotations for formalized statements about EML components. recommendedUsage: all datasets.	Utility Modules

eml - software	<p>The eml-software module - Software specific information. recommendedUsage: All datasets where software was used in the analysis or creation of the dataset.</p> <p>The eml-software module contains general information that describes software resources. This module is intended to fully document software that is needed in order to view a resource (such as a dataset) or to process a dataset. The software module is also imported into the eml-methods module in order to document what software was used to process or perform quality control procedures on a dataset.</p> <p>The eml-software module, like other modules, may be "referenced" via the <references> tag. This allows a software resource to be described once, and then used as a reference in other locations within the EML document via its ID.</p>	Core
eml - text	The eml-text module - Text field formatting. recommendedUsage: any module.	Utility Modules
eml - view	The eml-view module - Data tables resulting from a database query recommendedUsage: all datasets that contain one or more views.	Data Structure Module
eml - party	The eml-party module - People and organization information. recommendedUsage: all datasets	Discovery and Interpretation Modules
eml - physical	The eml-physical module - Physical file format. RecommendedUsage: Any data object that is being described by EML needs this information so the entities and attributes that reside with in the data object can be extracted.	Discovery and Interpretation Modules
eml - storedProcedure	The eml-storedProcedure module - Data tables resulting from procedures stored in a database. recommendedUsage: Use the storedProcedure module to document datasets that use storedProcedures to retrieve archived data.	Data Structure Module

ANEXO 4. Esquemas e elementos do padrão EML

MÓDULO	ELEMENTOS
eml	eml/access eml/additionalMetadata eml/additionalMetadata/describes eml/additionalMetadata/metadata eml/annotations eml/annotations/annotation eml/citation eml/dataset eml/protocol eml/software
eml-documentation	description example lineage module moduleDocs moduleDocs/moduleDescription moduleDocs/moduleName moduleDocs/recommendedUsage moduleDocs/standAlone summary tooltip
eml-text	ListType/listitem ListType/listitem/itemizedlist ListType/listitem/orderedlist ListType/listitem/para ParagraphType/emphasis ParagraphType/emphasis/value ParagraphType/itemizedlist ParagraphType/literalLayout ParagraphType/literalLayout/value ParagraphType/orderedlist ParagraphType/subscript ParagraphType/superscript ParagraphType/ulink ParagraphType/ulink/citetitle ParagraphType/value SectionType/para SectionType/section SectionType/title

MÓDULO	ELEMENTOS
	<u>SubSuperScriptType/subscript</u> <u>SubSuperScriptType/superscript</u> <u>SubSuperScriptType/value</u> <u>TextType/markdown</u> <u>TextType/para</u> <u>TextType/section</u> text
eml-resource	<u>ConnectionDefinitionType/description</u> <u>ConnectionDefinitionType/parameterDefinition</u> <u>ConnectionDefinitionType/parameterDefinition/defaultValue</u> <u>ConnectionDefinitionType/parameterDefinition/definition</u> <u>ConnectionDefinitionType/parameterDefinition/name</u> <u>ConnectionDefinitionType/schemeName</u> <u>ConnectionType/connectionDefinition</u> <u>ConnectionType/parameter</u> <u>ConnectionType/parameter/name</u> <u>ConnectionType/parameter/value</u> <u>DistributionType/inline</u> <u>DistributionType/offline</u> <u>DistributionType/online</u> <u>LicenseType/identifier</u> <u>LicenseType/licenseName</u> <u>LicenseType/url</u> <u>OfflineType/mediumDensity</u> <u>OfflineType/mediumDensityUnits</u> <u>OfflineType/mediumFormat</u> <u>OfflineType/mediumName</u> <u>OfflineType/mediumNote</u> <u>OfflineType/mediumVolume</u> <u>OnlineType/connection</u> <u>OnlineType/connectionDefinition</u> <u>OnlineType/onlineDescription</u> <u>OnlineType/url</u> <u>ReferencesGroup/references</u> <u>ResourceGroup/abstract</u> <u>ResourceGroup/additionalInfo</u> <u>ResourceGroup/alternateIdentifier</u> <u>ResourceGroup/annotation</u> <u>ResourceGroup/associatedParty</u> <u>ResourceGroup/associatedParty/role</u> <u>ResourceGroup/coverage</u> <u>ResourceGroup/creator</u> <u>ResourceGroup/distribution</u>

MÓDULO	ELEMENTOS
	<u>ResourceGroup/intellectualRights</u> <u>ResourceGroup/keywordSet</u> <u>ResourceGroup/keywordSet/keyword</u> <u>ResourceGroup/keywordSet/keywordThesaurus</u> <u>ResourceGroup/language</u> <u>ResourceGroup/licensed</u> <u>ResourceGroup/metadataProvider</u> <u>ResourceGroup/pubDate</u> <u>ResourceGroup/series</u> <u>ResourceGroup/shortName</u> <u>ResourceGroup/title</u> <u>i18nNonEmptyStringType/value</u>
eml-party	<u>Address/administrativeArea</u> <u>Address/city</u> <u>Address/country</u> <u>Address/deliveryPoint</u> <u>Address/postalCode</u> <u>Person/givenName</u> <u>Person/salutation</u> <u>Person/surName</u> <u>ResponsibleParty/address</u> <u>ResponsibleParty/electronicMailAddress</u> <u>ResponsibleParty/individualName</u> <u>ResponsibleParty/onlineUrl</u> <u>ResponsibleParty/organizationName</u> <u>ResponsibleParty/phone</u> <u>ResponsibleParty/positionName</u> <u>ResponsibleParty/userId</u> <u>party</u>
eml-coverage	<u>Coverage/geographicCoverage</u> <u>Coverage/taxonomicCoverage</u> <u>Coverage/temporalCoverage</u> <u>GRingPointType/gRingLatitude</u> <u>GRingPointType/gRingLongitude</u> <u>GeographicCoverage/boundingCoordinates</u> <u>GeographicCoverage/boundingCoordinates/boundingAltitudes</u> <u>GeographicCoverage/boundingCoordinates/boundingAltitudes/altitudeMaximum</u> <u>GeographicCoverage/boundingCoordinates/boundingAltitudes/altitudeMinimum</u> <u>GeographicCoverage/boundingCoordinates/boundingAltitudes/altitudeUnits</u> <u>GeographicCoverage/boundingCoordinates/eastBoundingCoordinate</u> <u>GeographicCoverage/boundingCoordinates/northBoundingCoordinate</u>

MÓDULO	ELEMENTOS
	<u>GeographicCoverage/boundingCoordinates/southBoundingCoordinate</u>
	<u>GeographicCoverage/boundingCoordinates/westBoundingCoordinate</u>
	<u>GeographicCoverage/datasetGPolygon</u>
	<u>GeographicCoverage/datasetGPolygon/datasetGPolygonExclusionGRing</u>
	<u>GeographicCoverage/datasetGPolygon/datasetGPolygonExclusionGRing/gRing</u>
	<u>GeographicCoverage/datasetGPolygon/datasetGPolygonExclusionGRing/gRingPoint</u>
	<u>GeographicCoverage/datasetGPolygon/datasetGPolygonOuterGRing</u>
	<u>GeographicCoverage/datasetGPolygon/datasetGPolygonOuterGRing/gRing</u>
	<u>GeographicCoverage/datasetGPolygon/datasetGPolygonOuterGRing/gRingPoint</u>
	<u>GeographicCoverage/geographicDescription</u>
	<u>SingleDateTimeType/alternativeTimeScale</u>
	<u>SingleDateTimeType/alternativeTimeScale/timeScaleAgeEstimate</u>
	<u>SingleDateTimeType/alternativeTimeScale/timeScaleAgeExplanation</u>
	<u>SingleDateTimeType/alternativeTimeScale/timeScaleAgeUncertainty</u>
	<u>SingleDateTimeType/alternativeTimeScale/timeScaleCitation</u>
	<u>SingleDateTimeType/alternativeTimeScale/timeScaleName</u>
	<u>SingleDateTimeType/calendarDate</u>
	<u>SingleDateTimeType/time</u>
	<u>TaxonomicClassificationType/commonName</u>
	<u>TaxonomicClassificationType/taxonId</u>
	<u>TaxonomicClassificationType/taxonRankName</u>
	<u>TaxonomicClassificationType/taxonRankValue</u>
	<u>TaxonomicClassificationType/taxonomicClassification</u>
	<u>TaxonomicCoverage/generalTaxonomicCoverage</u>
	<u>TaxonomicCoverage/taxonomicClassification</u>
	<u>TaxonomicCoverage/taxonomicSystem</u>
	<u>TaxonomicCoverage/taxonomicSystem/classificationSystem</u>
	<u>TaxonomicCoverage/taxonomicSystem/classificationSystem/classificationSystemCitation</u>
	<u>TaxonomicCoverage/taxonomicSystem/classificationSystem/classificationSystemModifications</u>
	<u>TaxonomicCoverage/taxonomicSystem/identificationReference</u>
	<u>TaxonomicCoverage/taxonomicSystem/identifierName</u>
	<u>TaxonomicCoverage/taxonomicSystem/taxonomicCompleteness</u>
	<u>TaxonomicCoverage/taxonomicSystem/taxonomicProcedures</u>
	<u>TaxonomicCoverage/taxonomicSystem/vouchers</u>
	<u>TaxonomicCoverage/taxonomicSystem/vouchers/repository</u>
	<u>TaxonomicCoverage/taxonomicSystem/vouchers/repository/originator</u>
	<u>TaxonomicCoverage/taxonomicSystem/vouchers/specimen</u>
	<u>TemporalCoverage/rangeOfDates</u>
	<u>TemporalCoverage/rangeOfDates/beginDate</u>

MÓDULO	ELEMENTOS
	<u>TemporalCoverage/rangeOfDates/endDate</u> <u>TemporalCoverage/singleDateTime</u>
eml-literature	<u>Article/ISSN</u> <u>Article/issue</u> <u>Article/journal</u> <u>Article/pageRange</u> <u>Article/publicationPlace</u> <u>Article/publisher</u> <u>Article/volume</u> <u>AudioVisual/ISBN</u> <u>AudioVisual/performer</u> <u>AudioVisual/publicationPlace</u> <u>AudioVisual/publisher</u> <u>Book/ISBN</u> <u>Book/edition</u> <u>Book/numberOfVolumes</u> <u>Book/publicationPlace</u> <u>Book/publisher</u> <u>Book/totalFigures</u> <u>Book/totalPages</u> <u>Book/totalTables</u> <u>Book/volume</u> <u>Chapter/bookTitle</u> <u>Chapter/chapterNumber</u> <u>Chapter/editor</u> <u>Chapter/pageRange</u> <u>CitationListType/bibtex</u> <u>CitationListType/citation</u> <u>CitationType/article</u> <u>CitationType/audioVisual</u> <u>CitationType/bibtex</u> <u>CitationType/book</u> <u>CitationType/chapter</u> <u>CitationType/conferenceProceedings</u> <u>CitationType/contact</u> <u>CitationType/editedBook</u> <u>CitationType/generic</u> <u>CitationType/manuscript</u> <u>CitationType/map</u> <u>CitationType/personalCommunication</u> <u>CitationType/presentation</u> <u>CitationType/report</u> <u>CitationType/thesis</u>

MÓDULO	ELEMENTOS
	<u>ConferenceProceedings/conferenceDate</u> <u>ConferenceProceedings/conferenceLocation</u> <u>ConferenceProceedings/conferenceName</u> <u>Generic/ISBN</u> <u>Generic/ISSN</u> <u>Generic/edition</u> <u>Generic/numberOfVolumes</u> <u>Generic/originalPublication</u> <u>Generic/publicationPlace</u> <u>Generic/publisher</u> <u>Generic/referenceType</u> <u>Generic/reprintEdition</u> <u>Generic/reviewedItem</u> <u>Generic/totalFigures</u> <u>Generic/totalPages</u> <u>Generic/totalTables</u> <u>Generic/volume</u> <u>Manuscript/institution</u> <u>Manuscript/totalPages</u> <u>Map/edition</u> <u>Map/geographicCoverage</u> <u>Map/publisher</u> <u>Map/scale</u> <u>PersonalCommunication/communicationType</u> <u>PersonalCommunication/publicationPlace</u> <u>PersonalCommunication/publisher</u> <u>PersonalCommunication/recipient</u> <u>Presentation/conferenceDate</u> <u>Presentation/conferenceLocation</u> <u>Presentation/conferenceName</u> <u>Report/publicationPlace</u> <u>Report/publisher</u> <u>Report/reportNumber</u> <u>Report/totalPages</u> <u>Thesis/degree</u> <u>Thesis/institution</u> <u>Thesis/totalPages</u> <u>citation</u>
eml-access	<u>AccessRule/permission</u> <u>AccessRule/principal</u> <u>AccessType/allow</u> <u>AccessType/deny</u> <u>access</u>

MÓDULO	ELEMENTOS
eml-project	<u>AwardType/awardNumber</u> <u>AwardType/awardUrl</u> <u>AwardType/funderIdentifier</u> <u>AwardType/funderName</u> <u>AwardType/title</u> <u>ResearchProjectType/abstract</u> <u>ResearchProjectType/award</u> <u>ResearchProjectType/designDescription</u> <u>ResearchProjectType/designDescription/citation</u> <u>ResearchProjectType/designDescription/description</u> <u>ResearchProjectType/funding</u> <u>ResearchProjectType/personnel</u> <u>ResearchProjectType/personnel/role</u> <u>ResearchProjectType/relatedProject</u> <u>ResearchProjectType/studyAreaDescription</u> <u>ResearchProjectType/studyAreaDescription/citation</u> <u>ResearchProjectType/studyAreaDescription/coverage</u> <u>ResearchProjectType/studyAreaDescription/descriptor</u> <u>ResearchProjectType/studyAreaDescription/descriptor/citation</u> <u>ResearchProjectType/studyAreaDescription/descriptor/descriptorValue</u> <u>ResearchProjectType/title</u> <u>researchProject</u>
eml-semantic	<u>SemanticAnnotation/propertyURI</u> <u>SemanticAnnotation/valueURI</u> <u>annotation</u>
eml-dataset	<u>DatasetType/acknowledgements</u> <u>DatasetType/contact</u> <u>DatasetType/dataTable</u> <u>DatasetType/gettingStarted</u> <u>DatasetType/introduction</u> <u>DatasetType/literatureCited</u> <u>DatasetType/maintenance</u> <u>DatasetType/methods</u> <u>DatasetType/otherEntity</u> <u>DatasetType/project</u> <u>DatasetType/pubPlace</u> <u>DatasetType/publisher</u> <u>DatasetType/purpose</u> <u>DatasetType/referencePublication</u> <u>DatasetType/spatialRaster</u> <u>DatasetType/spatialVector</u> <u>DatasetType/storedProcedure</u> <u>DatasetType/usageCitation</u>

MÓDULO	ELEMENTOS
	<u>DatasetType/view</u> <u>MaintenanceType/changeHistory</u> <u>MaintenanceType/changeHistory/changeDate</u> <u>MaintenanceType/changeHistory/changeScope</u> <u>MaintenanceType/changeHistory/comment</u> <u>MaintenanceType/changeHistory/oldValue</u> <u>MaintenanceType/description</u> <u>MaintenanceType/maintenanceUpdateFrequency</u> <u>dataset</u>
eml-entity	<u>EntityGroup/additionalInfo</u> <u>EntityGroup/alternateIdentifier</u> <u>EntityGroup/annotation</u> <u>EntityGroup/coverage</u> <u>EntityGroup/entityDescription</u> <u>EntityGroup/entityName</u> <u>EntityGroup/methods</u> <u>EntityGroup/physical</u> <u>OtherEntityType/attributeList</u> <u>OtherEntityType/constraint</u> <u>OtherEntityType/entityType</u> <u>otherEntity</u>
eml-physical	<u>PhysicalDistributionType/access</u> <u>PhysicalDistributionType/inline</u> <u>PhysicalDistributionType/offline</u> <u>PhysicalDistributionType/online</u> <u>PhysicalOnlineType/connection</u> <u>PhysicalOnlineType/onlineDescription</u> <u>PhysicalOnlineType/url</u> <u>PhysicalType/authentication</u> <u>PhysicalType/characterEncoding</u> <u>PhysicalType/compressionMethod</u> <u>PhysicalType/dataFormat</u> <u>PhysicalType/dataFormat/binaryRasterFormat</u> <u>PhysicalType/dataFormat/binaryRasterFormat/bandgapbytes</u> <u>PhysicalType/dataFormat/binaryRasterFormat/bandrowbytes</u> <u>PhysicalType/dataFormat/binaryRasterFormat/byteorder</u> <u>PhysicalType/dataFormat/binaryRasterFormat/multiBand</u> <u>PhysicalType/dataFormat/binaryRasterFormat/multiBand/layout</u> <u>PhysicalType/dataFormat/binaryRasterFormat/multiBand/nbands</u> <u>PhysicalType/dataFormat/binaryRasterFormat/nbits</u> <u>PhysicalType/dataFormat/binaryRasterFormat/rowColumnOrientation</u> <u>PhysicalType/dataFormat/binaryRasterFormat/skipbytes</u> <u>PhysicalType/dataFormat/binaryRasterFormat/totalrowbytes</u>

MÓDULO	ELEMENTOS
	<p><u>PhysicalType/dataFormat/externallyDefinedFormat</u></p> <p><u>PhysicalType/dataFormat/externallyDefinedFormat/citation</u></p> <p><u>PhysicalType/dataFormat/externallyDefinedFormat/formatName</u></p> <p><u>PhysicalType/dataFormat/externallyDefinedFormat/formatVersion</u></p> <p><u>PhysicalType/dataFormat/textFormat</u></p> <p><u>PhysicalType/dataFormat/textFormat/attributeOrientation</u></p> <p><u>PhysicalType/dataFormat/textFormat/complex</u></p> <p><u>PhysicalType/dataFormat/textFormat/complex/textDelimited</u></p> <p><u>PhysicalType/dataFormat/textFormat/complex/textDelimited/collapseDelimiters</u></p> <p><u>PhysicalType/dataFormat/textFormat/complex/textDelimited/fieldDelimiter</u></p> <p><u>PhysicalType/dataFormat/textFormat/complex/textDelimited/lineNumber</u></p> <p><u>PhysicalType/dataFormat/textFormat/complex/textDelimited/literalCharacter</u></p> <p><u>PhysicalType/dataFormat/textFormat/complex/textDelimited/quoteCharacter</u></p> <p><u>PhysicalType/dataFormat/textFormat/complex/textFixed</u></p> <p><u>PhysicalType/dataFormat/textFormat/complex/textFixed/fieldStartColumn</u></p> <p><u>PhysicalType/dataFormat/textFormat/complex/textFixed/fieldWidth</u></p> <p><u>PhysicalType/dataFormat/textFormat/complex/textFixed/lineNumber</u></p> <p><u>PhysicalType/dataFormat/textFormat/maxRecordLength</u></p> <p><u>PhysicalType/dataFormat/textFormat/numFooterLines</u></p> <p><u>PhysicalType/dataFormat/textFormat/numHeaderLines</u></p> <p><u>PhysicalType/dataFormat/textFormat/numPhysicalLinesPerRecord</u></p> <p><u>PhysicalType/dataFormat/textFormat/physicalLineDelimiter</u></p> <p><u>PhysicalType/dataFormat/textFormat/recordDelimiter</u></p> <p><u>PhysicalType/dataFormat/textFormat/simpleDelimited</u></p> <p><u>PhysicalType/dataFormat/textFormat/simpleDelimited/collapseDelimiters</u></p> <p><u>PhysicalType/dataFormat/textFormat/simpleDelimited/fieldDelimiter</u></p> <p><u>PhysicalType/dataFormat/textFormat/simpleDelimited/literalCharacter</u></p> <p><u>PhysicalType/dataFormat/textFormat/simpleDelimited/quoteCharacter</u></p> <p><u>PhysicalType/distribution</u></p> <p><u>PhysicalType/encodingMethod</u></p> <p><u>PhysicalType/objectName</u></p> <p><u>PhysicalType/size</u></p> <p><u>physical</u></p>
eml-methods	<p><u>MethodsType/methodStep</u></p> <p><u>MethodsType/methodStep/dataSource</u></p> <p><u>MethodsType/qualityControl</u></p> <p><u>MethodsType/sampling</u></p> <p><u>MethodsType/sampling/citation</u></p> <p><u>MethodsType/sampling/samplingDescription</u></p> <p><u>MethodsType/sampling/spatialSamplingUnits</u></p>

MÓDULO	ELEMENTOS
	MethodsType/sampling/spatialSamplingUnits/coverage MethodsType/sampling/spatialSamplingUnits/referencedEntityId MethodsType/sampling/studyExtent MethodsType/sampling/studyExtent/coverage MethodsType/sampling/studyExtent/description ProcedureStepType/citation ProcedureStepType/description ProcedureStepType/instrumentation ProcedureStepType/protocol ProcedureStepType/software ProcedureStepType/subStep methods
eml-software	DependencyType/action DependencyType/software SoftwareType/dependency SoftwareType/implementation SoftwareType/implementation/checksum SoftwareType/implementation/dependency SoftwareType/implementation/diskUsage SoftwareType/implementation/distribution SoftwareType/implementation/language SoftwareType/implementation/language/LanguageCodeStandard SoftwareType/implementation/language/LanguageValue SoftwareType/implementation/machineProcessor SoftwareType/implementation/operatingSystem SoftwareType/implementation/programmingLanguage SoftwareType/implementation/runtimeMemoryUsage SoftwareType/implementation/size SoftwareType/implementation/virtualMachine SoftwareType/license SoftwareType/licenseURL SoftwareType/project SoftwareType/version dependency software
eml-protocol	ProtocolType/proceduralStep protocol
eml-attribute	Accuracy/attributeAccuracyReport Accuracy/quantitativeAttributeAccuracyAssessment Accuracy/quantitativeAttributeAccuracyAssessment/attributeAccuracyExplanation Accuracy/quantitativeAttributeAccuracyAssessment/attributeAccuracyValue AttributeListType/attribute

MÓDULO	ELEMENTOS
	<u>AttributeType/accuracy</u>
	<u>AttributeType/annotation</u>
	<u>AttributeType/attributeDefinition</u>
	<u>AttributeType/attributeLabel</u>
	<u>AttributeType/attributeName</u>
	<u>AttributeType/coverage</u>
	<u>AttributeType/measurementScale</u>
	<u>AttributeType/measurementScale/dateTime</u>
	<u>AttributeType/measurementScale/dateTime/dateTimeDomain</u>
	<u>AttributeType/measurementScale/dateTime/dateTimePrecision</u>
	<u>AttributeType/measurementScale/dateTime/formatString</u>
	<u>AttributeType/measurementScale/interval</u>
	<u>AttributeType/measurementScale/interval/numericDomain</u>
	<u>AttributeType/measurementScale/interval/precision</u>
	<u>AttributeType/measurementScale/interval/unit</u>
	<u>AttributeType/measurementScale/nominal</u>
	<u>AttributeType/measurementScale/nominal/nonNumericDomain</u>
	<u>AttributeType/measurementScale/ordinal</u>
	<u>AttributeType/measurementScale/ordinal/nonNumericDomain</u>
	<u>AttributeType/measurementScale/ratio</u>
	<u>AttributeType/measurementScale/ratio/numericDomain</u>
	<u>AttributeType/measurementScale/ratio/precision</u>
	<u>AttributeType/measurementScale/ratio/unit</u>
	<u>AttributeType/methods</u>
	<u>AttributeType/missingValueCode</u>
	<u>AttributeType/missingValueCode/code</u>
	<u>AttributeType/missingValueCode/codeExplanation</u>
	<u>AttributeType/storageType</u>
	<u>BoundsDateGroup/bounds</u>
	<u>BoundsDateGroup/bounds/maximum</u>
	<u>BoundsDateGroup/bounds/minimum</u>
	<u>BoundsGroup/bounds</u>
	<u>BoundsGroup/bounds/maximum</u>
	<u>BoundsGroup/bounds/minimum</u>
	<u>NonNumericDomainType/enumeratedDomain</u>
	<u>NonNumericDomainType/enumeratedDomain/codeDefinition</u>
	<u>NonNumericDomainType/enumeratedDomain/codeDefinition/code</u>
	<u>NonNumericDomainType/enumeratedDomain/codeDefinition/definition</u>
	<u>NonNumericDomainType/enumeratedDomain/codeDefinition/source</u>
	<u>NonNumericDomainType/enumeratedDomain/entityCodeList</u>
	<u>NonNumericDomainType/enumeratedDomain/entityCodeList/definitionAttributeReference</u>
	<u>NonNumericDomainType/enumeratedDomain/entityCodeList/entityReference</u>

MÓDULO	ELEMENTOS
	<p>nce</p> <p>NonNumericDomainType/enumeratedDomain/entityCodeList/orderAttributeReference</p> <p>NonNumericDomainType/enumeratedDomain/entityCodeList/valueAttributeReference</p> <p>NonNumericDomainType/enumeratedDomain/externalCodeSet</p> <p>NonNumericDomainType/enumeratedDomain/externalCodeSet/citation</p> <p>NonNumericDomainType/enumeratedDomain/externalCodeSet/codesetName</p> <p>NonNumericDomainType/enumeratedDomain/externalCodeSet/codesetURL</p> <p>NonNumericDomainType/textDomain</p> <p>NonNumericDomainType/textDomain/definition</p> <p>NonNumericDomainType/textDomain/pattern</p> <p>NonNumericDomainType/textDomain/source</p> <p>NumericDomainType/numberType</p> <p>UnitType/customUnit</p> <p>UnitType/standardUnit</p> <p>attribute</p> <p>attributeList</p>
eml-constraint	<p>ConstraintBaseGroup/constraintDescription</p> <p>ConstraintBaseGroup/constraintName</p> <p>ConstraintType/checkConstraint</p> <p>ConstraintType/checkConstraint/checkCondition</p> <p>ConstraintType/foreignKey</p> <p>ConstraintType/joinCondition</p> <p>ConstraintType/joinCondition/referencedKey</p> <p>ConstraintType/joinCondition/referencedKey/attributeReference</p> <p>ConstraintType/notNullConstraint</p> <p>ConstraintType/notNullConstraint/key</p> <p>ConstraintType/notNullConstraint/key/attributeReference</p> <p>ConstraintType/primaryKey</p> <p>ConstraintType/primaryKey/key</p> <p>ConstraintType/primaryKey/key/attributeReference</p> <p>ConstraintType/uniqueKey</p> <p>ConstraintType/uniqueKey/key</p> <p>ConstraintType/uniqueKey/key/attributeReference</p> <p>ForeignKeyGroup/cardinality</p> <p>ForeignKeyGroup/cardinality/childOccurrences</p> <p>ForeignKeyGroup/cardinality/parentOccurrences</p> <p>ForeignKeyGroup/entityReference</p> <p>ForeignKeyGroup/key</p> <p>ForeignKeyGroup/key/attributeReference</p> <p>ForeignKeyGroup/relationshipType</p>

MÓDULO	ELEMENTOS
eml-dataTable	<u>DataTableType/attributeList</u> <u>DataTableType/caseSensitive</u> <u>DataTableType/constraint</u> <u>DataTableType/numberOfRecords</u> <u>dataTable</u>
eml-spatialRaster	<u>BandType/highWavelength</u> <u>BandType/lowWaveLength</u> <u>BandType/peakResponse</u> <u>BandType/sequenceIdentifier</u> <u>BandType/waveLengthUnits</u> <u>DataQuality/accuracyReport</u> <u>DataQuality/quantitativeAccuracyReport</u> <u>DataQuality/quantitativeAccuracyReport/quantitativeAccuracyMethod</u> <u>DataQuality/quantitativeAccuracyReport/quantitativeAccuracyValue</u> <u>SpatialRasterType/attributeList</u> <u>SpatialRasterType/cellGeometry</u> <u>SpatialRasterType/cellSizeXDirection</u> <u>SpatialRasterType/cellSizeYDirection</u> <u>SpatialRasterType/columns</u> <u>SpatialRasterType/constraint</u> <u>SpatialRasterType/georeferenceInfo</u> <u>SpatialRasterType/georeferenceInfo/bilinearFit</u> <u>SpatialRasterType/georeferenceInfo/bilinearFit/xIntercept</u> <u>SpatialRasterType/georeferenceInfo/bilinearFit/xSlope</u> <u>SpatialRasterType/georeferenceInfo/bilinearFit/yIntercept</u> <u>SpatialRasterType/georeferenceInfo/bilinearFit/ySlope</u> <u>SpatialRasterType/georeferenceInfo/controlPoint</u> <u>SpatialRasterType/georeferenceInfo/controlPoint/column</u> <u>SpatialRasterType/georeferenceInfo/controlPoint/pointInPixel</u> <u>SpatialRasterType/georeferenceInfo/controlPoint/row</u> <u>SpatialRasterType/georeferenceInfo/controlPoint/xCoordinate</u> <u>SpatialRasterType/georeferenceInfo/controlPoint/yCoordinate</u> <u>SpatialRasterType/georeferenceInfo/cornerPoint</u> <u>SpatialRasterType/georeferenceInfo/cornerPoint/corner</u> <u>SpatialRasterType/georeferenceInfo/cornerPoint/pointInPixel</u> <u>SpatialRasterType/georeferenceInfo/cornerPoint/xCoordinate</u> <u>SpatialRasterType/georeferenceInfo/cornerPoint/yCoordinate</u> <u>SpatialRasterType/horizontalAccuracy</u> <u>SpatialRasterType/imageDescription</u> <u>SpatialRasterType/imageDescription/bandDescription</u> <u>SpatialRasterType/imageDescription/cameraCalibrationInformationAvailability</u> <u>SpatialRasterType/imageDescription/cloudCoverPercentage</u>

MÓDULO	ELEMENTOS
	<p><u>SpatialRasterType/imageDescription/compressionGenerationQuality</u> <u>SpatialRasterType/imageDescription/filmDistortionInformationAvailability</u> <u>SpatialRasterType/imageDescription/illuminationAzimuthAngle</u> <u>SpatialRasterType/imageDescription/illuminationElevationAngle</u> <u>SpatialRasterType/imageDescription/imageOrientationAngle</u> <u>SpatialRasterType/imageDescription/imageQualityCode</u> <u>SpatialRasterType/imageDescription/imagingCondition</u> <u>SpatialRasterType/imageDescription/lensDistortionInformationAvailability</u> <u>SpatialRasterType/imageDescription/preProcessingTypeCode</u> <u>SpatialRasterType/imageDescription/radiometricDataAvailability</u> <u>SpatialRasterType/imageDescription/triangulationIndicator</u> <u>SpatialRasterType/numberOfBands</u> <u>SpatialRasterType/offset</u> <u>SpatialRasterType/rasterOrigin</u> <u>SpatialRasterType/rows</u> <u>SpatialRasterType/scaleFactor</u> <u>SpatialRasterType/spatialReference</u> <u>SpatialRasterType/toneGradation</u> <u>SpatialRasterType/verticalAccuracy</u> <u>SpatialRasterType/verticals</u> <u>spatialRaster</u></p>
emi-spatialReference	<p><u>SpatialReferenceType/horizCoordSysDef</u> <u>SpatialReferenceType/horizCoordSysName</u> <u>SpatialReferenceType/vertCoordSys</u> <u>SpatialReferenceType/vertCoordSys/altitudeSysDef</u> <u>SpatialReferenceType/vertCoordSys/altitudeSysDef/altitudeDatumName</u> <u>SpatialReferenceType/vertCoordSys/altitudeSysDef/altitudeDistanceUnits</u> <u>SpatialReferenceType/vertCoordSys/altitudeSysDef/altitudeEncodingMethod</u> <u>SpatialReferenceType/vertCoordSys/altitudeSysDef/altitudeResolution</u> <u>SpatialReferenceType/vertCoordSys/depthSysDef</u> <u>SpatialReferenceType/vertCoordSys/depthSysDef/depthDatumName</u> <u>SpatialReferenceType/vertCoordSys/depthSysDef/depthDistanceUnits</u> <u>SpatialReferenceType/vertCoordSys/depthSysDef/depthEncodingMethod</u> <u>SpatialReferenceType/vertCoordSys/depthSysDef/depthResolution</u> <u>geogCoordSysType/datum</u> <u>geogCoordSysType/primeMeridian</u> <u>geogCoordSysType/spheroid</u> <u>geogCoordSysType/unit</u> <u>horizCoordSysType/geogCoordSys</u> <u>horizCoordSysType/projCoordSys</u> <u>horizCoordSysType/projCoordSys/geogCoordSys</u> <u>horizCoordSysType/projCoordSys/projection</u></p>

MÓDULO	ELEMENTOS
	horizCoordSysType/projCoordSys/projection/parameter horizCoordSysType/projCoordSys/projection/unit projectionList projectionList/horizCoordSysDef spatialReference
eml-spatialVector	DataQuality/accuracyReport DataQuality/quantitativeAccuracyReport DataQuality/quantitativeAccuracyReport/quantitativeAccuracyMethod DataQuality/quantitativeAccuracyReport/quantitativeAccuracyValue SpatialVectorType/attributeList SpatialVectorType/constraint SpatialVectorType/geometricObjectCount SpatialVectorType/geometry SpatialVectorType/horizontalAccuracy SpatialVectorType/spatialReference SpatialVectorType/topologyLevel SpatialVectorType/verticalAccuracy spatialVector
eml-storedProcedure	ParameterType/domainDescription ParameterType/name ParameterType/repeats ParameterType/required StoredProcedureType/attributeList StoredProcedureType/constraint StoredProcedureType/parameter storedProcedure
eml-view	ViewType/attributeList ViewType/constraint ViewType/queryStatement view

ANEXO 5. Lista de TERMOS e CONCEITOS das propriedades apresentadas nos núcleos de um DwC-A.

