



Universidade de Brasília
Faculdade Ciência da Informação (FCI)
Programa de Pós-Graduação em Ciência da Informação (PPCInf)

Raquel Viana

**COMPARTILHAMENTO DE DADOS DE PESQUISA EM REPOSITÓRIOS
DIGITAIS : O CENÁRIO LATINO AMERICANO**

Orientador: Prof. Dr. Emir José Suaiden
Co-Orientadora: Profa. Dra. Marília Augusta de Freitas

Brasília
2019

RAQUEL VIANA

**COMPARTILHAMENTO DE DADOS DE PESQUISA EM REPOSITÓRIOS
DIGITAIS : O CENÁRIO LATINO AMERICANO**

Dissertação apresentada ao Programa de Pós-Graduação em Ciência da Informação (PPGCINF) da Faculdade de Ciência da Informação (FCI) da Universidade de Brasília (UnB), como requisito parcial para a obtenção do título de mestre em Ciência da Informação.

Orientador: Prof. Dr. Emir José Suaiden

Co-Orientadora: Profa. Dra. Marília
Augusta de Freitas

Brasília 2019

FOLHA DE APROVAÇÃO

Título: “Compartilhamento de dados de pesquisa em repositórios digitais: o cenário latino americano”

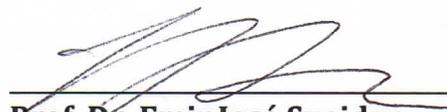
Autor (a): Raquel Viana Ferreira

Área de concentração: Gestão da Informação

Linha de pesquisa: Comunicação e Mediação 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.

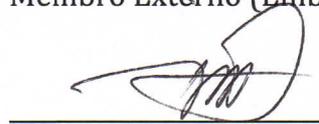
Dissertação aprovada em: 15 de março 2019.



Prof. Dr. Emir José Suaiden
Presidente (UnB/PPGCINF)



Profª Drª Patricia Rocha Bello Bertin
Membro Externo (Embrapa)



Prof. Dr. Fernando César Lima Leite
Membro Interno (UnB/PPGCINF)

Profª Drª Michelli Pereira da Costa
Suplente - (UnB/FCI)

DEDICATÓRIA

Dedico essa dissertação a todos e todas que acharam
que ela não seria possível...

EPÍGRAFE

*Viver no mundão tá ligado que é caso sério
Dando a cara a tapa
E às vezes sem ter critério
Levar o legado sendo parte do mistério
Trabalhar pra prosperar
Ascensão do império*

Glória Groove do Brasil

AGRADECIMENTOS

Há na mitologia africana uma crença que dita que nós não somos apenas um, nós somos também todos os nossos antepassados que se transformaram no barro que nos moldou, por isso uma pessoa consegue ser muitos. Hoje, se eu terminei essa dissertação, percebo que não consegui sozinha, compreendo que não trilhei esse caminho na solidão, estavam comigo as minhas amigas de trabalho Ruth, Thayse, Gaby, Patricia, Tania, Jacqueline e Luanna que forma além que colegas de serviço e me deram um apoio surreal na realização desse trabalho.

Me ajudaram a caminhar os meus amigos de curso Rafael Rosa, Dani D., Patrícia, Elton e Erika, com quem aprendi muito sobre os caminhos da Pós-Graduação.

Nessa trajetória eu aprendi que algumas pedras, que entram em nossas vidas apenas para atravancar o nosso caminho, também são importantes e necessárias porque, se as pedras te derrubam, amigos como o Rodrigo, a Tyara, o Felipe, a Janaina, Seu Antônio e a Rosa te ajudam a levantar, e a eles eu sou muito grata. As pedras passaram e eu passarinho.

Eu também agradeço a todas as minhas amigas do Instituto Brasileiro de Ciência e Tecnologia – IBICT – Mih, Taína, Fernanda, Luísa, Rayanne e Bianca pelo apoio e incentivo.

Especialmente eu agradecerei a Marília Freitas, que em um momento que mais precisei, ela, com um bebê recém-nascido, conseguiu arrumar tempo e disposição para me estender uma mão. A Marília conseguiu dar uma nova perspectiva a elaboração dessa dissertação e sem ela, creio que que todo esse caminho teria sido muito mais tortuoso. Não tenho palavras para expressar o quanto sou agradecida não somente pela ajuda na orientação na elaboração da dissertação, mas pela amizade que ela me ofereceu desde o momento que me conheceu. Agradeço também ao meu orientador, Professor Emir, pelo apoio e direcionamento.

Com eles, como orientadores e amigos, eu tive a relação que todos os orientandos deveriam ter com seus mentores: uma relação de amizade, apoio e direcionamento.

Eu não poderia deixar de ser grata aos meus dois braços fortes, a quem eu sempre posso contar, quem me apoia, me segura, me dá força, mesmo quando elas mesmo estão cansadas, a minha irmã Shirley Viana- que vem cuidando e zelando por meu caminho até mesmo antes que eu nascesse- e a minha verdadeira orientadora, genitora, a minha mãe, eu agradeço, verdadeiramente dedico esse trabalho.

A todos vocês, meu barro forte, o meu muito obrigado!

LISTA DE ABREVIATURAS E SIGLAS

| | |
|---------|---|
| CNPq | Conselho Nacional de Desenvolvimento Científico e Tecnológico |
| OECD | Organisation for Economic Co-Operation and Development |
| RD | Repositório de Dados de pesquisa |
| Re3data | Registry Of Research Data Repositories |
| RI | Repositórios Institucionais |
| RT | Repositórios Temáticos |
| TICs | Tecnologias de Informação e Comunicação |
| FOSTER | (Facilitate Open Science Training for European Research) |
| FIOCRUZ | Fundação Oswaldo Cruz |

RESUMO

O sistema de comunicação científica passou por um processo de reestruturação a partir do momento em que a internet e outras ferramentas tecnológicas foram percebidas como veículos muito eficazes para a disseminação da informação. A partir dessa percepção, diferentes áreas do conhecimento ressignificaram seus canais preferenciais para a troca dos elementos e resultados de suas pesquisas em diferentes dimensões. As mudanças ocorridas nas trocas de resultados também foram alavancadas pelo movimento de Acesso Aberto, que tem como prerrogativa a livre circulação das descobertas científicas. A adoção desse modelo de circulação da informação científica transformou os repositórios em importantes elementos não só para a disseminação da informação, mas também dos componentes basilares para a sua formação, os dados científicos.

O compartilhamento de dados científicos em repositórios é adotado com diferentes níveis e particularidades em diferentes áreas do conhecimento, no entanto é comum que todos sigam os preceitos estabelecidos pelo movimento que lhes deram origem. Sendo assim, tão importante quanto mensurar quais são os elementos cruciais que caracterizam os repositórios como pertencentes a esse movimento é estabelecer quais repositórios se encaixam nessa categoria e aferir suas particularidades. Neste contexto, esta pesquisa objetivou identificar as características e contribuições dos repositórios de dados de pesquisa científicas da América Latina para o desenvolvimento da ciência aberta. Para o alcance desse objetivo, foi realizado um estudo com metodologia mista e abordagem comparativa. Para a observação do comportamento dos repositórios foi utilizada uma lista de comparação baseada nos elementos essenciais do movimento de Acesso Aberto que foram traçados ao longo da revisão de literatura.

As preferências comportamentais dos repositórios quanto a sua organização e representação dos dados de pesquisa foram identificadas conforme a variação das áreas do conhecimento a qual pertencem. Os resultados que serão encontrados na presente pesquisa poderão sinalizar em

qual grau os repositórios digitais de dados científicos corroborarão com o movimento de Acesso Aberto, no âmbito geográfico da América Latina.

Palavras-chave: Comunicação científica; Compartilhamento de dados; Acesso Aberto; Diferenças disciplinares Repositório de dados.

ABSTRACT

The scientific communication system underwent a process of restructuring from the moment the Internet and others technological tools were perceived as very effective vehicles for the dissemination of information. From this perception, different areas of knowledge re-signified their preferred channels for the exchange of elements and results of their research in different dimensions. The changes that occurred in the exchange of results were also leveraged by the Open Access movement, which has as its prerogative the free circulation of scientific discoveries. The adoption of this model of circulation of scientific information has transformed the repositories into important elements not only for the dissemination of information, but also for the basic components for their formation, the scientific data.

The sharing of scientific data in repositories is adopted with different levels and particularities in different areas of knowledge, however it is common for all to follow the precepts established by the movement that gave rise to them. Thus, it is as important to measure what are the crucial elements that characterize them as belonging to this movement is to establish which repositories fit into that category and to gauge their particularities. In this context, this research aimed to identify the characteristics and contributions of scientific research data repositories in Latin America for the development of open science. To reach this objective, a study with mixed methodology and comparative approach was carried out. To observe the behavior of the repositories, a comparison list was used based on the essential elements of the Open Access movement that were traced throughout the literature review.

The behavioral preferences of the repositories regarding their organization and representation of the research data were identified according to the variation of the areas of knowledge to which they belong. The results that will be found in the present research will be able to indicate in which degree the digital repositories of scientific data will corroborate with the Open Access movement in the geographic scope of Latin America.

FIGURAS

| | |
|--|-----|
| Figura 1: Dimensões da Ciência Aberta | 25 |
| Figura 2: Principais iniciativas de acesso aberto no mundo e na América Latina..... | 48 |
| Figura 3: Arxiv.org..... | 50 |
| Figura 4: O processo de Acesso Aberto, uma visão geral..... | 56 |
| Figura 5: Paradigmas da ciência | 62 |
| Figura 6: Ciência aberta: um termo, cinco escolas de pensamento | 66 |
| Figura 7: Aspectos conceituais do Big Data. | 76 |
| Figura 8: Como os dados abertos se relacionam com outros tipos de dados..... | 78 |
| Figura 9: Como os dados são abertos ou fechados, com base com base em quatro características | 80 |
| Figura 10: Infográfico dos benefícios dos Dados de Pesquisa abertos . | 81 |
| Figura 11: Barreiras para o compartilhamento de dados..... | 82 |
| Figura 12: modelo de gerenciamento e preservação para dados de pesquisa..... | 84 |
| Figura 13: O ciclo dos dados científicos | 85 |
| Figura 14: Modelo de comunicação científica de Garvey e Griffith..... | 89 |
| Figura 15: Comunicação Científica - um modelo para 2020..... | 91 |
| Figura 16: Publicação no Ciclo de vida da Pesquisa..... | 92 |
| Figura 17: Diagrama do ciclo de vida da pesquisa | 97 |
| Figura 18: Relacionamento do ciclo de vida da pesquisa (A) com o ciclo de vida do dado (B)..... | 98 |
| Figura 19: Meios de compartilhamento..... | 112 |
| Figura 20: Motivação dos pesquisadores para compartilhar dados..... | 114 |
| Figura 21: Onde os cientistas compartilham dados | 115 |
| Figura 22: Esquema de arquivamento de dados científicos | 117 |
| Figura 23: O ciclo de vida dos repositórios de dados de pesquisa | 122 |
| Figura 24: Padrões de metadados dos repositórios..... | 129 |
| Figura 25: Estrutura modular do DDI..... | 131 |

| | |
|---|-----|
| Figura 26: Software utilizados pelos repositórios de dados de pesquisa | 134 |
| Figura 27: Dataverse / Biblioteca Digital do Texas (TDL) | 134 |
| Figura 28: Índice vertebrado inicial | 142 |
| Figura 29: Desenho da Pesquisa..... | 144 |
| Figura 30: Resumo da pesquisa | 148 |
| Figura 31: Mapa da América Latina..... | 151 |
| Figura 32: Esquema de análise das características dos RD's..... | 154 |
| Figura 33: Ficha para análise dos RDs..... | 155 |
| Figura 34: Quantidade de RD's Latino Americanos | 163 |
| Figura 35: Relação Tipo de dados - Área Geográfica – Área do Conhecimento – Conteúdo armazenado - Brasil..... | 165 |
| Figura 36: Relação Tipo de dados - Área Geográfica – Área do Conhecimento – Conteúdo armazenado - Chile..... | 166 |
| Figura 37: Relação Tipo de dados - Área Geográfica – Área do Conhecimento – Conteúdo armazenado - Colômbia..... | 167 |
| Figura 38: Relação Tipo de dados - Área Geográfica – Área do Conhecimento – Conteúdo armazenado - México | 168 |
| Figura 39: Relação Tipo de dados - Área Geográfica – Área do Conhecimento – Conteúdo armazenado - Panamá | 170 |
| Figura 40: Relação Tipo de dados - Área Geográfica – Área do Conhecimento – Conteúdo armazenado - Peru | 171 |
| Figura 41: Segmentação do tipo de dados por região geográfica | 172 |
| Figura 42: Vínculo institucional dos RD's..... | 173 |
| Figura 43: Relação Área Geográfica- Vínculo institucional..... | 174 |
| Figura 44: Relação Área Geográfica- Perfil acadêmico | 175 |
| Figura 45: Relações Área Geográfica – Tipo de licença adotada – Conteúdo completo – Brasil | 178 |
| Figura 46: Área Geográfica – Tipo de licença adotada – Conteúdo completo – Chile..... | 179 |
| Figura 47: Área Geográfica – Tipo de licença adotada – Conteúdo completo – México | 180 |
| Figura 48: Área Geográfica – Tipo de licença adotada – Conteúdo completo – Colômbia..... | 180 |

| | |
|---|-----|
| Figura 49: Área Geográfica – Tipo de licença adotada – Conteúdo completo – Panamá | 181 |
| Figura 50: Área Geográfica – Tipo de licença adotada – Conteúdo completo – Peru | 182 |
| Figura 51: Interação nos RD's | 183 |
| Figura 52: Relações Área Geográfica – Critérios de Persistência – Estatística – Brasil..... | 186 |
| Figura 53: Relações Área Geográfica – Critérios de Persistência – Estatística – Chile..... | 186 |
| Figura 54: Relações Área Geográfica – Critérios de Persistência – Estatística – Colômbia..... | 187 |
| Figura 55: Relações Área Geográfica – Critérios de Persistência – Estatística – México | 188 |
| Figura 56: Relações Área Geográfica – Critérios de Persistência – Estatística – Panamá..... | 188 |
| Figura 57: Relações Área Geográfica – Critérios de Persistência – Estatística – Peru | 189 |
| Figura 58: Relações Área Geográfica – Esquema de Software e metadados adotados - Brasil..... | 190 |
| Figura 59: Relações Área Geográfica – Esquema de Software e metadados adotados - Chile..... | 191 |
| Figura 60: Relações Área Geográfica – Esquema de Software e metadados adotados – Colômbia..... | 192 |
| Figura 61: Relações Área Geográfica – Esquema de Software e metadados adotados - México | 193 |
| Figura 62: Relações Área Geográfica – Esquema de Software e metadados adotados - Panamá | 194 |
| Figura 63: Relações Área Geográfica – Esquema de Software e metadados adotados - Peru | 194 |
| Figura 64: Relação Área Geográfica – Política de Gestão e Política de Gestão presente no Melibea - Brasil | 196 |
| Figura 65: Relação Área Geográfica – Política de Gestão e Política de Gestão presente no Melibea -Chile | 196 |

| | |
|--|-----|
| Figura 66: Relação Área Geográfica – Política de Gestão e Política de Gestão presente no Melibea - Colômbial | 197 |
| Figura 67: Relação Área Geográfica – Política de Gestão e Política de Gestão presente no Melibea - México | 198 |
| Figura 68: Relação Área Geográfica – Política de Gestão e Política de Gestão presente no Melibea - Panamá | 199 |
| Figura 69: Relação Área Geográfica – Política de Gestão e Política de Gestão presente no Melibea - Peru..... | 199 |
| Figura 70: Resumo do método | 208 |

QUADROS

| | |
|--|-----|
| Quadro 1: Características dos arquivos abertos..... | 54 |
| Quadro 2: Recomendações à Comunidade Científica..... | 60 |
| Quadro 3: O uso da comunicação informal nas áreas do conhecimento | 103 |
| Quadro 4: Análise Mueller 2005 | 107 |
| Quadro 5: Distribuição de pesquisadores por Área (n=226)..... | 107 |
| Quadro 6: Total de documentos por tipo e grande área do saber | 108 |
| Quadro 7: Tipos de compartilhamento de dados | 118 |
| Quadro 8: Softwares de repositórios de dados..... | 133 |
| Quadro 9: Sistematização da metodologia da pesquisa..... | 159 |
| Quadro 10: Amostra de RDs analisados..... | 164 |

TABELAS

| | |
|---|-----|
| Tabela 1: Comparação entre os textos das declarações de Budapeste, Bethesda e Berlim que regem o Acesso Livre ao conhecimento. | 58 |
| Tabela 2: Compartilhamento de dados por disciplina | 116 |
| Tabela 3: Quantidade de RD's Latino Americanos | 161 |
| Tabela 4: Caracterização dos RD's de acordo com as dimensões apontadas pela literatura | 202 |
| Tabela 5: Distribuição geral das áreas do conhecimento | 211 |

SUMÁRIO

| | |
|--|------------|
| 1. INTRODUÇÃO..... | 20 |
| 1.1. CONTEXTUALIZAÇÃO DO PROBLEMA..... | 21 |
| 1.1.1. <i>Objetivos</i> | 30 |
| 1.1.1.1. <i>Objetivo Geral</i> | 30 |
| 1.1.1.2. <i>Objetivos específicos</i> | 30 |
| 1.2. JUSTIFICATIVA..... | 30 |
| 1.2.1. <i>Natureza teórica</i> | 33 |
| 1.2.2. <i>Natureza prática</i> | 33 |
| 2. FUNDAMENTAÇÃO TEÓRICA | 35 |
| 2.1. COMUNICAÇÃO CIENTÍFICA | 36 |
| 2.2. MOVIMENTO DE ACESSO ABERTO A INFORMAÇÃO CIENTÍFICA | 40 |
| 2.2.1. <i>A Emergência do Acesso Aberto</i> | 45 |
| 2.3. CIÊNCIA ABERTA | 64 |
| 2.3.1. <i>Escola Pública</i> | 67 |
| 2.3.2. <i>Escola de medição</i> | 68 |
| 2.3.3. <i>Escola Pragmática</i> | 69 |
| 2.3.4. <i>Escola de Infraestrutura</i> | 70 |
| 2.3.5. <i>Escola Democrática</i> | 71 |
| 2.3.6. <i>Dados Abertos: conceituação e classificação</i> | 72 |
| 2.3.6.1. <i>Dados: conceituação</i> | 72 |
| 2.3.6.2. <i>Dados abertos: conceituação</i> | 77 |
| 2.3.6.3. <i>Ciclo de Vida dos Dados de Pesquisa</i> | 83 |
| 2.3.7. <i>Contestações a respeito da Ciência Aberta</i> | 86 |
| 2.4. MODELOS DE COMUNICAÇÃO CIENTÍFICA..... | 88 |
| 2.4.1. <i>O compartilhamento de dados e os novos modelos de Comunicação Científica</i> | 93 |
| 2.5. AS DIFERENÇAS CULTURAIS CIENTÍFICAS NA TRANSMISSÃO DO CONHECIMENTO | 101 |
| 2.5.1. <i>Diferenças disciplinares relacionadas a dados de pesquisa</i> | 111 |
| 2.6. REPOSITÓRIOS DE DADOS DE PESQUISA..... | 117 |
| 2.6.1. <i>A UTILIZAÇÃO DE METADADOS PARA A DESCRIÇÃO EM REPOSITÓRIOS DE DADOS</i> | 128 |
| 2.6.2. <i>A UTILIZAÇÃO DE SOFTWARES PARA A CRIAÇÃO DE REPOSITÓRIOS DE DADOS</i> | 132 |
| 3. PRINCIPIOS E RECOMENDAÇÕES SOBRE O ACESSO ABERTO EM REPOSITÓRIOS DE DADOS DE PESQUISA..... | 136 |

| | |
|--|------------|
| 4. METODOLOGIA..... | 138 |
| 4.1. PROCEDIMENTOS | 139 |
| 4.1.1. <i>Desenho da Pesquisa</i> | 144 |
| 4.1.2. <i>Delimitações e limites da pesquisa</i> | 148 |
| 4.1.2.1. Universos e fontes de dados..... | 148 |
| 4.1.2.1.1. Universo “a”..... | 149 |
| 4.1.2.1.2. Universo “b”..... | 151 |
| 4.1.3. <i>Método de coleta e Análise dos dados</i> | 152 |
| 5. ANÁLISE E DISCUSSÃO DOS DADOS | 160 |
| 5.1. ANÁLISE DOS CRITÉRIOS DE CARACTERÍSTICAS BÁSICAS DOS RDS..... | 165 |
| 5.1.1. <i>Relação Tipo de dados - Área Geográfica – Área do Conhecimento</i> | 165 |
| 5.1.1.1. Brasil..... | 165 |
| 5.1.1.2. Chile..... | 166 |
| 5.1.1.3. Colômbia | 167 |
| 5.1.1.4. México..... | 168 |
| 5.1.1.5. Panamá..... | 169 |
| 5.1.1.6. Peru..... | 171 |
| 5.2. RELAÇÃO ÁREA GEOGRÁFICA- VÍNCULO INSTITUCIONAL..... | 173 |
| 5.3. RELAÇÃO ÁREA GEOGRÁFICA- PERFIL ACADÊMICO..... | 175 |
| 5.4. ANÁLISE DOS CRITÉRIOS DE CARACTERÍSTICAS DE ACESSO ABERTO DOS RD'S | 177 |
| 5.4.1. <i>Relações Área Geográfica – Tipo de licença adotada – Conteúdo</i> | |
| <i>completo</i> 177 | |
| 5.4.1.1. Brasil..... | 177 |
| 5.4.1.2. Chile..... | 178 |
| 5.4.1.3. México..... | 179 |
| 5.4.1.4. Colômbia | 180 |
| 5.4.1.5. Panamá..... | 181 |
| 5.4.1.6. Peru..... | 181 |
| 5.4.2. <i>Colaboração científica</i> | 183 |
| 5.5. ANÁLISE DOS CRITÉRIOS DE CARACTERÍSTICAS DE GESTÃO E FUNCIONAMENTO | |
| DOS RD'S 185 | |
| 5.5.1. <i>Relações Área Geográfica – Critérios de Persistência – Estatística</i> | 185 |
| 5.5.1.1. Brasil..... | 185 |
| 5.5.1.2. Chile..... | 186 |
| 5.5.1.3. Colômbia | 187 |
| 5.5.1.4. México..... | 187 |
| 5.5.1.5. Panamá..... | 188 |
| 5.5.1.6. Peru..... | 189 |
| 5.5.1. <i>Relações Área Geográfica – Esquema de Software e metadados</i> | |
| <i>adotados</i> 190 | |
| 5.5.1.1. Brasil..... | 190 |
| 5.5.1.2. Chile..... | 191 |

| | |
|---|------------|
| 5.5.1.3. Colômbia | 191 |
| 5.5.1.4. México..... | 192 |
| 5.5.1.5. Panamá..... | 193 |
| 5.5.1.6. Peru..... | 194 |
| 5.5.2. <i>Relação Área Geográfica – Política de Gestão e Política de Gestão presente no Melibea</i> | 195 |
| 5.5.2.1. Brasil | 195 |
| 5.5.2.2. Chile..... | 196 |
| 5.5.2.3. Colômbia | 197 |
| 5.5.2.4. México..... | 197 |
| 5.5.2.5. Panamá..... | 198 |
| 5.5.2.6. Peru..... | 199 |
| 5.6. CARACTERIZAÇÃO DOS RD'S DE ACORDO COM AS DIMENSÕES APONTADAS PELA LITERATURA | 200 |
| 6. CONCLUSÕES..... | 207 |
| 7. REFERÊNCIAS BIBLIOGRÁFICAS | 212 |
| 8. ANEXOS..... | 227 |

1. INTRODUÇÃO

1.1. CONTEXTUALIZAÇÃO DO PROBLEMA

A construção de um conhecimento científico está ligada às atividades desenvolvidas pelos cientistas e pesquisadores envolvidos por um contexto de informações que baseiam e norteiam a produção de documentos. Le Coadic (2004) contextualiza a produção de conhecimento às chamadas comunidades científicas surgidas no século XIX, onde os próprios cientistas impulsionavam o desenvolvimento intelectual por meio da troca de ideias desvinculadas de ideologias políticas e econômicas. Segundo o autor, essas comunidades funcionavam mediante a troca de informações baseada na confirmação do pesquisador como real cientista. Nota-se que a real intenção das comunidades científicas, desde sua gênese, estava ligada aos princípios do compartilhamento do conhecimento entre cientistas.

Le Coadic (2004) também explica como o conhecimento é construído dentro dessas comunidades. O autor cita três elementos constitutivos desse processo:

- Construção;
- Comunicação e o
- Uso da informação.

Esses elementos formam o ciclo da informação, de onde surge o conhecimento. O processo de construção do conhecimento também pode ser observado no conceito de Lopes Yepes (1995) que considera que o processo informativo seja constituído por um sujeito ativo, que informa; um objeto da informação que é emitida por um meio de comunicação e um sujeito passivo que recebe a mensagem comunicada.

Meadows (1999) informa que a maneira como essa mensagem é comunicada pelo cientista depende do veículo empregado, da natureza das informações e do público-alvo, sendo que a forma mais importante de comunicar essa mensagem é através da fala ou da escrita. Essa informação segue então por um fluxo, que segundo Mueller (2000, p. 28), normalmente é composto por relatórios preliminares, seminários, publicações pré-prints e artigos científicos.

No entanto, Cardoso, Jacobertty e Duarte (2012) alertam que atualmente a sociedade está diante de uma série de novas práticas de investigação científica que alteraram a forma como as publicações científicas e técnicas são compartilhadas e disponibilizadas entre os cientistas.

Bartling e Friesike (2014, p.6) mencionam duas revoluções pelas quais as práticas científicas passaram. A primeira é chamada de “Profissionalização da Ciência”. Essa revolução criou as bases do trabalho científico que temos hoje ao requerer “meios para avaliar o valor de uma contribuição, de modo que os incentivos para uma pesquisa bem-sucedida pudessem ser fornecidos.” Ainda explicam que, pelo fato do controle da impressão e divulgações dos resultados terem sido concentrados nos editores, o impacto científico se desenvolveu em torno dos periódicos científicos.

A Segunda Revolução se caracteriza pela adoção da Internet pelos cientistas como meio alternativo de publicação em papel, com a certeza da correta citação dos seus trabalhos e como ferramenta de interação entre os pesquisadores, que não mais necessitam esperar pela conclusão de uma pesquisa para a sua disponibilização. Puschmann (2014, p. 91) relata que essas tecnologias tornam mais fácil, barato e rápido o processo de troca de informações entre os cientistas e seus pares em todo mundo.

Se tratando as atuais tecnologias utilizadas pelos pesquisadores, Castells (2010) acredita que a revolução tecnológica da informação é a causa das transformações ocorridas na geração, processamento e transmissão da informação. O autor explica que essa revolução está alicerçada nas inovações tecnológicas ocorridas principalmente em três campos: a microeletrônica, os computadores e as telecomunicações.

O autor ainda relata que esse novo paradigma foi organizado com base na tecnologia da informação e, apesar de ter se concretizado a partir do papel decisivo do financiamento militar e dos mercados¹ nos primeiros estágios da indústria eletrônica (1940-1960), foi na década de 70 que alcançou o seu grande progresso relacionando-se à cultura da liberdade.

1 Segundo Castells (2010) ambos os seguimentos, militar e de mercado, possuem bases e posturas conservadoras e tradicionalistas, o que dá maior destaque a postura libertária que as novas tecnologias de informação adotaram.

“A ênfase nos dispositivos personalizados, na interatividade, na formação de redes e na busca incansável de novas descobertas tecnológicas, mesmo quando não faziam muito sentido comercial, não combinava com a tradição, de certa forma cautelosa, do mundo corporativo. Meio inconscientemente, a revolução da tecnologia da informação difundiu pela cultura mais significativa de nossas sociedades o espírito libertário dos movimentos dos anos 60.” (CASTELLS, 2010, p. 43).

As transformações ocorridas no ciclo informacional devido as inovações tecnológicas possibilitaram o surgimento de um movimento social chamado de *Open Science*, ou Ciência Aberta, que possui uma identidade que imprime a liberdade à circulação e distribuição da informação. De acordo com os preceitos estabelecidos pelo *Open Definition*², para que qualquer obra seja considerada aberta deve satisfazer os seguintes critérios básicos:

- Ter uma licença aberta;
- Ter o acesso aberto e
- Ter um formato aberto.

Segundo o que foi enunciado por Boulton (2013), a Ciência Aberta permite a “replicação de observações ou experiências ou a sua refutação, garantindo que o conhecimento científico é cumulativo.”.

Os termos “Ciência 2.0”, “Ciência Aberta”, “Humanidades Digitais”, “eScience” ou “Pesquisa Aberta” são considerados, de acordo com Bartling e Friesike (2014), realçadores das mudanças científicas que ocorreram nos últimos anos e enfatizam vários aspectos da segunda revolução científica, sem, no entanto, ter uma delimitação conceitual consolidada.

Ciência Aberta³, em conformidade com o que foi definido pelos autores é “uma cultura científica que é caracterizada pela sua abertura. Os cientistas compartilham resultados quase que imediatamente e com um público muito amplo.”, o que também inclui o fornecimento de acesso aos dados que embasaram a pesquisa realizada.

De acordo com o que Cardoso, Jacobertty e Duarte (2012) consideram, a Ciência Aberta se define como algo além de:

² <https://opendefinition.org/od/2.0/pt-br/>

³ O tema “Ciência Aberta” será melhor aprofundado em um capítulo específico dessa dissertação.