NÚCLEO			CLASSE	PROPRIEDADES	DESCRIÇÃO
Taxon	Ocurrence	Event			
X	X		Taxon	taxonID	An identifier for the set of taxon information (data associated with the Taxon class). May be a global unique identifier or an identifier specific to the data set.
X	X		Taxon	scientificNameID	An identifier for the nomenclatural (not taxonomic) details of a scientific name.
X	X		Taxon	acceptedNameUsageID	An identifier for the name usage (documented meaning of the name according to a source) of the currently valid (zoological) or accepted (botanical) taxon.
X	X		Taxon	parentNameUsageID	An identifier for the name usage (documented meaning of the name according to a source) of the direct, most proximate higher-rank parent taxon (in a classification) of the most specific element of the scientificName.
X	X		Taxon	originalNameUsageID	An identifier for the name usage (documented meaning of the name according to a source) in which the terminal element of the scientificName was originally established under the rules of the associated nomenclaturalCode.
X	X		Taxon	nameAccordingToID	An identifier for the source in which the specific taxon concept circumscription is defined or implied. See nameAccordingTo.

NÚCLEO			CLASSE	PROPRIEDADES	DESCRIÇÃO
Taxon	Ocurrence	Event			
X	X		Taxon	namePublishedInID	An identifier for the publication in which the scientificName was originally established under the rules of the associated nomenclaturalCode.
X	X		Taxon	taxonConceptID	An identifier for the taxonomic concept to which the record refers - not for the nomenclatural details of a taxon.
X	X		Taxon	scientificName	The full scientific name, with authorship and date information if known. When forming part of an Identification, this should be the name in lowest level taxonomic rank that can be determined. This term should not contain identification qualifications, which should instead be supplied in the IdentificationQualifier term.
X	X		Taxon	acceptedNameUsage	The full name, with authorship and date information if known, of the currently valid (zoological) or accepted (botanical) taxon.
X	X		Taxon	parentNameUsage	The full name, with authorship and date information if known, of the direct, most proximate higher-rank parent taxon (in a classification) of the most specific element of the scientificName.
X	X		Taxon	originalNameUsage	The taxon name, with authorship and date information if known, as it originally appeared when first established under the rules of the associated nomenclaturalCode. The basionym (botany) or basonym (bacteriology) of the scientificName or the senior/earlier homonym for replaced names.
X	X		Taxon	nameAccordingTo	The reference to the source in which the specific taxon concept circumscription is defined or implied - traditionally signified by the Latin "sensu" or "sec." (from secundum, meaning "according to"). For taxa that result from identifications, a reference to the keys, monographs, experts and other sources should be given.
X	X		Taxon	namePublishedIn	A reference for the publication in which the scientificName was originally established under the rules of the associated nomenclaturalCode.
X	X		Taxon	namePublishedInYear	The four-digit year in which the scientificName was published.
X	X		Taxon	higherClassification	A list (concatenated and separated) of taxa names terminating at the rank immediately superior to the taxon referenced in the taxon record.
X	X		Taxon	kingdom	The full scientific name of the kingdom in which the taxon is classified.

NÚCLEO			CLASSE	PROPRIEDADES	DESCRIÇÃO
Taxon	Ocurrence	Event			
X	X		Taxon	phylum	The full scientific name of the phylum or division in which the taxon is classified.
X	X		Taxon	class	The full scientific name of the class in which the taxon is classified.
X	X		Taxon	order	The full scientific name of the order in which the taxon is classified.
X	X		Taxon	family	The full scientific name of the family in which the taxon is classified.
X	X		Taxon	genus	The full scientific name of the genus in which the taxon is classified.
X	X		Taxon	subgenus	The full scientific name of the subgenus in which the taxon is classified. Values should include the genus to avoid homonym confusion.
X	X		Taxon	specificEpithet	The name of the first or species epithet of the scientificName.
X	X		Taxon	infraspecificEpithet	The name of the lowest or terminal infraspecific epithet of the scientificName, excluding any rank designation.
X	X		Taxon	taxonRank	The taxonomic rank of the most specific name in the scientificName.
X	X		Taxon	verbatimTaxonRank	The taxonomic rank of the most specific name in the scientificName as it appears in the original record.
X	X		Taxon	scientificNameAuthorship	The authorship information for the scientificName formatted according to the conventions of the applicable nomenclaturalCode.
X	X		Taxon	vernacularName	A common or vernacular name.
X	X		Taxon	nomenclaturalCode	The nomenclatural code (or codes in the case of an ambiregnal name) under which the scientificName is constructed. Recommended best practice is to use a controlled vocabulary (e.g. ICN, ICZN, BC, ICNCP, BioCode).
X	X		Taxon	taxonomicStatus	The status of the use of the scientificName as a label for a taxon. Requires taxonomic opinion to define the scope of a taxon. Rules of priority then are used to define the taxonomic status of the nomenclature contained in that scope, combined with the experts opinion. It must be linked to a specific taxonomic reference that defines the concept.
X	X		Taxon	nomenclaturalStatus	The status related to the original publication of the name and its conformance to the relevant rules of nomenclature. It is based essentially on an algorithm according to the business rules of the code. It requires no taxonomic opinion.

NÚCLEO			CLASSE	PROPRIEDADES	DESCRIÇÃO
Taxon	Ocurrence	Event			
X	X		Taxon	taxonRemarks	Comments or notes about the taxon or name.
	X	X	Event	eventID	An identifier for the set of information associated with an Event (something that occurs at a place and time). May be a global unique identifier or an identifier specific to the data set. see also http://rs.tdwg.org/dwc/terms/index.htm#eventID
	X	X	Event	parentEventID	An event identifier for the super event which is composed of one or more sub-sampling events. The value must refer to an existing eventID. If the identifier is local it must exist within the given dataset. May be a globally unique identifier or an identifier specific to the data set. see also http://rs.tdwg.org/dwc/terms/index.htm#parentEventID
	X	X	Event	samplingProtocol	The name of, reference to, or description of the method or protocol used during an Event. see also http://rs.tdwg.org/dwc/terms/index.htm#samplingProtocol
	X	X	Event	sampleSizeValue	A numeric value for a measurement of the size (time duration, length, area, or volume) of a sample in a sampling event. see also http://rs.tdwg.org/dwc/terms/index.htm#sampleSizeValue
	X	X	Event	sampleSizeUnit	The unit of measurement of the size (time duration, length, area, or volume) of a sample in a sampling event. see also http://rs.tdwg.org/dwc/terms/index.htm#sampleSizeUnit
	X	X	Event	samplingEffort	The amount of effort expended during an Event. see also http://rs.tdwg.org/dwc/terms/index.htm#samplingEffort
	X	X	Event	eventDate	The date-time or interval during which an Event occurred. For occurrences, this is the date-time when the event was recorded. Not suitable for a time in a geological context. Recommended best practice is to use an encoding scheme, such as ISO 8601:2004(E). see also http://rs.tdwg.org/dwc/terms/index.htm#eventDate
	X	X	Event	eventTime	The time or interval during which an Event occurred. Recommended best practice is to use an encoding scheme, such as ISO 8601:2004(E). see also http://rs.tdwg.org/dwc/terms/index.htm#eventTime
	X	X	Event	startDayOfYear	The earliest ordinal day of the year on which the Event occurred (1 for January 1, 365 for December 31, except in a leap year, in which case it is 366). see also http://rs.tdwg.org/dwc/terms/index.htm#startDayOfYear

NÚCLEO			CLASSE	PROPRIEDADES	DESCRIÇÃO
Taxon	Ocurrence	Event			
	X	X	Event	endDayOfYear	The latest ordinal day of the year on which the Event occurred (1 for January 1, 365 for December 31, except in a leap year, in which case it is 366). see also http://rs.tdwg.org/dwc/terms/index.htm#endDayOfYear
	X	X	Event	year	The four-digit year in which the Event occurred, according to the Common Era Calendar. see also http://rs.tdwg.org/dwc/terms/index.htm#year
	X	X	Event	month	The ordinal month in which the Event occurred.
	X	X	Event	day	The integer day of the month on which the Event occurred.
	X	X	Event	verbatimEventDate	The verbatim original representation of the date and time information for an Event.
	X	X	Event	habitat	A category or description of the habitat in which the Event occurred. see also http://rs.tdwg.org/dwc/terms/index.htm#habitat
	X	X	Event	fieldNumber	An identifier given to the event in the field. Often serves as a link between field notes and the Event. see also http://rs.tdwg.org/dwc/terms/index.htm#fieldNumber
	X	X	Event	fieldNotes	One of a) an indicator of the existence of, b) a reference to (publication, URI), or c) the text of notes taken in the field about the Event. see also http://rs.tdwg.org/dwc/terms/index.htm#fieldNotes
	X	X	Event	eventRemarks	Comments or notes about the Event. see also http://rs.tdwg.org/dwc/terms/index.htm#eventRemarks
	X	X	GeologicalContext	geologicalContextID	An identifier for the set of information associated with a GeologicalContext (the location within a geological context, such as stratigraphy). May be a global unique identifier or an identifier specific to the data set. see also http://rs.tdwg.org/dwc/terms/index.htm#geologicalContextID
	X	X	GeologicalContext	earliestEonOrLowestEonothem	The full name of the earliest possible geochronologic eon or lowest chronostratigraphic eonothem or the informal name ("Precambrian") attributable to the stratigraphic horizon from which the cataloged item was collected. see also http://rs.tdwg.org/dwc/terms/index.htm#earliestEonOrLowestEonothem
	X	X	GeologicalContext	latestEonOrHighestEonothem	The full name of the latest possible geochronologic eon or highest chronostratigraphic eonothem or the informal name ("Precambrian") attributable to the stratigraphic horizon from which the cataloged item was collected. see also http://rs.tdwg.org/dwc/terms/index.htm#latestEonOrHighestEonothem

NÚCLEO			CLASSE	PROPRIEDADES	DESCRIÇÃO
Taxon	Ocurrence	Event			
	X	X	GeologicalContext	earliestEraOrLowestErathem	The full name of the earliest possible geochronologic era or lowest chronostratigraphic erathem attributable to the stratigraphic horizon from which the cataloged item was collected. see also http://rs.tdwg.org/dwc/terms/index.htm#earliestEraOrLowestErathem
	X	X	GeologicalContext	latestEraOrHighestErathem	The full name of the latest possible geochronologic era or highest chronostratigraphic erathem attributable to the stratigraphic horizon from which the cataloged item was collected. see also http://rs.tdwg.org/dwc/terms/index.htm#latestEraOrHighestErathem
	X	X	GeologicalContext	earliestPeriodOrLowestSystem	The full name of the earliest possible geochronologic period or lowest chronostratigraphic system attributable to the stratigraphic horizon from which the cataloged item was collected. see also http://rs.tdwg.org/dwc/terms/index.htm#earliestPeriodOrLowestSystem
	X	X	GeologicalContext	latestPeriodOrHighestSystem	The full name of the latest possible geochronologic period or highest chronostratigraphic system attributable to the stratigraphic horizon from which the cataloged item was collected. see also http://rs.tdwg.org/dwc/terms/index.htm#latestPeriodOrHighestSystem
	X	X	GeologicalContext	earliestEpochOrLowestSeries	The full name of the earliest possible geochronologic epoch or lowest chronostratigraphic series attributable to the stratigraphic horizon from which the cataloged item was collected. see also http://rs.tdwg.org/dwc/terms/index.htm#earliestEpochOrLowestSeries
	X	X	GeologicalContext	latestEpochOrHighestSeries	The full name of the latest possible geochronologic epoch or highest chronostratigraphic series attributable to the stratigraphic horizon from which the cataloged item was collected. see also http://rs.tdwg.org/dwc/terms/index.htm#latestEpochOrHighestSeries
	X	X	GeologicalContext	earliestAgeOrLowestStage	The full name of the earliest possible geochronologic age or lowest chronostratigraphic stage attributable to the stratigraphic horizon from which the cataloged item was collected. see also http://rs.tdwg.org/dwc/terms/index.htm#earliestAgeOrLowestStage
	X	X	GeologicalContext	latestAgeOrHighestStage	The full name of the latest possible geochronologic age or highest chronostratigraphic stage attributable to the stratigraphic horizon from which the cataloged item was collected. see also http://rs.tdwg.org/dwc/terms/index.htm#latestAgeOrHighestStage

NÚCLEO			CLASSE	PROPRIEDADES	DESCRIÇÃO
Taxon	Ocurrence	Event			
	X	X	GeologicalContext	lowestBiostratigraphicZone	The full name of the lowest possible geological biostratigraphic zone of the stratigraphic horizon from which the cataloged item was collected. see also http://rs.tdwg.org/dwc/terms/index.htm#lowestBiostratigraphicZone
	X	X	GeologicalContext	highestBiostratigraphicZone	The full name of the highest possible geological biostratigraphic zone of the stratigraphic horizon from which the cataloged item was collected. see also http://rs.tdwg.org/dwc/terms/index.htm#highestBiostratigraphicZone
	X	X	GeologicalContext	lithostratigraphicTerms	The combination of all litho-stratigraphic names for the rock from which the cataloged item was collected. see also http://rs.tdwg.org/dwc/terms/index.htm#lithostratigraphicTerms
	X	X	GeologicalContext	group	The full name of the lithostratigraphic group from which the cataloged item was collected. see also http://rs.tdwg.org/dwc/terms/index.htm#group
	X	X	GeologicalContext	formation	The full name of the lithostratigraphic formation from which the cataloged item was collected. see also http://rs.tdwg.org/dwc/terms/index.htm#formation
	X	X	GeologicalContext	member	The full name of the lithostratigraphic member from which the cataloged item was collected. see also http://rs.tdwg.org/dwc/terms/index.htm#member
	X	X	GeologicalContext	bed	The full name of the lithostratigraphic bed from which the cataloged item was collected. see also http://rs.tdwg.org/dwc/terms/index.htm#bed
	X		Identification	identificationID	An identifier for the Identification (the body of information associated with the assignment of a scientific name). May be a global unique identifier or an identifier specific to the data set. see also http://rs.tdwg.org/dwc/terms/index.htm#identificationID
	X		Identification	identifiedBy	A list (concatenated and separated) of names of people, groups, or organizations who assigned the Taxon to the subject. The recommended best practice is to separate the values with a vertical bar (' '). see also http://rs.tdwg.org/dwc/terms/index.htm#identifiedBy
	X		Identification	dateIdentified	The date on which the subject was identified as representing the Taxon. Recommended best practice is to use an encoding scheme, such as ISO 8601:2004(E). see also http://rs.tdwg.org/dwc/terms/index.htm#dateIdentified
	X		Identification	identificationReferences	A list (concatenated and separated) of references (publication, global unique identifier, URI) used in the Identification. The recommended best practice is to separate the values with a vertical bar (' '). see also http://rs.tdwg.org/dwc/terms/index.htm#identificationReferences

NÚCLEO			CLASSE	PROPRIEDADES	DESCRIÇÃO
Taxon	Ocurrence	Event			
	X		Identification	identificationRemarks	Comments or notes about the Identification. see also http://rs.tdwg.org/dwc/terms/index.htm#identificationRemarks
	X		Identification	identificationQualifier	A brief phrase or a standard term ("cf.", "aff.") to express the determiner's doubts about the Identification. see also http://rs.tdwg.org/dwc/terms/index.htm#identificationQualifier
	X		Identification	identificationVerificationStatus	A categorical indicator of the extent to which the taxonomic identification has been verified to be correct. Recommended best practice is to use a controlled vocabulary such as that used in HISPID/ABCD. see also http://rs.tdwg.org/dwc/terms/index.htm#identificationVerificationStatus
	X		Identification	typeStatus	A list (concatenated and separated) of nomenclatural types (type status, typified scientific name, publication) applied to the subject. The recommended best practice is to separate the values with a vertical bar (' '). see also http://rs.tdwg.org/dwc/terms/index.htm#typeStatus
	X	X	Location	locationID	An identifier for the set of location information (data associated with dcterms:Location). May be a global unique identifier or an identifier specific to the data set. see also http://rs.tdwg.org/dwc/terms/index.htm#locationID
	X	X	Location	higherGeographyID	An identifier for the geographic region within which the Location occurred. Recommended best practice is to use an persistent identifier from a controlled vocabulary such as the Getty Thesaurus of Geographic Names. see also http://rs.tdwg.org/dwc/terms/index.htm#higherGeographyID
	X	X	Location	higherGeography	A list (concatenated and separated) of geographic names less specific than the information captured in the locality term. The recommended best practice is to separate the values with a vertical bar (' '). see also http://rs.tdwg.org/dwc/terms/index.htm#higherGeography
	X	X	Location	continent	The name of the continent in which the Location occurs. Recommended best practice is to use a controlled vocabulary such as the Getty Thesaurus of Geographic Names or the ISO 3166 Continent code. see also http://rs.tdwg.org/dwc/terms/index.htm#continent
	X	X	Location	waterBody	The name of the water body in which the Location occurs. Recommended best practice is to use a controlled vocabulary such as the Getty Thesaurus of Geographic Names. see also http://rs.tdwg.org/dwc/terms/index.htm#waterBody

NÚCLEO			CLASSE	PROPRIEDADES	DESCRIÇÃO
Taxon	Ocurrence	Event			
	X	X	Location	islandGroup	The name of the island group in which the Location occurs. Recommended best practice is to use a controlled vocabulary such as the Getty Thesaurus of Geographic Names. see also http://rs.tdwg.org/dwc/terms/index.htm#islandGroup
	X	X	Location	island	The name of the island on or near which the Location occurs. Recommended best practice is to use a controlled vocabulary such as the Getty Thesaurus of Geographic Names. see also http://rs.tdwg.org/dwc/terms/index.htm#island
	X	X	Location	country	The name of the country or major administrative unit in which the Location occurs. Recommended best practice is to use a controlled vocabulary such as the Getty Thesaurus of Geographic Names. see also http://rs.tdwg.org/dwc/terms/index.htm#country
	X	X	Location	countryCode	The standard code for the country in which the Location occurs. Recommended best practice is to use ISO 3166-1-alpha-2 country codes. see also http://rs.tdwg.org/dwc/terms/index.htm#countryCode
	X	X	Location	stateProvince	The name of the next smaller administrative region than country (state, province, canton, department, region, etc.) in which the Location occurs. see also http://rs.tdwg.org/dwc/terms/index.htm#stateProvince
	X	X	Location	county	The full, unabbreviated name of the next smaller administrative region than stateProvince (county, shire, department, etc.) in which the Location occurs. see also http://rs.tdwg.org/dwc/terms/index.htm#county
	X	X	Location	municipality	The full, unabbreviated name of the next smaller administrative region than county (city, municipality, etc.) in which the Location occurs. Do not use this term for a nearby named place that does not contain the actual location. see also http://rs.tdwg.org/dwc/terms/index.htm#municipality
	X	X	Location	locality	The specific description of the place. Less specific geographic information can be provided in other geographic terms (higherGeography, continent, country, stateProvince, county, municipality, waterBody, island, islandGroup). This term may contain information modified from the original to correct perceived errors or standardize the description. see also http://rs.tdwg.org/dwc/terms/index.htm#locality
	X	X	Location	verbatimLocality	The original textual description of the place. see also http://rs.tdwg.org/dwc/terms/index.htm#verbatimLocality

NÚCLEO			CLASSE	PROPRIEDADES	DESCRIÇÃO
Taxon	Ocurrence	Event			
	X	X	Location	verbatimElevation	The original description of the elevation (altitude, usually above sea level) of the Location. see also http://rs.tdwg.org/dwc/terms/index.htm#verbatimElevation
	X	X	Location	minimumElevationInMeters	The lower limit of the range of elevation (altitude, usually above sea level), in meters. see also http://rs.tdwg.org/dwc/terms/index.htm#minimumElevationInMeters
	X	X	Location	maximumElevationInMeters	The upper limit of the range of elevation (altitude, usually above sea level), in meters. see also http://rs.tdwg.org/dwc/terms/index.htm#maximumElevationInMeters
	X	X	Location	verbatimDepth	The original description of the depth below the local surface. see also http://rs.tdwg.org/dwc/terms/index.htm#verbatimDepth
	X	X	Location	minimumDepthInMeters	The lesser depth of a range of depth below the local surface, in meters. see also http://rs.tdwg.org/dwc/terms/index.htm#minimumDepthInMeters
	X	X	Location	maximumDepthInMeters	The greater depth of a range of depth below the local surface, in meters. see also http://rs.tdwg.org/dwc/terms/index.htm#maximumDepthInMeters
	X	X	Location	minimumDistanceAboveSurfaceInMeters	The lesser distance in a range of distance from a reference surface in the vertical direction, in meters. Use positive values for locations above the surface, negative values for locations below. If depth measures are given, the reference surface is the location given by the depth, otherwise the reference surface is the location given by the elevation. see also http://rs.tdwg.org/dwc/terms/index.htm#minimumDistanceAboveSurfaceInMeters
	X	X	Location	maximumDistanceAboveSurfaceInMeters	The greater distance in a range of distance from a reference surface in the vertical direction, in meters. Use positive values for locations above the surface, negative values for locations below. If depth measures are given, the reference surface is the location given by the depth, otherwise the reference surface is the location given by the elevation. see also http://rs.tdwg.org/dwc/terms/index.htm#maximumDistanceAboveSurfaceInMeters
	X	X	Location	locationAccordingTo	Information about the source of this Location information. Could be a publication (gazetteer), institution, or team of individuals. see also http://rs.tdwg.org/dwc/terms/index.htm#locationAccordingTo
	X	X	Location	locationRemarks	Comments or notes about the Location. see also http://rs.tdwg.org/dwc/terms/index.htm#locationRemarks

NÚCLEO			CLASSE	PROPRIEDADES	DESCRIÇÃO
Taxon	Ocurrence	Event			
	X	X	Location	verbatimCoordinates	The verbatim original spatial coordinates of the Location. The coordinate ellipsoid, geodeticDatum, or full Spatial Reference System (SRS) for these coordinates should be stored in verbatimSRS and the coordinate system should be stored in verbatimCoordinateSystem. see also http://rs.tdwg.org/dwc/terms/index.htm#verbatimCoordinates
	X	X	Location	verbatimLatitude	The verbatim original latitude of the Location. The coordinate ellipsoid, geodeticDatum, or full Spatial Reference System (SRS) for these coordinates should be stored in verbatimSRS and the coordinate system should be stored in verbatimCoordinateSystem. see also http://rs.tdwg.org/dwc/terms/index.htm#verbatimLatitude
	X	X	Location	verbatimLongitude	The verbatim original longitude of the Location. The coordinate ellipsoid, geodeticDatum, or full Spatial Reference System (SRS) for these coordinates should be stored in verbatimSRS and the coordinate system should be stored in verbatimCoordinateSystem. see also http://rs.tdwg.org/dwc/terms/index.htm#verbatimLongitude
	X	X	Location	verbatimCoordinateSystem	The spatial coordinate system for the verbatimLatitude and verbatimLongitude or the verbatimCoordinates of the Location. Recommended best practice is to use a controlled vocabulary. see also http://rs.tdwg.org/dwc/terms/index.htm#verbatimCoordinateSystem
	X	X	Location	verbatimSRS	The ellipsoid, geodetic datum, or spatial reference system (SRS) upon which coordinates given in verbatimLatitude and verbatimLongitude, or verbatimCoordinates are based. Recommended best practice is use the EPSG code as a controlled vocabulary to provide an SRS, if known. Otherwise use a controlled vocabulary for the name or code of the geodetic datum, if known. Otherwise use a controlled vocabulary for the name or code of the ellipsoid, if known. If none of these is known, use the value "unknown". see also http://rs.tdwg.org/dwc/terms/index.htm#verbatimSRS
	X	X	Location	decimalLatitude	The geographic latitude (in decimal degrees, using the spatial reference system given in geodeticDatum) of the geographic center of a Location. Positive values are north of the Equator, negative values are south of it. Legal values lie between -90 and 90, inclusive. see also http://rs.tdwg.org/dwc/terms/index.htm#decimalLatitude

NÚCLEO			CLASSE	PROPRIEDADES	DESCRIÇÃO
Taxon	Ocurrence	Event			
	X	X	Location	decimalLongitude	The geographic longitude (in decimal degrees, using the spatial reference system given in geodeticDatum) of the geographic center of a Location. Positive values are east of the Greenwich Meridian, negative values are west of it. Legal values lie between -180 and 180, inclusive. see also http://rs.tdwg.org/dwc/terms/index.htm#decimalLongitude
	X	X	Location	geodeticDatum	The ellipsoid, geodetic datum, or spatial reference system (SRS) upon which the geographic coordinates given in decimalLatitude and decimalLongitude as based. Recommended best practice is use the EPSG code as a controlled vocabulary to provide an SRS, if known. Otherwise use a controlled vocabulary for the name or code of the geodetic datum, if known. Otherwise use a controlled vocabulary for the name or code of the ellipsoid, if known. If none of these is known, use the value "unknown". see also http://rs.tdwg.org/dwc/terms/index.htm#geodeticDatum
	X	X	Location	coordinateUncertaintyInMeters	The horizontal distance (in meters) from the given decimalLatitude and decimalLongitude describing the smallest circle containing the whole of the Location. Leave the value empty if the uncertainty is unknown, cannot be estimated, or is not applicable (because there are no coordinates). Zero is not a valid value for this term. see also http://rs.tdwg.org/dwc/terms/index.htm#coordinateUncertaintyInMeters
	X	X	Location	coordinatePrecision	A decimal representation of the precision of the coordinates given in the decimalLatitude and decimalLongitude. see also http://rs.tdwg.org/dwc/terms/index.htm#coordinatePrecision
	X	X	Location	pointRadiusSpatialFit	The ratio of the area of the point-radius (decimalLatitude, decimalLongitude, coordinateUncertaintyInMeters) to the area of the true (original, or most specific) spatial representation of the Location. Legal values are 0, greater than or equal to 1, or undefined. A value of 1 is an exact match or 100% overlap. A value of 0 should be used if the given point-radius does not completely contain the original representation. The pointRadiusSpatialFit is undefined (and should be left blank) if the original representation is a point without uncertainty and the given georeference is not that same point (without uncertainty). If both the original and the given georeference are the same point, the pointRadiusSpatialFit is 1. see also http://rs.tdwg.org/dwc/terms/index.htm#pointRadiusSpatialFit

NÚCLEO			CLASSE	PROPRIEDADES	DESCRIÇÃO
Taxon	Ocurrence	Event			
	X	X	Location	footprintWKT	A Well-Known Text (WKT) representation of the shape (footprint, geometry) that defines the Location. A Location may have both a point-radius representation (see decimalLatitude) and a footprint representation, and they may differ from each other. see also http://rs.tdwg.org/dwc/terms/index.htm#footprintWKT
	X	X	Location	footprintSRS	A Well-Known Text (WKT) representation of the Spatial Reference System (SRS) for the footprintWKT of the Location. Do not use this term to describe the SRS of the decimalLatitude and decimalLongitude, even if it is the same as for the footprintWKT - use the geodeticDatum instead. see also http://rs.tdwg.org/dwc/terms/index.htm#footprintSRS
	X	X	Location	footprintSpatialFit	The ratio of the area of the footprint (footprintWKT) to the area of the true (original, or most specific) spatial representation of the Location. Legal values are 0, greater than or equal to 1, or undefined. A value of 1 is an exact match or 100% overlap. A value of 0 should be used if the given footprint does not completely contain the original representation. The footprintSpatialFit is undefined (and should be left blank) if the original representation is a point and the given georeference is not that same point. If both the original and the given georeference are the same point, the footprintSpatialFit is 1. see also http://rs.tdwg.org/dwc/terms/index.htm#footprintSpatialFit
	X	X	Location	georeferencedBy	A list (concatenated and separated) of names of people, groups, or organizations who determined the georeference (spatial representation) the Location. The recommended best practice is to separate the values with a vertical bar (' '). see also http://rs.tdwg.org/dwc/terms/index.htm#georeferencedBy
	X	X	Location	georeferencedDate	The date on which the Location was georeferenced. Recommended best practice is to use an encoding scheme, such as ISO 8601:2004(E). see also http://rs.tdwg.org/dwc/terms/index.htm#georeferencedDate
	X	X	Location	georeferenceProtocol	A description or reference to the methods used to determine the spatial footprint, coordinates, and uncertainties. see also http://rs.tdwg.org/dwc/terms/index.htm#georeferenceProtocol

NÚCLEO			CLASSE	PROPRIEDADES	DESCRIÇÃO
Taxon	Occurrence	Event			
	X	X	Location	georeferenceSources	A list (concatenated and separated) of maps, gazetteers, or other resources used to georeference the Location, described specifically enough to allow anyone in the future to use the same resources. The recommended best practice is to separate the values with a vertical bar (' '). see also http://rs.tdwg.org/dwc/terms/index.htm#georeferenceSources
	X	X	Location	georeferenceVerificationStatus	A categorical description of the extent to which the georeference has been verified to represent the best possible spatial description. Recommended best practice is to use a controlled vocabulary. see also http://rs.tdwg.org/dwc/terms/index.htm#georeferenceVerificationStatus
	X	X	Location	georeferenceRemarks	Notes or comments about the spatial description determination, explaining assumptions made in addition or opposition to the those formalized in the method referred to in georeferenceProtocol. see also http://rs.tdwg.org/dwc/terms/index.htm#georeferenceRemarks
	X		MaterialSample	materialSampleID	An identifier for the MaterialSample (as opposed to a particular digital record of the material sample). In the absence of a persistent global unique identifier, construct one from a combination of identifiers in the record that will most closely make the materialSampleID globally unique. see also http://rs.tdwg.org/dwc/terms/index.htm#materialSampleID
	X		Occurrence	occurrenceID	An identifier for the Occurrence (as opposed to a particular digital record of the occurrence). In the absence of a persistent global unique identifier, construct one from a combination of identifiers in the record that will most closely make the occurrenceID globally unique. see also http://rs.tdwg.org/dwc/terms/index.htm#occurrenceID
	X		Occurrence	catalogNumber	An identifier (preferably unique) for the record within the data set or collection. see also http://rs.tdwg.org/dwc/terms/index.htm#catalogNumber
	X		Occurrence	occurrenceRemarks	Comments or notes about the Occurrence. see also http://rs.tdwg.org/dwc/terms/index.htm#occurrenceRemarks
	X		Occurrence	recordNumber	An identifier given to the Occurrence at the time it was recorded. Often serves as a link between field notes and an Occurrence record, such as a specimen collector's number. see also http://rs.tdwg.org/dwc/terms/index.htm#recordNumber