“[...] uma mera tendência editorial de acesso livre a publicações científicas, assiste-se atualmente e crescentemente, entre a comunidade científica, a um movimento social mais alargado de abertura da ciência, que se concretiza não apenas na acessibilidade às publicações científicas resultantes das investigações, mas também numa maior visibilidade e transparência nas metodologias desenvolvidas pelos investigadores no decurso das suas pesquisas, nas estratégias e ferramentas de observação e de recolha de informação empírica, bem como na divulgação e disseminação dos dados recolhidos em si, designadamente através da disponibilização *online* de bases de dados. Esta tendência mais generalizada de transparência e visibilidade do processo científico leva a acreditar que está em curso, entre a comunidade científica, um ideal de maior transparência, acessibilidade e difusão da informação. Ou seja, mais do que facilitar o acesso aos artigos científicos na Internet, o que se assiste é, cada vez mais, a outros movimentos paralelos e similares, não apenas respeitantes ao produto final das pesquisas, mas também aos instrumentos de observação e recolha de informação empírica, bem como aos dados recolhidos em si, o que reflete numa maior colaboração e cooperação entre os cientistas. (CARDOSO, JACOBERTTY e DUARTE, 2012, p. 18 e 19)

Em resumo, os autores definem Ciência Aberta como “a disponibilização *online* dos vários produtos do processo de pesquisa, desde a fase de observação e recolha de dados à sua forma final sob formato de publicação”. Eles ainda explicam que a Ciência Aberta, é uma prática que se insere em uma perspectiva que ultrapassa o modelo de compartilhamento de produção científica caracterizado pelo *open access* às publicações científicas, se estendendo assim, a várias etapas do processo científico, como é o caso do compartilhamento dos dados recolhidos no decorrer da elaboração da pesquisa.

No entanto, apesar de citarem as várias vantagens para a adoção das práticas de Ciência Aberta pela sociedade, Pontika et al. (2015) declararam que continua obscuro para alguns atores científicos como os financiadores da pesquisa, os próprios pesquisadores, os bibliotecários e os gerentes de repositórios como a Ciência Aberta pode ser alcançada.

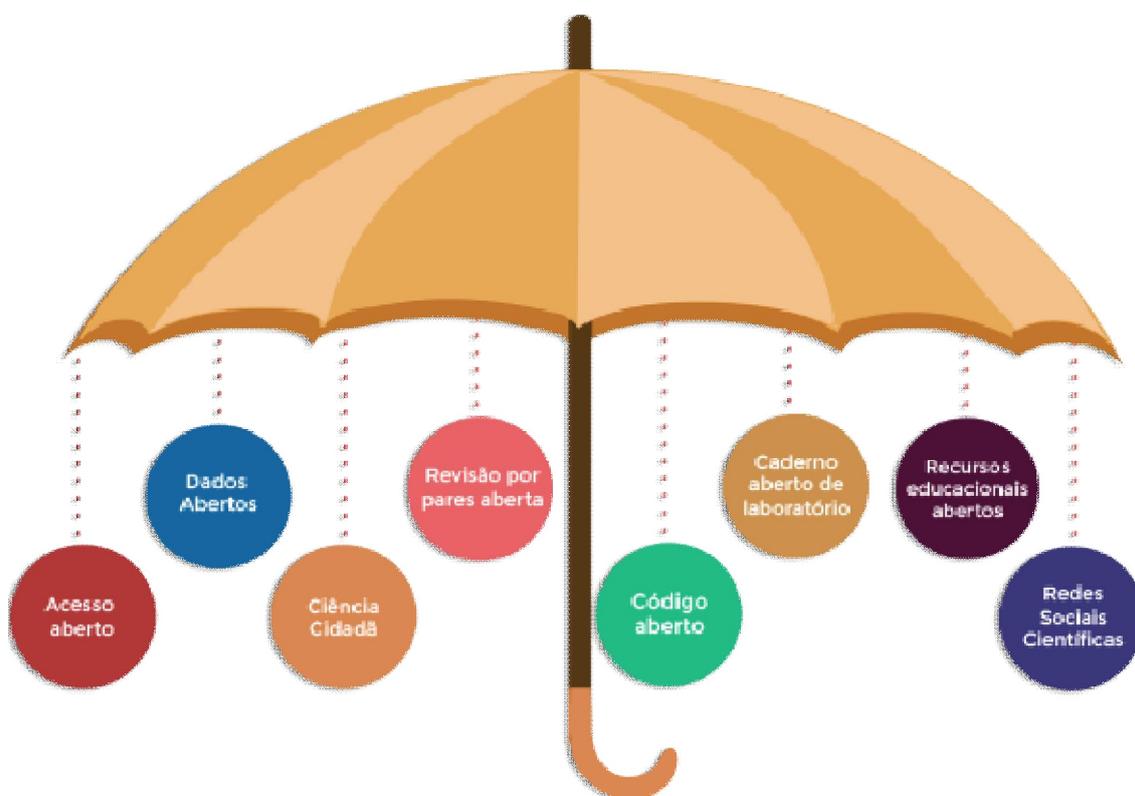
Além do clássico conceito que envolve o compartilhamento de forma aberta de códigos, notas e dados, Gezalter (2009) cita quatro pilares fundamentais para a consolidação da Ciência Aberta:

1. Transparência na metodologia experimental, observação e coleta de dados;
2. Acessibilidade pública e transparência da comunicação científica.

3. Uso de ferramentas baseadas na web para facilitar a colaboração científica;
4. Disponibilidade pública e reutilização de dados científicos.

A Fundação Oswaldo Cruz (Fiocruz) compartilha da linha de raciocínio dos autores Cardoso, Jacobertty e Duarte (2012) e também afirma que o termo “Ciência Aberta” faz referência a um movimento que se sobrepõe ao acesso aberto às publicações e dados de pesquisa. Para a fundação, fazer uma Ciência Aberta implica em que todo o processo de pesquisa seja aberto, “usando métodos, ferramentas e workflows que facilitem a partilha, reutilização e colaboração. Por isso, a Ciência Aberta é um conceito abrangente que se baseia em diversos pilares e inclui múltiplas dimensões.”, conforme é indicado na figura abaixo.

Figura 1: Dimensões da Ciência Aberta



Fonte: Fiocruz (O que é Ciência Aberta)

Dentre as múltiplas dimensões que abarcam o termo “Ciência Aberta”, a Fiocruz aponta:

- Acesso Aberto;
- Dados Abertos;
- Ciência Cidadã;
- Revisão por pares aberta;
- Código Aberto;
- Caderno aberto de laboratório;
- Recursos educacionais abertos e
- Redes Sociais Científicas.

Entre as vertentes da Ciência Aberta que são apontadas pela Fiocruz, essa dissertação foca apenas nos dois primeiros temas: “Acesso Aberto” e “Dados Abertos”, ainda que, segundo Pontika et al. (2015) componentes da Ciência Aberta como a Pesquisa Aberta e a sua reprodutibilidade careçam de mais atenção pela comunidade científica, já que ela aparenta estar restrita a ciências como a Ciência da Computação e as Engenharias da computação.

Pontika et al. (2015) estabelecem que o conceito de “Acesso Aberto” se refere aos “resultados acadêmicos on-line, revisados por pares, que são livres de ler, com restrições limitadas ou inexistentes de direitos autorais e licenciamento”. Para os autores, a Convenção de Budapeste consolidou esse termo em uma estratégia que induziu o aumento na quantidade de pesquisas acessíveis e reutilizáveis de forma gratuita. Eles ainda afirmam que desde então a agenda de comunicações científicas abertas cresceu, sendo que agora novos termos como “código aberto” e “dados abertos” foram incluídas nela.

Um dos conceitos atribuídos a palavra “dados⁴” é a tradução do Latim da palavra “datum” significando “coisas que podem ser dadas”, conforme explicação feita por Buckland (1991). O autor afirma que dados são classificados como informação-como-coisa por serem considerados informativos e terem a característica de poderem ser processados para um uso posterior.

⁴ O termo “dado” foi conceituado de formas diferentes por diferentes autores, sendo uma variante de definição o seu contexto e a atribuição do seu uso. Ao longo dessa dissertação, serão expostos alguns desses conceitos.

Pampel et al. (2013) aprofundam o entendimento do que sejam dados os relacionando a pesquisas. Para os autores o termo “dados de pesquisa” “é definido como dados digitais sendo uma parte (descritiva) ou o resultado de um processo de pesquisa.

Das pesquisas desenvolvidas pelas comunidades científicas resultam uma imensa quantidade de dados, que se constituem elementos básicos da construção de um conhecimento. Klump et. al. (2006) declaram que o conhecimento científico “é o último passo em um processo originado de dados científicos primários. Estes dados são analisados, sintetizados, interpretados e o resultado deste processo é publicado como um artigo científico”.

Os dados de pesquisa são obtidos por meio das atividades de investigação científica ou reutilizados no contexto científico, mesmo que tenham sido gerados com outro propósito que não o científico, como por exemplo, o administrativo.

Os eles encontram no movimento pela Ciência Aberta um cenário propício para o seu compartilhamento e reuso. Cardoso, Jacobertty e Duarte (2012) acreditam que, como um movimento social a Ciência Aberta torna a sociedade mais próxima das práticas científicas ao diminuir as restrições de acesso aos dados científicos, as ferramentas de pesquisas e as publicações dos resultados dessas pesquisas.

A disponibilização aberta desses dados necessita de curadorias que estimulem o seu compartilhamento e reuso por outros pesquisadores. A curadoria de dados de pesquisa é melhor realizada em sistemas de informação, que permitem a representação, a organização e a recuperação de informações processadas e possibilitam o compartilhamento dos dados coletados.

Os autores Kim e Stanton (2013) afirmam que o compartilhamento de dados se tornou essencial na maioria das atividades de pesquisa modernas. Este fato pode ser explicado pelo que foi exposto por Tenopir et al. (2011). Segundo eles, na medida em que a ciência se torna mais colaborativa, e mais informações são comunicadas, o compartilhamento de dados de pesquisa se torna mais importante, sendo que o seu uso e reuso possuem muitas vantagens, tais como a possibilidade de verificação dos resultados, as

diferentes interpretações e abordagens aos dados e a proteção contra condutas relacionadas à falsificação de resultados.

Entretanto, a prática de compartilhamento de dados abertos não se desenvolve de forma linear em todas as áreas do conhecimento, tampouco com hegemonia em todas as regiões geográficas do mundo, em especial na América Latina. As atividades relacionadas com o compartilhamento de dados ainda aparentam estar restritas a poucas áreas do conhecimento, de acordo com Borgman (2012).

Tenopir et al. (2011) possuem a mesma visão expressada por aquela autora (2012) já que, para eles o montante de dados que são compartilhados ou restritos parece variar de acordo com a área de atuação dos pesquisadores. Os autores ainda explicam que algumas disciplinas e subdisciplinas já possuem culturas próprias no que tange ao compartilhamento de dados, sendo que algumas áreas do conhecimento já expressam mais significativamente essa cultura, como, por exemplo, as áreas de Geofísica, Biodiversidade e Astronomia.

Pontika et al. (2015) demonstram partilhar da mesma constatação feita pelos autores ao afirmarem que:

[...] distribuir dados de pesquisa abertamente ainda não é um comportamento comum e as práticas diferem; por exemplo, embora haja uma grande aceitação em compartilhar abertamente os dados em astronomia e genômica, os métodos de compartilhamento variam entre os pesquisadores, mesmo nos mesmos campos de assunto. (PONTIKA et al., 2015, p. 2)

A variação no nível de compartilhamento de dados científicos entre as áreas do conhecimento, se junta à outras distinções encontradas ao se analisar as áreas científicas, como as já identificadas por Mueller (2005), e Meadows (1999) no que tange ao canal preferencial para publicação científica dos pesquisadores, o que já determina certas necessidades diferenciadas quanto a disseminação e preservação da informação nessas áreas.

A preservação de dados de pesquisa é tão importante quanto o seu compartilhamento, pois ambos são vitais para o desenvolvimento e manutenção da ciência. A perspectiva da preservação e organização dos dados torna-se particularmente relevante quando se trata de um grande volume

de dados, onde passa a ser necessária a capacidade de representação e recuperação da informação.

Uma das soluções que tem sido adotada pela comunidade científica quanto ao compartilhamento de dados de pesquisa é o depósito e manutenção desses dados em repositórios digitais de acesso aberto. Estes repositórios digitais não só permitem a preservação dos dados, mas também ampliam a visibilidade e permitem assegurar a integridade da pesquisa realizada.

No mundo, o diretório de repositórios de dados, Registry Of Research Data Repositories⁵ (Re3data), registra a existência de 2933 repositórios⁶, sendo que do total listados no diretório, apenas 31 fazem parte da chamada América Latina, o que revela um valor desproporcional ao restante do mundo.

Entretanto, antes de compreender o atual cenário de compartilhamento de dados é necessário considerar o histórico cultural científico de tais áreas e regiões, principalmente no que tange ao compartilhamento e disseminação de informações em acesso aberto, com a intenção de contextualizar suas particularidades científicas atuais.

Essa dissertação também visa complementar o estudo já realizado por Costa (2014) em relação aos depósitos de publicações científicas em acesso aberto na América Latina em repositórios institucionais, já que, conforme explica Sayão e Sales (2016), com a amplitude das exigências relacionadas à Ciência Aberta:

as instituições de pesquisa as instituições de pesquisa começam a expandir os seus serviços de repositório digital – que até então estavam voltados preferencialmente para publicações digitais (eprints) - para o domínio heterogêneo e complexo dos dados de pesquisa; como desdobramento desse fato, começam a planejar e implantar sistemas de repositório de dados de pesquisa nas mais diversas configurações, plataformas tecnológicas e modelos de gestão. (SAYÃO e SALES, 2016, p. 93)

Diante do exposto, esta dissertação analisa o compartilhamento de dados de pesquisa por meio dos repositórios digitais caracterizados como de acesso aberto com delimitação para a área geográfica da América Latina, levando em consideração as diferenças que são encontradas entre as áreas do

⁵ <https://www.re3data.org/>

⁶ Dados coletados em 17 fev. 2019

conhecimento. O objetivo principal deste trabalho consiste em identificar o cenário atual do compartilhamento de dados de pesquisa, visando responder à seguinte pergunta: Quais as características e contribuições dos repositórios de dados de pesquisa científicas da América Latina para o desenvolvimento da ciência aberta?

1.1.1. Objetivos

1.1.1.1. Objetivo Geral

Identificar as características e contribuições dos repositórios de dados de pesquisa científicas da América Latina para o desenvolvimento da Ciência Aberta.

1.1.1.2. Objetivos específicos

São quatro objetivos específicos, a saber:

1. Identificar, com base na literatura, princípios e definições para Ciência Aberta e repositórios de dados.
2. Mapear os Repositórios de Dados Científicos da América Latina.
3. Caracterizar os Repositórios de dados de Científicos da América Latina de acordo com as dimensões apontadas pela literatura da área.
4. Identificar a organização e representação dos dados de pesquisa nos repositórios digitais de dados científicos.

1.2. JUSTIFICATIVA

Essa dissertação de mestrado tem o propósito de retratar qual é o atual cenário, bem como as principais características infraestruturais⁷ dos repositórios de dados científicos da América Latina, expondo assim suas principais características em relação as práticas de compartilhamento de dados alinhadas com os preceitos da Ciência Aberta e diferenças culturais.

A Biblioteconomia tem que lidar com muitos domínios e, ao lidar com áreas diferentes deve-se considerar a visão analítica de cada domínio, não podendo tratar todos as áreas como se fossem fundamentalmente semelhantes, conforme explica Hjørland (2002, p. 422). Para o autor, os profissionais de informação devem identificar os recursos de informação, descrevê-los, organizá-los e os comunicar de forma a atender os objetivos específicos de cada domínio.

Vilarinho (2015, p. 20) parte do pressuposto que as “diferentes comunidades científicas têm hábitos de informação e publicação que variam conforme a sua natureza e que o conhecimento de tais hábitos permite uma melhor caracterização de cada disciplina ou comunidade.”. A autora também acredita que o profissional bibliotecário possui a função social de intermediador da informação e quem a procura e, por essa razão, entender as especificidades de cada área do conhecimento é fundamental para que os bibliotecários desempenhem o seu papel de forma eficaz.

Se relacionando ao compartilhamento de dados científicos, objeto central da pesquisa proposta nessa dissertação, Sayão e Sales (2015) além de reconhecerem a heterogeneidade dos tipos de dados, alertam para a necessidade de formular estratégias para a gestão de amplo espectro que englobe os vários tipos de dados. Ainda de afirmam que:

O reconhecimento dessas diferenças torna-se crucial quando se estabelecem as opções gerenciais e tecnológicas para o arquivamento persistente e para a curadoria digital das coleções de dados de pesquisa. (SAYÃO e SALES, 2015).

A análise dos repositórios de dados científicos, responsáveis pela disseminação, arquivamento e curadoria dos dados sob seu domínio, que está

⁷ O termo “infraestrutura” nessa dissertação é adotado com a significação do conjunto de elementos que possibilitam e suportam o repositório de dados científicos, entre esses elementos podem ser entendidos como software, esquema de metadados e estrutura organizacional das comunidades e coleções.

sendo proposta por esse estudo, espera proporcionar uma visão do atual cenário de compartilhamento de dados científicos em repositórios no âmbito da América Latina em diferentes áreas do conhecimento.

Essa ação irá permitir as macro e micro-visões dos cosmos dos países da América Latina em relação ao compartilhamento de dados de pesquisas científicas de forma aberta, apontando dessa forma, quais são as contribuições que os repositórios de dados científicos latino americanos fornecem à Ciência Aberta.

Para a realização do estudo proposto nessa dissertação é importante compreender que o atual cenário de cooperação científica mundial criou um novo paradigma causado pelas novas práticas de acesso livre à informação iniciado, segundo Andrade (2015, p. 270), com base nos termos da Declaração de Budapeste (2002) e na Declaração de Berlim (2003): a Ciência Aberta. Essas novas práticas alteraram a forma como as disciplinas acadêmicas realizam as suas pesquisas, compartilham os seus produtos científicos e principalmente, a destinação dos dados que embasaram esse produto.

Sendo assim, se torna importante realizar uma investigação sobre o atual cenário de práticas científicas acadêmicas. Meadows (2000, p. 23) informa que existem duas maneiras básicas de se estudar as atividades que ocorrem em uma comunidade científica. A primeira seria avaliar como as atividades de informação mudaram ao longo do tempo, sua sequência evolutiva, sendo que nesses casos se caracterizam como estudos “longitudinais”. A outra maneira seria fornecer as relações de atividades que estão acontecendo naquele momento, como uma fotografia, nesses casos as pesquisas se caracterizam como estudos “latitudinais”. O autor ainda afirma que quando a evolução dos fenômenos se dá de uma forma lenta, é possível realizar a combinação de vários estudos latitudinais realizados em épocas diferentes.

Com base nas observações de Meadows (2000, p.23) acima citadas nota-se que essa pesquisa de mestrado possui a intenção de retratar, como uma fotografia, o cenário de compartilhamento de dados de pesquisa na América Latina, para que no futuro, possa ser uma peça que possibilite uma análise macro da evolução das práticas de compartilhamento em diferentes áreas do conhecimento por outros pesquisadores.

Sendo assim, esse estudo de cunho latitudinal possui dupla justificativa de acordo com sua essência: uma com natureza teórica, a outra de natureza prática.

1.2.1. Natureza teórica

As atividades de pesquisa estão cada vez mais dependentes de materiais digitais e, para que haja a evolução do conhecimento sem a duplicação de esforços é necessário que sejam adotadas estratégias que, em longo prazo, garantam que os dados que fundamentam as pesquisas realizadas no presente possam ser acessados, interpretados e reutilizados no futuro de forma aberta, de acordo com o que relatam Sayão e Sales (2012).

Sendo assim, em uma perspectiva teórica, o estudo dos repositórios de dados de pesquisa latino americanos representa para a Ciência da Informação um elemento que retratará com maior precisão a difusão do modelo e das práticas da Ciência Aberta pelos repositórios digitais no tocante ao compartilhamento de dados na região.

Portanto, visto que o armazenamento e a preservação dos dados de pesquisas se tornam essenciais para o progresso científico, a investigação do alinhamento das práticas de compartilhamento de dados com os preceitos estabelecidos pela Ciência Aberta, além do estabelecimento de critérios que respeitem as individualidades disciplinares na criação de ferramentas que possibilitem o arquivamento e acesso a esses dados, é imperativo.

1.2.2. Natureza prática

A construção de repositórios digitais de acesso aberto foi solução encontrada pelos gestores para a disponibilização de trabalhos através de sistemas interoperáveis que poderiam ser acessados com maior facilidade em um âmbito mundial pelo usuário através da internet.

Especificamente, os repositórios de dados de pesquisa ainda não são amplamente adotados pelas comunidades científicas quando comparados com o uso dos repositórios destinados a disseminação e produtos científicos pela mesma comunidade. Entretanto, a literatura já aponta a sua importância ímpar para um desenvolvimento científico acelerado, além de permitir o “escrutínio da lógica do argumento e replicação de observações ou experiências ou a sua refutação, garantindo que o conhecimento científico é cumulativo”, de acordo com o que afirma Boulton (2013, p. 239).

Sendo assim, esta pesquisa científica justifica a sua realização de forma prática por meio da investigação representação e organização da informação de dados de pesquisa com a finalidade maior de descrever o atual cenário de compartilhamento desses dados na América Latina, apresentando assim, o atual estado da arte do assunto na região de acordo com os pilares que caracterizam a Ciência Aberta.

A apuração de concretização da pesquisa possibilitará a visão específica das características do compartilhamento de forma aberta em repositórios digitais na América Latina. Essa compreensão possibilitará ações específicas por parte dos gestores de repositórios de dados que visem à criação de condições mais favoráveis as atividades ligadas à disseminação e a coleta de trabalhos que são o alvo do estudo de suas áreas.

2. FUNDAMENTAÇÃO TEÓRICA

2.1. COMUNICAÇÃO CIENTÍFICA

Comunicar-se cientificamente denota seguir uma série de protocolos para que se haja, de fato, o compartilhamento de determinadas informações ou conhecimentos que foram descobertos ao longo do processo de investigação estabelecido pelos pesquisadores. Não cabe a ciência indicar a evidência da verdade absoluta, mas sim, o saber científico que direciona a interpretação e compreensão dos processos que nos cercam nos determinados contextos que estão inseridos.

Etimologicamente o termo “ciência” deriva do latim significando “aprender” ou “conhecer”, segundo definição atribuída por Trujillo Ferrari (1982, p. 1). O autor aprofunda sua reflexão sobre o real significado das atividades que envolvem o verdadeiro objetivo da ciência concluindo que a “ciência é todo um conjunto de atitudes e de atividades racionais, dirigido ao sistemático conhecimento com o objetivo limitado, capaz de ser submetido a verificação” observando fenômenos que são encontrados no mundo natural.

De acordo com Kerlinger (1979, p. 1) a finalidade da ciência é chegar a uma teoria, descobrir explicações válidas para fenômenos naturais. Conforme o autor explana, a ciência se desenvolveu a partir da necessidade de um método de conhecimento que fosse mais seguro e válido para conhecer o mundo do que, por exemplo, os derivados da autoridade e os observacionais, pois eles nem sempre são dignos de confiança.

A palavra ciência possui uma ampla variedade de significação, de acordo com o que Melton (1974, p.38) apontou:

É usada geralmente para indicar: 1) um conjunto de métodos característicos por meio dos quais os conhecimentos são comprovados; 2) um acervo de conhecimentos acumulados, provenientes da aplicação desses métodos; 3) um conjunto de valores e costumes culturais que governam as atividades chamadas científicas; ou 4) qualquer combinação dos itens anteriores. (MELTON, 1974, p.38)

Seguindo a corrente de pensamentos do mesmo autor (1974, p.39) obtém-se que a ciência ainda pode ser entendida como uma instituição que não possui apenas métodos, mas também costumes que a circundam.

Bourdieu (1983, p. 89) percebe esse mesmo conceito, porém o atribui a denominação de “*habitus*”, e o resume como um ofício, um conjunto de técnicas e referências dos cientistas de determinados campos.

Ancorada nos ensinamentos de Bourdieu, Garcia (1996, p.65) elucida que a noção de “*habitus*” desenvolvida pelo autor também se refere a “um sistema de disposições duráveis e socialmente constituídas que, orientam e dão significado às ações e representações” dos cientistas durante o processo de produção do conhecimento científico.

Em relação ao conhecimento científico gerado pelas práticas de investigação, um dos componentes indissociáveis da sua natureza é o seu elemento operacional. Ele atribui a confiabilidade necessária para que os resultados obtidos se tornem válidos de fato. A confiabilidade é uma das características mais importantes da ciência, sendo que para que ela seja obtida, é necessário que a pesquisa tenha seguido uma rigorosa metodologia científica e seus resultados possam ser avaliados por outros cientistas, conforme entende Mueller (2007, p. 21).

A ciência, por ser mais que um conhecimento pessoal, consiste naquilo que pode ser comunicado de uma pessoa a outra, devendo ser capaz de se expressar em uma linguagem universal e inequívoca, segundo destacou Ziman (1996, p. 25). As autoras Chan, Okune e Sambuli (2015, p. 93) manifestam pensamento complementar ao afirmarem que a característica fundamental da ciência é tornar público os resultados das pesquisas científicas para permitir futuras construções do conhecimento, já que o progresso científico depende do acesso a contribuições anteriores.

Costa (2000, p. 87) ainda acrescenta que “pesquisadores de qualquer área do conhecimento precisam comunicar os seus resultados por meio de um sistema de comunicação e no mais breve espaço de tempo possível”. O que significa que os pesquisadores devem se utilizar do sistema de comunicação mais apto que a sua área científica oferece para que os resultados de suas pesquisas atinjam a sua comunidade acadêmica, demais pares e interessados.

Sobre a importância da comunicação científica, Puschmann (2014) afirma que:

A comunicação bem-sucedida é essencial para a erudição porque permite que o conhecimento científico se prolifere, permita aplicações práticas e torne-se um conhecimento social enraizado, mas também porque frequentemente os resultados da pesquisa científica têm implicações sociais abrangentes e são altamente controversos (por exemplo, pesquisa climática, nuclear energia, genética). (PUSCHAMANN, 2014, p. 91, tradução nossa).

Tanto a exposição do trabalho intelectual dos pesquisadores quanto a obtenção dos resultados alcançados por outros cientistas passam por um sistema de comunicação que compreende os canais formais e informais de comunicação, de acordo com a explanação feita por Mueller (2007, p. 22).

Meadows (1999, p. 7) orienta que a distinção entre os canais informal e formal de informação é que o primeiro se dedica a comunicação que, em geral, é “efêmera, sendo posta à disposição apenas de um público limitado” e a segunda “encontra-se disponível por longos períodos de tempo para um público amplo”.

Sobre o mesmo assunto, Rosa e Gomes (2010, p. 20) elucidam que entre os elementos que compõe o canal informal de publicação estão as conversas, os telefonemas, palestras e outras discussões interpessoais. Já Russel (2001, p. 271) acredita que os canais informais se diferenciam dos formais pela sua possibilidade de permitir uma interação imediata entre os transmissores e receptores da informação.

Segundo o que informa Bonetta (2007), canais informais de comunicação científica, como os blogs, “fornecem um fórum de discussão on-line para assuntos de interesse atual e são atualizados regularmente com novos artigos curtos sobre os quais os leitores podem comentar”. A autora ainda argumenta que essas ferramentas científicas propiciam opiniões confiáveis com o objetivo de envolver os demais cientistas em discussões sobre a literatura científica ou a política científica. Essas opiniões, ainda conforme a autora, fornecem, muitas vezes, um contexto mais rico que os próprios artigos de notícias.

Os canais informais de publicação oferecem um ambiente relativamente irrestrito para a troca e informações, conforme destaca Eason et al. (1997). Na visão dos mesmos autores, os canais formais possuem convenções e procedimentos que, por muitas vezes se tornam altamente restritivos.

Os elementos que compõem o canal formal são as comunicações impressas, dentre eles o periódico científico, o qual Mueller (2007, p. 23) afirma ser um importante veículo de divulgação para a ciência, sustentando a ideia de Barbalho (2005, p. 2) que assegura o periódico científico como um agente que “desempenha papel fundamental no processo da comunicação científica por se constituir na principal via de veiculação de novos conhecimentos e possibilitar a sua difusão para determinado público”. Bartling e Friesike (2014, p.7) ainda argumentam que “a publicação de artigos em revistas científicas se tornou o bloco de construção fundamental da ciência moderna”.

Sobre a importância do periódico científico para as áreas do conhecimento, Baptista et al. (2007) afirmam que:

O periódico científico tem sido considerado, para as ciências exatas e naturais, assim como para parte significativa das ciências humanas e sociais e para parcela menor das artes e humanidades, o veículo mais importante de comunicação da pesquisa. Representa, assim, um dos veículos que compõem o produto final, formal, consolidado da disseminação de resultados de pesquisas realizadas por estudiosos de todo o mundo. (BAPTISTA et al. 2007).

Derivado do sistema de troca de informações por cartas, o surgimento desses periódicos científicos se deu devido a várias razões: “Algumas eram específicas (como a expectativa de seus editores de que teriam lucro); algumas, gerais (como a crença de que para fazer novos descobrimentos era preciso que houvesse um debate coletivo) ”, segundo afirmação oriunda de Meadows (1999, 7).