NÚCLEO			CLASSE	PROPRIEDADES	DESCRIÇÃO
Taxon	Occurrence	Event			
	X		Occurrence	recordedBy	A list (concatenated and separated) of names of people, groups, or organizations responsible for recording the original Occurrence. The primary collector or observer, especially one who applies a personal identifier (recordNumber), should be listed first. The recommended best practice is to separate the values with a vertical bar (' '). see also http://rs.tdwg.org/dwc/terms/index.htm#recordedBy
	X		Occurrence	individualCount	The number of individuals represented present at the time of the Occurrence. see also http://rs.tdwg.org/dwc/terms/index.htm#individualCount
	X		Occurrence	organismQuantity	A number or enumeration value for the quantity of organisms. see also http://rs.tdwg.org/dwc/terms/index.htm#organismQuantity
	X		Occurrence	organismQuantityType	The type of quantification system used for the quantity of organisms. see also http://rs.tdwg.org/dwc/terms/index.htm#organismQuantityType
	X		Occurrence	sex	The sex of the biological individual(s) represented in the Occurrence. Recommended best practice is to use a controlled vocabulary. see also http://rs.tdwg.org/dwc/terms/index.htm#sex
	X		Occurrence	lifeStage	The age class or life stage of the biological individual(s) at the time the Occurrence was recorded. Recommended best practice is to use a controlled vocabulary. see also http://rs.tdwg.org/dwc/terms/index.htm#lifeStage
	X		Occurrence	reproductiveCondition	The reproductive condition of the biological individual(s) represented in the Occurrence. Recommended best practice is to use a controlled vocabulary. see also http://rs.tdwg.org/dwc/terms/index.htm#reproductiveCondition
	X		Occurrence	behavior	A description of the behavior shown by the subject at the time the Occurrence was recorded. Recommended best practice is to use a controlled vocabulary. see also http://rs.tdwg.org/dwc/terms/index.htm#behavior
	X		Occurrence	establishmentMeans	The process by which the biological individual(s) represented in the Occurrence became established at the location. Recommended best practice is to use a controlled vocabulary. see also http://rs.tdwg.org/dwc/terms/index.htm#establishmentMeans
	X		Occurrence	occurrenceStatus	A statement about the presence or absence of a Taxon at a Location. Recommended best practice is to use a controlled vocabulary. see also http://rs.tdwg.org/dwc/terms/index.htm#occurrenceStatus

NÚCLEO			CLASSE	PROPRIEDADES	DESCRIÇÃO
Taxon	Occurrence	Event			
	X		Occurrence	preparations	A list (concatenated and separated) of preparations and preservation methods for a specimen. The recommended best practice is to separate the values with a vertical bar (' '). see also http://rs.tdwg.org/dwc/terms/index.htm#preparations
	X		Occurrence	disposition	The current state of a specimen with respect to the collection identified in collectionCode or collectionID. Recommended best practice is to use a controlled vocabulary. see also http://rs.tdwg.org/dwc/terms/index.htm#disposition
	X		Occurrence	otherCatalogNumbers	A list (concatenated and separated) of previous or alternate fully qualified catalog numbers or other human-used identifiers for the same Occurrence, whether in the current or any other data set or collection. The recommended best practice is to separate the values with a vertical bar (' '). see also http://rs.tdwg.org/dwc/terms/index.htm#otherCatalogNumbers
	X		Occurrence	associatedMedia	A list (concatenated and separated) of identifiers (publication, global unique identifier, URI) of media associated with the Occurrence. The recommended best practice is to separate the values with a vertical bar (' '). see also http://rs.tdwg.org/dwc/terms/index.htm#associatedMedia
	X		Occurrence	associatedReferences	A list (concatenated and separated) of identifiers (publication, bibliographic reference, global unique identifier, URI) of literature associated with the Occurrence. The recommended best practice is to separate the values with a vertical bar (' '). see also http://rs.tdwg.org/dwc/terms/index.htm#associatedReferences
	X		Occurrence	associatedSequences	A list (concatenated and separated) of identifiers (publication, global unique identifier, URI) of genetic sequence information associated with the Occurrence. The recommended best practice is to separate the values with a vertical bar (' '). see also http://rs.tdwg.org/dwc/terms/index.htm#associatedSequences
	X		Occurrence	associatedTaxa	A list (concatenated and separated) of identifiers or names of taxa and their associations with the Occurrence. The recommended best practice is to separate the values with a vertical bar (' '). see also http://rs.tdwg.org/dwc/terms/index.htm#associatedTaxa

NÚCLEO			CLASSE	PROPRIEDADES	DESCRIÇÃO
Taxon	Ocurrence	Event			
	X		Organism	organismID	An identifier for the Organism instance (as opposed to a particular digital record of the Organism). May be a globally unique identifier or an identifier specific to the data set. see also http://rs.tdwg.org/dwc/terms/index.htm#organismID
	X		Organism	organismName	A textual name or label assigned to an Organism instance. see also http://rs.tdwg.org/dwc/terms/index.htm#organismName
	X		Organism	organismScope	A description of the kind of Organism instance. Can be used to indicate whether the Organism instance represents a discrete organism or if it represents a particular type of aggregation. Recommended best practice is to use a controlled vocabulary. see also http://rs.tdwg.org/dwc/terms/index.htm#organismScope
	X		Organism	associatedOccurrences	A list (concatenated and separated) of identifiers of other Occurrence records and their associations to this Occurrence. The recommended best practice is to separate the values with a vertical bar (' '). see also http://rs.tdwg.org/dwc/terms/index.htm#associatedOccurrences
	X		Organism	associatedOrganisms	A list (concatenated and separated) of identifiers of other Organisms and their associations to this Organism. The recommended best practice is to separate the values with a vertical bar (' '). see also http://rs.tdwg.org/dwc/terms/index.htm#associatedOrganisms
	X		Organism	previousIdentifications	A list (concatenated and separated) of previous assignments of names to the Occurrence. The recommended best practice is to separate the values with a vertical bar (' '). see also http://rs.tdwg.org/dwc/terms/index.htm#previousIdentifications
	X		Organism	organismRemarks	Comments or notes about the Organism instance. see also http://rs.tdwg.org/dwc/terms/index.htm#organismRemarks
	X	X	Record-level	type	The nature or genre of the resource. For Darwin Core, recommended best practice is to use the name of the class that defines the root of the record. see also http://rs.tdwg.org/dwc/terms/index.htm#dcterms:type
X	X	X	Record-level	modified	Data da última atualização do registro, em conformidade com a ISO 8601:2004(E).
X	X	X	Record-level	language	Idioma do recurso. Recomendável como boa prática o uso de vocabulário controlado, como o RFC 4646 [RFC4646].
X	X	X	Record-level	license	Documento concedendo permissão oficial para utilizar o recurso.

NÚCLEO			CLASSE	PROPRIEDADES	DESCRIÇÃO
Taxon	Ocurrence	Event			
X	X	X	Record-level	rightsHolder	Uma pessoa ou organização que possui ou gerencia os direitos sobre o recurso.
X	X	X	Record-level	accessRights	Informação sobre quem pode acessar o recurso ou indicação do status de segurança. Direitos de acesso podem incluir informações referentes acesso ou restrições baseadas em privacidade, segurança ou outras políticas.
X	X	X	Record-level	bibliographicCitation	Informações de citação especificadas pelo editor de dados. Uma referência bibliográfica para o recurso como uma declaração indicando como esse registro deve ser citado (atribuído) quando usado. A prática recomendada é incluir detalhes bibliográficos suficientes para identificar o recurso da forma mais inequívoca possível. As informações de citação são herdadas por todos os táxons filhos, se nenhuma outra citação for incluída. As informações de citação não são acumuladas de recursos hierarquicamente superiores. Por exemplo, uma citação pode estar vinculada a uma entrada de <i>Mammalia</i> e geralmente se aplica a todas as espécies de mamíferos, mas uma citação diferente para um táxon filho, <i>Primates</i> , se aplica a todos os táxons filhos de primatas.
X	X	X	Record-level	references	Uma URL para um recurso relacionado que é referenciado, citado ou apontado de alguma forma pelo recurso descrito. Muitas vezes, outra página da web mostra o mesmo recurso detalhado.
	X	X	Record-level	institutionID	An identifier for the institution having custody of the object(s) or information referred to in the record. see also http://rs.tdwg.org/dwc/terms/index.htm#institutionID
	X		Record-level	collectionID	An identifier for the collection or dataset from which the record was derived. For physical specimens, the recommended best practice is to use the identifier in a collections registry such as the Biodiversity Collections Index (http://www.biodiversitycollectionsindex.org/). see also http://rs.tdwg.org/dwc/terms/index.htm#collectionID
X	X	X	Record-level	datasetID	Identificador para um (sub) conjunto de dados – (sub) dataset. Preferencialmente que seja global e único, mas qualquer identificador (id) é permitido.
	X	X	Record-level	institutionCode	The name (or acronym) in use by the institution having custody of the object(s) or information referred to in the record. see also http://rs.tdwg.org/dwc/terms/index.htm#institutionCode
	X		Record-level	collectionCode	The name, acronym, coden, or initialism identifying the collection or data set from which the record was derived. see

NÚCLEO			CLASSE	PROPRIEDADES	DESCRIÇÃO
Taxon	Ocurrence	Event			
					also http://rs.tdwg.org/dwc/terms/index.htm#collectionCode
X	X	X	Record-level	datasetName	O título do (sub) conjunto de dados – (sub) dataset – também referenciado via datasetID . O nome que identifica o conjunto de dados do qual o registro foi derivado.
X	X	X	Record-level	informationWithheld	Informação adicional que existe, mas que não foi compartilhada no próprio recurso.
	X		Record-level	basisOfRecord	The specific nature of the data record - a subtype of the dcterms:type. Recommended best practice is to use a controlled vocabulary such as the Darwin Core Type Vocabulary (http://rs.tdwg.org/dwc/terms/type-vocabulary/index.htm). see also http://rs.tdwg.org/dwc/terms/index.htm#basisOfRecord
	X	X	Record-level	ownerInstitutionCode	The name (or acronym) in use by the institution having ownership of the object(s) or information referred to in the record.
	X	X	Record-level	dataGeneralizations	Actions taken to make the shared data less specific or complete than in its original form. Suggests that alternative data of higher quality may be available on request. see also http://rs.tdwg.org/dwc/terms/index.htm#dataGeneralizations
	X	X	Record-level	dynamicProperties	A list (concatenated and separated) of additional measurements, facts, characteristics, or assertions about the record. Meant to provide a mechanism for structured content such as key-value pairs. The recommended best practice is to use a key:value encoding schema such as JSON. see also http://rs.tdwg.org/dwc/terms/index.htm#dynamicProperties

ANEXO 6. Lista de categorias e elementos que integram o GMP.

CATEGORIA	Dataset (Resource)	The dataset field has elements relating to a single dataset (resource).
		DEFINIÇÃO
Dataset (Resource)	alternateIdentifier	It is a Universally Unique Identifier (UUID) for the EML document and not for the dataset. This term is optional. A list of different identifiers can be supplied. E.g., 619a4b95-1a82-4006-be6a-7dbe3c9b33c5.
	title	A description of the resource that is being documented that is long enough to differentiate it from other similar resources. Multiple titles may be provided, particularly when trying to express the title in more than one language (use the "xml:lang" attribute to indicate the language if not English/en). E.g. Vernal pool amphibian density data, Isla Vista, 1990-1996.
	creator	The resource creator is the person or organization responsible for creating the resource itself.
	metadataProvider	The metadataProvider is the person or organization responsible for providing documentation for the resource. See section "People and Organisations" for more details.
	associatedParty	An associatedParty is another person or organisation that is associated with the resource. These parties might play various roles in the creation or maintenance of the resource, and these roles should be indicated in the "role" element. See section "People and Organisations" for more details.
	contact	The contact field contains contact information for this dataset. This is the person or institution to contact with questions about the use, interpretation of a data set.
	pubDate	The date that the resource was published. The format should be represented as: CCYY, which represents a 4 digit year, or as CCYY-MM-DD, which denotes the full year, month, and day. Note that month and day are optional components. Formats must conform to ISO 8601. E.g. 2010-09-20.

	language	The language in which the resource (not the metadata document) is written. This can be a well-known language name, or one of the ISO language codes to be more precise. GBIF recommendation is to use the ISO language code (http://vocabularies.gbif.org/vocabularies/lang). E.g., English.
	additionalInfo	Information regarding omissions, instructions or other annotations that resource managers may wish to include with a dataset. Basically, any information that is not characterized by the other resource metadata fields.
	url	The URL of the resource that is available online.
	abstract	A brief overview of the resource that is being documented.
	Project	The project field contains information on the project in which this dataset was collected. It includes information such as project personnel, funding, study area, project design and related projects.
Project	NOME DO ELEMENTO	DEFINIÇÃO
	title	A descriptive title for the research project. E.g., Species diversity in Tennessee riparian habitats
	personnel	The personnel field is used to document people involved in a research project by providing contact information and their role in the project.
	funding	The funding field is used to provide information about funding sources for the project such as: grant and contract numbers; names and addresses of funding sources.
	studyAreaDescription	The studyAreaDescription field documents the physical area associated with the research project. It can include descriptions of the geographic, temporal, and taxonomic coverage of the research location and descriptions of domains (themes) of interest such as climate, geology, soils or disturbances.
	designDescription	The field designDescription contains general textual descriptions of research design. It can include detailed accounts of goals, motivations, theory, hypotheses, strategy, statistical design, and actual work. Literature citations may also be used to describe the research design.
People and Organisations	People and Organisations	There are several fields that could represent either a person or an organisation. Below is a list of the various fields used to describe a person or organisation.
	NOME DO ELEMENTO	DEFINIÇÃO

givenName

Subfield of individualName field. The given name field can be used for the first name of the individual associated with the resource, or for any other names that are not intended to be alphabetized (as appropriate). E.g., Jonny

surName

Subfield of individualName field. The surname field is used for the last name of the individual associated with the resource. This is typically the family name of an individual, for example, the name by which s/he is referred to in citations. E.g. Carson

organizationName

The full name of the organization that is associated with the resource. This field is intended to describe which institution or overall organization is associated with the resource being described. E.g., National Center for Ecological Analysis and Synthesis
This field is intended to be used instead of a particular person or full organization name. If the associated person that holds the role changes frequently, then Position Name would be used for consistency. Note that this field, used in conjunction with 'organizationName' and 'individualName' make up a single logical originator. Because of this, an originator with only the individualName of 'Joe Smith' is NOT the same as an originator with the name of 'Joe Smith' and the organizationName of 'NSF'. Also, the positionName should not be used in conjunction with individualName unless only that individual at that position would be considered an originator for the data package. If a positionName is used in conjunction with an organizationName, then that implies that any person who currently occupies said positionName at organizationName is the originator of the data package. E.g., HAST herbarium data manager

positionName**electronicMailAddress**

The electronic mail address is the email address for the party. It is intended to be an Internet SMTP email address, which should consist of a username followed by the @ symbol, followed by the email server domain name address. E.g. jcuadra@gbif.org

deliveryPoint

Subfield of the address field that describes the physical or electronic address of the responsible party for a resource. The delivery point field is used for the physical address for postal communication. E.g., GBIF Secretariat, Universitetsparken 15

role

Use this field to describe the role the party played with respect to the resource. E.g. technician, reviewer, principal investigator, etc.

phone	The phone field describes information about the responsible party's telephone, be it a voice phone, fax. E.g. +4530102040
postalCode	Subfield of the address field that describes the physical or electronic address of the responsible party for a resource. The postal code is equivalent to a U.S. zip code, or the number used for routing to an international address. E.g., 52000.
city	Subfield of the address field that describes the physical or electronic address of the responsible party for a resource. The city field is used for the city name of the contact associated with a particular resource. E.g. San Diego.
country	Subfield of the address field that describes the physical or electronic address of the responsible party for a resource. The country field is used for the name of the contact's country. The country name is most often derived from the ISO 3166 country code list. E.g., Japan.
onlineUrl	A link to associated online information, usually a web site. When the party represents an organization, this is the URL to a website or other online information about the organization. If the party is an individual, it might be their personal web site or other related online information about the party. E.g., http://www.yourdomain.edu/~doe .
KeywordSet (General Keywords)	The keywordSet field is a wrapper for the keyword and keywordThesaurus elements, both of which are required together.
NOME DO ELEMENTO	DEFINIÇÃO
keyword	A keyword or key phrase that concisely describes the resource or is related to the resource. Each keyword field should contain one and only one keyword (i.e., keywords should not be separated by commas or other delimiters). E.g., biodiversity.
keywordThesaurus	The name of the official keyword thesaurus from which keyword was derived. If an official thesaurus name does not exist, please keep a placeholder value such as "N/A" instead of removing this element as it is required together with the keyword element to constitute a keywordSet. E.g., IRIS keyword thesaurus.
Coverage	Describes the extent of the coverage of the resource in terms of its spatial extent, temporal extent, and taxonomic extent.

Taxonomic Coverage A container for taxonomic information about a resource. It includes a list of species names (or higher level ranks) from one or more classification systems. Please note the taxonomic classifications should not be nested, just listed one after the other.

NOME DO ELEMENTO

DEFINIÇÃO

generalTaxonomicCoverage

Taxonomic Coverage is a container for taxonomic information about a resource. It includes a list of species names (or higher level ranks) from one or more classification systems. A description of the range of taxa addressed in the data set or collection. Use a simple comma separated list of taxa. E.g., "All vascular plants were identified to family or species, mosses and lichens were identified as moss or lichen."

taxonomicClassification

Information about the range of taxa addressed in the dataset or collection.

taxonRankName

The name of the taxonomic rank for which the Taxon rank value is provided. E.g., phylum, class, genus, species.

taxonRankValue

The name representing the taxonomic rank of the taxon being described. E.g. Acer would be an example of a genus rank value, and rubrum would be an example of a species rank value, together indicating the common name of red maple. It is recommended to start with Kingdom and include ranks down to the most detailed level possible.

commonName

Applicable common names; these common names may be general descriptions of a group of organisms if appropriate. E.g., invertebrates, waterfowl.

Geographic Coverage

A container for spatial information about a resource; allows a bounding box for the overall coverage (in lat long), and also allows description of arbitrary polygons with exclusions.

NOME DO ELEMENTO

DEFINIÇÃO

geographicDescription

A short text description of a dataset's geographic areal domain. A text description is especially important to provide a geographic setting when the extent of the dataset cannot be well described by the "boundingCoordinates". E.g., "Manistee River watershed", "extent of 7 1/2 minute quads containing any property belonging to Yellowstone National Park"

westBoundingCoordinate

Subfield of boundingCoordinates field covering the W margin of a bounding box. The longitude in decimal degrees of the western-most point of the bounding box that is being described. E.g., -18.25, +25, 45.24755.

eastBoundingCoordinate

Subfield of boundingCoordinates field covering the E margin of a bounding box. The longitude in decimal degrees of the eastern-most point of the bounding box that is being described. E.g., -18.25, +25, 45.24755.

northBoundingCoordinate

Subfield of boundingCoordinates field covering the N margin of a bounding box. The longitude in decimal degrees of the northern-most point of the bounding box that is being described. E.g., -18.25, +25, 65.24755.

southBoundingCoordinate

Subfield of boundingCoordinates field covering the S margin of a bounding box. The longitude in decimal degrees of the southern-most point of the bounding box that is being described. E.g., -118.25, +25, 84.24755.

Temporal Coverage

This container allows coverage to be a single point in time, multiple points in time, or a range of dates.

NOME DO ELEMENTO

DEFINIÇÃO

beginDate

Subfield of rangeOfDates field: It may be used multiple times with a endDate field to document multiple date ranges. A single time stamp signifying the beginning of some time period. The calendar date field is used to express a date, giving the year, month, and day. The format should be one that complies with the International Standards Organization's standard 8601. The recommended format for EML is YYYY-MM-DD, where Y is the four digit year, M is the two digit month code (01 - 12, where January = 01), and D is the two digit day of the month (01 - 31). This field can also be used to enter just the year portion of a date. E.g. 2010-09-20

endDate

Subfield of rangeOfDates field: It may be used multiple times with a beginDate field to document multiple date ranges. A single time stamp signifying the end of some time period. The calendar date field is used to express a date, giving the year, month, and day. The format should be one that complies with the International Standards Organization's standard 8601. The recommended format for EML is YYYY-MM-DD, where Y is the four digit year, M is the two digit month code (01 - 12, where January = 01), and D is the two digit day of the month (01 - 31). This field can also be used to enter just the year portion of a date. E.g. 2010-09-20.

singleDateTime

The SingleDateTime field is intended to describe a single date and time for an event.

Methods

This field documents scientific methods used in the collection of the resource. It includes information on items such as tools, instrument calibration and software.

NOME DO ELEMENTO**DEFINIÇÃO****methodStep**

The methodStep field allows for repeated sets of elements that document a series of procedures followed to produce a data object. These include text descriptions of the procedures, relevant literature, software, instrumentation, source data and any quality control measures taken.

qualityControl

The qualityControl field provides a location for the description of actions taken to either control or assess the quality of data resulting from the associated method step.

sampling

Description of sampling procedures including the geographic, temporal and taxonomic coverage of the study.

studyExtent

Subfield of the sampling field. The coverage field allows for a textual description of the specific sampling area, the sampling frequency (temporal boundaries, frequency of occurrence), and groups of living organisms sampled (taxonomic coverage). The field studyExtent represents both a specific sampling area and the sampling frequency (temporal boundaries, frequency of occurrence). The geographic studyExtent is usually a surrogate (representative area of) for the larger area documented in the "studyAreaDescription".

samplingDescription

Subfield of the sampling field. The samplingDescription field allows for a text-based/human readable description of the sampling procedures used in the research project. The content of this element would be similar to a description of sampling procedures found in the methods section of a journal article.

Intellectual Property Rights

Contain a rights management statement for the resource, or a reference to a service providing such information.

NOME DO ELEMENTO**DEFINIÇÃO****purpose**

A description of the purpose of this dataset.

intellectualRights

A rights management statement for the resource, or reference a service providing such information. Rights information encompasses Intellectual Property Rights (IPR), Copyright, and various Property Rights. In the case of a data set, rights might include requirements for use, requirements for attribution, or other requirements the owner would like to impose. E.g., Copyright 2001 Regents of the University of California Santa Barbara. Free for use by all individuals provided that the owners are acknowledged in any use or publication.

**Additional Metadata + Natural
Collections Description Data (NCD)
Related**

The **additionalMetadata** field is a container for any other relevant metadata that pertains to the resource being described. This field allows EML to be extensible in that any XML-based metadata can be included in this element. The elements provided here in the GMP include those required for conformance with ISO 19139 and a subset of NCD (Natural Collections Descriptions) elements.

NOME DO ELEMENTO**DEFINIÇÃO****dateStamp**

The dateTime the metadata document was created or modified. E.g., 2002-10-23T18:13:51.235+01:00

metadataLanguage

The language in which the metadata document (as opposed to the resource being described by the metadata) is written. Composed of an ISO639-2/T three-letter language code and an ISO3166-1 three-letter country code. E.g., en_UK

citation

The citation for the work itself. See eml

bibliography

A list of citations (see below) that form a bibliography on literature related / used in the dataset

resourceLogoUrl

URL of the logo associated with a resource. E.g., <http://www.gbif.org/logo.jpg>

parentCollectionIdentifier

Subfield of collection field. Is an optional field. Identifier for the parent collection for this sub-collection. Enables a hierarchy of collections and sub collections to be built.

collectionIdentifier

Subfield of collection field. Is an optional field. The URI (LSID or URL) of the collection. In RDF, used as URI of the collection resource.

formationPeriod

Text description of the time period during which the collection was assembled. E.g., "Victorian", or "1922 - 1932", or "c. 1750".

livingTimePeriod

Time period during which biological material was alive (for palaeontological collections).

**Additional Metadata +
Natural Collections
Description Data (NCD)
Related**



specimenPreservationMethod

jgtiCuratorialUnit

Picklist keyword indicating the process or technique used to prevent physical deterioration of non-living collections. Expected to contain an instance from the Specimen Preservation Method Type Term vocabulary. E.g., formaldehyde.

A quantitative descriptor (number of specimens, samples or batches). The actual quantification could be covered by 1) an exact number of "JGI-units" in the collection plus a measure of uncertainty (+/- x); 2) a range of numbers (x to x), with the lower value representing an exact number, when the higher value is omitted. The discussion concluded that the quantification should encompass all specimens, not only those that have not yet been digitised. This is to avoid having to update the numbers too often. The number of non-public data (not digitised or not accessible) can be calculated from the GBIF numbers as opposed to the JGTI-data.

Fonte: <https://github.com/gbif/ipt/wiki/GMPHowToGuide>

ANEXO 7. Artigos resultantes das pesquisas por “metadata use” e “metadata standard use”

ANO	TÍTULO	PERIÓDICO	IDENTIFICADOR
1	2017	Interoperability and FAIRness through a novel combination of Web technologies	PeerJ Computer Science 3:e110 DOI: 10.7717/peerj-cs.110
2	2015	Public data archiving in ecology and evolution: how well are we doing?	PLoS Biology, vol. 13, no. 11, 2015. Gale Academic OneFile. Accessed 18 May 2020. DOI: 10.1371/journal.pbio.1002295
3	2015	Towards an Interoperable Field Spectroscopy Metadata Standard with Extended Support for Marine Specific Applications	Remote Sens. 2015, 7, 15668-15701; ISSN 2072-4292; doi:10.3390/rs71115668
4	2014	Issues in the re-use of non-forestry specific spatial data sources for national environmental reporting on planted forests in New Zealand	Journal of Spatial Science, Volume 59, 2014 - Issue 2, Pages 237-252 ISSN: 1449-8596 Online ISSN: 1836-5655, DOI:10.1080/14498596.2014.913508
5	2014	The anArchive taxonomic Checklist for Italian botanical data banking and vegetation analysis: Theoretical basis and advantages	Plant Biosystems - An International Journal Dealing with all Aspects of Plant Biology, 149:6, 958-965 DOI: 10.1080/11263504.2014.984010
6	2011	Towards mainstreaming of biodiversity data publishing: recommendations of the GBIF Data Publishing Framework Task Group	BMC Bioinformatics. 2011;12(Suppl 15):S1 DOI: 10.1186/1471-2105-12-S15-S1 ISSN: 1471-2105 (Online)
7	2004	Biological Information Standards	Bulletin of the American Society for Information Science and Technology, Dec 2003/Jan 2004, Vol.30(2), pp.15-17 ISSN: 19316550 ; E-ISSN: 15508366

ANEXO 8. Lista final de artigos utilizados na coleta de dados sobre o uso de metadados

ANO	TÍTULO	PERIÓDICO	IDENTIFICADOR	
1	2019	BioModelos: A collaborative online system to map species distributions	PLoS One, Mar 2019, Vol.14(3), p.e0214522	E-ISSN: 19326203 , DOI: 10.1371/journal.pone.0214522
2	2019	Feeding essential biodiversity variables (EBVs): Actual and potential contributions from LTER-Italy	Nature Conservation, 3 May 2019, Vol.34, pp.477-503	ISSN: 13146947 , E-ISSN: 13143301 , DOI: 10.3897/natureconservation.34.30735
3	2018	A survey of digitized data from U.S. fish collections in the iDigBio data aggregator	PLoS One, Dec 2018, Vol.13(12), p.e0207636	E-ISSN: 19326203 , DOI: 10.1371/journal.pone.0207636
4	2018	Building essential biodiversity variables (EBVs) of species distribution and abundance at a global scale	Biological Reviews, February 2018, Vol.93(1), pp.600-625	ISSN: 1464-7931 , E-ISSN: 1469-185X , DOI: 10.1111/brv.12359
5	2018	Evaluating and benchmarking biodiversity monitoring: Metadata-based indicators for sampling design, sampling effort and data analysis.(Report)	Ecological Indicators, 2018, Vol.85, p.624	ISSN: 1470-160X , DOI: 10.1016/j.ecolind.2017.11.012
6	2018	The influence of community recommendations on metadata completeness	Ecological Informatics, January 2018, Vol.43, pp.38-51	ISSN: 1574-9541 , DOI: 10.1016/j.ecoinf.2017.09.005
7	2018	Using collections data to infer biogeographic, environmental, and host structure in communities of endophytic fungi	Mycologia, 2018, Vol.110(1), pp.47-62	E-ISSN: 1557-2536 , PMID: 29863996 Version:1 , DOI: 10.1080/00275514.2018.1442078
8	2017	Blueprints of Effective Biodiversity and Conservation Knowledge Products That Support Marine Policy	Frontiers in Marine Science, Volume 4, pp.96-112	ISSN=2296-7745 doi.org/10.3389/fmars.2017.00096
9	2017	Building a Natural and Cultural Heritage Repository for the Storage and Dissemination of Knowledge: The Algarium Veneticum and the Archivio di Studi Adriatici Case Study	Journal of Library Metadata, 03 April 2017, Vol.17(2), pp.111-125	ISSN: 1938-6389 , E-ISSN: 1937-5034 , DOI: 10.1080/19386389.2017.1355165
10	2017	Contribution of citizen science towards international biodiversity monitoring	Biological Conservation, September 2017, Vol.213, pp.280-294	ISSN: 0006-3207 , DOI: 10.1016/j.biocon.2016.09.004
11	2017	Digital Herbarium of Moscow State University: The Largest Russian Biodiversity Database.(Report)	Biology Bulletin, 2017, Vol.44(6), p.584(7)	ISSN: 1062-3590 , DOI: 10.1134/S1062359017060103
12	2017	IRBAS : An online database to collate, analyze, and synthesize data on the biodiversity and ecology of intermittent rivers worldwide	Ecology and Evolution, Feb 2017, pp.815-823	E-ISSN: 20457758 , DOI: 10.1002/ece3.2679
13	2017	Temporal degradation of data limits biodiversity research	Ecology and Evolution, Sep 2017, pp.6863-6870	E-ISSN: 20457758 , DOI: 10.1002/ece3.3259

ANO	TÍTULO	PERIÓDICO	IDENTIFICADOR
14	2017	The database of the PREDICTS (Projecting Responses of Ecological Diversity In Changing Terrestrial Systems) project	Ecology and Evolution, Jan 2017, pp.145-188 E-ISSN: 20457758 , DOI: 10.1002/ece3.2579
15	2017	Toward a new data standard for combined marine biological and environmental datasets - expanding OBIS beyond species occurrences	Biodiversity Data Journal, 2017(5) E-ISSN: 1314-2828 , DOI: 10.3897/BDJ.5.e10989 , PMCID: 5345125 , PMID: 28325978
16	2016	Aggregate control of scientific data	Archives and Records, 02 January 2016, Vol.37(1), pp.53-64 ISSN: 2325-7962 , E-ISSN: 2325-7989 , DOI: 10.1080/23257962.2016.1145578
17	2016	LifeWatch Greece data-services: Discovering Biodiversity Data using Semantic Web Technologies	Biodiversity Data Journal, 2016(4) E-ISSN: 1314-2828 , DOI: 10.3897/BDJ.4.e8443 , PMCID: 5136645 , PMID: 27932908
18	2016	Organisation and description of datasets	Archives and Manuscripts, 03 May 2016, Vol.44(2), pp.73-85 ISSN: 0157-6895 , E-ISSN: 2164-6058 , DOI: 10.1080/01576895.2016.1179585
19	2016	The Global Registry of Biodiversity Repositories: A Call for Community Curation	Biodiversity data journal, 2016(4), pp.e10293 ISSN: 1314-2828 , PMID: 27660523 Version:1 , DOI: 10.3897/BDJ.4.e10293
20	2015	Community Next Steps for Making Globally Unique Identifiers Work for Biocollections Data	ZooKeys 494: 133–154 (2015) doi: 10.3897/zookeys.494.9352 http://zookeys.pensoft.net
21	2015	Fostering ecological data sharing: collaborations in the International Long Term Ecological Research Network	Ecosphere, 2015 Oct, Vol.6(10) ISSN: 2150-8925 , DOI: 10.1890/ES14-00281.1
22	2015	Rapid multi-nation distribution assessment of a charismatic conservation species using open access ensemble model GIS predictions: Red panda (<i>Ailurus fulgens</i>) in the Hindu-Kush Himalaya region	Biological Conservation, January 2015, Vol.181, pp.150-161 ISSN: 0006-3207 , DOI: 10.1016/j.biocon.2014.10.007
23	2015	rBEFdata: documenting data exchange and analysis for a collaborative data management platform	Ecology And Evolution, 2015 Jul(14), pp.2890-2897 ISSN: 2045-7758 , DOI: 10.1002/ece3.1547
24	2014	A crowdsourcing framework for the management of mobile multimedia nature observations	International Journal of Pervasive Computing and Communications, 2014, Vol.10(3), pp.216-238 ISSN: 1742-7371
25	2014	Documenting, storing, and executing models in Ecology: A conceptual framework and real implementation in a global change monitoring program	Environmental Modelling & Software, 2014 Feb, Vol.52, pp.192-199 ISSN: 1364-8152 , DOI: 10.1016/j.envsoft.2013.10.027
26	2014	Evaluating the fitness for use of spatial data sets to promote quality in ecological assessment and monitoring	International Journal Of Geographical Information Science, 2014 Nov 2, Vol.28(11), pp.2356-2371 ISSN: 1365-8816 , DOI: 10.1080/13658816.2014.924627