2.2. MOVIMENTO DE ACESSO ABERTO A INFORMAÇÃO CIENTÍFICA

As “expectativas financeiras” sobre a criação de periódicos científicos se desenvolveram exponencialmente desde o século XVII, fazendo com que o alto custo relacionado à comercialização de periódicos se tornasse desproporcional ao orçamento das instituições, que haviam de realizar um duplo pagamento sobre as pesquisas científicas: um relacionado à sua elaboração e o outro relacionado ao seu acesso posterior a publicação nessas revistas.

No entanto, em 1974 Merton já possuía a opinião que, como um dos imperativos institucionais da ciência, ela deva possuir uma propriedade comum de seus bens. Isso se deve ao fato que, segundo o autor, a ciência já em sua origem é um bem comum por ser produto da colaboração social e, sendo assim, seus frutos deveriam ser destinados à comunidade, haja vista que suas descobertas devam ser uma herança comum e os lucros do produtor devam ser limitados, ou até mesmo reduzidos ao mínimo.

Canessa e Zennaro (2008) explicam que desde o início as publicações científicas no formato de artigos não são feitas com o intuito de retorno monetário aos cientistas, e sim com a finalidade da permuta de benefícios intangíveis. Como exemplo, os autores citam que os cientistas almejavam “deixar sua marca no tempo, algo que lhes daria prioridade sobre outros cientistas que trabalham na problemática, bem como prestígio, compromissos e o impacto que impulsiona suas carreiras”.

Ao desvendar a noção e o funcionamento do campo científico⁸, Bourdieu (1983) explica que a representação do capital ou poder simbólico dentro deste campo se dá por meio da autoridade, legitimidade ou competência científica que fornecem aos pesquisadores a propriedade de falar e agir em nome da

⁸ Bourdieu (1983) define campo científico como um “sistema de relações objetivas entre posições adquiridas (em lutas anteriores), é o lugar, o espaço de jogo de uma luta concorrencial. O que está em jogo especificamente nessa luta é o monopólio da autoridade científica definida, de maneira inseparável, como capacidade técnica e poder social; ou, se quisermos, o monopólio da competência científica, compreendida enquanto capacidade de falar e de agir legitimamente (isto é, de maneira autorizada e com autoridade), que é socialmente outorgada a um agente determinado.”.

ciência, o que os diferencia e os compreende como elementos além dos recursos materiais estabelecidos como parâmetro para os demais campos.

É opinião de Melton (1974) que “o direito do cientista à sua propriedade intelectual limita-se à gratidão e à estima que, se a instituição funciona com o mínimo de eficácia, são mais ou menos proporcionais aos aumentos trazidos ao fundo de conhecimento”.

Os pesquisadores, como autores, não estão, na maioria das vezes, conscientes dos preços monetários pagos pelo acesso ao material científico que lhes é oferecido, pois as assinaturas desses materiais são realizadas por meio do orçamento das bibliotecas, o que faz com que eles não cobrem das editoras para a publicação dos artigos com suas pesquisas, conforme evidenciam Terry e Kiley (2006).

Terry (2005) informa que a Wellcome Trust, uma instituição de pesquisa biomédica, encomendou um estudo sobre a economia do setor editorial científico. Nesse estudo foi descrito que, em termos econômicos, o setor tradicional de publicações de pesquisas científicas é deficiente, visto que os pesquisadores, como produtores da informação, cedem os direitos autorais de seus trabalhos gratuitamente aos editores, que “vendem para bibliotecas a preços que variam de suficiente para cobrir seus custos até alguns lucros bastante altos - alguns acima de 30%”.

Suber (2012) demonstra compartilhar da mesma linha de argumentos que Terry e Kiley (2006) ao declarar que os autores de artigos científicos não são pagos pelas publicações de seus trabalhos, o que os distingue dos músicos, cineastas e da maior parte de outros autores. Segundo seu raciocínio, a motivação dos pesquisadores é “cultivar a maior audiência e impacto”, e de certo, não é tornar os editores ainda mais ricos, o que reafirma a coerência dos pensamentos traçados por Merton em 1974 e Bourdieu em 1983.

Conforme explica Terry (2005), esse esquema economicamente deficiente de publicação editorial tem como uma das consequências da sua desigualdade o monopólio de poder dos editores que controlam os preços desses periódicos onde eram publicados os artigos científicos. O custo de manutenção da assinatura desses periódicos cresceu de forma muito superior ao orçamento das bibliotecas, fazendo com que o número de títulos

disponibilizados realmente descaísse, de acordo com que Terry e Kiley (2006) informam.

O pensamento de Russel (2008, p. 272) corrobora essa informação ao ser sustentado que “as bibliotecas acadêmicas foram forçadas a diminuir a quantidade de assinaturas de periódicos, mesmo que a produção de informações acadêmicas cresça exponencialmente”. O autor estabelece que essa é a consequência da diminuição do orçamento aliado ao alto custo das revistas especializadas.

Sarmiento et al. (2005) defendem que a especialização cada vez mais crescente das disciplinas acadêmicas fez com que as universidades de institutos de pesquisas se vissem obrigados a adquirir uma quantidade mais vasta de “artefatos de comunicação científica formal, em particular dos periódicos.”

Como, em sua maioria, as pesquisas científicas são fruto do financiamento público do Estado, do ponto de vista ético essas pesquisas deveriam ser de livres para aqueles que desejassem ter acesso aos seus conteúdos, o que de fato não acontece, fazendo com que o cidadão tenha que pagar para acessar aquilo que em teoria já está pago, conforme relata Kuramoto (2006).

Sarmiento et al. (2005) sustentam que esse fenômeno se trata da mercantização da informação pelos grandes editores:

[...] os pesquisadores entregam gratuitamente os resultados do seu trabalho, financiados com verbas das universidades ou de agências de financiamento, sendo que os mesmos resultados são posteriormente vendidos para as universidades a preços elevados. (SARMENTO et al., 2005.)

Sobre o assunto, Rabelo (2012) expressa que as editoras “não são mais necessárias e nos dão mais dor de cabeça do que ajudam”. O autor expõe que elas agem mais como verdadeiras “inimigas da ciência” e por isso deveriam sofrer um processo de boicote pelos pesquisadores científicos. Para ele “a melhor estratégia em curto prazo para enfraquecer as editoras é não submeter artigos a revistas publicadas por elas.”. Entretanto, Barros (2012, p. 366) relembra que “a carreira dos acadêmicos depende da publicação em periódicos de maior fator de impacto possível” e que grande parte dos periódicos de

acesso aberto têm seus fatores de impacto muito abaixo dos títulos já estabelecidos.

No entanto, Barros (2012, p. 367) também tem a opinião que o acesso restrito que é imposto pelas editoras comerciais científicas atrapalha outras descobertas científicas, além de impor barreiras para o público.

Os editores científicos justificam que o preço elevado se dá pelo processo de manutenção e revisão por pares de alta qualidade. Também é dado como argumento que os lucros das editoras possibilitam a inovação no sistema de publicação. Se junta a essa alegação a afirmativa de que é por meio da taxação desses lucros que o Estado consegue arrecadar volumes financeiros que posteriormente servirão como incentivo à pesquisa, conforme especifica Barros (2012).

Ao abordar o boicote feito a Elsevier como forma de protesto aos altos preços cobrados associados as pressões legais feitas pela editora em favor do acesso restrito e pago às publicações científicas, Fischmann (2012) publica a explicação de Alice Wise (diretora de Acesso Universal da Elsevier). A diretora argumenta que a editora melhora o acesso aos periódicos em vez de o impedir, como exemplo cita que “os downloads da Internet de alguns periódicos aumentaram em até 40% quando a Elsevier os adicionou às coleções vendidas a bibliotecas”:

“No ano passado, nossos preços foram mais baixos do que os nossos concorrentes. Não sei por que estamos no foco desse boicote, mas estou muito preocupada com um cientista insatisfeito, e estou preocupada com 2000.” (Alice Wise em entrevista dada a FISCHMANN, 2012, [s.p], tradução nossa).

Nessa entrevista, Fischmann (2012) também reproduz que Alice Wise contrapõe que a empresa disponibiliza para as bibliotecas uma ampla variedade de pacotes além de negociar descontos baseados no tamanho, tipo e padrões de uso da instituição. Wise sustenta que os aumentos nos preços das assinaturas se concentraram nos anos 80 e 90 se estabilizando desde então. Essa estabilidade, ainda segundo a diretora da Elsevier, se deu devido a conscientização da prática errônea da editora. Wise assegura que “melhoramos e nos tornamos bons cidadãos”.

Entretanto, Barros (2012) reargumenta que esse processo está “aquém das expectativas, simplesmente reunindo artigos em periódicos e empurrando-os para as bibliotecas apertadas em orçamentos”.

É importante salientar que antes mesmo do cenário de boicote promovido a editora Elsevier, já existia um panorama onde a comercialização da produção acadêmica impedia a fluida circulação da informação, o que pressionou a comunidade científica a procurar meios alternativos ao periódico impresso para disseminar os resultados de suas pesquisas científicas. Essa circunstância acaba por impulsionar mudanças cada vez mais significativas no sistema de comunicação científica em direção ao acesso aberto onde as tecnologias de informação e comunicação, como computadores e redes de pesquisas, são utilizadas como ferramentas chave para a comunicação da informação científica.

Castells (2010, p. 67) chama a atenção para um momento histórico em que um novo paradigma tecnológico se organizou em torno da tecnologia da informação. Segundo o que afirma o autor:

A tecnologia da informação é para essa revolução o que as novas fontes de energia foram para as revoluções industriais sucessivas, do vapor à eletricidade, aos combustíveis fósseis e até mesmo à energia nuclear, visto que a geração e distribuição de energia foi o elemento principal na base da sociedade industrial. (CASTELLS, 2010, p. 67).

Russel (2001, p. 271) afirma que as novas tecnologias de computação em rede causaram fortes mudanças nos sistemas de comunicação científica. Ele compara tais mudanças ao surgimento da ciência global, essa nova ciência conecta, por meio das novas tecnologias, diferentes pessoas, em diferentes pesquisas que estão em diferentes partes do mundo.

Tais mudanças foram mais elencadas e explicadas por Luce (2002) que explana que anteriormente as comunicações acadêmicas envolviam correspondências, fax, FTP e correios eletrônicos, o que requeria uma grande quantidade de tempo, materiais e dinheiro, enquanto as comunicações eletrônicas, na visão do autor, quase não envolvem custos de distribuição e zeram os custos relacionados à publicação.

Para Luce (2002) algumas tecnologias computacionais “fornecem uma maneira conveniente para os cientistas compartilharem seus resultados com

seus colegas muito rapidamente, na forma de artigos individuais”, o que corrobora a afirmação de Sarmiento et al. (2015) que elucidam que os novos meios de comunicação aceleraram mais o desenvolvimento da produção científica pois as atividades informatizadas aumentaram a produtividade científica permitindo também o crescimento da literatura científica.

A prática aberta que foi desencadeada pelas mudanças tecnológicas, possui tanto suas raízes nos antigos valores da ciência quanto nos recentes processos econômicos e tecnológicos que abalaram a estrutura da ciência, como a ““crise dos periódicos”, que tomou as bibliotecas universitárias a partir do final dos anos 1980, e o advento da World Wide Web”, segundo Ortellado (2008, p. 187),.

A seção seguinte busca abordar quais foram as raízes do Acesso Aberto sobre dupla perspectiva: a internacional e a partir dos movimentos ocorridos no âmbito da América Latina.

2.2.1. A Emergência do Acesso Aberto

O gatilho que desencadeou a crise em que resultou o surgimento do Movimento de Acesso Aberto foi a impossibilidade de as bibliotecas universitárias americanas continuarem a adquirir os periódicos científicos para manterem as suas coleções em decorrência da alta na inflação dos preços de suas assinaturas, segundo orienta Mueller (20016, p. 31).

Canessa e Zennaro (2008, p. 15) também comentam o surgimento do acesso aberto. Eles argumentam que por mais de 300 anos o conhecimento científico circulou por meio de artigos de periódicos em edições que tinham seus custos cobertos pelas taxas de assinatura. Entretanto, para os autores, o surgimento da internet possibilitasse a distribuição gratuita desses artigos aos leitores.

Seguindo o raciocínio estabelecido por Canessa e Zennaro (2008, p. 15) sobre a crise de periódicos, é de interpretação dos autores que os preços das revistas científicas dispararam concomitantemente ao nascimento da internet.

O resultado é uma crise de acesso em que as instituições não podem cobrir os custos para acessar a ampla gama de publicações. Bibliotecários foram forçados a cancelar assinaturas e fazer cortes em seus orçamentos para comprar livros. Os cientistas reagiram procurando formas alternativas de compartilhar suas pesquisas. (CANESSA e ZENNARO, 2008, p. 15, tradução nossa).

Cardoso, Jacobertty e Duarte (2012) também relacionam o surgimento do acesso aberto com o advento da internet e com as novas tecnologias de informação e comunicação. Conforme os autores argumentam, esses elementos influenciaram a mutação no sistema tradicional de publicação científica, bem como no sistema de difusão de modo que as publicações fossem mais rápidas e com um custo reduzido, o que torna questionável o aumento nas taxas de assinaturas cobradas pelas grandes editoras científicas.

Este panorama histórico é analisado por Sarmiento et al. (2015) que expõem que mesmo com as facilidades criadas pelo aumento exponencial do número de cientistas e desenvolvimento das tecnologias de informação, houve a crescente mercantilização da informação científica pelas grandes editoras, o que fez surgir o movimento de Acesso Livre.

Sarmiento et al. (2015) relacionam o Acesso Aberto com o aumento dos custos de acesso aos produtos da comunicação científica aliado ao sentimento de injustiça provocado “pela necessidade de as universidades e institutos de investigação comprarem de volta os resultados da sua própria investigação, ou a consciência de que esses resultados não estariam à disposição de todos os cientistas em igualdade de circunstâncias”.

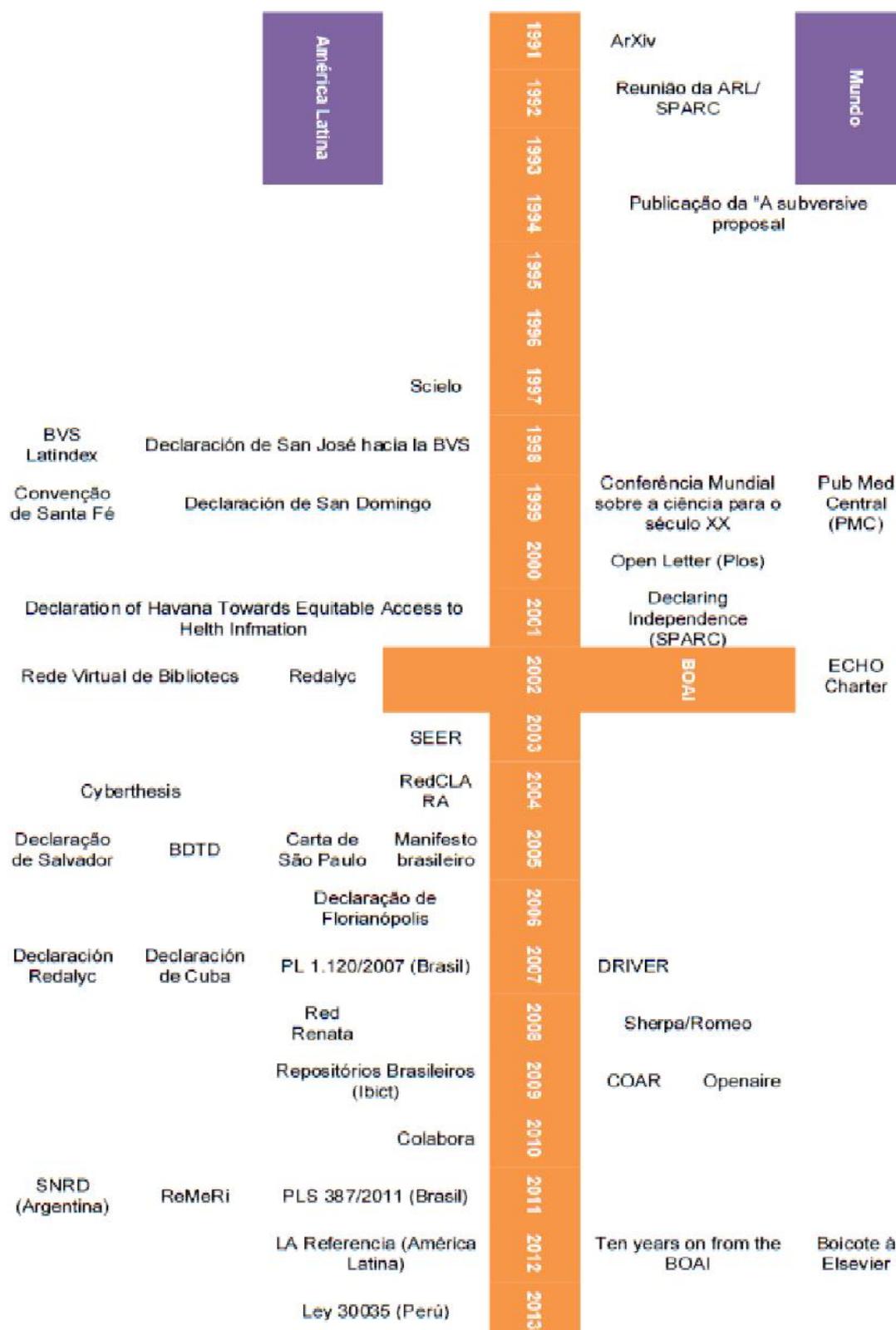
Se tratando da crise no acesso a informação científica, Terry e Kiley (2006) relembram que o custo de acesso aos periódicos estava muito superior ao orçamento das bibliotecas, o que fez com que esses centros de informação se vissem obrigados a cancelar as assinaturas. Como consequência, os autores informam que o número de pesquisadores com acesso a informação caiu.

Ainda seguindo a corrente de pensamento de Mueller (20016, p. 31), essa crise já vinha há muito sendo sentida em países como o Brasil, no entanto, foi detonada quando atingiu as universidades norte americanas. Isso fez com que ambos os seguimentos adotassem posturas e correntes complementares em seus próprios âmbitos a favor do acesso aberto a

comunicação científica, como apontado na pesquisa feita por Costa (2014), (figura 2), com um paralelo regional e mundial.

A constatação de Mueller (20016) expõe que quando países como os Estados Unidos começaram a sentir as inquietações promovidas pelos altos custos relacionados à manutenção das assinaturas dos periódicos, há muito elas já eram sentidas por países com economias menos fortes, como é o caso dos que se localizam na América Latina.

Figura 2: Principais iniciativas de acesso aberto no mundo e na América Latina



Fonte: Costa (2014, p. 59)

Em uma época aonde a concentração da propriedade dos periódicos científicos chegou a ser centralizada basicamente em três grandes indexadores (Reed Elsevier, Taylor and Francis e Springer) a internet tornou possível o

acesso massivo, por um baixo custo, ao conteúdo científico por meio digital, conforme demonstra Ortellado (2008, p. 188).

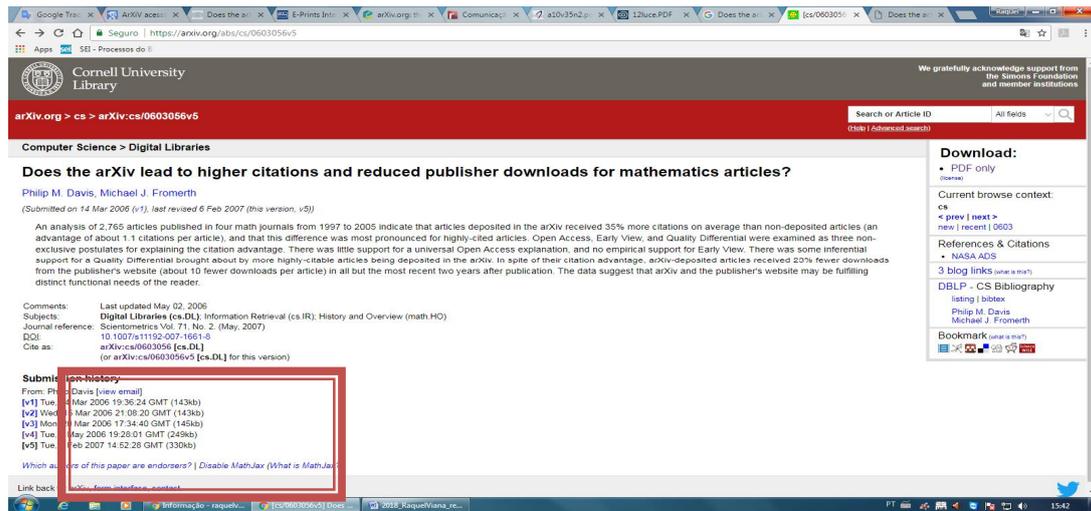
A internet possibilitou que diversas iniciativas em todo o mundo surgissem em prol do acesso aberto e irrestrito aos frutos das pesquisas científicas realizadas. Uma linha histórica desenvolvida pela Scielo aponta que o primeiro marco na trajetória do Acesso Aberto aconteceu em 1971 com a criação do Projeto Gutenberg por Michael Hart. Sendo considerada a primeira biblioteca digital do mundo, objetivo central do projeto era fazer a livre disponibilização de obras por meio da digitalização de textos em domínio público. No entanto, a literatura aponta o ArXiv⁹ como o primeiro lançamento bem sucedido na disseminação da informação por meio de pré-prints.

Segundo informações fornecidas pelo próprio ArXiv, ele é um servidor eletrônico automatizado de arquivo e distribuição para artigos de pesquisa em áreas como Física, Matemática, Ciência da computação, Biologia Estatística, Engenharia elétrica e Economia.

O princípio do ArXiv é funcionar como uma ferramenta onde os autores de artigos científicos possam disponibilizar os seus trabalhos para que os usuários possam ter acesso, desde os documentos mais antigos a versões mais atualizadas fornecendo o histórico de submissão dos trabalhos e contato direto com o autor, conforme indicado na figura 2, sem sequer pretender ter o processo de revisão por pares que é tradicional das revistas científicas. Na opinião de Luce (2002), essa ferramenta age como um repositório onde os cientistas podem promover o rápido acesso dos resultados das suas pesquisas à sua comunidade científica.

⁹ <https://arxiv.org/>

Figura 3: Arxiv.org



Fonte: Elaboração própria

Ainda segundo Luce (2002), o ArXiv foi o primeiro arquivo eletrônico que foi um sucesso nas áreas do conhecimento que abrange, mudando a forma como os cientistas se comunicavam.

O arquivo online funciona como um repositório de versões eletrônicas de documentos, fornecendo uma maneira conveniente para os cientistas compartilharem seus resultados com os colegas sem ter que esperar que o artigo apareça impresso. (LUCE, 2002).

A ferramenta forneceu meios abertos, universais e sem custo para o acesso a informação que era paga e restrita. Esse fator contrariou a expectativa negativa dos editores de periódicos pagos, ao apresentar índices de citações e leituras muito superiores aos dos periódicos de acesso restrito, porém já consagrados pelo meio acadêmico.

Kurtz et al. (2005) identificaram que os artigos da área de Astronomia que eram postados livremente na internet eram amplamente mais citados do que os que não eram. Então, se utilizando de dados da NASA Astrophysics Data System (ads.harvard.edu) e do ArXiv os pesquisadores buscaram investigar as razões desse fenômeno e chegaram aos seguintes possíveis postulados:

1. Como o acesso aos artigos é irrestrito por qualquer mecanismo de pagamento, os autores podem lê-los mais facilmente

e, assim, citam-nos com mais frequência; o postulado do *Open Access* (OA).¹⁰

2. Porque o artigo aparece mais cedo ganha primazia e tempo adicional na imprensa, e é assim mais citado; o postulado do *Early Access* (EA).

3. Autores preferencialmente tendem a promover (neste caso, postando na internet) os artigos mais importantes e, portanto, mais citáveis; o postulado da *Self-selection Bias* (SB). (Kurtz et al.,2005, tradução nossa).

Já tendo em foco o espaço geográfico Latino Americano, a primeira iniciativa em prol do acesso aberto a comunicação científica que a literatura aponta foi o Scielo ¹¹(Scientific Electronic Library Online), um projeto brasileiro desenvolvido pela Fundação de Apoio à Pesquisa do Estado de São Paulo (FAPESP) em parceria com a BIREME (Centro Latino-americano de Informação em Ciências da Saúde). Esse projeto foi iniciado em 1997 e lançado oficialmente em 1998 com o objetivo de, segundo o próprio Scielo, desenvolver “uma metodologia comum para a preparação, armazenamento, disseminação e avaliação da produção científica em formato eletrônico.”

O Scielo, por ter sido lançado quatro anos antes Declaração de Budapeste, a principal declaração sobre o Acesso aberto até o presente momento, foi considerado um pioneiro no tema dando uma força importante no movimento internacional, conforme Packer e Meneghini (2014).

Segundo afirma Greene (2000) o desenvolvimento do Scielo ocorreu com base nas demandas de muitas revistas brasileiras que enfrentavam o problema da falta de visibilidade e de acessibilidade aos textos científicos já que, em sua maioria, as revistas brasileiras não estavam indexadas nas bases de dados internacionais. Ainda segundo o autor, “mesmo aquelas que estão incluídas em bancos de dados têm enfrentado sérios problemas na distribuição de resumos e separatas em nível mundial.”

Packer e Meneghini (2014, p. 20) explicam que o fato da maioria dos periódicos latino americanos não estarem indexados internacionalmente os faziam não ter visibilidade ou serem reconhecidos por sua qualidade. Os autores informam que:

¹⁰ Kurtz et al acabam por concluir no final de suas pesquisas que o Postulado do Open Access não exerce influência no campo da Astronomia.

¹¹ <http://www.scielo.org/php/index.php>

Apenas 14 periódicos brasileiros, de diferentes disciplinas, foram indexados internacionalmente em 1997 na base de dados do ISI (atualmente Thomson & Reuters Web of Science), enquanto dezenas de outros periódicos eram publicados, com pequenas tiragens, geralmente restritos a bibliotecas e membros de associações científica. Poucos conseguiam assinaturas suficientes para cobrir uma parte significativa dos seus custos. (PACKER e MENEHINI, 2014, p. 20).

Funcionando em quatorze¹² países, entre eles Brasil, Chile, Costa Rica e Cuba, o Scielo melhora a visibilidade os periódicos científicos desses países, além de medir os indicadores de desempenho dos periódicos por meio do número de downloads, conforme orientam Packer e Meneghini (2014).

Conforme Greene (2000) compreende, o Scielo possui um modelo para sua existência, que consiste em três objetivos básicos:

o desenvolvimento de metodologia para editar, armazenar, criar Internet hyperlinks, publicar, divulgar e avaliar revistas científicas. O segundo componente é a aplicação da metodologia para operacionalizar os SciELO sites, que são coleções de revistas eletrônicas. O terceiro componente é o desenvolvimento da rede de SciELO sites através da promoção de parcerias e comunicações científicas – autores, editores científicos e técnicos, instituições, agências financiadoras, com conseqüente melhoramento da comunicação científica nos países em desenvolvimento. (GREENE, 2000).

O Scielo utiliza esse modelo, juntamente com as facilidades da internet, para “fornecer acesso aberto aos periódicos e facilitar a navegação e pesquisa em conteúdos de coleções, periódicos, fascículos e artigos.”, conforme afirmam Packer e Meneghini (2014).

Outro documento que ressaltou a necessidade de disponibilizar para todos os resultados das pesquisas científicas ocorreu em março 1999 na 1ª Reunião Regional de Consulta da América Latina e do Caribe com relação a Conferência mundial sobre a ciência para o século xxi: uma visão nova e uma base de ação, realizada na República Dominicana, a Declaração de Santo Domingo.

¹² Atualmente a Rede Scielo é composta por: África do Sul, Argentina, Bolívia, Brasil, Chile, Colômbia, Costa Rica, Cuba, Espanha, México, Peru, Portugal, Uruguai e Venezuela, sendo que o Equador e Paraguai se encontram em fase de desenvolvimento.

Essa declaração destaca a necessidade do aumento dos esforços em prol das atividades científicas e tecnológicas na região da América Latina e Caribe, já que, segundo a Declaração de Santo Domingo (1999), os benefícios oriundos do conhecimento científico produzido na região “não estão equitativamente distribuídos, aumentando a distância entre os países industrializados e os países em via de desenvolvimento.”.

Segundo destaca Costa (2017, p. 46) esse documento destaca a necessidade de uma agenda de pesquisa voltada para uma ciência que esteja disponível toda a sociedade, e, portanto, para que essa necessidade seja sanada é necessário o aumento das infraestruturas de tecnologias de informação e comunicação da ciência.

Segundo o que a própria Convenção afirma:

- A democratização da ciência permeia três grandes metas:
- (i) a ampliação do conjunto de seres humanos que se beneficiam diretamente dos avanços da pesquisa científica e tecnológica, que deve privilegiar os problemas da população afetada pela pobreza;
 - (ii) uma maior abertura no acesso à ciência, entendida como um componente central da cultura;
 - (iii) o controle social da ciência e da tecnologia, e da sua orientação, a partir o exercício de opções morais e políticas coletivas e explícitas. Isso enfatiza a importância da educação e da popularização da ciência e da tecnologia para o conjunto da sociedade. (DECLARAÇÃO DE SANTO DOMINGO, 1999)

Em outubro do mesmo ano ocorreu em Santa Fé, Novo México, uma convenção onde o ArXiv.org já havia sido consolidado como um veículo crucial na comunicação de pesquisas na área de Física, no entanto o número de servidores com a mesma finalidade de disponibilizar documentos eletrônicos era muito reduzido.