ANO	TÍTULO	PERIÓDICO	IDENTIFICADOR	
27	2014	Facilitating Access to Biodiversity Information: A Survey of Users' Needs and Practices	Environmental Management, March 2014, Vol.53(3), pp.690-701	ISSN: 0364-152X , E-ISSN: 1432-1009 , DOI: 10.1007/s00267-014-0229-7
28	2014	Meeting report: advancing practical applications of biodiversity ontologies	Standards in Genomic Sciences, 2014, Vol.9, p.17-17	E-ISSN: 1944-3277 , DOI: 10.1186/1944-3277-9-17 , PMCID: 4334987
29	2014	Metadata management, interoperability and Linked Data publishing support for Natural History Museums	International Journal on Digital Libraries, 2014, Vol.14(3), pp.127-140	ISSN: 1432-5012 , E-ISSN: 1432-1300 , DOI: 10.1007/s00799-014-0114-2
30	2014	The Edaphobase project of GBIF-Germany-A new online soil-zoological data warehouse	Applied Soil Ecology, 2014 Nov, Vol.83, pp.3-12	ISSN: 0929-1393 , DOI: 10.1016/j.apsoil.2014.03.021 , E-ISSN: 1873-0272
31	2014	The GBIF Integrated Publishing Toolkit: Facilitating the Efficient Publishing of Biodiversity Data on the Internet	PLoS One, Aug 2014, Vol.9(8), p.e102623	E-ISSN: 19326203 , DOI: 10.1371/journal.pone.0102623
32	2014	Vesper: Visualising species archives	Ecological Informatics, November 2014, Vol.24, pp.132-147	ISSN: 1574-9541 , DOI: 10.1016/j.ecoinf.2014.08.004
33	2013	Assessing the Primary Data Hosted by the Spanish Node of the Global Biodiversity Information Facility (GBIF)	PLoS One, Jan 2013, Vol.8(1), p.e55144	E-ISSN: 19326203 , DOI: 10.1371/journal.pone.0055144
34	2013	The WISER metadatabase: the key to more than 100 ecological datasets from European rivers, lakes and coastal waters	Hydrobiologia, Mar 2013, Vol.704(1), pp.29-38	ISSN: 0018-8158 , E-ISSN: 1573-5117 , DOI: 10.1007/s10750-012-1295-6
35	2012	A semantically integrated, user-friendly data model for species observation data	Ecological Informatics, 2012 Mar, Vol.8, pp.1-9	ISSN: 1574-9541 , DOI: 10.1016/j.ecoinf.2011.11.002
36	2012	Data platforms in integrative biodiversity research	Ecological Informatics, 2012 Sep, Vol.11, pp.1-4	ISSN: 1574-9541 , DOI: 10.1016/j.ecoinf.2012.04.001
37	2012	Efficient rescue of threatened biodiversity data using reBiND workflows	Plant Biosystems - An International Journal Dealing with all Aspects of Plant Biology, 01 December 2012, Vol.146(4), pp.752-755	ISSN: 1126-3504 , E-ISSN: 1724-5575 , DOI: 10.1080/11263504.2012.740086
38	2012	From documents to datasets: A MediaWiki-based method of annotating and extracting species observations in century-old field notebooks	ZooKeys, 2012(209), p.235-253	ISSN: 1313-2989 , E-ISSN: 1313-2970 , DOI: 10.3897/zookeys.209.3247 , PMCID: 3406479 , PMID: 22859891
39	2012	From text to structured data: Converting a word-processed floristic checklist into Darwin Core Archive format	PhytoKeys, 2012(9), pp.1-13	E-ISSN: 1314-2003 , PMID: 22371687 Version:1 , DOI: 10.3897/phytokeys.9.2770
40	2012	Improving access to biodiversity data for, and from, EIAs - a data publishing framework built to global standards	Impact Assessment and Project Appraisal, 01 September 2012, Vol.30(3), pp.148-156	ISSN: 1461-5517 , E-ISSN: 1471-5465 , DOI: 10.1080/14615517.2012.705068
41	2012	No specimen left behind: Industrial scale digitization of natural history collections	ZooKeys, 2012, Vol.209, pp.133-146	ISSN: 13132989 , E-ISSN: 13132970 , DOI: 10.3897/zookeys.209.3178

ANO		TÍTULO	PERIÓDICO	IDENTIFICADOR
42	2012	The ABCD of primary biodiversity data access	Plant Biosystems - An International Journal Dealing with all Aspects of Plant Biology, 01 December 2012, Vol.146(4), pp.771-779	ISSN: 1126-3504 , E-ISSN: 1724-5575 , DOI: 10.1080/11263504.2012.740085
43	2011	Data sharing and retrieval using OAI-PMH	Earth Science Informatics, Mar 2011, pp.1-5	ISSN: 18650473 , E-ISSN: 18650481 , DOI: 10.1007/s12145-010-0073-0
44	2011	Predictions of 27 Arctic pelagic seabird distributions using public environmental variables, assessed with colony data: a first digital IPY and GBIF open access synthesis platform	Marine Biodiversity, 2011, Vol.41(1), pp.141-179	ISSN: 1867-1616 , E-ISSN: 1867-1624 , DOI: 10.1007/s12526-011-0083-2
45	2011	Supporting Red List threat assessments with GeoCAT: geospatial conservation assessment tool	ZooKeys, 2011(150), p.117-126	ISSN: 1313-2989 , E-ISSN: 1313-2970 , DOI: 10.3897/zookeys.150.2109 , PMCID: 3234434 , PMID: 22207809
46	2011	The future of the past in the present: biodiversity informatics and geological time	ZooKeys, 2011(150), p.397-405	ISSN: 1313-2989 , E-ISSN: 1313-2970 , DOI: 10.3897/zookeys.150.2350 , PMCID: 3234446 , PMID: 22207819
47	2011	The Global Index of Vegetation-Plot Databases (GIVD): a new resource for vegetation science	Journal Of Vegetation Science, 2011 Aug, Vol.22(4), pp.582-597	ISSN: 1100-9233 , DOI: 10.1111/j.1654-1103.2011.01265.x
48	2010	A flexible online metadata editing and management system	Ecological Informatics, 2010, Vol.5(1), pp.26-31	ISSN: 1574-9541 , DOI: 10.1016/j.ecoinf.2009.08.012
49	2010	Leveraging the fullest potential of scientific collections through digitisation	Biodiversity Informatics, 01 October 2010, Vol.7(2)	E-ISSN: 1546-9735 , DOI: 10.17161/bi.v7i2.3987
50	2010	Mercury: reusable metadata management, data discovery and access system	Earth Science Informatics, Jun 2010, Vol.3(1-2), pp.87-94	ISSN: 18650473 , E-ISSN: 18650481 , DOI: 10.1007/s12145-010-0050-7
51	2010	Natural History Specimen Digitization: Challenges and Concerns	Biodiversity Informatics, 01 October 2010, Vol.7(2)	E-ISSN: 1546-9735 , DOI: 10.17161/bi.v7i2.3992
52	2010	Optimizing Workflow through Metadata Repurposing and Batch Processing	Journal of Library Metadata, 17 November 2010, Vol.10(4), pp.219-237	ISSN: 1938-6389 , E-ISSN: 1937-5034 , DOI: 10.1080/19386389.2010.524862
53	2010	Summary of Recommendations of the GBIF Task Group on the Global Strategy and Action Plan for the Digitisation of Natural History Collections	Biodiversity Informatics, 01 October 2010, Vol.7(2)	E-ISSN: 1546-9735 , DOI: 10.17161/bi.v7i2.3989
54	2010	Using geographical and taxonomic metadata to set priorities in specimen digitization	Biodiversity Informatics, 01 October 2010, Vol.7(2)	E-ISSN: 1546-9735 , DOI: 10.17161/bi.v7i2.3988

ANO	TÍTULO	PERIÓDICO	IDENTIFICADOR	
55	2009	Beyond data management: how ecoinformatics can benefit environmental monitoring programs	Environmental Monitoring and Assessment, Mar 2009, Vol.150(1-4), pp.227-35	ISSN: 0167-6369 , E-ISSN: 1573-2959 , DOI: 10.1007/s10661-008-0675-x
56	2009	Multi-level discrepancies with sharing data on protected areas: What we have and what we need for the global village	Journal Of Environmental Management, 2009 Jan, Vol.90(1), pp.8-24	ISSN: 0301-4797 , DOI: 10.1016/j.jenvman.2007.11.001
57	2009	SFMN GeoSearch: An interactive approach to the visualization and exchange of point-based ecological data	Ecological Informatics, 2009 Sep, Vol.4(4), pp.196-205	ISSN: 1574-9541 , DOI: 10.1016/j.ecoinf.2009.07.007
58	2009	Weaving a New Fabric of Natural History	Interdisciplinary Science Reviews, Sept. 2009, Vol.34(2-3), pp.206-223	ISSN: 0308-0188 , DOI: 10.1179/174327909X441117
59	2008	A digital library environment for integrating, disseminating and exploring ecological data	Ecological Informatics, Volume 3, Issues 4–5, October 2008, Pages 295-308	https://doi.org/10.1016/j.ecoinf.2008.09.006
60	2008	Advancing ecological research with ontologies	Trends In Ecology & Evolution, 2008 Mar(3), pp.159-168	ISSN: 0169-5347 , DOI: 10.1016/j.tree.2007.11.007
61	2007	Integration of Biodiversity Databases in Taiwan and Linkage to Global Databases	Data Science Journal, 2007, Vol.6, pp.S2-S10	ISSN: 1683-1470 , DOI: 10.2481/dsj.6.S2
62	2007	Species at risk: Data and knowledge management within the WILDSPACEa"cents Decision Support System.(Report)	Environmental Modelling and Software, April, 2007, Vol.22(4), p.423(8)	ISSN: 1364-8152
63	2006	A digital library framework for biodiversity information systems	International Journal on Digital Libraries (2006) 6(1): 3–17	DOI 10.1007/s00799-005-0124-1
64	2006	Analytic Webs Support the Synthesis of Ecological Data Sets	Analytic Webs Support the Synthesis of Ecological Data Sets. Ecology 87: 1345-1358.	ISSN: 0012-9658
65	2006	The New Bioinformatics: Integrating Ecological Data from the Gene to the Biosphere	2006, Vol.37, p.519-544	ISSN: 1543-592X , DOI: 10.1146/annurev.ecolsys.37.091305.110031
66	2005	The European Forest Information System -- an Internet based interface between information providers and the user community.(Report)	Computers and Electronics in Agriculture, June, 2005, Vol.47(3), p.185(22)	ISSN: 0168-1699
67	2004	The establishment of China's core biodiversity metadata standard	Biodiversity and Conservation, May 2004, Vol.13(5), pp.1009-1022	ISSN: 09603115 , DOI: 10.1023/B:BIOC.0000014466.91355.0b
68	2003	Darwin and MARC: a voyage of metadata discovery	Library Collections, Acquisitions and Technical Services, 2003, Vol.27(3), pp.291-304	ISSN: 1464-9055 , E-ISSN: 1873-1821 , DOI: 10.1016/S1464-9055(03)00071-X
69	2003	Geospatially structured biodiversity information as a component of a regional biodiversity clearing house	Biodiversity and Conservation 12: 103–120, 2003.	DOI: 10.1023/A:1021238212962

ANO	TÍTULO	PERIÓDICO	IDENTIFICADOR
70	2000	Biodiversity Clearing-House Mechanism in China: present status and future needs	Biodiversity and Conservation 9: 361–378, 2000. https://doi.org/10.1023/A:1008958902306
71	1999	China National Biodiversity Information Query System	Journal of Environmental Management, May 1999, Vol.56(1), pp.45-59 ISSN: 03014797 , DOI: 10.1006/jema.1999.0266
72	2017	Interoperability and FAIRness through a novel combination of Web technologies	PeerJ Computer Science 3:e110 DOI: 10.7717/peerj-cs.110
73	2015	Public data archiving in ecology and evolution: how well are we doing?	PLoS Biology, vol. 13, no. 11, 2015. Gale Academic OneFile. Accessed 18 May 2020. DOI: 10.1371/journal.pbio.1002295
74	2015	Towards an Interoperable Field Spectroscopy Metadata Standard with Extended Support for Marine Specific Applications	Remote Sens. 2015, 7, 15668-15701; ISSN 2072-4292; doi:10.3390/rs71115668
75	2014	Issues in the re-use of non-forestry specific spatial data sources for national environmental reporting on planted forests in New Zealand	Journal of Spatial Science, Volume 59, 2014 - Issue 2, Pages 237-252 ISSN: 1449-8596 Online ISSN: 1836-5655, DOI:10.1080/14498596.2014.913508
76	2014	The anArchive taxonomic Checklist for Italian botanical data banking and vegetation analysis: Theoretical basis and advantages	Plant Biosystems - An International Journal Dealing with all Aspects of Plant Biology, 149:6, 958-965 DOI: 10.1080/11263504.2014.984010
77	2011	Towards mainstreaming of biodiversity data publishing: recommendations of the GBIF Data Publishing Framework Task Group	BMC Bioinformatics. 2011;12(Suppl 15):S1 DOI: 10.1186/1471-2105-12-S15-S1 ISSN: 1471-2105 (Online)
78	2004	Biological Information Standards	Bulletin of the American Society for Information Science and Technology, Dec 2003/Jan 2004, Vol.30(2), pp.15-17 ISSN: 19316550 ; E-ISSN: 15508366

ANEXO 9. Lista de conjuntos de dados vinculados a instituições globais

Título do conjunto de dados	classe	Nome da instituição responsável
Arctic Ocean Diversity	ocorrência	Alaska Ocean Observing System
Biodiversity Survey of Echuya forest, Year 2012	ocorrência	Albertine Rift Conservation Society (ARCOS)
Biodiversity survey of Rugezi wetland, Year 2013	ocorrência	Albertine Rift Conservation Society (ARCOS)
Echuya Bird Species Survey 2018	ocorrência	Albertine Rift Conservation Society (ARCOS)
Data from ILAM surveys conducted by ARCOS in Bugesera landscape for year 2016	evento	Albertine Rift Conservation Society (ARCOS)
Data from ILAM surveys conducted by ARCOS in Bugesera landscape for year 2017	evento	Albertine Rift Conservation Society (ARCOS)
Data from ILAM surveys conducted by ARCOS in Mukura landscape for year 2016	evento	Albertine Rift Conservation Society (ARCOS)
Data from ILAM surveys conducted by ARCOS in Mukura landscape for year 2017	evento	Albertine Rift Conservation Society (ARCOS)
Data from ILAM surveys conducted by ARCOS in Rusumo landscape for year 2016	evento	Albertine Rift Conservation Society (ARCOS)
Data from ILAM surveys conducted by ARCOS in Rusumo landscape for year 2017	evento	Albertine Rift Conservation Society (ARCOS)
JRS Wetlands data Rwanda, 2019	evento	Albertine Rift Conservation Society (ARCOS)
The Register of Antarctic Marine Species	metadado de recurso	Antarctic Biodiversity Information Facility (ANTABIF)
amundsen_sea_molluscs	ocorrência	Antarctic Biodiversity Information Facility (ANTABIF)
Antarctic macrobenthic communities: A compilation of circumpolar information	ocorrência	Antarctic Biodiversity Information Facility (ANTABIF)
Antarctic, Sub-Antarctic and cold temperate echinoid database	ocorrência	Antarctic Biodiversity Information Facility (ANTABIF)
Bacteria from Penguin Guano, Antarctica	ocorrência	Antarctic Biodiversity Information Facility (ANTABIF)
Circumpolar dataset of sequenced specimens of <i>Promachocrinus kerguelensis</i> (Echinodermata, Crinoidea)	ocorrência	Antarctic Biodiversity Information Facility (ANTABIF)
East Weddell Sea echinoids from the JR275 expedition	ocorrência	Antarctic Biodiversity Information Facility (ANTABIF)
Photodocumentation_of_Antarctic_Peninsula_Lichens	ocorrência	Antarctic Biodiversity Information Facility (ANTABIF)
PoE At Sea Observations of Birds and Marine Mammals during PS 81 (ANT-XXIX/4)	ocorrência	Antarctic Biodiversity Information Facility (ANTABIF)
The First Comprehensive Description of the Biodiversity and Biogeography of Antarctic and Sub-Antarctic Intertidal Communities	ocorrência	Antarctic Biodiversity Information Facility (ANTABIF)
10 year trend of levels of organochlorine pollutants in Antarctic seabirds 2003/04	ocorrência	Australian Antarctic Data Centre
2007-08 Voyage 3 of the Aurora Australis, CEAMARC-CASO Krill Data	ocorrência	Australian Antarctic Data Centre

Título do conjunto de dados	classe	Nome da instituição responsável
2016 Aerial survey data of southern right whales (<i>Eubalaena australis</i>) off southern Australia	ocorrência	Australian Antarctic Data Centre
Abatus Microsatellites data set 2009/13	ocorrência	Australian Antarctic Data Centre
ABBBS Bird Banding records from the Australian Antarctic Territory and Heard Island.	ocorrência	Australian Antarctic Data Centre
Abundances of benthic invertebrate species in the CEAMARC region 2007/08	ocorrência	Australian Antarctic Data Centre
Annotated Atlas of the Vascular Flora of Macquarie Island - 1979	ocorrência	Australian Antarctic Data Centre
Antarctic Fur Seal Populations on Heard Island Summer 1987-1988	ocorrência	Australian Antarctic Data Centre
Antarctic Marine Protists species checklist	ocorrência	Australian Antarctic Data Centre
APIS - Antarctic Pack Ice Seals 1994-1999, plus historical data from the 1980s	ocorrência	Australian Antarctic Data Centre
ARGOS Satellite Tracking of animals	ocorrência	Australian Antarctic Data Centre
Australian Antarctic Division Herbarium	ocorrência	Australian Antarctic Data Centre
Australias Census of Antarctic Marine Life project - IYGPT Data collected on the CEAMARC cruise 2007/08	ocorrência	Australian Antarctic Data Centre
Biology and population studies of two endemic orchid species on sub-Antarctic Macquarie Island	ocorrência	Australian Antarctic Data Centre
Biology of Southern Ocean squid, ecological importance and potential commercial implications - a preliminary assessment 1995/2006	ocorrência	Australian Antarctic Data Centre
Bunger Hills, 1999/2000 survey - nest sites of snow petrels <i>Pagodroma nivea</i>	ocorrência	Australian Antarctic Data Centre
Cape Darnley Early-Autumn Phytoplankton Bloom, March 2012	ocorrência	Australian Antarctic Data Centre
Cetacean Sightings Survey and Southern Ocean cetacean program	ocorrência	Australian Antarctic Data Centre
Cetacean Sightings Survey and Southern Ocean Cetacean Program - BROKE-West	ocorrência	Australian Antarctic Data Centre
Diatom abundance from CEAMARC coretop samples 2007/08	ocorrência	Australian Antarctic Data Centre
Diatom and foraminiferal samples from surficial sediments of Prydz Bay, Antarctica	ocorrência	Australian Antarctic Data Centre
Diatoms from SAZ Sediment traps	ocorrência	Australian Antarctic Data Centre
Diatoms from the Vestfold Hills Lakes	ocorrência	Australian Antarctic Data Centre
Diet results from Adelie penguins at Bechervaise Island and Whitney Point, 2012/13	ocorrência	Australian Antarctic Data Centre
Diversity and evolution of Australian Antarctic sea spiders: Understanding species diversification and distribution patterns	ocorrência	Australian Antarctic Data Centre
Duck sightings per month, Macquarie Island	ocorrência	Australian Antarctic Data Centre

Título do conjunto de dados	classe	Nome da instituição responsável
Ecology and Zoogeography of Parasites of Antarctic Fish	ocorrência	Australian Antarctic Data Centre
Electron Micrograph Database - Marine Specimens	ocorrência	Australian Antarctic Data Centre
Electron Micrograph Database - Terrestrial and Limnetic Specimens	ocorrência	Australian Antarctic Data Centre
Elephant Seal Sightings, Heard Island	ocorrência	Australian Antarctic Data Centre
Elephant Seal Sightings, Macquarie Island	ocorrência	Australian Antarctic Data Centre
Exotic terrestrial flatworms, Macquarie Island	ocorrência	Australian Antarctic Data Centre
Fish and zooplankton from RMT-8 net hauls on the BROKE voyage	ocorrência	Australian Antarctic Data Centre
Fish catch from 1996/97 Voyage 1 WASTE (WOCE Antarctic Southern Transect Expedition)	ocorrência	Australian Antarctic Data Centre
Foraging zones of Macaroni Penguins breeding at Heard Island 2000	ocorrência	Australian Antarctic Data Centre
Further investigations of the effects of the Nella Dan oil spill 1988/94	ocorrência	Australian Antarctic Data Centre
Gross body measurements of Antarctic Rock Cod (<i>Trematomus bernacchii</i>) collected at sites adjacent to Casey Station 2012/13	ocorrência	Australian Antarctic Data Centre
Gross body measurements of <i>Laternula elliptica</i> collected from sites adjacent to Casey Station 2014/15	ocorrência	Australian Antarctic Data Centre
Heard Island Shag <i>Phalacrocorax nivalis</i> census, Heard Island 1992	ocorrência	Australian Antarctic Data Centre
High occurrence of jellyfish predation by black-browed and Campbell albatross identified by DNA metabarcoding 2017	ocorrência	Australian Antarctic Data Centre
Hydroids of the BANZARE Antarctic expeditions 1929 - 1931	ocorrência	Australian Antarctic Data Centre
Ice core meiofauna during the SIPEX and SIPEX II voyages	ocorrência	Australian Antarctic Data Centre
Inventory of Antarctic seabird breeding sites	ocorrência	Australian Antarctic Data Centre
Invertebrate survey, Heard Island 2001	ocorrência	Australian Antarctic Data Centre
Invertebrates compiled by W.Block	ocorrência	Australian Antarctic Data Centre
Kelp rafts in the Southern Ocean	ocorrência	Australian Antarctic Data Centre
Kelp rafts in the Southern Ocean: intercontinental travel for sessile and semi-sessile organisms 2010	ocorrência	Australian Antarctic Data Centre
King penguin census data, Gadget Gully, Macquarie Island (1993-2008)	ocorrência	Australian Antarctic Data Centre
Length and weight of Antarctic Rock Cod (<i>Trematomus bernacchii</i>) from sites adjacent to Davis Station 2012/13	ocorrência	Australian Antarctic Data Centre
Leopard and Weddell seal program 1999/2002	ocorrência	Australian Antarctic Data Centre

Título do conjunto de dados	classe	Nome da instituição responsável
Leopard Seal census, Heard Island 1987/88	ocorrência	Australian Antarctic Data Centre
Macquarie Island Fur Seal Database	ocorrência	Australian Antarctic Data Centre
Marine fauna survey of the Vestfold Hills and Rauer Island, 1981-82	ocorrência	Australian Antarctic Data Centre
Microalgal Biodiversity at Antarctica 2001/04	ocorrência	Australian Antarctic Data Centre
Morphological Responses of Abatus heart urchins to Sewage Contamination at Davis Station 2012/13	ocorrência	Australian Antarctic Data Centre
Moss sampling in the Windmill Islands and the Vestfold Hills	ocorrência	Australian Antarctic Data Centre
Moss species from Heard Island, 1987 and 2000	ocorrência	Australian Antarctic Data Centre
National Whale and Dolphin Sightings and Strandings Database	ocorrência	Australian Antarctic Data Centre
Nest census, Windmill Islands 2002/03	ocorrência	Australian Antarctic Data Centre
Orca observations from the shores of Macquarie Island	ocorrência	Australian Antarctic Data Centre
Patterns of population genetic structure among Australian and South Pacific humpback whales 2007/11	ocorrência	Australian Antarctic Data Centre
Pelagic Fish Observations 1968-1999	ocorrência	Australian Antarctic Data Centre
Phytoplankton samples - 2001/02 V3 CLIVAR	ocorrência	Australian Antarctic Data Centre
Phytoplankton samples - 2001/02 V7 LOSS	ocorrência	Australian Antarctic Data Centre
Phytoplankton samples - 2005/06 V3 BROKE-West	ocorrência	Australian Antarctic Data Centre
Phytoplankton samples - 2006/07 V3 SAZ-Sense	ocorrência	Australian Antarctic Data Centre
Presence of Disease in the Penguins and Skuas of Macquarie Island 2006	ocorrência	Australian Antarctic Data Centre
Resilience of Antarctic marine benthic invertebrates and the ecological consequences of environmental change - Amphipod Data 2002/10	ocorrência	Australian Antarctic Data Centre
RMT Trawl catch from 1985/86 V1 ADBEX III voyage	ocorrência	Australian Antarctic Data Centre
RMT Trawl catch from the 1980/81 V5 FIBEX voyage	ocorrência	Australian Antarctic Data Centre
RMT Trawl catch from the 1982/83 V2 ADBEX I voyage	ocorrência	Australian Antarctic Data Centre
RMT Trawl catch from the 1983/84 V5 ADBEX II voyage	ocorrência	Australian Antarctic Data Centre
RMT Trawl catch from the 1984/85 V5 SIBEX2 voyage	ocorrência	Australian Antarctic Data Centre
RMT Trawl catch from the 1984/85 V5 SIBEX2 voyage	ocorrência	Australian Antarctic Data Centre
RMT Trawl catch from the 1986/87 V7 AAMBER voyage	ocorrência	Australian Antarctic Data Centre

Título do conjunto de dados	classe	Nome da instituição responsável
RMT Trawl catch from the 1990/91 V6 AAMBER2 voyage	ocorrência	Australian Antarctic Data Centre
RMT Trawl catch from the 1992/93 V6 KROCK voyage	ocorrência	Australian Antarctic Data Centre
RMT Trawl catch from the 1995/96 V4 BROKE voyage	ocorrência	Australian Antarctic Data Centre
RMT Trawl catch from the 2005/06 V3 BROKE-West voyage - Fish	ocorrência	Australian Antarctic Data Centre
RMT Trawl catch from the 2005/06 V3 BROKE-West voyage - Zooplankton	ocorrência	Australian Antarctic Data Centre
Rotifers of Macquarie Island	ocorrência	Australian Antarctic Data Centre
Royal penguin Eudyptes schlegeli census and observations at North Head, Macquarie Island 1952/53	ocorrência	Australian Antarctic Data Centre
Royal penguin Eudyptes schlegeli census, Macquarie Island, 1984	ocorrência	Australian Antarctic Data Centre
Sample summary and analysis of biopsy samples collected during the Antarctic blue whale voyage 2013	ocorrência	Australian Antarctic Data Centre
Sample summary and analysis of biopsy samples collected during the New Zealand Australian Antarctic Ecosystems Voyage 2015	ocorrência	Australian Antarctic Data Centre
Seabirds of the Southern and South Indian Ocean	ocorrência	Australian Antarctic Data Centre
Snow Petrel census, Reeve Hill, Windmill Islands, East Antarctica (1984-2003)	ocorrência	Australian Antarctic Data Centre
Snow Petrel nest census, Mawson region 2004/05	ocorrência	Australian Antarctic Data Centre
Soft sediment assemblages at Casey Station: Brown Bay Grid - samples taken along a pollution gradient.	ocorrência	Australian Antarctic Data Centre
Southern Giant Petrel census data within the Australian Antarctic Territory	ocorrência	Australian Antarctic Data Centre
Southern Ocean Continuous Zooplankton Recorder (SO-CPR) Survey	ocorrência	Australian Antarctic Data Centre
Spring Phytoplankton Assemblages in the Southern Ocean Between Australia and Antarctica (CLIVAR-SR3)	ocorrência	Australian Antarctic Data Centre
Squid collection from 1995/96 V4 BROKE	ocorrência	Australian Antarctic Data Centre
Squid specimens from the 1991/92 V6 AAMBER2 voyage	ocorrência	Australian Antarctic Data Centre
Stable isotope data from marine sediment and invertebrates from Davis Station 2009/10	ocorrência	Australian Antarctic Data Centre
Subantarctic Fur seals at Heard Island, 1987/88	ocorrência	Australian Antarctic Data Centre
Survey of benthic and other marine invertebrates of Prydz Bay, 1990/91 Voyage 6	ocorrência	Australian Antarctic Data Centre
Systematic list of benthic invertebrates identified in the Heard Island and McDonald Islands region from 129 beam trawl or epibenthic sled samples 2003/08	ocorrência	Australian Antarctic Data Centre

Título do conjunto de dados	classe	Nome da instituição responsável
The Ecology and Biogeography of Heard Island Marine Benthos 1987/88	ocorrência	Australian Antarctic Data Centre
Toxicity of copper to three Macquarie Island marine gastropods 2013/15	ocorrência	Australian Antarctic Data Centre
Toxicity of copper, cadmium and zinc to Macquarie Island marine invertebrates 2012/15	ocorrência	Australian Antarctic Data Centre
Type Localities for organisms and fossils first described from Antarctica	ocorrência	Australian Antarctic Data Centre
Venom on ice: Novel venoms from Antarctic marine animals 2007/08	ocorrência	Australian Antarctic Data Centre
Weddell Seal census, Vestfold Hills, Antarctica	ocorrência	Australian Antarctic Data Centre
Weddell Seal Sightings, Vestfold Hills, Antarctica	ocorrência	Australian Antarctic Data Centre
Whale catches in the Southern Ocean	ocorrência	Australian Antarctic Data Centre
Whale log - observations from ANARE voyages	ocorrência	Australian Antarctic Data Centre
Whale Observations from the British, Australian and New Zealand Antarctic Research Expedition (BANZARE) voyages 1929-30 and 1930-31	ocorrência	Australian Antarctic Data Centre
Winter Foraging Locations Of Southern Ocean Predators - sub-Antarctic islands 2008/09	ocorrência	Australian Antarctic Data Centre
Zooplankton abundance in the Kerguelen Axis region 2016	ocorrência	Australian Antarctic Data Centre
Zooplankton Abundance_Umbrella net_sea ice zone 2007/12	ocorrência	Australian Antarctic Data Centre
Amphibia.my	checklist	Bioinformatics Research Group, Biological Science Department, Faculty of Bioscience & Bioengineering
izOO fauna dataset	checklist	Bioinformatics Research Group, Biological Science Department, Faculty of Bioscience & Bioengineering
Malaysian Fern	checklist	Bioinformatics Research Group, Biological Science Department, Faculty of Bioscience & Bioengineering
Malaysian Mangrove	checklist	Bioinformatics Research Group, Biological Science Department, Faculty of Bioscience & Bioengineering
Malaysian Mangrove	checklist	Bioinformatics Research Group, Biological Science Department, Faculty of Bioscience & Bioengineering
Malaysian Moss	checklist	Bioinformatics Research Group, Biological Science Department, Faculty of Bioscience & Bioengineering
Reptilia.my	checklist	Bioinformatics Research Group, Biological Science Department, Faculty of Bioscience & Bioengineering
izOO Belum Forest Species Occurrences	ocorrência	Bioinformatics Research Group, Biological Science Department, Faculty of Bioscience & Bioengineering