Segundo o que foi previsto por essa convenção, haveria a criação de muitas outras iniciativas como a de Paul Ginsparg nos próximos anos, o que abriria a demanda por orientações no sentido de estabelecer protocolos de interoperabilidade entre eles.

Prevemos a criação de muitos outros arquivos e-prints nos próximos anos. Esses arquivos serão distribuídos pela Internet; alguns serão orientados para determinadas disciplinas acadêmicas, outras

baseadas na afiliação institucional. (CONVENÇÃO DE SANTA FÉ, 1999).

Conforme afirmam Triska e Café (2001, p. 92), durante a convenção foi consolidado o conceito de arquivo aberto e a criação de princípios para a filosofia aberta: o auto-arquivamento, a revisão pela comunidade e a interoperabilidade.

Por arquivo aberto, a Convenção de Santa Fé elegeu duas características básicas:

Quadro 1: Características dos arquivos abertos

| |
|---|
| Um mecanismo para submissão; Um sistema de armazenamento de longo prazo. |
|---|

Fonte: Elaboração própria conforme orientações da Convenção de Santa Fé

Com o mesmo propósito de discutir a inevitabilidade da disseminação universal e democratização da informação e conhecimento, os participantes da Segunda Reunião de Coordenação Regional da Biblioteca Virtual em Saúde (BVS) e do V Congresso Regional de Informação em Ciências da Saúde (CRICS V) na América Latina e Caribe, se reuniram em Havana em 2001, que teve como resultado a Declaração de Havana que destacava o acesso desigual à informação e conhecimento em saúde.

Segundo Costa (2016, p. 5):

Em 2001, a Declaration of Havana Towards Equitable Access to Health Information destaca novamente a responsabilidade do Estado em relação ao acesso a informação. Nesta declaração os participantes de um encontro de coordenação da BVS definiram a informação científica da área como um bem público global e destacaram a necessidade de políticas nacionais e internacionais para garantir seu acesso e disseminação. Assim, reiteraram seu compromisso com o desenvolvimento da BVS. (COSTA, 2016, p. 5).

Em âmbito internacional, Sarmento et al. (2015) consideram a existência de três grandes declarações que exerceram grande influência sobre o Acesso Aberto: Budapest, Bethesda e Berlin.

A Budapest Open Access Initiative (BOAI), em 2002, se derivou da na reunião promovida pelo Open Society Institute (OSI), realizada em Budapeste, Hungria, em dezembro de 2001. Costa (2014) acredita que a BOAI dita que os pesquisadores devem tornar acessível gratuitamente online, sem esperar pagamento, o resultado de suas pesquisas, principalmente as que estão sob a forma de artigos publicados em periódicos, mas engloba também qualquer trabalho não publicado e não revisado que tenha a intenção de disponibilizar online para alertar colegas sobre importantes descobertas científicas.

Segundo foi definido em BOAI (2002), pode-se entender “acesso aberto” à literatura como a:

“disponibilidade gratuita na internet, permitindo a qualquer usuário a ler, baixar, copiar, distribuir, imprimir, buscar ou usar desta literatura com qualquer propósito legal, sem nenhuma barreira financeira, legal ou técnica que não o simples acesso à internet A única limitação quanto à reprodução e distribuição, e o único papel do copyright neste domínio sendo o controle por parte dos autores sobre a integridade de seu trabalho e o direito de ser propriamente reconhecido e citado. ” (BUDAPEST OPEN ACCESS INITIATIVE, 2002).

Suber (2012, p. 8) afirma que acesso aberto tem a ideia básica de “disponibilizar a literatura científica on-line, sem barreiras de preços e sem a maioria das barreiras de permissão.”¹³. Conforme afirmam Harnad et al. (2004), Morrison (2009), Freitas (2015, p. 1) e Andrade (2015, p. 270) a disponibilização dos documentos abertos pode ser feita de duas maneiras distintas, mas não excludentes: pela via dourada (revistas de acesso aberto) e/ou pela via verde (arquivos abertos).

A via dourada se refere a periódicos que, por si mesmo são de acesso aberto, por isso, os artigos a eles submetidos, tornam-se imediatamente de acesso aberto, de acordo com as afirmações feitas por Morrison (2009).

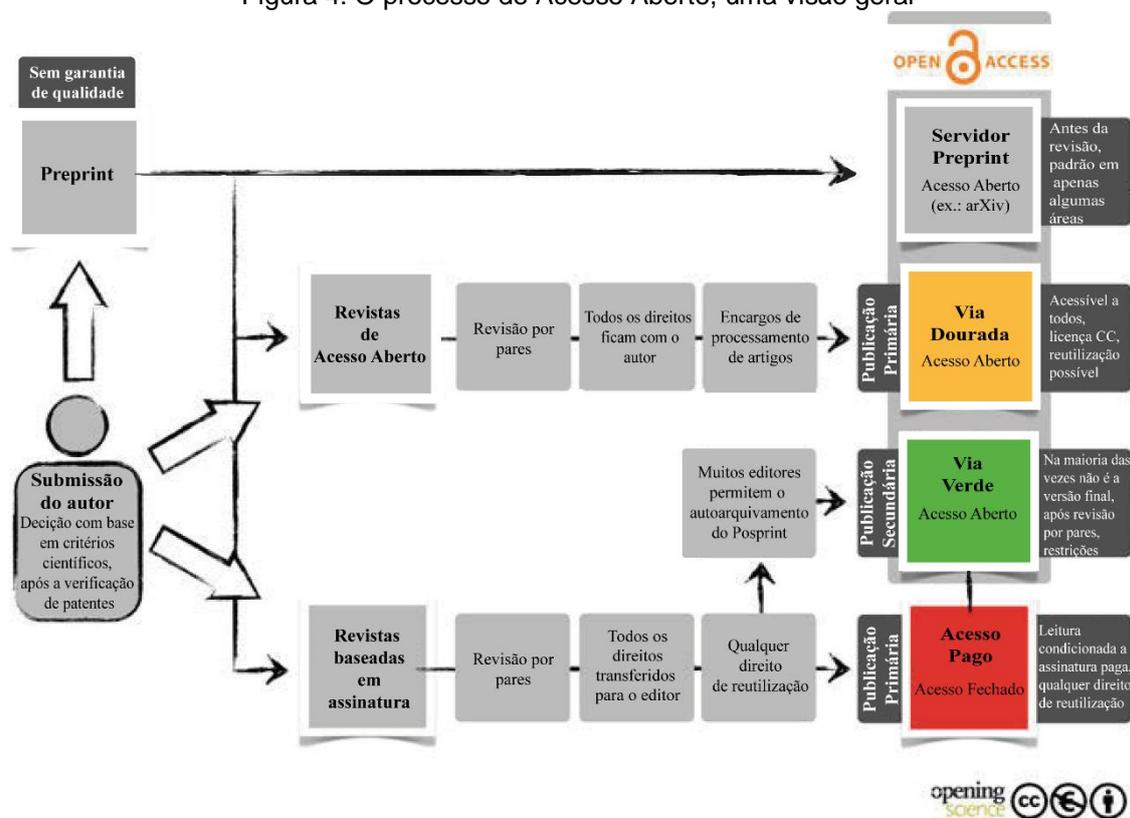
Conforme explica Andrade (2015, p. 270), a Via Verde incentiva a criação de repositórios institucionais, onde as instituições de pesquisa depositam publicações produzidas por seus integrantes. Entre os tipos de

¹³ Suber (2012) explica que remover barreiras de permissão significa que “os pesquisadores estão livres para usar ou reusar a literatura para propósitos acadêmicos. O que inclui leitura e pesquisa, tradução, mineração de texto, migração de mídias, arquivamento por longos prazos, e incontáveis formas de pesquisas, análises e processamentos que nós ainda não imaginamos”.

trabalho são listados: artigos de periódicos, teses, dissertações, conferências e relatórios.

Essas duas vias, segundo orientam Sitek e Bertelmann (2014), oferecem várias vantagens para melhorar os processos de comunicação científica. De acordo com o raciocínio estabelecido por eles, os autores possuem diversas possibilidades de publicar os resultados de suas pesquisas, sendo que as vias do Acesso Aberto maximalizam a visibilidade e o alcance das publicações.

Figura 4: O processo de Acesso Aberto, uma visão geral



Fonte: Sitek e Bertelmann (2014, p. 141)

Segundo Sitek e Bertelmann (2014) explicam, a Via Dourada descreve um modelo de publicação em periódicos que oferece acesso gratuito aos seus artigos e buscam formas alternativas para cobrirem os custos de publicação. Enquanto, paralelamente, o modelo de publicação por Via Verde, na explicação dos mesmos autores, representa a possibilidade que os autores têm de “depositarem seus artigos, que são publicados principalmente nos sites do editor, em outro servidor. Geralmente, um repositório institucional ou baseado em assunto é usado para essa finalidade.”.

A segunda declaração considerada por Sarmiento et al. (2015) como parte fundamental para o Acesso Aberto, a Declaração de Bethesda. Conforme afirma Kuramoto (2006) esta declaração foi o resultado do encontro ocorrido em 2003 onde foi estabelecida a definição do termo “publicação de acesso livre”.

Apesar de essa declaração ter seu cerne voltado para a área de Ciências da Saúde, ela ampliou a proposta de publicação em acesso aberto aos materiais suplementares de pesquisa, conforme afirma Kuramoto (2006, p. 97).

Segundo a declaração, uma publicação de Acesso Aberto deve seguir no mínimo duas condições básicas:

1. O(s) autor(es) e o(s) detentor(es) dos direitos autorais concedem a todos os usuários um direito irrevogável, gratuito e irrevogável de acesso e a licença para copiar, usar, distribuir, transmitir e exibir publicamente o trabalho e para torná-lo público. e distribuir trabalhos derivados, em qualquer meio digital, para qualquer propósito responsável, sujeito à devida atribuição de autoria, bem como o direito de fazer um pequeno número de cópias impressas para seu uso pessoal.
2. Uma versão completa do trabalho e todos os materiais suplementares, incluindo uma cópia da permissão acima, em um formato eletrônico padrão adequado, são depositados imediatamente após a publicação inicial em pelo menos um repositório online que seja apoiado por uma instituição acadêmica, sociedade acadêmica, agência governamental, ou outra organização bem estabelecida que busque permitir acesso aberto, distribuição irrestrita, interoperabilidade e arquivamento a longo prazo (para as ciências biomédicas, o PubMed Central é um exemplo desses repositórios). (DECLARAÇÃO DE BETHESDA, 2003).

Em 2003, a Declaração de Berlim sobre Acesso ao Conhecimento nas Ciências e Humanidades foi a terceira a estabelecer diretrizes essenciais ao Movimento em prol do Acesso Aberto às publicações científicas. Em seus objetivos a declaração sustenta que a informação deve ser tornada disponível em larga escala para a sociedade reforçando a corrente que os frutos científicos são um patrimônio cultural de todos, conforme Merton (1974) destacava.

Dessa forma em Berlim, o acesso livre foi definido como “uma fonte universal do conhecimento humano e do patrimônio cultural que foi aprovada pela comunidade científica.”. Segundo Guedes (2012, p. 36), a Declaração de

Berlim continuou a encorajar os pesquisadores a publicarem livremente os seus artigos, além de reforçar a necessidade da revisão do que venha ser definido como qualidade e reconhecimento nos processos de avaliação e progressão acadêmica.

Conforme explica Barros (2012, p. 366), a “carreira dos acadêmicos depende da publicação em periódicos de maior fator de impacto possível”, sendo que a publicação feita em periódicos menos conceituados pela Qualis é considerada um “suicídio profissional”.

Resumindo as contribuições que as principais declarações sobre acesso aberto trouxeram em suas individualidades, Sarmiento et al. (2015) elaboraram a seguinte tabela explicativa:

Tabela 1: Comparação entre os textos das declarações de Budapeste, Bethesda e Berlim que regem o Acesso Livre ao conhecimento.

| | Declaração de Budapeste Área: Genérica | Declaração de Bethesda Área: Biomédica | Declaração de Berlim Área: Ciências e Humanidades |
|--------------------------|---|--|--|
| Qualidade (peer review) | - Artigos de periódicos revisados - pre-prints sem revisão | - Artigos revisados | - Aprovação pela comunidade científica (revisado) |
| Copyright | - Controle sobre a integridade do trabalho - Direito de propriedade intelectual - Direito a citação (atribuição da autoria) | - Atribuição da autoria | - Atribuição da autoria |
| Finalidade da utilização | - Várias, incluindo qualquer outro propósito legal | - Várias, incluindo qualquer propósito responsável | - Várias, incluindo qualquer propósito responsável |
| Obras derivadas | - Não menciona | - Não menciona | - Não menciona |
| Acesso | - Gratuito - Público (mundial) | - Gratuito - Mundial - Irrevogável - Perpétuo (necessidade de preservação) | - Gratuito - Mundial - Irrevogável |

Fonte: Sarmiento et al. (2015)

Destaca-se também no contexto de Acesso Aberto a Declaração de Valparaíso, Chile, que em 2004 teve o objetivo central de melhorar a comunicação e a publicação no meio acadêmico. Em sugestão, os especialistas envolvidos indicaram que os periódicos científicos deveriam melhorar os seus processos de produção por meio do uso de tecnologias on-line para agilizar as atividades de publicação, já que pesquisas de hábitos de leitura já apontavam a internet como um meio preferido para a transmissão do conhecimento.

Durante a convenção que deu origem a essa declaração também foi destacado a imprescindibilidade da criação de um modelo alternativo para avaliação da produção científica, de modo que a literatura escrita fora da língua inglesa receba suas devidas considerações no contexto global.

Segundo o que informa Abbott (2009) a estrutura de recompensas das universidades, assim como o sistema de revisão por pares, privilegia pesquisas que interessam as comunidades científicas, mas que não necessariamente são objeto de interesse para aqueles que não fazem parte do ambiente acadêmico.

Em âmbito nacional, se salienta o Manifesto Brasileiro de Apoio ao Acesso Livre à Informação Científica lançado em 2005 pelo Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT) em apoio ao acesso livre à informação.

A OAI constitui, portanto, um marco na área do tratamento e disseminação da informação em geral e na área da comunicação científica em especial. Essa iniciativa proporcionou a construção, implantação e manutenção de diversos repositórios de acesso livre, assim como o surgimento de diversas ferramentas de software para a construção e manutenção de repositórios, como o E-Prints, o Open Journal Systems (OJS), o DSPACE, entre outros. (MANIFESTO BRASILEIRO DE APOIO AO ACESSO LIVRE À INFORMAÇÃO CIENTÍFICA, 2005).