Título do conjunto de dados	classe	Nome da instituição responsável
Biodiversity Collecting Mission Database	ocorrência	Biodiversity International
EURISCO, The European Genetic Resources Search Catalogue	ocorrência	Biodiversity International
SINGER Coordinator	ocorrência	Biodiversity International
The System-wide Information Network for Genetic Resources (SINGER)	ocorrência	Biodiversity International
CABI Africa Invasive and Alien Species data	ocorrência	CABI (Centre for Agriculture and Biosciences International)
CABI Bioscience Fungus Collection	ocorrência	CABI (Centre for Agriculture and Biosciences International)
CABI Bioscience Genetic Resource Collection	ocorrência	CABI (Centre for Agriculture and Biosciences International)
CABI Bioscience Nematode Collection	ocorrência	CABI (Centre for Agriculture and Biosciences International)
Abundance and diversity of the Amphipoda (Crustacea) from the Greenlandic shelf	ocorrência	Conservation of Arctic Flora and Fauna
Archives of the Arctic Seas Zooplankton	ocorrência	Conservation of Arctic Flora and Fauna
Arctic benthic invertebrate collection of the Zoological Institute of the Russian Academy of Science	ocorrência	Conservation of Arctic Flora and Fauna
Arctic Marine Fish Museum Specimens	ocorrência	Conservation of Arctic Flora and Fauna
Arctic Species Trend Index (ASTI) : Freshwater	ocorrência	Conservation of Arctic Flora and Fauna
Arctic Species Trend Index (ASTI) : Marine	ocorrência	Conservation of Arctic Flora and Fauna
Arctic Species Trend Index (ASTI) : Terrestrial	ocorrência	Conservation of Arctic Flora and Fauna
Barents Sea Zooplankton " March and May 1998 and July 1999. Day/night stratified zooplankton sampling onboard the Norwegian R/V Jan Mayen. Program ALV	ocorrência	Conservation of Arctic Flora and Fauna
Barents Sea Zooplankton " March and May 1998 and July 1999. Zooplankton sampling on board the Norwegian R/V Jan Mayen, program ALV	ocorrência	Conservation of Arctic Flora and Fauna
Benthic Microalgae Chukchi Sea, 1972	ocorrência	Conservation of Arctic Flora and Fauna
Benthos Chirnov Basin 1986 and 2002	ocorrência	Conservation of Arctic Flora and Fauna
BenthosBarentsSeaPolarstern1991	ocorrência	Conservation of Arctic Flora and Fauna
BenthosChukchiFN762_1976_Falk5	ocorrência	Conservation of Arctic Flora and Fauna
Biology and Ecology of Cryopelagic Amphipods from Arctic Sea Ice Collected near Franz Josef Land in the summer of 1994	ocorrência	Conservation of Arctic Flora and Fauna

Título do conjunto de dados	classe	Nome da instituição responsável
Canada_BasinMegabenthos2002	ocorrência	Conservation of Arctic Flora and Fauna
CASES: The pre-winter assemblages of southeastern Beaufort Sea	ocorrência	Conservation of Arctic Flora and Fauna
CASES2003_2004	ocorrência	Conservation of Arctic Flora and Fauna
Chucki_Seaooplankton1953/4	ocorrência	Conservation of Arctic Flora and Fauna
Circumpolar Seabird Monitoring Plan	ocorrência	Conservation of Arctic Flora and Fauna
Composition and distribution of the biomass of zooplankton in the central Arctic Basin 1975, 1976, 1977	ocorrência	Conservation of Arctic Flora and Fauna
Database of the Distributions of Marine Fishes of Arctic Canada	ocorrência	Conservation of Arctic Flora and Fauna
Diet, Tidal and Vertical Variations of Phytoplankton and Its Environment in Frobisher Bay, Canada, 1980 and 1981	ocorrência	Conservation of Arctic Flora and Fauna
Disko Fan Conservation Area in the Davis Strait (Eastern Arctic) - Epibenthic Megafauna Database	ocorrência	Conservation of Arctic Flora and Fauna
Distribution and diversity of sipunculan fauna in high Arctic fjords (west Svalbard)	ocorrência	Conservation of Arctic Flora and Fauna
Distribution data of Arctic species of genus Microporella and Pseudoflustra gathered from museum collections	ocorrência	Conservation of Arctic Flora and Fauna
Greenland macrobenthos 2006	ocorrência	Conservation of Arctic Flora and Fauna
Greenland Macrobenthos Young Sound, 1996	ocorrência	Conservation of Arctic Flora and Fauna
Ice Algae Barents Sea	ocorrência	Conservation of Arctic Flora and Fauna
Ice Amphipods Canada Basin	ocorrência	Conservation of Arctic Flora and Fauna
Ice Amphipods Svalbard, 2000	ocorrência	Conservation of Arctic Flora and Fauna
Ice Diatoms Arctic Basin 1977, 1980	ocorrência	Conservation of Arctic Flora and Fauna
Ice Zooplankton Beaufort Sea	ocorrência	Conservation of Arctic Flora and Fauna
IPOE_Benthos_Steffens	ocorrência	Conservation of Arctic Flora and Fauna
IPOE_Schnack_Polychaeta	ocorrência	Conservation of Arctic Flora and Fauna
Kittlitzs_murrelet	ocorrência	Conservation of Arctic Flora and Fauna
Laptev Sea and Nansen Basin Zooplankton, 1993	ocorrência	Conservation of Arctic Flora and Fauna
Macrobenthos and Meiobenthos Tuktoyaktuk Harbor and Mason Bay 1985-1988 NOGAP	ocorrência	Conservation of Arctic Flora and Fauna
Macrobenthos Chukchi Sea, 1986	ocorrência	Conservation of Arctic Flora and Fauna

Título do conjunto de dados	classe	Nome da instituição responsável
Macrozoobenthos composition, abundance and biomass in the Arctic Ocean along a transect between Svalbard and the Makarov Basin 1991	ocorrência	Conservation of Arctic Flora and Fauna
Megabenthos-Epibenthos Greenland	ocorrência	Conservation of Arctic Flora and Fauna
Muskox Abundance	ocorrência	Conservation of Arctic Flora and Fauna
Oceanexploration2002_vers4	ocorrência	Conservation of Arctic Flora and Fauna
On the composition of the benthic fauna of the western Fram Strait	ocorrência	Conservation of Arctic Flora and Fauna
Phytoplankton from the White Sea, Barents Sea, Norwegian Sea and Arctic Basin 1993-2003	ocorrência	Conservation of Arctic Flora and Fauna
Sea ice meiofauna abundance in coastal fast ice off Barrow, Alaska, with a focus on <i>Scolecipis squamata</i> (Polychaeta), 2005-2006	ocorrência	Conservation of Arctic Flora and Fauna
Sea Ice Nematodes	ocorrência	Conservation of Arctic Flora and Fauna
SealceMeiofaunaCanadaBasin2002	ocorrência	Conservation of Arctic Flora and Fauna
Seasonal dynamics of sub-ice fauna below pack ice in the Arctic (Fram Strait)	ocorrência	Conservation of Arctic Flora and Fauna
Shebaplankton_Falk3	ocorrência	Conservation of Arctic Flora and Fauna
SIO_FAMIZ	ocorrência	Conservation of Arctic Flora and Fauna
Species Composition and Abundance of Zooplankton in the Nearshore Beaufort Sea in Winter-Spring 1978-1980	ocorrência	Conservation of Arctic Flora and Fauna
Spider species diversity from wet and dry habitats	ocorrência	Conservation of Arctic Flora and Fauna
Structure and function of contemporary food webs on Arctic shelves: A panarctic comparison. The pelagic system of the Kara Sea- communities and components of carbon flow	ocorrência	Conservation of Arctic Flora and Fauna
Structures and Nutrition Requirements of Macrozoobenthic Communities in the area of the Lomonossov Ridge, 1995-1998	ocorrência	Conservation of Arctic Flora and Fauna
Taxonomy and zoogeography boundaries of pelagic ostracods in Svalbard waters, 2001-2006	ocorrência	Conservation of Arctic Flora and Fauna
The Ecology of the Inshore Marine Zooplankton of the Chukchi Sea near Point Barrow, Alaska	ocorrência	Conservation of Arctic Flora and Fauna
The pre-winter 2007 vertical distribution of zooplankton in the Cape Bathurst and North Water polynyas, and Lancaster Sound, Canadian Arctic	ocorrência	Conservation of Arctic Flora and Fauna
The Sea Ice Fauna of Frobisher Bay, Arctic Canada 1981 and 1982	ocorrência	Conservation of Arctic Flora and Fauna
Under-ice Amphipods in the Greenland Sea and Fram Strait (Arctic): Environmental Controls and Seasonal Patterns Below the Pack Ice	ocorrência	Conservation of Arctic Flora and Fauna

Título do conjunto de dados	classe	Nome da instituição responsável
Variability of benthic Foraminifera north and south of the Denmark Strait	ocorrência	Conservation of Arctic Flora and Fauna
WhiteSeaPlankton	ocorrência	Conservation of Arctic Flora and Fauna
ZINRAS_Arctic_Benthos	ocorrência	Conservation of Arctic Flora and Fauna
Zoobenthos data from the Southern Beaufort Sea, 1971-1975	ocorrência	Conservation of Arctic Flora and Fauna
Zooplankton Abundancies White Sea, 1972	ocorrência	Conservation of Arctic Flora and Fauna
Zooplankton and Ichthyoplankton Data Collected from the Chukchi and Beaufort Seas during the R/V Mirai Cruise, September 2002	ocorrência	Conservation of Arctic Flora and Fauna
Zooplankton Bering Strait Tiglax 1991	ocorrência	Conservation of Arctic Flora and Fauna
Zooplankton Eastern Arctic Ocean Polarstern, 1995-1998	ocorrência	Conservation of Arctic Flora and Fauna
ZooplanktonBeaufortSeaNOGAP1	ocorrência	Conservation of Arctic Flora and Fauna
ZooplanktonBeaufortSeaNOGAP2	ocorrência	Conservation of Arctic Flora and Fauna
ZooplanktonNOGAP32b1986	ocorrência	Conservation of Arctic Flora and Fauna
CZE National PGR Inventory	ocorrência	Crop Research Institute (CRI)
Abyssal fauna of the UK-1 polymetallic nodule exploration area, Clarion-Clipperton Zone, central Pacific Ocean: Cnidaria	ocorrência	Deep-sea OBIS node
Abyssal fauna of the UK-1 polymetallic nodule exploration claim, Clarion-Clipperton Zone, central Pacific Ocean: Echinodermata	ocorrência	Deep-sea OBIS node
Deep Atlantic Bivalves	ocorrência	Deep-sea OBIS node
Global Distribution and Diversity of Solenidae Species	ocorrência	Deep-sea OBIS node
Macroinvertebrate groups on deep-sea volcanic habitats in the Galapagos Marine Reserve in the Eastern Tropical Pacific Ocean: Mollusca	ocorrência	Deep-sea OBIS node
Megafauna of the UKSRL exploration contract area and eastern Clarion-Clipperton Zone in the Pacific Ocean: Echinodermata	ocorrência	Deep-sea OBIS node
REVIZEE Central Coast Deep Ocean	ocorrência	Deep-sea OBIS node
The Deepwater Program: Northern Gulf of Mexico Continental Slope Habitat and Benthic Ecology - DgoMB: Fishs	ocorrência	Deep-sea OBIS node
The Deepwater Program: Northern Gulf of Mexico Continental Slope Habitat and Benthic Ecology - DgoMB: Inverts	ocorrência	Deep-sea OBIS node
The Deepwater Program: Northern Gulf of Mexico Continental Slope Habitat and Benthic Ecology - DgoMB: Isopods	ocorrência	Deep-sea OBIS node

Título do conjunto de dados	classe	Nome da instituição responsável
The Deepwater Program: Northern Gulf of Mexico Continental Slope Habitat and Benthic Ecology - DgoMB: Macros	ocorrência	Deep-sea OBIS node
The Deepwater Program: Northern Gulf of Mexico Continental Slope Habitat and Benthic Ecology - DgoMB: Meios	ocorrência	Deep-sea OBIS node
The Deepwater Program: Northern Gulf of Mexico Continental Slope Habitat and Benthic Ecology - DgoMB: Polys	ocorrência	Deep-sea OBIS node
The Deepwater Program: Northern Gulf of Mexico Continental Slope Habitat and Benthic Ecology - DgoMB: Trawls	ocorrência	Deep-sea OBIS node
Abyssal fauna of the UK-1 polymetallic nodule exploration area, Clarion-Clipperton Zone, central Pacific Ocean: Mollusca	evento	Deep-sea OBIS node
Nematode genera abundance from Portuguese canyons HERMES HERMIONE	evento	Deep-sea OBIS node
NW Pacific Deep-sea Benthos Biodiversity (Beneficial Project)	evento	Deep-sea OBIS node
NW pacific deep-sea benthos distribution and abundance (Beneficial Project)	evento	Deep-sea OBIS node
Anura Classification	checklist	Encyclopedia of Life (EOL)
Cyphophthalmi Classification	checklist	Encyclopedia of Life (EOL)
Encyclopedia of Life	checklist	Encyclopedia of Life (EOL)
The Diatoms Classification	checklist	Encyclopedia of Life (EOL)
The global Choreutidae classification	checklist	Encyclopedia of Life (EOL)
EUNIS Biodiversity Database	checklist	European Environment Agency
Fauna Europaea	checklist	Fauna Europaea Consortium
Fauna Europaea - Lepidoptera	checklist	Fauna Europaea Consortium
Backbone Family Classification Patch	checklist	GBIF Secretariat
Backbone Nomenclator	checklist	GBIF Secretariat
Families of Living Organisms (FALO)	checklist	GBIF Secretariat
GBIF Algae Classification	checklist	GBIF Secretariat
GBIF Backbone Patch	checklist	GBIF Secretariat
GBIF Backbone Taxonomy	checklist	GBIF Secretariat
GBIF Type Specimen Names	checklist	GBIF Secretariat
ION Taxonomic Hierarchy	checklist	GBIF Secretariat

Título do conjunto de dados	classe	Nome da instituição responsável
Zeigerwerte von Pflanzen & Flechten in Mitteleuropa	checklist	GBIF Secretariat
Index Fungorum	checklist	Index Fungorum Partnership
Dermaptera in the icipe collection	ocorrência	International Centre for Insect Physiology and Ecology
Sand flies of Eastern Africa	ocorrência	International Centre for Insect Physiology and Ecology
Tsetse flies of East Africa update	ocorrência	International Centre for Insect Physiology and Ecology
Biodiversity of the Api Nampa Conservation Area	checklist	International Centre for Integrated Mountain Development
Kangchenjunga Corridor	checklist	International Centre for Integrated Mountain Development
Mahananda Neora Valley Corridor	checklist	International Centre for Integrated Mountain Development
Neora Valley Toorsa Corridor	checklist	International Centre for Integrated Mountain Development
Seed Plants of Barsey Rhododendron Sanctuary	checklist	International Centre for Integrated Mountain Development
Seed Plants of Kangchendzonga Biosphere Reserve	checklist	International Centre for Integrated Mountain Development
Seed Plants of Kangchenjunga Conservation Area	checklist	International Centre for Integrated Mountain Development
Seed Plants of Mahananda Wildlife Sanctuary	checklist	International Centre for Integrated Mountain Development
Seed Plants of Neora Valley National Park	checklist	International Centre for Integrated Mountain Development
Seed Plants of Senchal Wildlife Sanctuary	checklist	International Centre for Integrated Mountain Development
Seed Plants of Singhalila National Park	checklist	International Centre for Integrated Mountain Development
Seed Plants of Toorsa Strict Nature Reserve	checklist	International Centre for Integrated Mountain Development
Senchal Mahananda Corridor	checklist	International Centre for Integrated Mountain Development
Singhalila Senchal Corridor	checklist	International Centre for Integrated Mountain Development

Título do conjunto de dados	classe	Nome da instituição responsável
		Development
Toorsa Jigme Dorji Corridor	checklist	International Centre for Integrated Mountain Development
Official Lists and Indexes of Names in Zoology	checklist	International Commission on Zoological Nomenclature
ZooBank	checklist	International Commission on Zoological Nomenclature
ICTV Master Species List 2019 v2	checklist	International Committee on Taxonomy of Viruses
Global Compositae Checklist (GCC)	checklist	International Compositae Alliance
IOC World Bird List, v10.2	checklist	International Ornithologists Union (IOU)
Multilingual IOC World Bird List, vnull	checklist	International Ornithologists Union (IOU)
Global Invasive Species Database	checklist	International Union for Conservation of Nature
IUCN Red List of Threatened Species	checklist	International Union for Conservation of Nature
IUCN Red List assessment occurrence data for freshwater species native to the Lake Malawi/Nyasa/Niassa Catchment	ocorrência	International Union for Conservation of Nature
MICROBIS database	ocorrência	Marine Biology Laboratory
Nordic Crop Wild Relative (CWR) Checklist	checklist	Nordic Genetic Resource Center (NORDGEN)
Nordic Genetic Resources	ocorrência	Nordic Genetic Resource Center (NORDGEN)
Fauna Atlas N.T.	ocorrência	Northern Territory Department of Environment and Natural Resources
Flora Atlas N.T.	ocorrência	Northern Territory Department of Environment and Natural Resources
Species named after famous people	checklist	Public Domain Datasets
Reef Life Survey: Cryptic Fish	ocorrência	Reef Life Survey
Airborne Bacteria from Miers Valley, Antarctica	metadado de recurso	SCAR - Microbial Antarctic Resource System
Antarctic cryptoendolithic fungal communities ITS amplicon sequencing	metadado de recurso	SCAR - Microbial Antarctic Resource System
Antarctic pack ice Bacterial (16S) communities	metadado de recurso	SCAR - Microbial Antarctic Resource System
Antarctic Peninsula Bacterioplankton 16S rRNA gene surveys and metagenomes from Winter 2002 and Summer 2006.	metadado de recurso	SCAR - Microbial Antarctic Resource System

Título do conjunto de dados	classe	Nome da instituição responsável
Antarctic polar desert surface soil microbiome	metadado de recurso	SCAR - Microbial Antarctic Resource System
Antarctic snow algae communities	metadado de recurso	SCAR - Microbial Antarctic Resource System
Antarctic Surface Snow Bacterial Communities	metadado de recurso	SCAR - Microbial Antarctic Resource System
Arctic multi-year sea ice Bacteria	metadado de recurso	SCAR - Microbial Antarctic Resource System
Arctic Ocean microbial metagenomes sampled aboard CGC Healy during the 2015 GEOTRACES Arctic research cruise	metadado de recurso	SCAR - Microbial Antarctic Resource System
Associated microbiome (Bacteria and Fungi) in Antarctic sea stars (healthy versus exhibiting epidermal disease)	metadado de recurso	SCAR - Microbial Antarctic Resource System
Bacteria (16S ssu rRNA) in an Antarctic snow sample	metadado de recurso	SCAR - Microbial Antarctic Resource System
Bacteria (16S) and Eukaryotes (18S) from the Sor Rondane Mountains (East Antarctica)	metadado de recurso	SCAR - Microbial Antarctic Resource System
Bacteria (16S) in Antarctic terrestrial soils of the Sor Rondane Mountains	metadado de recurso	SCAR - Microbial Antarctic Resource System
Bacteria (16S) in growth laminae of a large conical mats from Lake Untersee, East Antarctic	metadado de recurso	SCAR - Microbial Antarctic Resource System
Bacteria and Archaea biodiversity in Arctic and Subarctic terrestrial ecosystems in Alaska	metadado de recurso	SCAR - Microbial Antarctic Resource System
Bacteria and Archaea biodiversity in Arctic terrestrial ecosystems affected by climate change in Northern Siberia	metadado de recurso	SCAR - Microbial Antarctic Resource System
Bacteria and Archaea biodiversity in terrestrial ecosystems of Chilean Patagonia	metadado de recurso	SCAR - Microbial Antarctic Resource System
Bacteria and Archaea in different soil types on King George Island (South Shetland Islands, Antarctica)	metadado de recurso	SCAR - Microbial Antarctic Resource System
Bacteria in Antarctic glacial foreland soils	metadado de recurso	SCAR - Microbial Antarctic Resource System
Bacteria on mummified seals in the Antarctic Dry Valleys	metadado de recurso	SCAR - Microbial Antarctic Resource System
Bacterial community during <i>Phaeocystis antarctica</i> blooms (Amundsen Sea polynya)	metadado de	SCAR - Microbial Antarctic Resource System

Título do conjunto de dados	classe	Nome da instituição responsável
	recurso	
Bacterial Diversity Assessment in Antarctic Terrestrial and Aquatic Microbial Mats	metadado de recurso	SCAR - Microbial Antarctic Resource System
Bacterial diversity in closed cryoconite holes in Southern Victoria Land (Antarctica)	metadado de recurso	SCAR - Microbial Antarctic Resource System
Bacterial in the Amundsen Sea Polynya (Southern Ocean): community composition in environmental samples and mesocosm experiment	metadado de recurso	SCAR - Microbial Antarctic Resource System
benthic communities bacterial in Antarctica and the arctic	metadado de recurso	SCAR - Microbial Antarctic Resource System
Benthic microbial communities (Bacteria, 16S) of coastal terrestrial and ice shelf Antarctic meltwater ponds	metadado de recurso	SCAR - Microbial Antarctic Resource System
Community diversity in microbial eukaryotes from lakes in the the Vestfold Hills, Antarctica	metadado de recurso	SCAR - Microbial Antarctic Resource System
Cyanobacteria in microbial mats from Antarctic lakes	metadado de recurso	SCAR - Microbial Antarctic Resource System
fumarolic microbial communities at Tramway Ridge, Mt. Erebus, Antarctica	metadado de recurso	SCAR - Microbial Antarctic Resource System
Fungal communities (ITS) in Antarctic brines.	metadado de recurso	SCAR - Microbial Antarctic Resource System
Fungi (LSU) in soils from the South Shetland island, Antarctic Peninsula island, and Union Glacier	metadado de recurso	SCAR - Microbial Antarctic Resource System
Geomicrobiology of Antarctic Subglacial Environments - Subglacial Lake Whillans	metadado de recurso	SCAR - Microbial Antarctic Resource System
Global biogeography of desert cyanobacteria	metadado de recurso	SCAR - Microbial Antarctic Resource System
Hypolithic and soil bacteria in Miers_Valley, Antarctica	metadado de recurso	SCAR - Microbial Antarctic Resource System
Inter-Valley Soil Comparative Survey of the McMurdo Dry Valleys	metadado de recurso	SCAR - Microbial Antarctic Resource System
Marine bacterial, archaeal and eukaryotic microbial communities on the continental shelf of the western Antarctic Peninsula	metadado de recurso	SCAR - Microbial Antarctic Resource System
Marine bacterioplankton community structure in the vicinity of Antarctic icebergs	metadado de recurso	SCAR - Microbial Antarctic Resource System

Título do conjunto de dados	classe	Nome da instituição responsável
Marine_ microbial eukaryotes (18S) from the Ross Sea (Antarctica)	metadado de recurso	SCAR - Microbial Antarctic Resource System
Maritime and Sub-Antarctic microbial soil fungi communities	metadado de recurso	SCAR - Microbial Antarctic Resource System
Metagenome assembly of an Amundsen Sea (Antarctica, Southern Ocean) water sample	metadado de recurso	SCAR - Microbial Antarctic Resource System
Metatranscriptome from microbial mats in Arctic and Antarctic polar tundra environments	metadado de recurso	SCAR - Microbial Antarctic Resource System
Microbes (16S and 18S rRNA genes) from benthic Antarctic sediments	metadado de recurso	SCAR - Microbial Antarctic Resource System
Microbes (Eukaryotes and Archaea) in sea water from Fildes Peninsula (King George Island, Antarctica)	metadado de recurso	SCAR - Microbial Antarctic Resource System
Microbial (16S) diversity in sediments of the former subglacial Hodgson lake (Antarctica)	metadado de recurso	SCAR - Microbial Antarctic Resource System
Microbial (Bacteria, 16S) Diversity in Antarctic marine sediments (Admiralty Bay and Bransfield Strait)	metadado de recurso	SCAR - Microbial Antarctic Resource System
Microbial communities (Bacteria and Archaea) in Lake Fryxell (Antarctica) along an oxygen gradient	metadado de recurso	SCAR - Microbial Antarctic Resource System
Microbial communities (Bacteria, Eukaryotes and Fungi) in Arctic, Antarctic and Sub-Antarctic lacustrine biofilms	metadado de recurso	SCAR - Microbial Antarctic Resource System
Microbial communities (Eukaryotes and Bacteria) of transiently wetted vs. arid Antarctic Dry Valley soils	metadado de recurso	SCAR - Microbial Antarctic Resource System
Microbial diversity (Bacteria and Archaea 16S rRNA gene) in geothermal sites of Deception Island volcano, Antarctica	metadado de recurso	SCAR - Microbial Antarctic Resource System
microbial Eukaryotes in lakes along an Argentinian-Antarctic geographical gradient	metadado de recurso	SCAR - Microbial Antarctic Resource System
Microbial fungal communities (18S) of Antarctic Dry Valley lakes	metadado de recurso	SCAR - Microbial Antarctic Resource System
Microbial Fungi in soils on different Sub-Antarctic islands	metadado de recurso	SCAR - Microbial Antarctic Resource System
Microbial metabarcoding surveys (Bacteria, Archaea and Eukaryota) of the arctic marine environment	metadado de recurso	SCAR - Microbial Antarctic Resource System
Microbial population dynamics along a terrestrial Antarctic moisture gradient	metadado de	SCAR - Microbial Antarctic Resource System

Título do conjunto de dados	classe	Nome da instituição responsável
	recurso	
Microbial soil Fungi (ITS2) diversity from Maritime Antarctica	metadado de recurso	SCAR - Microbial Antarctic Resource System
Microbiome (Archaea, Bacteria and Eukaryota) of lake Fryxell and lake Bonney (McMurdo Dry Valleys, Antarctica)	metadado de recurso	SCAR - Microbial Antarctic Resource System
Microbiome (Archaea, Bacteria and Fungi) in soils from King George Island (Antarctica)	metadado de recurso	SCAR - Microbial Antarctic Resource System
Microbiome (Bacteria, Archaea and fungi) from University Valley Permafrost cores (Antarctica)	metadado de recurso	SCAR - Microbial Antarctic Resource System
Microorganisms (Bacteria 16S; Eukaryota 18S) from soil crusts in polar (Arctic and Antarctic) ecosystems	metadado de recurso	SCAR - Microbial Antarctic Resource System
Microorganisms (Bacteria and Archaea) in the marine cavity beneath the McMurdo Ice Shelf, Antarctica	metadado de recurso	SCAR - Microbial Antarctic Resource System
Microorganisms (Bacteria, Archaea and phototroph eukaryotes) from Fildes Bay, King George Island, Antarctica	metadado de recurso	SCAR - Microbial Antarctic Resource System
Microorganisms (eukaryote and bacteria) in Antarctic cryoconite holes.	metadado de recurso	SCAR - Microbial Antarctic Resource System
Microorganisms in frost flowers on young Arctic sea ice, comparison between different ice types	metadado de recurso	SCAR - Microbial Antarctic Resource System
Microorganisms in Paleomats at the shores of lake Vanda, Antarctica	metadado de recurso	SCAR - Microbial Antarctic Resource System
Near-shore microbial communities (Eukaryotes, Bacteria and Archaea) of the sub-Antarctic Prince Edward Islands	metadado de recurso	SCAR - Microbial Antarctic Resource System
New Zealand Terrestrial Biocomplexity Survey	metadado de recurso	SCAR - Microbial Antarctic Resource System
Palmer LTER 2014 cruise microbial Eukaryote (18s) plankton	metadado de recurso	SCAR - Microbial Antarctic Resource System
Pelagic microbes (Eukaryote 18S and Material 16S amplicons) from sites in the Southern, Indian and Arctic Oceans	metadado de recurso	SCAR - Microbial Antarctic Resource System
RNA-Virome (RNA shotgun sequencing) from lake Limnopolar on Livingston Island (Antarctica)	metadado de recurso	SCAR - Microbial Antarctic Resource System
Sequenced genes (ureC gene) and a metagenome from Archaea in Arctic and Antarctic marine environments	metadado de recurso	SCAR - Microbial Antarctic Resource System

Título do conjunto de dados	classe	Nome da instituição responsável
Soil microbiome (Bacteria) on Keller Peninsula (Antarctica)	metadado de recurso	SCAR - Microbial Antarctic Resource System
Southern Ocean Experimental Metatranscriptome to Investigate Micronutrient Colimitation	metadado de recurso	SCAR - Microbial Antarctic Resource System
Southern Ocean microbial dataset	metadado de recurso	SCAR - Microbial Antarctic Resource System
Surface and deep marine bacterial communities in the Arctic and Antarctic	metadado de recurso	SCAR - Microbial Antarctic Resource System
Synoicum_adareanum_microbiome	metadado de recurso	SCAR - Microbial Antarctic Resource System
The skin microbiome (based on 16S) of Antarctic Humpback whales (Megaptera novaeangliae)	metadado de recurso	SCAR - Microbial Antarctic Resource System
V3-V4 16S rDNA of soil bacteria in King George Island and Deception Island	metadado de recurso	SCAR - Microbial Antarctic Resource System
Western Antarctic Peninsula Bacteria (Mesocosm exp.)	metadado de recurso	SCAR - Microbial Antarctic Resource System
Winter Picoplankton Diversity and Distribution in the US Antarctic Marine Living Resources Study Area - Northern Antarctic Peninsula	metadado de recurso	SCAR - Microbial Antarctic Resource System
BCCM/ULC Cyanobacteria culture collection at the University of Liège (ISO9001 certified)	ocorrência	SCAR - Microbial Antarctic Resource System
DNA metabarcoding of the prey and microbiome of museum specimen Antarctic trematomid fishes	ocorrência	SCAR - Microbial Antarctic Resource System
Geothermal Fumarole Subsurface Mt. Erebus, Antarctica	ocorrência	SCAR - Microbial Antarctic Resource System
Heterotrophic bacterial diversity in aquatic microbial mat communities from Antarctica	evento	SCAR - Microbial Antarctic Resource System
Senckenberg - CeDAMar Resource	ocorrência	Senckenberg - CeDAMar Provider
SA Fauna (BDBSA)	ocorrência	South Australia, Department for Environment and Water
SA Flora (BDBSA)	ocorrência	South Australia, Department for Environment and Water
3i taxonomic databases, Curculionidae, subfamily Entiminae in the Catalogue of Life	checklist	The Catalogue of Life Partnership
A Global Information System for Lichenized and Non-Lichenized Ascomycetes in the Catalogue of Life	checklist	The Catalogue of Life Partnership
A World Catalogue of Centipedes (Chilopoda) for the web in the Catalogue of Life	checklist	The Catalogue of Life Partnership
An online nomenclatural information system of Eumycetozoa in the Catalogue of Life	checklist	The Catalogue of Life Partnership

Título do conjunto de dados	classe	Nome da instituição responsável
Annonaceae GSD in the Catalogue of Life	checklist	The Catalogue of Life Partnership
Annotated checklist of the rotifers (Phylum Rotifera) in the Catalogue of Life	checklist	The Catalogue of Life Partnership
Aphid Species File in the Catalogue of Life	checklist	The Catalogue of Life Partnership
Ascidiacea World Database in the Catalogue of Life	checklist	The Catalogue of Life Partnership
Bdellid & Cunaxid Databases in the Catalogue of Life	checklist	The Catalogue of Life Partnership
Brachiopoda Database in the Catalogue of Life	checklist	The Catalogue of Life Partnership
Brassicaceae species checklist and database in the Catalogue of Life	checklist	The Catalogue of Life Partnership
Brentidae of the World in the Catalogue of Life	checklist	The Catalogue of Life Partnership
Catalogue of Craneflies of the World in the Catalogue of Life	checklist	The Catalogue of Life Partnership
Catalogue of Life	checklist	The Catalogue of Life Partnership
Catalogue of Life	checklist	The Catalogue of Life Partnership
Catalogue of Life China in the Catalogue of Life	checklist	The Catalogue of Life Partnership
Catalogue of the Pterophoroidea of the World in the Catalogue of Life	checklist	The Catalogue of Life Partnership
Cerambycidae database in the Catalogue of Life	checklist	The Catalogue of Life Partnership
Cercopoidea Organised On Line in the Catalogue of Life	checklist	The Catalogue of Life Partnership
Checklist of Ferns and Lycophytes of the World in the Catalogue of Life	checklist	The Catalogue of Life Partnership
Chrysididae Species File in the Catalogue of Life	checklist	The Catalogue of Life Partnership
Cockroach Species File in the Catalogue of Life	checklist	The Catalogue of Life Partnership
Computer Aided Identification of Phlebotomine sandflies of Americas in the Catalogue of Life	checklist	The Catalogue of Life Partnership
Conifer Database in the Catalogue of Life	checklist	The Catalogue of Life Partnership
Coreoidea Species File in the Catalogue of Life	checklist	The Catalogue of Life Partnership
Dermaptera Species File in the Catalogue of Life	checklist	The Catalogue of Life Partnership
Dothideomycetes in the Catalogue of Life	checklist	The Catalogue of Life Partnership
Droseraceae Database in the Catalogue of Life	checklist	The Catalogue of Life Partnership
Early Land Plants Today in the Catalogue of Life	checklist	The Catalogue of Life Partnership
Electronic Catalogue of Weevil names (Curculionoidea) in the Catalogue of Life	checklist	The Catalogue of Life Partnership

Título do conjunto de dados	classe	Nome da instituição responsável
Embioptera Species File in the Catalogue of Life	checklist	The Catalogue of Life Partnership
ETI WBD: World Biodiversity Database (Euphausiacea) in the Catalogue of Life	checklist	The Catalogue of Life Partnership
FishBase in the Catalogue of Life	checklist	The Catalogue of Life Partnership
Fossil Ginkgoales in the Catalogue of Life	checklist	The Catalogue of Life Partnership
Fulgoromorpha Lists On the WEB in the Catalogue of Life	checklist	The Catalogue of Life Partnership
Geranium Taxonomic Information System in the Catalogue of Life	checklist	The Catalogue of Life Partnership
Global Butterfly Information System in the Catalogue of Life	checklist	The Catalogue of Life Partnership
Global database of ground beetles in the Catalogue of Life	checklist	The Catalogue of Life Partnership
Global Species Database of Odonata in the Catalogue of Life	checklist	The Catalogue of Life Partnership
Global Strepsiptera Database in the Catalogue of Life	checklist	The Catalogue of Life Partnership
Global Taxonomic Database of Gracillariidae in the Catalogue of Life	checklist	The Catalogue of Life Partnership
Global Taxonomic Database of Tineidae (Lepidoptera) in the Catalogue of Life	checklist	The Catalogue of Life Partnership
Glomeromycota in the Catalogue of Life	checklist	The Catalogue of Life Partnership
Grylloblattodea Species File in the Catalogue of Life	checklist	The Catalogue of Life Partnership
Hexacorallians (Actiniaria) of the World in the Catalogue of Life	checklist	The Catalogue of Life Partnership
Home of the Xylariaceae in the Catalogue of Life	checklist	The Catalogue of Life Partnership
Hymenoptera Information System in the Catalogue of Life	checklist	The Catalogue of Life Partnership
ICN Appendix II-VII	checklist	The Catalogue of Life Partnership
ICTV/MSL: International Committee on Taxonomy of Viruses / Master Species List in the Catalogue of Life	checklist	The Catalogue of Life Partnership
ILDIS World Database of Legumes in the Catalogue of Life	checklist	The Catalogue of Life Partnership
Illustrated catalog of Tessaratomidae in the Catalogue of Life	checklist	The Catalogue of Life Partnership
Isoptera Species File in the Catalogue of Life	checklist	The Catalogue of Life Partnership
Lace Bugs Database (Hemiptera: Tingidae) in the Catalogue of Life	checklist	The Catalogue of Life Partnership
LDL Neuropterida Species of the World in the Catalogue of Life	checklist	The Catalogue of Life Partnership
Lygaeoidea Species File in the Catalogue of Life	checklist	The Catalogue of Life Partnership
Mantodea Species File in the Catalogue of Life	checklist	The Catalogue of Life Partnership

Título do conjunto de dados	classe	Nome da instituição responsável
Mantophasmatodea Species File in the Catalogue of Life	checklist	The Catalogue of Life Partnership
Membracoidea of the World Database in the Catalogue of Life	checklist	The Catalogue of Life Partnership
MilliBase in the Catalogue of Life	checklist	The Catalogue of Life Partnership
Mites GSDs: OlogamasidBase in the Catalogue of Life	checklist	The Catalogue of Life Partnership
Mites GSDs: PhytoseiidBase in the Catalogue of Life	checklist	The Catalogue of Life Partnership
Mites GSDs: RhodacaridBase in the Catalogue of Life	checklist	The Catalogue of Life Partnership
Mites GSDs: TenuipalpidBase in the Catalogue of Life	checklist	The Catalogue of Life Partnership
MolluscaBase in the Catalogue of Life	checklist	The Catalogue of Life Partnership
Moss Bug Base in the Catalogue of Life	checklist	The Catalogue of Life Partnership
Moss TROPICOS Database in the Catalogue of Life	checklist	The Catalogue of Life Partnership
Nepticulidae and Opostegidae of the World in the Catalogue of Life	checklist	The Catalogue of Life Partnership
New Zealand Inventory of Biodiversity in the Catalogue of Life	checklist	The Catalogue of Life Partnership
On-line Systematic Catalog of Plant Bugs (Insecta: Heteroptera: Miridae) in the Catalogue of Life	checklist	The Catalogue of Life Partnership
Orthoptera Species File in the Catalogue of Life	checklist	The Catalogue of Life Partnership
Phasmida Species File in the Catalogue of Life	checklist	The Catalogue of Life Partnership
Phoronida Database in the Catalogue of Life	checklist	The Catalogue of Life Partnership
Phyllachorales in the Catalogue of Life	checklist	The Catalogue of Life Partnership
Phylum Ctenophora, a list of all valid species names harvested from the Internet between 1998 - present in the Catalogue of Life	checklist	The Catalogue of Life Partnership
Plecoptera Species File in the Catalogue of Life	checklist	The Catalogue of Life Partnership
Psocodea Species File in the Catalogue of Life	checklist	The Catalogue of Life Partnership
Psyllist: Psylloidea database in the Catalogue of Life	checklist	The Catalogue of Life Partnership
Rhytismatales in the Catalogue of Life	checklist	The Catalogue of Life Partnership
Saccharomycetes in the Catalogue of Life	checklist	The Catalogue of Life Partnership
SF Coleorrhyncha: Coleorrhyncha Species File in the Catalogue of Life	checklist	The Catalogue of Life Partnership
Species Fungorum in the Catalogue of Life	checklist	The Catalogue of Life Partnership

Título do conjunto de dados	classe	Nome da instituição responsável
Spider Mites Web in the Catalogue of Life	checklist	The Catalogue of Life Partnership
Staphyliniformia world catalog database in the Catalogue of Life	checklist	The Catalogue of Life Partnership
Synonymic Checklists of the Vascular Plants of the World in the Catalogue of Life	checklist	The Catalogue of Life Partnership
Systema Dipteriorum in the Catalogue of Life	checklist	The Catalogue of Life Partnership
Systematic Database of the Scale Insects of the World in the Catalogue of Life	checklist	The Catalogue of Life Partnership
Taxapad Ichneumonoidea in the Catalogue of Life	checklist	The Catalogue of Life Partnership
Taxonomic checklist of the world's whiteflies (Insecta: Hemiptera: Aleyrodidae) in the Catalogue of Life	checklist	The Catalogue of Life Partnership
The dinoflagellate genus Gymnodinium checklist in the Catalogue of Life	checklist	The Catalogue of Life Partnership
The Global Lepidoptera Names Index in the Catalogue of Life	checklist	The Catalogue of Life Partnership
The Integrated Taxonomic Information System in the Catalogue of Life	checklist	The Catalogue of Life Partnership
The Integrated Taxonomic Information System in the Catalogue of Life	checklist	The Catalogue of Life Partnership
The Paleobiology Database in the Catalogue of Life	checklist	The Catalogue of Life Partnership
The Reptile Database in the Catalogue of Life	checklist	The Catalogue of Life Partnership
The Scorpion Files in the Catalogue of Life	checklist	The Catalogue of Life Partnership
The World Ciliate Catalog in the Catalogue of Life	checklist	The Catalogue of Life Partnership
The World List of Cycads in the Catalogue of Life	checklist	The Catalogue of Life Partnership
The World of Jewel Beetles in the Catalogue of Life	checklist	The Catalogue of Life Partnership
ThripsWiki - providing information on the Worlds thrips in the Catalogue of Life	checklist	The Catalogue of Life Partnership
TicksBase in the Catalogue of Life	checklist	The Catalogue of Life Partnership
Trichomycetes in the Catalogue of Life	checklist	The Catalogue of Life Partnership
Unicellular spore-forming protozoan parasites in the Catalogue of Life	checklist	The Catalogue of Life Partnership
Universal Chalcidoidea Database in the Catalogue of Life	checklist	The Catalogue of Life Partnership
World Amphipoda Database in the Catalogue of Life	checklist	The Catalogue of Life Partnership
World Asteroidea Database in the Catalogue of Life	checklist	The Catalogue of Life Partnership
World Auchenorrhyncha Database in the Catalogue of Life	checklist	The Catalogue of Life Partnership
World Bee Checklist in the Catalogue of Life	checklist	The Catalogue of Life Partnership

Título do conjunto de dados	classe	Nome da instituição responsável
World checklist of freshwater Cladocera species in the Catalogue of Life	checklist	The Catalogue of Life Partnership
World checklist of freshwater Ephemeroptera species in the Catalogue of Life	checklist	The Catalogue of Life Partnership
World checklist of freshwater Halacaridae species in the Catalogue of Life	checklist	The Catalogue of Life Partnership
World checklist of freshwater Nematomorpha species in the Catalogue of Life	checklist	The Catalogue of Life Partnership
World Checklist of Selected Plant Families in the Catalogue of Life	checklist	The Catalogue of Life Partnership
World Cumacea Database in the Catalogue of Life	checklist	The Catalogue of Life Partnership
World Database of Fleas in the Catalogue of Life	checklist	The Catalogue of Life Partnership
World Echinoidea Database in the Catalogue of Life	checklist	The Catalogue of Life Partnership
World Foraminifera Database in the Catalogue of Life	checklist	The Catalogue of Life Partnership
World Gastrotricha Database in the Catalogue of Life	checklist	The Catalogue of Life Partnership
World Hydrozoa Database in the Catalogue of Life	checklist	The Catalogue of Life Partnership
World List of Antipatharia in the Catalogue of Life	checklist	The Catalogue of Life Partnership
World List of Appendicularia in the Catalogue of Life	checklist	The Catalogue of Life Partnership
World List of Bochusacea in the Catalogue of Life	checklist	The Catalogue of Life Partnership
World List of Brachypoda in the Catalogue of Life	checklist	The Catalogue of Life Partnership
World List of Bryozoa in the Catalogue of Life	checklist	The Catalogue of Life Partnership
World List of Cephalochordata in the Catalogue of Life	checklist	The Catalogue of Life Partnership
World list of Ceriantharia in the Catalogue of Life	checklist	The Catalogue of Life Partnership
World List of Cestoda in the Catalogue of Life	checklist	The Catalogue of Life Partnership
World List of Chaetognatha in the Catalogue of Life	checklist	The Catalogue of Life Partnership
World List of Corallimorpharia in the Catalogue of Life	checklist	The Catalogue of Life Partnership
World list of Cubozoa in the Catalogue of Life	checklist	The Catalogue of Life Partnership
World List of Gnathostomulida in the Catalogue of Life	checklist	The Catalogue of Life Partnership
World List of Holothuroidea in the Catalogue of Life	checklist	The Catalogue of Life Partnership
World List of Kinorhyncha in the Catalogue of Life	checklist	The Catalogue of Life Partnership
World List of Leptostraca in the Catalogue of Life	checklist	The Catalogue of Life Partnership

Título do conjunto de dados	classe	Nome da instituição responsável
World list of Loricifera in the Catalogue of Life	checklist	The Catalogue of Life Partnership
World List of marine Brachyura in the Catalogue of Life	checklist	The Catalogue of Life Partnership
World List of Marine Oligochaeta in the Catalogue of Life	checklist	The Catalogue of Life Partnership
World List of Merostomata in the Catalogue of Life	checklist	The Catalogue of Life Partnership
World List of Monogenea in the Catalogue of Life	checklist	The Catalogue of Life Partnership
World List of Mystacocarida in the Catalogue of Life	checklist	The Catalogue of Life Partnership
World list of Myxozoa in the Catalogue of Life	checklist	The Catalogue of Life Partnership
World List of Octocorallia in the Catalogue of Life	checklist	The Catalogue of Life Partnership
World List of Orthonectida in the Catalogue of Life	checklist	The Catalogue of Life Partnership
World List of Polycystina (Radiolaria) in the Catalogue of Life	checklist	The Catalogue of Life Partnership
World List of Priapulida in the Catalogue of Life	checklist	The Catalogue of Life Partnership
World List of Rhombozoa in the Catalogue of Life	checklist	The Catalogue of Life Partnership
World List of Scleractinia in the Catalogue of Life	checklist	The Catalogue of Life Partnership
World list of Scyphozoa in the Catalogue of Life	checklist	The Catalogue of Life Partnership
World list of Staurozoa in the Catalogue of Life	checklist	The Catalogue of Life Partnership
World List of Tanaidacea in the Catalogue of Life	checklist	The Catalogue of Life Partnership
World List of Tantulocarida in the Catalogue of Life	checklist	The Catalogue of Life Partnership
World list of Thaliacea in the Catalogue of Life	checklist	The Catalogue of Life Partnership
World List of Thermosbaenacea in the Catalogue of Life	checklist	The Catalogue of Life Partnership
World List of Trematoda in the Catalogue of Life	checklist	The Catalogue of Life Partnership
World List of turbellarian worms (Acoelomorpha, Catenulida, Rhabditophora) in the Catalogue of Life	checklist	The Catalogue of Life Partnership
World List of Xenoturbellida in the Catalogue of Life	checklist	The Catalogue of Life Partnership
World List of Zoantharia in the Catalogue of Life	checklist	The Catalogue of Life Partnership
World Marine, Freshwater and Terrestrial Isopod Crustaceans database in the Catalogue of Life	checklist	The Catalogue of Life Partnership
World Nemertea Database in the Catalogue of Life	checklist	The Catalogue of Life Partnership

Título do conjunto de dados	classe	Nome da instituição responsável
World of Copepods database in the Catalogue of Life	checklist	The Catalogue of Life Partnership
World Ophiuroidea database in the Catalogue of Life	checklist	The Catalogue of Life Partnership
World Ostracoda Database in the Catalogue of Life	checklist	The Catalogue of Life Partnership
World Placozoa Database in the Catalogue of Life	checklist	The Catalogue of Life Partnership
World Polychaeta database in the Catalogue of Life	checklist	The Catalogue of Life Partnership
World Porifera database in the Catalogue of Life	checklist	The Catalogue of Life Partnership
World Pycnogonida Database in the Catalogue of Life	checklist	The Catalogue of Life Partnership
World Remipedia Database in the Catalogue of Life	checklist	The Catalogue of Life Partnership
World Scarabaeidae Database in the Catalogue of Life	checklist	The Catalogue of Life Partnership
World Spider Catalog in the Catalogue of Life	checklist	The Catalogue of Life Partnership
World Wide Wattle in the Catalogue of Life	checklist	The Catalogue of Life Partnership
ZOBODAT: Zoological-Botanical Database (Vespoidea) in the Catalogue of Life	checklist	The Catalogue of Life Partnership
Zoraptera Species File in the Catalogue of Life	checklist	The Catalogue of Life Partnership
Zygomycetes in the Catalogue of Life	checklist	The Catalogue of Life Partnership
APG IV: Angiosperm Phylogeny Group classification for the orders and families of flowering plants	ocorrência	The Catalogue of Life Partnership
International Barcode of Life project (iBOL) Barcode Index Numbers (BINs)	checklist	The International Barcode of Life Consortium
International Barcode of Life project (iBOL)	ocorrência	The International Barcode of Life Consortium
International Plant Names Index	checklist	The International Plant Names Index Collaborators
Tree of Life Web Project Classification	checklist	The Tree of Life Web Project

ANEXO 10. Lista de conjuntos de dados e instituições vinculados ao Brasil

	Título do <i>dataset</i>	Classe <i>dataset</i>	ano*	Instituição publicadora	DOI
1	Composition of two campo rupestre communities from the Itacolomi State Park, Calais (present dataset), (related dataset: Lagoa Seca), MG, Brazil	checklist	2014	Laboratory of Ecology and Evolution of Plants, at Universidade Federal de Vicosa	10.5539/enrr.v2n3p115
2	Composition of two campo rupestre communities from the Itacolomi State Park, Calais (present dataset), (related dataset: Lagoa Seca), MG, Brazil	checklist	2014	Laboratory of Ecology and Evolution of Plants, at Universidade Federal de Vicosa	10.5539/enrr.v2n3p115
3	Composition of two campo rupestre communities from the Itacolomi State Park, Calais (present dataset), (related dataset: Lagoa Seca), MG, Brazil	checklist	2014	Laboratory of Ecology and Evolution of Plants, at Universidade Federal de Vicosa	10.5539/enrr.v2n3p115
4	Composition of two campo rupestre communities from the Itacolomi State Park, Calais (present dataset), (related dataset: Lagoa Seca), MG, Brazil	checklist	2014	Laboratory of Ecology and Evolution of Plants, at Universidade Federal de Vicosa	10.5539/enrr.v2n3p115
5	Composition of two campo rupestre communities from the Itacolomi State Park, Calais (present dataset), (related dataset: Lagoa Seca), MG, Brazil	checklist	2014	Laboratory of Ecology and Evolution of Plants, at Universidade Federal de Vicosa	10.5539/enrr.v2n3p115
6	Composition of two campo rupestre communities from the Itacolomi State Park, Calais (present dataset), (related dataset: Lagoa Seca), MG, Brazil	checklist	2014	Laboratory of Ecology and Evolution of Plants, at Universidade Federal de Vicosa	10.5539/enrr.v2n3p115
7	Composition of two campo rupestre communities from the Itacolomi State Park, Lagoa Seca (present dataset), (related dataset: Calais), MG, Brazil	checklist	2014	Laboratory of Ecology and Evolution of Plants, at Universidade Federal de Vicosa	10.15468/2tug9e
8	Composition of two campo rupestre communities from the Itacolomi State Park, Lagoa Seca (present dataset), (related dataset: Calais), MG, Brazil	checklist	2014	Laboratory of Ecology and Evolution of Plants, at Universidade Federal de Vicosa	10.15468/rucmx9

9	Composition of two campo rupestre communities from the Itacolomi State Park, Lagoa Seca (present dataset), (related dataset: Calais), MG, Brazil	checklist	2014	Laboratory of Ecology and Evolution of Plants, at Universidade Federal de Vicosa	10.15468/wxbkr6
10	Composition of two campo rupestre communities from the Itacolomi State Park, Lagoa Seca (present dataset), (related dataset: Calais), MG, Brazil	checklist	2014	Laboratory of Ecology and Evolution of Plants, at Universidade Federal de Vicosa	10.15468/zqulv0
11	Composition of two campo rupestre communities from the Itacolomi State Park, Lagoa Seca (present dataset), (related dataset: Calais), MG, Brazil	checklist	2014	Laboratory of Ecology and Evolution of Plants, at Universidade Federal de Vicosa	10.15468/6kwtaw
12	Composition of two campo rupestre communities from the Itacolomi State Park, Lagoa Seca (present dataset), (related dataset: Calais), MG, Brazil	checklist	2014	Laboratory of Ecology and Evolution of Plants, at Universidade Federal de Vicosa	10.15468/iccjvs
13	Composition of two campo rupestre communities from the Itacolomi State Park, Lagoa Seca (present dataset), (related dataset: Calais), MG, Brazil	checklist	2014	Laboratory of Ecology and Evolution of Plants, at Universidade Federal de Vicosa	10.15468/tgp496
14	Brazilian Flora 2020 project - Projeto Flora do Brasil 2020	checklist	2015	Instituto de Pesquisas Jardim Botânico do Rio de Janeiro	10.15468/1mtkaw
15	Catálogo de Plantas das Unidades de Conservação do Brasil - Parque Nacional do Itatiaia (PNI)	checklist	2019	Instituto de Pesquisas Jardim Botânico do Rio de Janeiro	10.15468/afdpuk
16	Levantamento de Arbóreas presentes na campina do RDSU	metadado de recurso	2015	Instituto Nacional de Pesquisas da Amazônia - INPA	10.15468/pic5qy
17	Biologia reprodutiva e dieta de <i>Amerotyphlops brongersmianus</i> (Vanzolini, 1976) da Restinga de Jurubatiba, Rio de Janeiro, Brasil (Serpentes: Typhlopidae)	metadado de recurso	2018	Sistema de Informação sobre a Biodiversidade Brasileira - SiBBr	10.15468/jtctxl
18	Dinâmica populacional e crescimento de <i>Tropidurus torquatus</i> (Wied-Neuwied, 1820) no Parque Nacional da Restinga de Jurubatiba, Rio de Janeiro, Brasil (Squamata: Tropiduridae)	metadado de recurso	2018	Sistema de Informação sobre a Biodiversidade Brasileira - SiBBr	10.15468/hr2x1n

19	Estrutura e composição das comunidades de peixes em lagoas do Parque Nacional da Restinga de Jurubatiba	metadado de recurso	2018	Sistema de Informação sobre a Biodiversidade Brasileira - SiBBr	10.15468/ouiphk
20	Phytoplankton and zooplankton diversity of 18 lakes of the lacustrine system of the Middle Rio Doce	metadado de recurso	2019	Sistema de Informação sobre a Biodiversidade Brasileira - SiBBr	10.15468/udwets
21	records-p-procurrans-and-volcanense	metadado de recurso	2019	Check List	10.15468/csjszx
22	Ecologia_energetica_na_planicie_de_inundacao_do_alto_rio_parana_estudos_de_longo_prazo	metadado de recurso	2020	Sistema de Informação sobre a Biodiversidade Brasileira - SiBBr	10.15468/nrkski
23	Estrutura_e_dinamica_da_comunidade_de_peixes_na_planicie_de_inundacao_do_alto_rio_parana_estudos_de_longo_prazo	metadado de recurso	2020	Sistema de Informação sobre a Biodiversidade Brasileira - SiBBr	10.15468/moppvg
24	Estrutura_e_dinamica_da_comunidade_de_protozoarios_planctonicos_na_planicie_de_inundacao_do_alto_rio_parana_estudos_de_longo_prazo	metadado de recurso	2020	Sistema de Informação sobre a Biodiversidade Brasileira - SiBBr	10.15468/szpjqq
25	estrutura_e_dinamica_da_comunidade_fitoplanctonica_na_planicie_de_inundacao_do_alto_rio_parana_estudos_de_longo_prazo	metadado de recurso	2020	Sistema de Informação sobre a Biodiversidade Brasileira - SiBBr	10.15468/12llnk
26	Estrutura_e_dinamica_da_comunidade_ictioplanctonica_na_planicie_de_inundacao_do_alto_rio_parana_estudos_de_longo_prazo	metadado de recurso	2020	Sistema de Informação sobre a Biodiversidade Brasileira - SiBBr	10.15468/x0mzik
27	Estrutura_e_dinamica_da_comunidade_perifitica_na_planicie_de_inundacao_do_alto_rio_parana_estudos_de_longo_prazo	metadado de recurso	2020	Sistema de Informação sobre a Biodiversidade Brasileira - SiBBr	10.15468/zkcbm1
28	Estrutura_e_dinamica_da_comunidade_zoobentonica_na_planicie_de_inundacao_do_alto_rio_parana_estudos_de_longo_prazo	metadado de recurso	2020	Sistema de Informação sobre a Biodiversidade Brasileira - SiBBr	10.15468/inezdy
29	Estrutura_e_dinamica_da_comunidade_zooplanctonica_na_planicie_de_inundacao_do_alto_rio_Parana_estudos_de_longo_prazo	metadado de recurso	2020	Sistema de Informação sobre a Biodiversidade Brasileira - SiBBr	10.15468/n4icpw

30	Ictioparasitologia_na_planicie_de_inundacao_do_alto_rio_parana_estudos_de_longo_prazo	metadado de recurso	2020	Sistema de Informação sobre a Biodiversidade Brasileira - SiBBr	10.15468/5sdwqi
31	macrofitas_rio_parana	metadado de recurso	2020	Sistema de Informação sobre a Biodiversidade Brasileira - SiBBr	10.15468/ld9bfo
32	CE-UFPE - Coleção Entomológica da UFPE	ocorrência	2007	Universidade Federal de Pernambuco	10.15468/sitdjm
33	CEPANN - Coleção Entomológica Paulo Nogueira-Neto - IB/USP	ocorrência	2007	Universidade de São Paulo	10.15468/bb5nnj
34	Coleção Entomológica Moure e Costa	ocorrência	2007	Centro de Referência em Informação Ambiental	10.15468/4lh1q1
35	Laboratório de Ecologia e Biogeografia de Insetos da Caatinga	ocorrência	2007	Centro de Referência em Informação Ambiental	10.15468/yzgzxk
36	RB - Rio de Janeiro Botanical Garden Herbarium Collection	ocorrência	2013	Instituto de Pesquisas Jardim Botânico do Rio de Janeiro	10.15468/7ep9i2
37	Arachnida Araneae do Museu Nacional (MNRJ ARAC ARAN)	ocorrência	2014	Museu Nacional / UFRJ	10.15468/lpysgc
38	Coleção de Polychaeta do Museu Nacional	ocorrência	2014	Museu Nacional / UFRJ	10.15468/ahcote
39	Community dynamics in a species-rich patch in the Brazilian Atlantic Forest near Viçosa/MG	ocorrência	2014	Laboratory of Ecology and Evolution of Plants, at Universidade Federal de Vicosa	10.15468/6nga6m
40	Fish Collection - Instituto Nacional de Pesquisas da Amazônia (INPA)	ocorrência	2014	Instituto Nacional de Pesquisas da Amazônia - INPA	10.15468/tmlike
41	Herbarium - Instituto Nacional de Pesquisas da Amazônia (INPA)	ocorrência	2014	Instituto Nacional de Pesquisas da Amazônia - INPA	10.15468/5ictpz
42	Museu Paraense Emilio Goeldi - Acari Collection	ocorrência	2014	Museu Paraense Emílio Goeldi	10.15468/hbpbii
43	Museu Paraense Emilio Goeldi - Amblypygi Collection	ocorrência	2014	Museu Paraense Emílio Goeldi	10.15468/jn2xgd
44	Museu Paraense Emilio Goeldi - Araneae Collection	ocorrência	2014	Museu Paraense Emílio Goeldi	10.15468/9dqolw
45	Museu Paraense Emilio Goeldi - Coleoptera Collection	ocorrência	2014	Museu Paraense Emílio Goeldi	10.15468/i7hmns
46	Museu Paraense Emilio Goeldi - Lepidoptera Collection	ocorrência	2014	Museu Paraense Emílio Goeldi	10.15468/oxbgtv
47	Museu Paraense Emílio Goeldi - Ophidia Collection	ocorrência	2014	Museu Paraense Emílio Goeldi	10.15468/lt0wet
48	Museu Paraense Emílio Goeldi - Opiliones Collection	ocorrência	2014	Museu Paraense Emílio Goeldi	10.15468/0qehdl

49	Museu Paraense Emílio Goeldi - Ornithology Collection - Feathers_and_Skin	ocorrência	2014	Museu Paraense Emílio Goeldi	10.15468/54sd1x
50	Museu Paraense Emilio Goeldi - Ricinulei Collection	ocorrência	2014	Museu Paraense Emílio Goeldi	10.15468/cjddos
51	Museu Paraense Emílio Goeldi - Schizomida Collection	ocorrência	2014	Museu Paraense Emílio Goeldi	10.15468/gikgjs
52	Museu Paraense Emilio Goeldi - Scorpiones Collection	ocorrência	2014	Museu Paraense Emílio Goeldi	10.15468/iewwh2
53	Museu Paraense Emílio Goeldi - Solifugae Collection	ocorrência	2014	Museu Paraense Emílio Goeldi	10.15468/7s50hy
54	Museu Paraense Emilio Goeldi - Uropygi Collection	ocorrência	2014	Museu Paraense Emílio Goeldi	10.15468/xt4kfa
55	Myxozoa Collection - Instituto Nacional de Pesquisas da Amazônia	ocorrência	2014	Instituto Nacional de Pesquisas da Amazônia - INPA	10.15468/gvqloo
56	Porifera Collection - Instituto Nacional de Pesquisas da Amazônia	ocorrência	2014	Instituto Nacional de Pesquisas da Amazônia - INPA	10.15468/wcmfnf
57	Porifera Collection - Instituto Nacional de Pesquisas da Amazônia	ocorrência	2014	Instituto Nacional de Pesquisas da Amazônia - INPA	10.15468/a6d2dx
58	Projeto Abrece o Boto Cinza	ocorrência	2014	Instituto Boto Cinza	10.15468/xnen6f
59	The publication of primary biological events in the international GBIF (Global Biodiversity Information Facility) network: A case study with a small dataset of field collected ocorrência samples of Buriti palms (<i>Mauritia flexuosa</i>).	ocorrência	2014	Laboratory of Ecology and Evolution of Plants, at Universidade Federal de Vicosa	10.15468/26yssi
60	The publication of primary biological events in the international GBIF (Global Biodiversity Information Facility) network: A case study with a small dataset of field collected ocorrência samples of Buriti palms (<i>Mauritia flexuosa</i>).	ocorrência	2014	Laboratory of Ecology and Evolution of Plants, at Universidade Federal de Vicosa	10.15468/auz3cf
61	The publication of primary biological events in the international GBIF (Global Biodiversity Information Facility) network: A case study with a small dataset of field collected ocorrência samples of Buriti palms (<i>Mauritia flexuosa</i>).	ocorrência	2014	Laboratory of Ecology and Evolution of Plants, at Universidade Federal de Vicosa	10.15468/lsl8vm

62	The publication of primary biological events in the international GBIF (Global Biodiversity Information Facility) network: A case study with a small dataset of field collected ocorrência samples of Buriti palms (<i>Mauritia flexuosa</i>).	ocorrência	2014	Laboratory of Ecology and Evolution of Plants, at Universidade Federal de Vicosa	10.15468/p3atx9
63	Acari Collection - Instituto Nacional de Pesquisas da Amazônia	ocorrência	2015	Instituto Nacional de Pesquisas da Amazônia - INPA	10.15468/ahzakc
64	ALCB - Herbário Alexandre Leal Costa	ocorrência	2015	Universidade Federal da Bahia	10.15468/wogetb
65	Amblypygi Collection - Instituto Nacional de Pesquisas da Amazônia (INPA)	ocorrência	2015	Instituto Nacional de Pesquisas da Amazônia - INPA	10.15468/hmlx0r
66	ASE - Herbário da Universidade Federal de Sergipe	ocorrência	2015	Universidade Federal de Sergipe	10.15468/urlzfy
67	BHCB-SL - Herbário UFMG - Samambaias e Licófitas	ocorrência	2015	Universidade Federal de Minas Gerais	10.15468/9idm7p
68	BHCB - Herbário da Universidade Federal de Minas Gerais	ocorrência	2015	Universidade Federal de Minas Gerais	10.15468/qgm9ch
69	BHZB - Herbário do Jardim Botânico da Fundação Zoo-Botânica de Belo Horizonte	ocorrência	2015	Fundação Zoo-Botânica de Belo Horizonte/Jardim Botânico	10.15468/eqotat
70	BOTU - Herbário Irina Delanova Gemtchújnicov	ocorrência	2015	Universidade Estadual Paulista - IBB	10.15468/vsjsdc
71	BOTUw - Xiloteca "Profa. Dra. Maria Aparecida Mourão Brasil"	ocorrência	2015	Universidade Estadual Paulista - IBB	10.15468/yd42ud
72	CEN - Herbário da Embrapa Recursos Genéticos e Biotecnologia	ocorrência	2015	Embrapa Cenargen	10.15468/enkiul
73	CEPEC-Fungi - Coleção Micológica do CEPEC	ocorrência	2015	Comissão Executiva do Plano da Lavoura Cacaueira – Centro de Pesquisas do Cacau	10.15468/gx9dd1
74	CEPEC - Herbário do Centro de Pesquisas do Cacau	ocorrência	2015	Comissão Executiva do Plano da Lavoura Cacaueira – Centro de Pesquisas do Cacau	10.15468/wvh0dx
75	CESJ - Herbário Leopoldo Krieger	ocorrência	2015	Universidade Federal de Juiz de Fora	10.15468/9pwrpu
76	CGMS - Herbário da Fundação Universidade Federal de Mato Grosso do Sul	ocorrência	2015	Universidade Federal de Mato Grosso do Sul	10.15468/sdsazf

77	Chilopoda Collection - Instituto Nacional de Pesquisas da Amazônia (INPA)	ocorrência	2015	Instituto Nacional de Pesquisas da Amazônia - INPA	10.15468/6ei1za
78	CNMT - Herbário Centro Norte Mato Grossense	ocorrência	2015	Universidade Federal de Mato Grosso - Campus Universitário de Sinop	10.15468/sx61uz
79	Coleção de Ictiologia do Museu de Ciências Naturais da Universidade do Vale do Taquari - Univates (ZIUMCN-UNIVATES)	ocorrência	2015	Universidade do Vale do Taquari - UNIVATES	10.15468/szixch
80	CPAP - Herbário CPAP	ocorrência	2015	Embrapa Pantanal	10.15468/wdmnid
81	CRI - Herbário Pe. Dr. Raulino Reitz	ocorrência	2015	Universidade do Extremo Sul Catarinense	10.15468/wvAMPL
82	CVRD - Herbário da Reserva Natural Vale	ocorrência	2015	CVRD - Reserva Natural Vale	10.15468/vmsmgr
83	Diplopoda Collection - Instituto Nacional de Pesquisas da Amazônia (INPA)	ocorrência	2015	Instituto Nacional de Pesquisas da Amazônia - INPA	10.15468/kdzsr0
84	Diptera Collection - Instituto Nacional de Pesquisas da Amazônia (INPA)	ocorrência	2015	Instituto Nacional de Pesquisas da Amazônia - INPA	10.15468/602rbt
85	DVPR - Herbário da Universidade Tecnológica Federal do Paraná - Dois Vizinhos	ocorrência	2015	Universidade Tecnológica Federal do Paraná - Campus Dois Vizinhos	10.15468/utbo2g
86	EAC - Herbário Prisco Bezerra	ocorrência	2015	Universidade Federal do Ceará	10.15468/mgeah1
87	EAFM - Herbário do Instituto Federal de Educação, Ciência e Tecnologia do Amazonas	ocorrência	2015	Instituto Federal de Educação, Ciência e Tecnologia do Amazonas	10.15468/vbjcw1
88	EAN - Herbário Jaime Coelho de Moraes	ocorrência	2015	Universidade Federal da Paraíba	10.15468/vaiw6e
89	EFC - Herbário Escola de Florestas Curitiba	ocorrência	2015	Universidade Federal do Paraná	10.15468/ucgdyq
90	ESA - Herbário da Escola Superior de Agricultura Luiz de Queiroz	ocorrência	2015	Escola Superior de Agricultura Luiz de Queiroz	10.15468/6e9ry8
91	FCAB - Herbário Friburguense	ocorrência	2015	Pontifícia Universidade Católica do Rio de Janeiro	10.15468/vd8xjc
92	Fiocruz/CCAMP - Coleção de Campylobacter	ocorrência	2015	FIOCRUZ - Oswaldo Cruz Foundation	10.15468/yh9zld
93	Fish biodiversity of the Vitória-Trindade Seamount Chain, Southwestern Atlantic: an updated database	ocorrência	2015	Brazilian Marine Biodiversity Database	10.15468/o5jdnr

94	FLOR - Herbário do Departamento de Botânica da Universidade Federal de Santa Catarina	ocorrência	2015	Universidade Federal de Santa Catarina	10.15468/yaie53
95	FUEL - Herbário da Universidade Estadual de Londrina	ocorrência	2015	Universidade Estadual de Londrina	10.15468/6b6axv
96	Funed-Pol - Coleção de lâminas de grãos de pólen	ocorrência	2015	Fundação Ezequiel Dias	10.15468/0mw2hm
97	FURB - Herbário Dr. Roberto Miguel Klein	ocorrência	2015	Universidade Regional de Blumenau	10.15468/9sagov
98	HAMAB - Herbário Amapaense	ocorrência	2015	Instituto de Pesquisas Científicas e Tecnológicas do Estado do Amapá	10.15468/ou76ng
99	HAS-Algae-Cyanobacteria - Coleção Ficológica do HAS	ocorrência	2015	Fundação Zoobotânica do Rio Grande do Sul	10.15468/8httul
100	HAS - Herbário Alarich Rudolf Holger Schultz	ocorrência	2015	Fundação Zoobotânica do Rio Grande do Sul	10.15468/3f1k1r
101	HBRA - Herbário do Instituto de Estudos Costeiros da Universidade Federal do Pará	ocorrência	2015	Instituto de Estudos Costeiros (IECOS)	10.15468/3w9k3k
102	HCDAL - Herbário Caririense Dárdano de Andrade-Lima	ocorrência	2015	Universidade Regional do Cariri	10.15468/lstihc
103	HCF - Herbário da Universidade Tecnológica Federal do Paraná Campus Campo Mourão	ocorrência	2015	Universidade Tecnológica Federal do Paraná - Campus Campo Mourão	10.15468/sfdea2
104	HDJF - Herbário Dendrológico Jeanine Felfili	ocorrência	2015	Universidade Federal dos Vales do Jequitinhonha e Mucuri	10.15468/jovi18
105	HEPH - Herbário Ezechias Paulo Heringer	ocorrência	2015	Jardim Botânico de Brasília	10.15468/ouq1mm
106	HERBAM - Herbário da Amazônia Meridional	ocorrência	2015	Universidade do Estado de Mato Grosso	10.15468/dlorqh
107	Herpetology Collection - Instituto Nacional de Pesquisas da Amazônia	ocorrência	2015	Instituto Nacional de Pesquisas da Amazônia - INPA	10.15468/n8h37m
108	HFSL-Fungos - Herbário Dr. Ary Tupinambá Penna Pinheiro - Fungos	ocorrência	2015	Centro de Ensino São Lucas	10.15468/sx994n
109	HFSL - Herbário Dr. Ary Tupinambá Penna Pinheiro	ocorrência	2015	Centro de Ensino São Lucas	10.15468/6j2hwj
110	HJ - Herbario Jataiense	ocorrência	2015	Universidade Federal de Goiás	10.15468/tnfep3
111	HPAN - Herbário do Pantanal "Vali Joana Pott"	ocorrência	2015	Universidade do Estado de Mato Grosso	10.15468/dhtcib

112	HPBR - Herbário Padre Balduino Rambo	ocorrência	2015	Universidade Regional Integrada do Alto Uruguai e das Missões	10.15468/6cvp75
113	HPL - Herbário do Jardim Botânico Plantarum	ocorrência	2015	Jardim Botânico Plantarum	10.15468/ymks0x
114	HRCB - Herbário Rioclarense	ocorrência	2015	Universidade Estadual Paulista - Rio Claro	10.15468/nprbvk
115	HST - Herbário Sérgio Tavares	ocorrência	2015	Universidade Federal Rural de Pernambuco	10.15468/bnu6s9
116	HTSA-Carpoteca - Carpoteca do Trópico Semiárido	ocorrência	2015	Embrapa Semiárido	10.15468/utviiq
117	HTSA - Herbário do Trópico Semiárido	ocorrência	2015	Embrapa Semiárido	10.15468/9tmqsk
118	HTSAw - Xiloteca do Trópico Semiárido	ocorrência	2015	Embrapa Semiárido	10.15468/41hp4t
119	HUCO - Herbário da Universidade Estadual do Centro-Oeste	ocorrência	2015	Universidade Estadual do Centro-Oeste	10.15468/4ulj0e
120	HUCPE - Herbário da Universidade Católica de Pernambuco	ocorrência	2015	Universidade Católica de Pernambuco	10.15468/iagcfe
121	HUCS - Herbário da Universidade de Caxias do Sul	ocorrência	2015	Universidade de Caxias do Sul	10.15468/nfw0hr
122	HUEFS - Herbario da Universidade Estadual de Feira de Santana	ocorrência	2015	Universidade Estadual de Feira de Santana	10.15468/6ymx97
123	HUEG - Herbário da Universidade Estadual de Goiás	ocorrência	2015	Universidade Estadual de Goiás	10.15468/hyyg9s
124	HUEM - Herbário da Universidade Estadual de Maringá	ocorrência	2015	Universidade Estadual de Maringá	10.15468/35v8cn
125	HUESB - Herbário da Universidade Estadual do Sudoeste da Bahia	ocorrência	2015	Universidade Estadual do Sudoeste da Bahia	10.15468/yipzhz
126	HUFSJ - Herbário da Universidade Federal de São João del-Rei	ocorrência	2015	Universidade Federal de São João del-Rei	10.15468/0bw8yf
127	HUFU - Herbarium Uberlandense	ocorrência	2015	Universidade Federal de Uberlândia	10.15468/m2murq
128	HURB - Herbário do Recôncavo da Bahia	ocorrência	2015	Universidade Federal do Recôncavo da Bahia	10.15468/f08gej
129	HUSC - Herbário Unisanta	ocorrência	2015	Universidade Santa Cecília	10.15468/ew5ksp
130	HUTO - Herbário da Universidade de Tocantins	ocorrência	2015	Fundação Universidade do Tocantins	10.15468/zhlmlh
131	HVASF - Herbário Vale do São Francisco	ocorrência	2015	Universidade Federal do Vale do São Francisco	10.15468/ebf7xz

132	HVAT - Herbário do Vale do Taquari	ocorrência	2015	Universidade do Vale do Taquari - UNIVATES	10.15468/cmo532
133	Hymenoptera Collection - Instituto Nacional de Pesquisas da Amazônia (INPA)	ocorrência	2015	Instituto Nacional de Pesquisas da Amazônia - INPA	10.15468/eupzne
134	IAC - Herbário do Instituto Agronômico de Campinas	ocorrência	2015	Instituto Agronômico (IAC)	10.15468/w48pii
135	ICN - Herbário do Instituto de Ciências Naturais	ocorrência	2015	Universidade Federal do Rio Grande do Sul	10.15468/gwuezt
136	IPA - Herbário - IPA Dárdano de Andrade Lima	ocorrência	2015	Instituto Agronômico de Pernambuco	10.15468/hxrell
137	IRAI - Herbário do Parque da Ciência Newton Freire Maia	ocorrência	2015	Parque da Ciência Newton Freire Maia	10.15468/nr4sgo
138	ISE - Herbário ISE	ocorrência	2015	Universidade Federal de Sergipe	10.15468/er457a
139	JOI - Herbário Joinvillea	ocorrência	2015	Universidade da Região de Joinville	10.15468/xykprp
140	JOIw - Xiloteca Joinvillea	ocorrência	2015	Universidade da Região de Joinville	10.15468/dgbx1n
141	JPB - Herbário Lauro Pires Xavier	ocorrência	2015	Universidade Federal da Paraíba	10.15468/uxvuhx
142	Levantamento da vegetação arbustiva do Parque Nacional do Jaú	ocorrência	2015	Instituto Nacional de Pesquisas da Amazônia - INPA	10.15468/xhqz9l
143	Levantamento de Arbóreas do Parque Nacional do Jaú	ocorrência	2015	Instituto Nacional de Pesquisas da Amazônia - INPA	10.15468/r7rs14
144	Levantamento de Herbáceas da RDS Uatumã	ocorrência	2015	Instituto Nacional de Pesquisas da Amazônia - INPA	10.15468/cwjhhg
145	Levantamento de Herbáceas do PARNA Jaú	ocorrência	2015	Instituto Nacional de Pesquisas da Amazônia - INPA	10.15468/j1zoip
146	Levantamento do estrato arbustivo da vegetação de campina do RDS Uatumã	ocorrência	2015	Instituto Nacional de Pesquisas da Amazônia - INPA	10.15468/cgk3tq
147	LUSC - Herbário de Lages da Universidade do Estado de Santa Catarina	ocorrência	2015	Universidade do Estado de Santa Catarina	10.15468/xtjdw
148	MAC - Herbário do Instituto do Meio Ambiente do Estado de Alagoas	ocorrência	2015	Instituto do Meio Ambiente do Estado de Alagoas	10.15468/kpsugm
149	MAR - Herbário do Maranhão	ocorrência	2015	Universidade Federal do Maranhão	10.15468/9ql44s

150	MBM - Herbário do Museu Botânico Municipal	ocorrência	2015	Museu Botânico Municipal	10.15468/g6ppmt
151	MBML-Herbario - Herbário Mello Leitão	ocorrência	2015	Instituto Nacional da Mata Atlântica	10.15468/z8djaf
152	MFS - Herbário Profª. Drª. Marlene Freitas da Silva	ocorrência	2015	Universidade do Estado do Pará	10.15468/bzidql
153	MOSS - Herbário Dárdano de Andrade Lima	ocorrência	2015	Universidade Federal Rural do Semi-Árido	10.15468/ha82ay
154	MPUC - Herbário do Museu da Pontifícia Universidade Católica do Rio Grande do Sul	ocorrência	2015	Pontifícia Universidade Católica do Rio Grande do Sul	10.15468/yjiad0
155	Nematoda Collection - Instituto Nacional de Pesquisas da Amazônia (INPA)	ocorrência	2015	Instituto Nacional de Pesquisas da Amazônia - INPA	10.15468/ylaxpt
156	NX-Criptogamos - Herbário Nova Xavantina	ocorrência	2015	Universidade do Estado de Mato Grosso	10.15468/gqhp0q
157	ocorrências in SinBIOTA	ocorrência	2015	Programa BIOTA/FAPESP - The Virtual Institute of Biodiversity	10.15468/j4zyyt
158	Opiliones Collection - Instituto Nacional de Pesquisas da Amazônia (INPA)	ocorrência	2015	Instituto Nacional de Pesquisas da Amazônia - INPA	10.15468/3cwlwa
159	OUPR - Herbário "Professor José Badini"	ocorrência	2015	Universidade Federal de Ouro Preto	10.15468/gg0o1b
160	PACA-AGP - Herbarium Anchieta	ocorrência	2015	Instituto Anchietano de Pesquisas/UNISINOS	10.15468/qchxtg
161	PACA-Fungi - Herbarium Anchieta - Fungi Rickiani	ocorrência	2015	Instituto Anchietano de Pesquisas/UNISINOS	10.15468/prbbyt
162	Palpigradi Collection - Instituto Nacional de Pesquisas da Amazônia (INPA)	ocorrência	2015	Instituto Nacional de Pesquisas da Amazônia - INPA	10.15468/ztijuq
163	Pauropoda Collection - Instituto Nacional de Pesquisas da Amazônia (INPA)	ocorrência	2015	Instituto Nacional de Pesquisas da Amazônia - INPA	10.15468/b3v18s
164	PEUFR - Herbário Professor Vasconcelos Sobrinho	ocorrência	2015	Universidade Federal Rural de Pernambuco	10.15468/o6lt9x
165	Pseudoscorpiones Collection - Instituto Nacional de Pesquisas da Amazônia (INPA)	ocorrência	2015	Instituto Nacional de Pesquisas da Amazônia - INPA	10.15468/xayxra
166	R-Tipos - Herbário do Museu Nacional - Tipos	ocorrência	2015	Museu Nacional / UFRJ	10.15468/mcbocy

167	R - Herbário do Museu Nacional	ocorrência	2015	Museu Nacional / UFRJ	10.15468/3qpd4g
168	RBspirit - Rio de Janeiro Botanical Garden fluid-preserved specimens.	ocorrência	2015	Instituto de Pesquisas Jardim Botânico do Rio de Janeiro	10.15468/wtozyg
169	Ricinulei Collection - Instituto Nacional de Pesquisas da Amazônia (INPA)	ocorrência	2015	Instituto Nacional de Pesquisas da Amazônia - INPA	10.15468/lzgexd
170	Schizomida Collection - Instituto Nacional de Pesquisas da Amazônia (INPA)	ocorrência	2015	Instituto Nacional de Pesquisas da Amazônia - INPA	10.15468/2znhye
171	Scorpiones Collection - Instituto Nacional de Pesquisas da Amazônia (INPA)	ocorrência	2015	Instituto Nacional de Pesquisas da Amazônia - INPA	10.15468/srul0s
172	SJRP-Algae - Herbário de algas de São José do Rio Preto	ocorrência	2015	Universidade Estadual Paulista - São José do Rio Preto	10.15468/xhikte
173	SJRP-Bryophyta - Herbário de Bryophyta de São José do Rio Preto	ocorrência	2015	Universidade Estadual Paulista - São José do Rio Preto	10.15468/p11o9y
174	SJRP-Pteridophyta - Herbário de Pteridophyta de São José do Rio Preto	ocorrência	2015	Universidade Estadual Paulista - São José do Rio Preto	10.15468/jzkk1b
175	SJRP - Herbário de São José do Rio Preto	ocorrência	2015	Universidade Estadual Paulista - São José do Rio Preto	10.15468/rdtyfv
176	SLUI - Herbário Rosa Mochel	ocorrência	2015	Universidade Estadual do Maranhão	10.15468/3sbsos
177	SMDB - Santa Maria Departamento de Biologia	ocorrência	2015	Universidade Federal de Santa Maria	10.15468/rkuop1
178	Solifugae Collection - Instituto Nacional de Pesquisas da Amazônia (INPA)	ocorrência	2015	Instituto Nacional de Pesquisas da Amazônia - INPA	10.15468/ogbfpj
179	SORO - Herbário do Centro de Ciências e Tecnologias para a Sustentabilidade	ocorrência	2015	Universidade Federal de São Carlos - Campus Sorocaba	10.15468/8catgj
180	SP-Algae - Herbário do Estado "Maria Eneyda P. Kaufmann Fidalgo - Coleção de Algas	ocorrência	2015	Instituto de Botânica, São Paulo	10.15468/md6zcx
181	SP-Bryophyta - Maria Eneyda P. Kauffman Fidalgo	ocorrência	2015	Instituto de Botânica, São Paulo	10.15468/vub9hl
182	SP-Fungi - Maria Eneyda Pacheco Kauffmann Fidalgo	ocorrência	2015	Instituto de Botânica, São Paulo	10.15468/ztl1a
183	SP - Herbário do Estado "Maria Eneyda P. Kaufmann Fidalgo" - Coleção de Fanerógamas	ocorrência	2015	Instituto de Botânica, São Paulo	10.15468/axlcjr

184	SPF-Algae - Herbário da Universidade de São Paulo - Coleção de Algas	ocorrência	2015	Universidade de São Paulo	10.15468/xhpmge
185	SPF - Herbário da Universidade de São Paulo	ocorrência	2015	Universidade de São Paulo	10.15468/hnlkxx
186	SPFw - Xiloteca do Instituto de Biociências da Universidade de São Paulo	ocorrência	2015	Universidade de São Paulo	10.15468/sihvr7
187	SPSF - Herbário Dom Bento José Pickel	ocorrência	2015	Instituto Florestal	10.15468/rw7kke
188	Symphyla Collection - Instituto Nacional de Pesquisas da Amazônia (INPA)	ocorrência	2015	Instituto Nacional de Pesquisas da Amazônia - INPA	10.15468/qjqpq8
189	TEPB - Herbário Graziela Barroso	ocorrência	2015	Universidade Federal do Piauí	10.15468/gjf70b
190	Thelyphonida Collection - Instituto Nacional de Pesquisas da Amazônia (INPA)	ocorrência	2015	Instituto Nacional de Pesquisas da Amazônia - INPA	10.15468/zobifu
191	Tree Diversity and Dynamics of the Forest of Seu Nico (Brazil/MG)	ocorrência	2015	Laboratory of Ecology and Evolution of Plants, at Universidade Federal de Vicosa	10.15468/4ckfsb
192	UB - Herbário da Universidade de Brasília	ocorrência	2015	Universidade de Brasília	10.15468/caq5no
193	UEC - Herbário da Universidade Estadual de Campinas	ocorrência	2015	Universidade Estadual de Campinas - Instituto de Biologia	10.15468/4hnsz8
194	UESC - Herbário Universidade Estadual de Santa Cruz	ocorrência	2015	Universidade Estadual de Santa Cruz	10.15468/vrwzob
195	UFG - Herbário da Universidade Federal de Goiás	ocorrência	2015	Universidade Federal de Goiás	10.15468/jn3v2g
196	UFMT - Herbário UFMT	ocorrência	2015	Universidade Federal do Mato Grosso	10.15468/glgkti
197	UFP - Herbário UFP - Geraldo Mariz	ocorrência	2015	Universidade Federal de Pernambuco	10.15468/k0gqg4
198	UFRN-Fungos - Herbário UFRN - Fungos	ocorrência	2015	Universidade Federal do Rio Grande do Norte	10.15468/kdr1fr
199	UPCB - Herbário do Departamento de Botânica	ocorrência	2015	Universidade Federal do Paraná	10.15468/udpvi1
200	URM - Herbário Pe. Camille Torrand	ocorrência	2015	Universidade Federal de Pernambuco	10.15468/gdvlxf
201	VIC - Herbário da Universidade Federal de Viçosa	ocorrência	2015	Universidade Federal de Viçosa	10.15468/8ujflh
202	VIES - Herbário Central da Universidade Federal do Espírito Santo	ocorrência	2015	Universidade Federal do Espírito Santo	10.15468/rt7ybs

203	An Online Database of the Immatures of Coleoptera (Arthropoda, Insecta) Described from Brazil	ocorrência	2016	Sistema de Informação sobre a Biodiversidade Brasileira - SiBBr	10.15468/sjdx5a
204	Blattaria Collection - Instituto Nacional de Pesquisas da Amazônia (INPA)	ocorrência	2016	Instituto Nacional de Pesquisas da Amazônia - INPA	10.15468/covrcl
205	Coleção de Ascidiacea do Museu Nacional - UFRJ	ocorrência	2016	Museu Nacional / UFRJ	10.15468/bltpyq
206	Coleção de Aves do Museu Nacional / UFRJ	ocorrência	2016	Museu Nacional / UFRJ	10.15468/ga2jww
207	Coleção de Brachiopoda do Museu Nacional - UFRJ	ocorrência	2016	Museu Nacional / UFRJ	10.15468/h9lcqq
208	Coleção de Bryozoa do Museu Nacional - UFRJ	ocorrência	2016	Museu Nacional / UFRJ	10.15468/ldxinl
209	Coleção de Echinodermata do Museu Nacional - MNRJ-ECHINO	ocorrência	2016	Museu Nacional / UFRJ	10.15468/wlblsg
210	Coleção de Herpetologia do Museu de Ciências Naturais da Universidade do Vale do Taquari - Univates (ZHUMCN-UNIVATES)	ocorrência	2016	Universidade do Vale do Taquari - UNIVATES	10.15468/za1pvg
211	Coleção de Nematoda do Museu Nacional - UFRJ	ocorrência	2016	Museu Nacional / UFRJ	10.15468/q6eu7k
212	Coleção de Nematomorpha do Museu Nacional - UFRJ	ocorrência	2016	Museu Nacional / UFRJ	10.15468/7dimki
213	Coleção de Nemertea do Museu Nacional - UFRJ	ocorrência	2016	Museu Nacional / UFRJ	10.15468/c2h7qa
214	Coleção de Platyhelminthes do Museu Nacional - UFRJ	ocorrência	2016	Museu Nacional / UFRJ	10.15468/snq1oq
215	Coleção de Porifera do Museu Nacional - UFRJ	ocorrência	2016	Museu Nacional / UFRJ	10.15468/ehye3n
216	Coleção de Sipuncula do Museu Nacional - UFRJ	ocorrência	2016	Museu Nacional / UFRJ	10.15468/enfwdx
217	Coleção de Tardigrada do Museu Nacional - UFRJ	ocorrência	2016	Museu Nacional / UFRJ	10.15468/0jp8lp
218	Coleção Ictiológica (MNRJ), Museu Nacional (MN), Universidade Federal do Rio de Janeiro(UFRJ)	ocorrência	2016	Museu Nacional / UFRJ	10.15468/lluzfl
219	corrego_fazendinha-gallery-forest_v2	ocorrência	2016	Laboratory of Ecology and Evolution of Plants, at Universidade Federal de Vicosa	10.15468/judbjja
220	Crustacea Collection - Instituto Nacional de Pesquisas da Amazônia (INPA)	ocorrência	2016	Instituto Nacional de Pesquisas da Amazônia - INPA	10.15468/dihynq
221	Ephemeroptera Collection - Instituto Nacional de Pesquisas da Amazônia (INPA)	ocorrência	2016	Instituto Nacional de Pesquisas da Amazônia - INPA	10.15468/xzfi0y

222	Fiocruz/CBAM - Coleção de Bactérias da Amazônia	ocorrência	2016	FIOCRUZ - Oswaldo Cruz Foundation	10.15468/r5sxxwz
223	Fiocruz/CCBH - Coleção de Culturas de Bactérias de Origem Hospitalar	ocorrência	2016	FIOCRUZ - Oswaldo Cruz Foundation	10.15468/jb9tx8
224	Fiocruz/CCFF - Coleção de Culturas de Fungos Filamentosos	ocorrência	2016	FIOCRUZ - Oswaldo Cruz Foundation	10.15468/esrihr
225	Fiocruz/CCGB - Coleção de Culturas do Gênero Bacillus e Gêneros Correlatos	ocorrência	2016	FIOCRUZ - Oswaldo Cruz Foundation	10.15468/wlytrp
226	Fiocruz/CFAM - Coleção de Fungos da Amazônia	ocorrência	2016	FIOCRUZ - Oswaldo Cruz Foundation	10.15468/xxxs2q
227	Fiocruz/CFP - Coleção de Fungos Patogênicos	ocorrência	2016	FIOCRUZ - Oswaldo Cruz Foundation	10.15468/1vptxf
228	Fiocruz/CLEP - Coleção de Leptospira	ocorrência	2016	FIOCRUZ - Oswaldo Cruz Foundation	10.15468/in79vz
229	Fiocruz/CLIOC - Coleção de Leishmania	ocorrência	2016	FIOCRUZ - Oswaldo Cruz Foundation	10.15468/hdmflp
230	Fiocruz/CLIST - Coleção de Listeria	ocorrência	2016	FIOCRUZ - Oswaldo Cruz Foundation	10.15468/uwsiii
231	Fiocruz/COLPROT - Coleção de Protozoários	ocorrência	2016	FIOCRUZ - Oswaldo Cruz Foundation	10.15468/j8grwf
232	Fiocruz/COLTRYP - Coleção de Trypanosoma de Mamíferos Silvestres, Domésticos e Vetores	ocorrência	2016	FIOCRUZ - Oswaldo Cruz Foundation	10.15468/thfjkg
233	Fiocruz/CYP - Coleção de Yersinia pestis	ocorrência	2016	FIOCRUZ - Oswaldo Cruz Foundation	10.15468/gdqktb
234	FNJV - Fonoteca Neotropical Jacques Viellard	ocorrência	2016	Universidade Estadual de Campinas - Instituto de Biologia	10.15468/fr6q2k
235	fundo	ocorrência	2016	Laboratory of Ecology and Evolution of Plants, at Universidade Federal de Vicosa	10.15468/wz2l5r
236	Hemiptera Collection - Instituto Nacional de Pesquisas da Amazônia (INPA)	ocorrência	2016	Instituto Nacional de Pesquisas da Amazônia - INPA	10.15468/3flouq
237	Ictiologia Collection - Instituto Nacional de Pesquisas da Amazônia (INPA)	ocorrência	2016	Instituto Nacional de Pesquisas da Amazônia - INPA	10.15468/edjk09
238	MBML-Anfíbios - Coleção de Anfíbios MBML	ocorrência	2016	Instituto Nacional da Mata Atlântica	10.15468/og6via
239	MBML-Aves - Coleção de Aves MBML	ocorrência	2016	Instituto Nacional da Mata Atlântica	10.15468/soqusy
240	MBML-Mamíferos - Coleção de Mamíferos	ocorrência	2016	Instituto Nacional da Mata Atlântica	10.15468/nmpnlr
241	MBML-Peixes - Coleção de Peixes	ocorrência	2016	Instituto Nacional da Mata Atlântica	10.15468/l4q0fn
242	MBML-Repteis - Coleção de Répteis	ocorrência	2016	Instituto Nacional da Mata Atlântica	10.15468/bt80dm

243	Museu Paraense Emílio Goeldi - Amphibia Collection	ocorrência	2016	Museu Paraense Emílio Goeldi	10.15468/Orerdu
244	Museu Paraense Emílio Goeldi - Chelonia Collection	ocorrência	2016	Museu Paraense Emílio Goeldi	10.15468/batjpc
245	Museu Paraense Emílio Goeldi - Crocodylia Collection	ocorrência	2016	Museu Paraense Emílio Goeldi	10.15468/f6emsr
246	Museu Paraense Emílio Goeldi - Ichthyology Collection	ocorrência	2016	Museu Paraense Emílio Goeldi	10.15468/njmykk
247	Museu Paraense Emílio Goeldi - Lacertilia Collection	ocorrência	2016	Museu Paraense Emílio Goeldi	10.15468/u1szyx
248	Museu Paraense Emílio Goeldi - Mastozoology Collection	ocorrência	2016	Museu Paraense Emílio Goeldi	10.15468/18polp
249	Museu Paraense Emilio Goeldi Herbarium	ocorrência	2016	Museu Paraense Emílio Goeldi	10.15468/igjr8k
250	Onychophora Collection - Instituto Nacional de Pesquisas da Amazônia (INPA)	ocorrência	2016	Instituto Nacional de Pesquisas da Amazônia - INPA	10.15468/3zckv9
251	Orthoptera Collection - Instituto Nacional de Pesquisas da Amazônia (INPA)	ocorrência	2016	Instituto Nacional de Pesquisas da Amazônia - INPA	10.15468/wyi68n
252	PACA-Bryophytes - Herbarium Anchieta - Aloysio Sehnem	ocorrência	2016	Instituto Anchietano de Pesquisas/UNISINOS	10.15468/2a93vs
253	Phasmatodea Collection - Instituto Nacional de Pesquisas da Amazônia (INPA)	ocorrência	2016	Instituto Nacional de Pesquisas da Amazônia - INPA	10.15468/ps135o
254	R-Bryophyta - Herbário do Museu Nacional - Bryophyta	ocorrência	2016	Museu Nacional / UFRJ	10.15468/d8h472
255	UFACPZ - Herbário da Universidade Federal do Acre	ocorrência	2016	Universidade Federal do Acre	10.15468/z51ipd
256	UNIVATES - Coleção de Ornitologia do Museu de Ciências Naturais da Universidade do Vale do Taquari - Univates (ZOUHCN-UNIVATES)	ocorrência	2016	Universidade do Vale do Taquari - UNIVATES	10.15468/ntcyur
257	ZUEC-AMP - Coleção de Anfíbios do Museu de Zoologia da UNICAMP	ocorrência	2016	Universidade Estadual de Campinas - Instituto de Biologia	10.15468/1hsrdr
258	ZUEC-API - Coleção de Apicomplexa do Museu de Zoologia da UNICAMP	ocorrência	2016	Universidade Estadual de Campinas - Instituto de Biologia	10.15468/xqaim7
259	ZUEC-APL - Coleção de Aplacophora do Museu de Zoologia da UNICAMP	ocorrência	2016	Universidade Estadual de Campinas - Instituto de Biologia	10.15468/hnoewv
260	ZUEC-ARA - Coleção de Aracnídeos do Museu de Zoologia da UNICAMP	ocorrência	2016	Universidade Estadual de Campinas - Instituto de Biologia	10.15468/xihyn4

261	ZUEC-ASC - Coleção de Ascidiacea do Museu de Zoologia da UNICAMP	ocorrência	2016	Universidade Estadual de Campinas - Instituto de Biologia	10.15468/35rudc
262	ZUEC-AST - Coleção de Asteroidea do Museu de Zoologia da UNICAMP	ocorrência	2016	Universidade Estadual de Campinas - Instituto de Biologia	10.15468/rh6dar
263	ZUEC-AVE - Coleção de Aves do Museu de Zoologia da UNICAMP	ocorrência	2016	Universidade Estadual de Campinas - Instituto de Biologia	10.15468/skokpl
264	ZUEC-BIV - Coleção de Bivalvia do Museu de Zoologia da UNICAMP	ocorrência	2016	Universidade Estadual de Campinas - Instituto de Biologia	10.15468/alez7z
265	ZUEC-BRA - Coleção de Brachiopoda do Museu de Zoologia da UNICAMP	ocorrência	2016	Universidade Estadual de Campinas - Instituto de Biologia	10.15468/alcqoe
266	ZUEC-BRY - Coleção de Bryozoa do Museu de Zoologia da UNICAMP	ocorrência	2016	Universidade Estadual de Campinas - Instituto de Biologia	10.15468/tiqruw
267	ZUEC-CEP - Coleção de Cephalopoda do Museu de Zoologia da UNICAMP	ocorrência	2016	Universidade Estadual de Campinas - Instituto de Biologia	10.15468/umanea
268	ZUEC-CLB - Coleção de Collembola do Museu de Zoologia da UNICAMP	ocorrência	2016	Universidade Estadual de Campinas - Instituto de Biologia	10.15468/trrgzp
269	ZUEC-CNI - Coleção de Cnidaria do Museu de Zoologia da UNICAMP	ocorrência	2016	Universidade Estadual de Campinas - Instituto de Biologia	10.15468/q9jo5y
270	ZUEC-COL - Coleção de Coleoptera do Museu de Zoologia da UNICAMP	ocorrência	2016	Universidade Estadual de Campinas - Instituto de Biologia	10.15468/nidzk8
271	ZUEC-CPH - Coleção de Cephalochordata do Museu de Zoologia da UNICAMP	ocorrência	2016	Universidade Estadual de Campinas - Instituto de Biologia	10.15468/3ij6x6
272	ZUEC-CRU - Coleção de Crustacea do Museu de Zoologia da UNICAMP	ocorrência	2016	Universidade Estadual de Campinas - Instituto de Biologia	10.15468/ovnu86
273	ZUEC-DIP - Coleção de Diptera do Museu de Zoologia da UNICAMP	ocorrência	2016	Universidade Estadual de Campinas - Instituto de Biologia	10.15468/yh6k6h
274	ZUEC-ECH - Coleção de Echinoidea do Museu de Zoologia da UNICAMP	ocorrência	2016	Universidade Estadual de Campinas - Instituto de Biologia	10.15468/7cl4yv

275	ZUEC-EPH - Coleção de Ephemeroptera do Museu de Zoologia da UNICAMP	ocorrência	2016	Universidade Estadual de Campinas - Instituto de Biologia	10.15468/i5sygx
276	ZUEC-GAS - Coleção de Gastropoda do Museu de Zoologia da UNICAMP	ocorrência	2016	Universidade Estadual de Campinas - Instituto de Biologia	10.15468/73mksc
277	ZUEC-GCH - Coleção de Gastrotricha do Museu de Zoologia da UNICAMP	ocorrência	2016	Universidade Estadual de Campinas - Instituto de Biologia	10.15468/c0wfhc
278	ZUEC-HEM - Coleção de Hemiptera do Museu de Zoologia da UNICAMP	ocorrência	2016	Universidade Estadual de Campinas - Instituto de Biologia	10.15468/r6xpl1
279	ZUEC-HOL - Coleção de Holothuroidea do Museu de Zoologia da UNICAMP	ocorrência	2016	Universidade Estadual de Campinas - Instituto de Biologia	10.15468/syenw6
280	ZUEC-HYM - Coleção de Hymenoptera do Museu de Zoologia da UNICAMP	ocorrência	2016	Universidade Estadual de Campinas - Instituto de Biologia	10.15468/fnppos
281	ZUEC-LEP - Coleção de Lepidoptera do Museu de Zoologia da UNICAMP	ocorrência	2016	Universidade Estadual de Campinas - Instituto de Biologia	10.15468/aclolh
282	ZUEC-MAM - Coleção de Mamíferos do Museu de Zoologia da UNICAMP	ocorrência	2016	Universidade Estadual de Campinas - Instituto de Biologia	10.15468/yqhyqe
283	ZUEC-MYX - Coleção de Myxozoa do Museu de Zoologia da UNICAMP	ocorrência	2016	Universidade Estadual de Campinas - Instituto de Biologia	10.15468/nioqsd
284	ZUEC-NEM - Coleção de Nemertea do Museu de Zoologia da UNICAMP	ocorrência	2016	Universidade Estadual de Campinas - Instituto de Biologia	10.15468/daxuws
285	ZUEC-NEU - Coleção de Neuroptera do Museu de Zoologia da UNICAMP	ocorrência	2016	Universidade Estadual de Campinas - Instituto de Biologia	10.15468/utljpp
286	ZUEC-NMA - Coleção de Nematoda do Museu de Zoologia da UNICAMP	ocorrência	2016	Universidade Estadual de Campinas - Instituto de Biologia	10.15468/umreyv
287	ZUEC-OPH - Coleção de Ophiuroidea do Museu de Zoologia da UNICAMP	ocorrência	2016	Universidade Estadual de Campinas - Instituto de Biologia	10.15468/sklinw2
288	ZUEC-ORT - Coleção de Orthoptera do Museu de Zoologia da UNICAMP	ocorrência	2016	Universidade Estadual de Campinas - Instituto de Biologia	10.15468/qsyo43

289	ZUEC-PIS - Coleção de Peixes do Museu de Zoologia da UNICAMP	ocorrência	2016	Universidade Estadual de Campinas - Instituto de Biologia	10.15468/jahatf
290	ZUEC-PLA - Coleção de Platyhelminthes do Museu de Zoologia da UNICAMP	ocorrência	2016	Universidade Estadual de Campinas - Instituto de Biologia	10.15468/eaitwv
291	ZUEC-POL - Coleção de Polychaeta do Museu de Zoologia da UNICAMP	ocorrência	2016	Universidade Estadual de Campinas - Instituto de Biologia	10.15468/6legyj
292	ZUEC-POP - Coleção de Polyplacophora do Museu de Zoologia da UNICAMP	ocorrência	2016	Universidade Estadual de Campinas - Instituto de Biologia	10.15468/a9vazy
293	ZUEC-POR - Coleção de Porifera do Museu de Zoologia da UNICAMP	ocorrência	2016	Universidade Estadual de Campinas - Instituto de Biologia	10.15468/ytfg0z
294	ZUEC-REP - Coleção de Répteis do Museu de Zoologia da UNICAMP	ocorrência	2016	Universidade Estadual de Campinas - Instituto de Biologia	10.15468/gjzorz
295	Coleção Entomológica do Museu Nacional / UFRJ	ocorrência	2017	Museu Nacional / UFRJ	10.15468/7lklen
296	Embrapa Pecuária Sul - Herbário CNPO	ocorrência	2017	Sistema de Informação sobre a Biodiversidade Brasileira - SiBBr	10.15468/mjkasx
297	Fiocruz/CAVAISC - Coleção de Artrópodes Vetores Ápteros de Importância em Saúde das Comunidades	ocorrência	2017	FIOCRUZ - Oswaldo Cruz Foundation	10.15468/vrioee
298	Fiocruz/CBAS - Coleção de Bactérias do Ambiente e Saúde	ocorrência	2017	FIOCRUZ - Oswaldo Cruz Foundation	10.15468/itf2dj
299	Fiocruz/CCER - Coleção de Ceratopogonidae	ocorrência	2017	FIOCRUZ - Oswaldo Cruz Foundation	10.15468/pn1zdi
300	Fiocruz/CCULI - Coleção de Culicidae	ocorrência	2017	FIOCRUZ - Oswaldo Cruz Foundation	10.15468/doevid
301	Fiocruz/CEIOC - Coleção Entomológica do Instituto Oswaldo Cruz	ocorrência	2017	FIOCRUZ - Oswaldo Cruz Foundation	10.15468/bdv6wn
302	Fiocruz/CMN - Coleção de Mosquitos Neotropicais	ocorrência	2017	FIOCRUZ - Oswaldo Cruz Foundation	10.15468/yxluxr
303	Fiocruz/COLFLEB - Coleção de Flebotomíneos	ocorrência	2017	FIOCRUZ - Oswaldo Cruz Foundation	10.15468/sxcpfp
304	Fiocruz/COLVEC - Coleção de Vetores da Doença de Chagas	ocorrência	2017	FIOCRUZ - Oswaldo Cruz Foundation	10.15468/s2fh6t
305	Fiocruz/CSIOC - Coleção de Simulídeos do Instituto Oswaldo Cruz	ocorrência	2017	FIOCRUZ - Oswaldo Cruz Foundation	10.15468/0cx9nu
306	Fiocruz/CTIOC - Coleção de Triatomíneos do Instituto Oswaldo Cruz	ocorrência	2017	FIOCRUZ - Oswaldo Cruz Foundation	10.15468/5auteh

307	Herbário da Pontifícia Universidade Católica do Paraná - HUCP	ocorrência	2017	Sistema de Informação sobre a Biodiversidade Brasileira - SiBBr	10.15468/o8q0jz
308	Inventário florístico de 20 parcelas permanentes estabelecidas em áreas de savana do norte da Amazônia brasileira	ocorrência	2017	Sistema de Informação sobre a Biodiversidade Brasileira - SiBBr	10.15468/wn5vzn
309	Mollusca Collection - Museu Nacional/UFRJ	ocorrência	2017	Museu Nacional / UFRJ	10.15468/g5ykee
310	Museu Paraense Emílio Goeldi - Acanthocephala Collection	ocorrência	2017	Museu Paraense Emílio Goeldi	10.15468/nd5zfr
311	Museu Paraense Emilio Goeldi - Annelida Collection	ocorrência	2017	Museu Paraense Emílio Goeldi	10.15468/eiymy1
312	Museu Paraense Emílio Goeldi - Carcinológica Collection	ocorrência	2017	Museu Paraense Emílio Goeldi	10.15468/tmqcgv
313	Museu Paraense Emílio Goeldi - Diptera Collection	ocorrência	2017	Museu Paraense Emílio Goeldi	10.15468/i3xuvn
314	Museu Paraense Emílio Goeldi - Hexapoda Collection	ocorrência	2017	Museu Paraense Emílio Goeldi	10.15468/5bnkz8
315	Museu Paraense Emilio Goeldi - Hymenoptera Collection	ocorrência	2017	Museu Paraense Emílio Goeldi	10.15468/efkpxu
316	Museu Paraense Emílio Goeldi - Mollusca Collection	ocorrência	2017	Museu Paraense Emílio Goeldi	10.15468/wsdlgq
317	Museu Paraense Emílio Goeldi - Myriapoda Collection	ocorrência	2017	Museu Paraense Emílio Goeldi	10.15468/tnt5nq
318	Museu Paraense Emílio Goeldi - Nematoda Collection	ocorrência	2017	Museu Paraense Emílio Goeldi	10.15468/tcxwrb
319	Museu Paraense Emílio Goeldi - Paleoinvertebrate Collection	ocorrência	2017	Museu Paraense Emílio Goeldi	10.15468/t1rwon
320	Museu Paraense Emílio Goeldi - Paleovertebrate Collection	ocorrência	2017	Museu Paraense Emílio Goeldi	10.15468/bgiwpp
321	Museu Paraense Emílio Goeldi - Platyhelminthes Collection	ocorrência	2017	Museu Paraense Emílio Goeldi	10.15468/banb6o
322	Tree species composition in different habitats of savanna used by indigenous in the Northern Brazilian Amazonia	ocorrência	2017	Sistema de Informação sobre a Biodiversidade Brasileira - SiBBr	10.15468/oby3vu
323	UEL - Coleção de Peixes - Museu de Zoologia da Universidade Estadual de Londrina	ocorrência	2017	Universidade Estadual de Londrina (UEL)	10.15468/cd6x6z
324	UEL - Coleção Ficológica da Universidade Estadual de Londrina (FUEL-Algae)	ocorrência	2017	Universidade Estadual de Londrina (UEL)	10.15468/iwriib
325	Coleção de Cultura de Bactérias Resistentes de Origem Alimentar - CCBROA	ocorrência	2018	Sistema de Informação sobre a Biodiversidade Brasileira - SiBBr	10.15468/cin0ca
326	Coleção Zoológica de Referência da Universidade Federal de Mato Grosso do Sul - Amphibia (ZUFMS-AMP)	ocorrência	2018	Sistema de Informação sobre a Biodiversidade Brasileira - SiBBr	10.15468/wahc6e

327	Coleção Zoológica de Referência da Universidade Federal de Mato Grosso do Sul - Chelicerata (ZUFMS-CHE)	ocorrência	2018	Sistema de Informação sobre a Biodiversidade Brasileira - SiBBr	10.15468/dujb8i
328	Coleção Zoológica de Referência da Universidade Federal de Mato Grosso do Sul - Chiroptera (ZUFMS-CHI)	ocorrência	2018	Sistema de Informação sobre a Biodiversidade Brasileira - SiBBr	10.15468/tbulch
329	Coleção Zoológica de Referência da Universidade Federal de Mato Grosso do Sul - Crustacea (ZUFMS-CRU)	ocorrência	2018	Sistema de Informação sobre a Biodiversidade Brasileira - SiBBr	10.15468/yflrpl
330	Coleção Zoológica de Referência da Universidade Federal de Mato Grosso do Sul - Diptera (ZUFMS-DIP)	ocorrência	2018	Sistema de Informação sobre a Biodiversidade Brasileira - SiBBr	10.15468/9qccn3
331	Coleção Zoológica de Referência da Universidade Federal de Mato Grosso do Sul - Fósseis (ZUFMS-FOS)	ocorrência	2018	Sistema de Informação sobre a Biodiversidade Brasileira - SiBBr	10.15468/orq3uh
332	Coleção Zoológica de Referência da Universidade Federal de Mato Grosso do Sul - Hemiptera (ZUFMS-HEM)	ocorrência	2018	Sistema de Informação sobre a Biodiversidade Brasileira - SiBBr	10.15468/dy9fx0
333	Coleção Zoológica de Referência da Universidade Federal de Mato Grosso do Sul - Hymenoptera (ZUFMS-HYM)	ocorrência	2018	Sistema de Informação sobre a Biodiversidade Brasileira - SiBBr	10.15468/y3h9ab
334	Coleção Zoológica de Referência da Universidade Federal de Mato Grosso do Sul - Mollusca (ZUFMS-MOL)	ocorrência	2018	Sistema de Informação sobre a Biodiversidade Brasileira - SiBBr	10.15468/u5r2vz
335	Coleção Zoológica de Referência da Universidade Federal de Mato Grosso do Sul - Nematoda (ZUFMS-NEM)	ocorrência	2018	Sistema de Informação sobre a Biodiversidade Brasileira - SiBBr	10.15468/1i1lrf
336	Coleção Zoológica de Referência da Universidade Federal de Mato Grosso do Sul - Orthoptera (ZUFMS-ORT)	ocorrência	2018	Sistema de Informação sobre a Biodiversidade Brasileira - SiBBr	10.15468/tjwvtz
337	Coleção Zoológica de Referência da Universidade Federal de Mato Grosso do Sul - Pisces (ZUFMS-PIS)	ocorrência	2018	Sistema de Informação sobre a Biodiversidade Brasileira - SiBBr	10.15468/oupodb
338	Coleção Zoológica de Referência da Universidade Federal de Mato Grosso do Sul - Reptilia (ZUFMS-REP)	ocorrência	2018	Sistema de Informação sobre a Biodiversidade Brasileira - SiBBr	10.15468/nkpsx
339	Coleção Zoológica de Referência da Universidade Federal de Mato Grosso do Sul - Tecidos (ZUFMS-TEC)	ocorrência	2018	Sistema de Informação sobre a Biodiversidade Brasileira - SiBBr	10.15468/yrnc1y
340	DGGE analysis of the structure of fungal communities associated to leaf litter and the gut of larvae of Triplectides (Trichoptera: Leptoceridae) in Atlantic Forest streams	ocorrência	2018	Sistema de Informação sobre a Biodiversidade Brasileira - SiBBr	10.15468/ugwdlo

341	DSEC - Coleção Entomológica do Depto. de Sistemática e Ecologia	ocorrência	2018	Universidade Federal da Paraíba	10.15468/zbzqz8
342	DZUP-Hymenoptera - Coleção Entomológica Pe. Jesus Santiago Moure (Hymenoptera)	ocorrência	2018	Universidade Federal do Paraná	10.15468/cwlwvm
343	fauna_do_municipio_de_araraquara-sp	ocorrência	2018	Sistema de Informação sobre a Biodiversidade Brasileira - SiBBR	10.15468/k6ezkv
344	Material-tipo das espécies brasileiras de Histeridae (Insecta: Coleoptera) depositado no Muséum National d'Histoire Naturelle, Paris, França	ocorrência	2018	Sistema de Informação sobre a Biodiversidade Brasileira - SiBBR	10.15468/yz2lfv
345	MCP - Coleção de Abelhas	ocorrência	2018	Pontifícia Universidade Católica do Rio Grande do Sul	10.15468/e5pqi
346	Museu Paraense Emilio Goeldi - Ornithology Collection - Eggs	ocorrência	2018	Museu Paraense Emílio Goeldi	10.15468/fonyee
347	Noctuidae brasileiros de importancia economica depositados na coleção MNHUK	ocorrência	2018	Sistema de Informação sobre a Biodiversidade Brasileira - SiBBR	10.15468/cx8pei
348	rojasianthe_ocurrence_guatemala	ocorrência	2018	Check List	10.15468/0fcvmp
349	RPSP - Coleção Entomológica Prof. J.M.F. Camargo, FFCLRP/USP	ocorrência	2018	Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto (FFCLRP), Universidade de São Paulo (USP)	10.15468/fkx59c
350	BHCB-Bryophyta - Herbário da Universidade Federal de Minas Gerais - Bryophyta	ocorrência	2019	Universidade Federal de Minas Gerais	10.15468/jsjhtj
351	BOTU-Fungi - Herbário Irina Delanova Gemtchújnicov - Coleção de Fungos	ocorrência	2019	Universidade Estadual Paulista - IBB	10.15468/swspua
352	BRBA herbarium - Universidade Federal do Oeste da Bahia - Herbário Virtual Re flora	ocorrência	2019	Instituto de Pesquisas Jardim Botânico do Rio de Janeiro	10.15468/sptcl6
353	Caatinga Biome - RB - Rio de Janeiro Botanical Garden Herbarium Collection	ocorrência	2019	Instituto de Pesquisas Jardim Botânico do Rio de Janeiro	10.15468/bbsqoa
354	COR herbarium - Universidade Federal de Mato Grosso do Sul / Campus Pantanal - Herbário Virtual Re flora	ocorrência	2019	Instituto de Pesquisas Jardim Botânico do Rio de Janeiro	10.15468/ogkonb

355	ECT herbarium - Embrapa Clima Temperado - Herbário Virtual REFLOA	ocorrência	2019	Instituto de Pesquisas Jardim Botânico do Rio de Janeiro	10.15468/a2auot
356	EVB herbarium - Herbário Evaldo Buttura - Herbário Virtual REFLOA	ocorrência	2019	Instituto de Pesquisas Jardim Botânico do Rio de Janeiro	10.15468/ywmowr
357	FUEL-Algae - Coleção Ficológica do Herbário da Universidade Estadual de Londrina	ocorrência	2019	Universidade Estadual de Londrina	10.15468/fdhvxc
358	HDCF herbarium - Universidade Federal de Santa Maria - Herbário Virtual REFLOA	ocorrência	2019	Instituto de Pesquisas Jardim Botânico do Rio de Janeiro	10.15468/0lynoc
359	HSTM herbarium - Universidade Federal do Oeste do Pará - Herbário Virtual REFLOA	ocorrência	2019	Instituto de Pesquisas Jardim Botânico do Rio de Janeiro	10.15468/ztkde
360	HUCS-Liquenoteca - Liquenoteca do Herbário da Universidade de Caxias do Sul	ocorrência	2019	Universidade de Caxias do Sul	10.15468/uv0tvc
361	HUCS-Micoteca - Micoteca do Herbário da Universidade de Caxias do Sul	ocorrência	2019	Universidade de Caxias do Sul	10.15468/pg17e2
362	HUEMG herbarium - Universidade do Estado de Minas Gerais - Campus Carangola - Herbário Virtual REFLOA	ocorrência	2019	Instituto de Pesquisas Jardim Botânico do Rio de Janeiro	10.15468/xvtji6
363	HVASFw - Xiloteca do Herbário Vale do São Francisco	ocorrência	2019	Universidade Federal do Vale do São Francisco	10.15468/vnIr0d
364	IAN herbarium - Embrapa Amazônia Oriental - Herbário Virtual REFLOA	ocorrência	2019	Instituto de Pesquisas Jardim Botânico do Rio de Janeiro	10.15468/cv2dmt
365	MCP-Anfibios - Coleção de Anfíbios	ocorrência	2019	Pontifícia Universidade Católica do Rio Grande do Sul	10.15468/g1e9ms
366	MCP-Aves - Coleção de Aves	ocorrência	2019	Pontifícia Universidade Católica do Rio Grande do Sul	10.15468/z9wv95
367	MCP-Mamiferos - Coleção de Mamíferos	ocorrência	2019	Pontifícia Universidade Católica do Rio Grande do Sul	10.15468/wxyu48
368	MCP-Repteis - Coleção de Répteis	ocorrência	2019	Pontifícia Universidade Católica do Rio Grande do Sul	10.15468/2lm0qf

369	MFS-Flores - Herbário Prof ^a . Dr ^a . Marlene Freitas da Silva - Coleção de Flores	ocorrência	2019	Universidade do Estado do Pará	10.15468/kq5pbw
370	MFS-Frutos - Herbário Prof ^a . Dr ^a . Marlene Freitas da Silva - Coleção de Frutos	ocorrência	2019	Universidade do Estado do Pará	10.15468/ca06tq
371	MFS-Plantulas - Herbário Prof ^a . Dr ^a . Marlene Freitas da Silva - Coleção de Plântulas	ocorrência	2019	Universidade do Estado do Pará	10.15468/vudz6z
372	MUFAL herbarium - Museu de História Natural da Universidade Federal de Alagoas - Herbário Virtual REFLOA	ocorrência	2019	Instituto de Pesquisas Jardim Botânico do Rio de Janeiro	10.15468/viuv6v
373	Museu Paraense Emilio Goeldi - Ornithology Collection - Anatomic	ocorrência	2019	Museu Paraense Emílio Goeldi	10.15468/wbhlel
374	Museu Paraense Emílio Goeldi - Ornithology Collection - Osteological	ocorrência	2019	Museu Paraense Emílio Goeldi	10.15468/5yzcll
375	Museu Paraense Emilio Goeldi - Xiloteca Collection	ocorrência	2019	Museu Paraense Emílio Goeldi	10.15468/dkqjs7
376	Museu_Paraense_Emilio_Goeldi-Ornithology_Collection_Nests	ocorrência	2019	Museu Paraense Emílio Goeldi	10.15468/8enga9
377	Museu_Paraense_Emilio_Goeldi_Microfosseis_Collection	ocorrência	2019	Museu Paraense Emílio Goeldi	10.15468/papqvn
378	Museu_Paraense_Emilio_Goeldi_Ornithology_Collection-Skin	ocorrência	2019	Museu Paraense Emílio Goeldi	10.15468/q1iarh
379	PACA-Algas - Herbarium Anchieta	ocorrência	2019	Instituto Anchieta de Pesquisas/UNISINOS	10.15468/u5bzet
380	PACA-Liquens - Herbarium Anchieta	ocorrência	2019	Instituto Anchieta de Pesquisas/UNISINOS	10.15468/9evnva
381	PMSP herbarium - Prefeitura do Município de São Paulo - Herbário Virtual REFLOA	ocorrência	2019	Instituto de Pesquisas Jardim Botânico do Rio de Janeiro	10.15468/scqmlm
382	RBR herbarium - Universidade Federal Rural do Rio de Janeiro - Herbário Virtual REFLOA	ocorrência	2019	Instituto de Pesquisas Jardim Botânico do Rio de Janeiro	10.15468/0svt7m
383	REAL herbarium - Universidade Federal da Fronteira Sul - Herbário Virtual REFLOA	ocorrência	2019	Instituto de Pesquisas Jardim Botânico do Rio de Janeiro	10.15468/1xuay5

384	Registros de mastofauna do PPG-Ciências Ambientais (Universidade de Brasília)	ocorrência	2019	Sistema de Informação sobre a Biodiversidade Brasileira - SiBBr	10.15468/hs3sod
385	SAMES herbarium - Herbário São Mateus - Herbário Virtual REFLORA	ocorrência	2019	Instituto de Pesquisas Jardim Botânico do Rio de Janeiro	10.15468/l0hdtn
386	SJRP-Fungi - Herbário de fungos de São José do Rio Preto	ocorrência	2019	Universidade Estadual Paulista - São José do Rio Preto	10.15468/lx7oqv
387	Tree species composition in ecotone forests of the eastern Maracá Island, Roraima, northern Brazilian Amazonia: preliminary data	ocorrência	2019	Sistema de Informação sobre a Biodiversidade Brasileira - SiBBr	10.15468/xa5lrb
388	Tree species composition of natural forest islands in a savanna matrix in the northern Brazilian Amazonia	ocorrência	2019	Sistema de Informação sobre a Biodiversidade Brasileira - SiBBr	10.15468/n8yolk
389	UFRN herbarium - Universidade Federal do Rio Grande do Norte - Herbário Virtual REFLORA	ocorrência	2019	Instituto de Pesquisas Jardim Botânico do Rio de Janeiro	10.15468/gtxawd
390	UNOP Herbarium - Universidade Estadual do Oeste do Paraná - Herbário Virtual Re flora	ocorrência	2019	Instituto de Pesquisas Jardim Botânico do Rio de Janeiro	10.15468/eqp1dr
391	VIES-Fungi - Herbário Central da Universidade Federal do Espírito Santo - Coleção de Fungos	ocorrência	2019	Universidade Federal do Espírito Santo	10.15468/k6spx1
392	Base de dados da Coleção Ictiológica do LEP-UFRRJ	ocorrência	2020	Sistema de Informação sobre a Biodiversidade Brasileira - SiBBr	10.15468/srsucy
393	Coleção de Aves Heretiano Zenaide	ocorrência	2020	Sistema de Informação sobre a Biodiversidade Brasileira - SiBBr	10.15468/3uxv7u
394	Coleção de Coleoptera da Universidade de Brasília	ocorrência	2020	Sistema de Informação sobre a Biodiversidade Brasileira - SiBBr	10.15468/gd8kdr
395	Coleção de Crustacea do Museu Nacional (MNRJ - CARCINO)	ocorrência	2020	Museu Nacional / UFRJ	10.15468/zv0lvr
396	Dung beetles (Scarabaeidae, Scarabaeinae) of the Foothills–Andean Forest strip of Villavicencio, Colombia	ocorrência	2020	Check List	10.15468/jc5uy2
397	R-Algae - Herbário do Museu Nacional	ocorrência	2020	Museu Nacional / UFRJ	10.15468/894rku
398	R-Fungi - Herbário do Museu Nacional	ocorrência	2020	Museu Nacional / UFRJ	10.15468/2w9h32

399	Dados da ocorrência de plantas lenhosas em um gradiente de perturbação/precipitação em área de Caatinga, Parque Nacional do Catimbau, Buíque, PE, Brasil	evento	2016	Sistema de Informação sobre a Biodiversidade Brasileira - SiBBR	10.15468/q3afgc
400	Benthic Macroinvertebrate Diversity in the middle Doce river basin, Brazil	evento	2017	Sistema de Informação sobre a Biodiversidade Brasileira - SiBBR	10.15468/cev8wb
401	Environmental monitoring of ants (Hymenoptera: Formicidae) in the influence areas of Santo Antônio Hydroelectric Power-Plant in Rondônia, Brazil.	evento	2017	Sistema de Informação sobre a Biodiversidade Brasileira - SiBBR	10.15468/gz2tsv
402	Continuous monitoring of the micro and mesozooplankton of the Patos Lagoon estuary and adjacent coastal area	evento	2019	Sistema de Informação sobre a Biodiversidade Brasileira - SiBBR	10.15468/1xkowr
403	Dynamics of Submerged Aquatic Vegetation - DIVAS	evento	2019	Sistema de Informação sobre a Biodiversidade Brasileira - SiBBR	10.15468/bjzlnb
404	Ecology of Lahilles bottlenose dolphin <i>Tursiops truncatus gephyreus</i> in the Patos Lagoon estuary and adjacent marine coast	evento	2019	Sistema de Informação sobre a Biodiversidade Brasileira - SiBBR	10.15468/4nh9ng
405	Ecology of the pink-shrimp <i>Penaeus paulensis</i> in Patos Lagoon estuary	evento	2019	Sistema de Informação sobre a Biodiversidade Brasileira - SiBBR	10.15468/ovayhc
406	Interannual variability of ichthyoplankton diversity in the Patos Lagoon estuary Southern Brazil	evento	2019	Sistema de Informação sobre a Biodiversidade Brasileira - SiBBR	10.15468/noeqwa
407	Phytoplankton and water quality parameters in the Patos Lagoon estuary and adjacent marine coast	evento	2019	Sistema de Informação sobre a Biodiversidade Brasileira - SiBBR	10.15468/xmlvxm
408	Species composition and abundance patterns of fish assemblages at shallow waters of Patos Lagoon estuary	evento	2019	Sistema de Informação sobre a Biodiversidade Brasileira - SiBBR	10.15468/kci8zb
409	Temporal data series of Benthic macrofauna abundance and composition from the Patos Lagoon estuary	evento	2019	Sistema de Informação sobre a Biodiversidade Brasileira - SiBBR	10.15468/lsoc2v

ANEXO 11. Dados coletados na rede GBIF: metadados dos conjuntos de dados

Devido ao grande volume de dados coletados, este anexo é apresentado no formato *xlsx nos arquivos do tipo planilha descritos a seguir:

- **11.1 - Lista de *datasets*** Classe evento (*sampling event* – 1.590 *datasets*) e metadados de recurso (*resource metadata* – 308 *datasets*)
- **11.2 - Lista de *datasets*** Classe ocorrência (*occurrence* – 19.680 *datasets*)
- **11.3 - Lista de *datasets*** Classe checklist (31.151 *datasets*)
- **11.4 - Quantidade de elementos de metadados por *datasets*** - Lista de *datasets* com dados das classes evento e ocorrência (total de 20.913 *datasets*)
- **11.5 - Quantidade de elementos de metadados por *datasets*** - Lista de *datasets* com dados das classes checklist (31.890 *datasets*)

*Observação: O ANEXO 11 é composto por cinco arquivos do tipo planilha eletrônica disponibilizados no Repositório Institucional da UnB, associadas à referência desta dissertação:

CARDOSO, Isabela Pereira. Dados científicos e metadados: estudo sobre o uso dos padrões de metadados no fluxo da informação científica sobre biodiversidade. Brasília: Universidade de Brasília, 2020.