Segundo o que é exposto nesse manifesto, o Brasil tinha a carência de uma política nacional que norteasse a comunidade científica ao aumento da visibilidade de suas pesquisas por meio de do paradigma de acesso livre a informação. O manifesto sintetiza as suas ações em quatro grandes objetivos:

- promover o registro da produção científica brasileira em consonância com o paradigma do acesso livre à informação;

- promover a disseminação da produção científica brasileira em consonância com o paradigma do acesso livre à informação;
- estabelecer uma política nacional de acesso livre à informação científica;
- buscar apoio da comunidade científica em prol do acesso livre à informação científica.

Para tanto, o Manifesto elenca ser imperativo que as instâncias acadêmicas brasileiras se comprometam a:

Quadro 2: Recomendações à Comunidade Científica

| | Instituições acadêmicas | Pesquisadores |
|---|--|--|
| 1 | Criar repositórios institucionais e temáticos, observando o paradigma do acesso livre; | Contribuam para o incremento de conteúdos em repositórios institucionais ou temáticos, depositando o maior número possível de seus trabalhos, publicados ou não, inclusive pré e post prints, material de aula, quando for o caso, entre outros materiais. |
| 2 | Requerer que seus pesquisadores depositem uma cópia de todos os seus trabalhos publicados em pelo menos um repositório de acesso livre; | Depositar, obrigatoriamente, em um repositório de acesso livre publicações que envolvam resultados de pesquisas financiadas com recursos públicos. |
| 3 | Encorajar seus pesquisadores a publicar seus resultados de pesquisa em periódicos de acesso livre, onde houver um periódico apropriado para isso. deve-se, além disso, prover o apoio necessário para que isso ocorra; | |
| 4 | Reconhecer a publicação em ambiente de acesso livre para efeito de avaliação e progressão acadêmica; | |
| 5 | Ter disponíveis, em ambiente de acesso livre, os periódicos editados pela instituição ou seus órgãos subordinados. | |

Fonte: Manifesto Brasileiro de Apoio ao Acesso Livre à Informação Científica com adaptações, 2005.

Percebe-se no manifesto do IBICT um realce nos repositórios como veículos de comunicação no Movimento de Acesso Aberto, mas Suber (2012), no entanto, reflete que existem muitas maneiras pelas quais o acesso aberto pode fluir. Entre elas o autor cita blogs, wikis, bancos de dados, e-books,

vídeos, áudios, fóruns e tantas outras formas que o pensamento e a criatividade humana possam imaginar. Não obstante, o pensamento do autor vai ao encontro dos demais autores reafirmando os periódicos e os repositórios como principais veículos de transmissão da informação.

As estratégias traçadas pelo acesso aberto demonstram o anseio social por instrumentos que possibilitem um novo regime de circulação de produtos intelectuais, em que não sejam privilegiados os fatores financeiros, mas sim a o empoderamento dos cidadãos com o conhecimento.

Uma ciência aberta, em sua teoria possibilita uma circulação do conhecimento mais fluida por meio do compartilhamento de dados e informações. De acordo com o que informa Boulton (2012), a “ciência aberta é definida aqui como dados abertos (dados disponíveis, inteligíveis, calculáveis e utilizáveis) combinados com acesso aberto a publicações científicas e comunicação efetiva de seus conteúdos.”

Pinheiro (2014) declara que as o avanço das Tecnologias de Informação e Comunicação (TICS), o acesso livre à informação e a complexidade das ciências, com equipes cada vez mais numerosas e cada vez mais trabalhando em colaboração, propiciaram o desenvolvimento de um fenômeno científico, político e social chamado de E-Science.

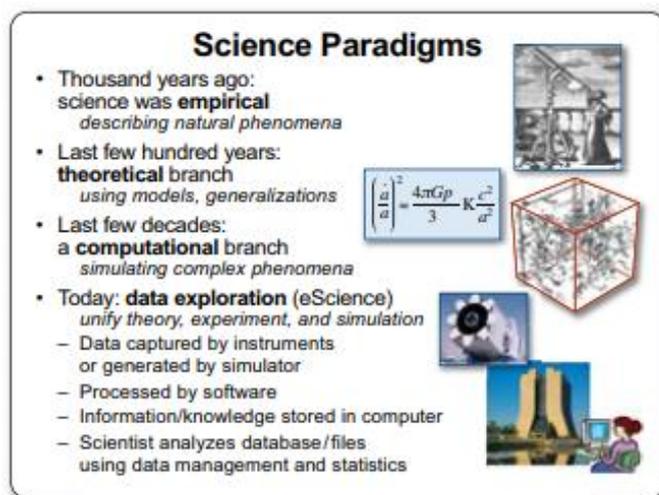
Segundo Appel (2014):

O uso do termo e-Science diz respeito a um movimento que prevê a concepção de uma ciência atuante no uso intensivo de dados e na colaboração por meio do uso de plataformas de pesquisa baseadas em computação avançada. (APPEL, [s.p], 2014).

De acordo com os autores Hey e Hey (2006, p. 516) a “e-Science” ou ciência baseada em dados e em rede é uma prática que está prestes a transformar a forma como os cientistas realizam as suas pesquisas por meio da grande quantidade de dados disseminados.

Segundo o que foi observado por Gray (2007), o paradigma científico evoluiu de uma ciência empírica que descrevia fenômenos naturais para uma etapa onde a teoria, experimentos e a simulação se unificaram e os dados são capturados por instrumentos ou simulação.

Figura 5: Paradigmas da ciência



Fonte: Gray (2007)

Ainda de acordo com Gray (2007), as novas tecnologias que exploram dados são tão diferentes que “vale a pena distinguir a ciência intensiva em dados da ciência computacional como um novo quarto paradigma para a exploração científica”.

Desde que as tecnologias que possibilitaram o Acesso Aberto e suas ferramentas operacionais também modificaram as formas pelas quais o conhecimento científico fluiu. Tais evoluções e particularidades serão o foco do estudo do capítulo seguinte.

É possível observar atualmente, de acordo com Bartling e Friesike (2014, p. 7) uma alteração na transmissão do conhecimento que foi desencadeada pela internet, que outrora era sequer imaginável. Eles ainda explicam que esses novos métodos surgiram em todas as disciplinas, entretanto, nas humanidades o termo usado foi "humanidades digitais" e de um “ângulo” da Web 2.0 sob o termo "Ciência 2.0", ou daqueles que lutam pelo conhecimento livre sob o termo "Pesquisa aberta" e “Ciência Aberta”.

O uso das tecnologias eletrônicas que beneficiaram e agilizaram o processo de comunicação científica foi destacado por Leite (2009, p.14). O autor entende que essas tecnologias possibilitaram o surgimento dos periódicos eletrônicos em 1990 e fizeram com que a hegemonia dos editores

científicos fosse examinada, chegando à conclusão que o sistema de comunicação nos moldes tradicionais estava limitando o conhecimento, assim como a sua circulação.

2.3. CIÊNCIA ABERTA

Suber (2012) chama o processo descrito por Leite (2009) de revolução digital. Em sua teoria, os pesquisadores que visam obter um maior impacto com seus trabalhos em relação aos lucros advindos das vendas começaram a ter vantagens a partir desta revolução.

Este fenômeno desencadeou uma reação de movimentos de acesso livre que tiveram início em 1999 e tinham como principal foco a “comunicação ampla e irrestrita dos resultados de pesquisas financiadas com recursos públicos”, segundo Andrade (2015, p. 269). Essa iniciativa acadêmica não inclui apenas o acesso a literatura científica, mas também os meios, procedimentos e ferramentas que são necessários para a elaboração do conhecimento.

Na visão de Gonzales (2006, p. 329) o movimento de livre circulação do conhecimento científico se derivou dos ideais propostos pelos movimentos de softwares não proprietários que ocorreu na área da computação. Para o autor, a incorporação dos princípios dos softwares de código aberto na área de pesquisa científica foi chamada de movimento de “Ciência Aberta”. Ele o define mais precisamente como “a aplicação de princípios e cláusulas de licenciamento de código aberto para proteger e distribuir os frutos da pesquisa científica.”

Albagli (2015) demonstra estar consoante com a proposta de Gonzales (2006) ao afirmar que:

O movimento pela ciência aberta se insere nesse quadro de tensão entre, por um lado, novas formas de produção colaborativa, interativa e compartilhada da informação, do conhecimento, da cultura. E, por outro, mecanismos de captura e privatização desse conhecimento que é coletiva e socialmente produzido. (ALBAGLI, 2015, p. 13).

Apesar da forte corrente de práticas científica que vem se formando ao redor das atividades relacionadas ao Acesso Aberto e a Ciência Aberta, Molloy (2011, p. 2) acredita que ainda existem muitas dúvidas em relação às reais “práticas abertas”. A autora acredita que o termo “grátis” se confunde com o termo “livre” e que, por não possuírem licenças de uso depreendidas de

cláusulas comerciais ou não derivativas, os trabalhos compartilhados nessas situações não colocam realmente os leitores em frente a uma “Ciência que seja Aberta”.

Fecher e Friesike (2014, p. 17) argumentam que o termo “Ciência Aberta” “engloba uma infinidade de pressupostos sobre o futuro da criação e disseminação do conhecimento.”, sendo composta por diversas variáveis relacionadas às mudanças que ocorrem no campo científico, seu único e aparente denominador comum é que “ciência no futuro próximo de alguma forma precisa se abrir mais”.

Maurer (2013, p. 4) indica estar de acordo com Fecher e Friesike (2014) ao também demonstrar acreditar que o termo “Ciência Aberta” possui várias definições, entretanto, o autor chega à conclusão que em suma o termo tende a conotar:

- a) uma publicação completa, franca e rápida resultados;
- b) a ausência de restrições de propriedade intelectual, e
- c) um aumento da transparência nas fases pré e pós-publicação de dados, atividades, e decisões dentro de grupos de pesquisa.

Por meio de uma nova mentalidade da prática científica os pesquisadores foram influenciados a compartilhar os procedimentos científicos com outros pesquisadores, essa prática é definida pela literatura como Ciência Aberta, que é o compartilhamento de informações abertas entre pesquisadores. Segundo Andrade (2015) um dos argumentos mais utilizados para justificar a Ciência Aberta é o posicionamento intelectual altruísta dos cientistas que visam o desenvolvimento da humanidade.

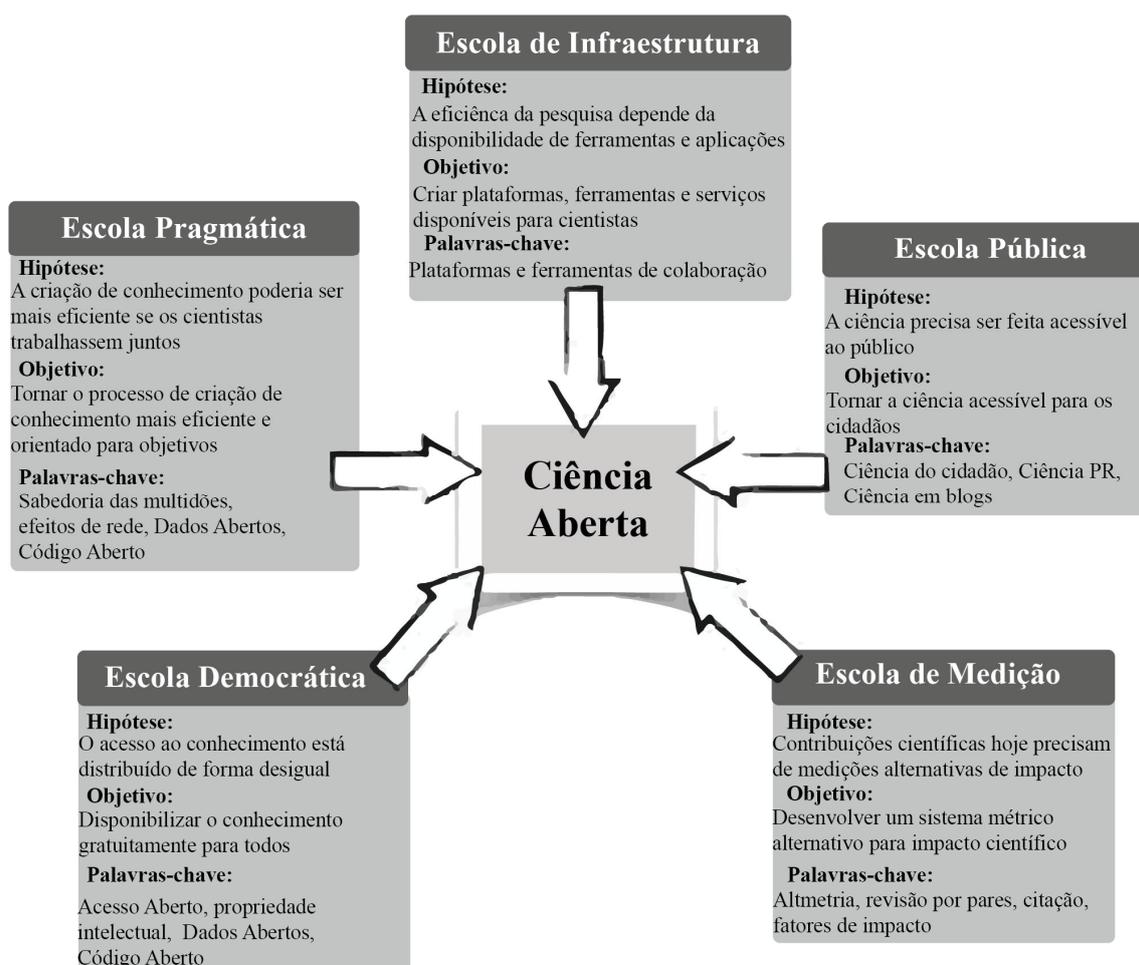
“Na verdade, o mesmo termo evoca entendimentos bem diferentes e abre uma infinidade de campos de batalha, desde o direito democrático de acesso ao conhecimento financiado publicamente (por exemplo, Acesso Aberto a publicações) ou a demanda por uma melhor ponte entre pesquisa e sociedade (por exemplo, ciência cidadã) para o desenvolvimento de ferramentas livremente disponíveis para colaboração (por exemplo, plataformas de mídia social para cientistas). Desse ponto de vista, a abertura poderia se referir a praticamente qualquer coisa: o processo de criação do conhecimento, seu resultado, o indivíduo pesquisador, ou a relação entre a pesquisa e o resto da sociedade.” (FECHER; FRIESIKE, 2014, p. 17-18)

De acordo com Andrade (2015) a Ciência Aberta possui uma série de recomendações a comunidade científica, como:

- A abertura de dados científicos de forma que estes possam ser acessados e reutilizados;
- A estimulação da produção colaborativa de dados de forma que a ciência se torne mais global e com um menor custo.

Por meio de uma ampla análise da literatura científica relevante sobre o tema, os cientistas Fecher e Friesike (2014) identificaram cinco escolas de pensamento com padrões de argumentação e pensamento sobre a Ciência Aberta.

Figura 6: Ciência aberta: um termo, cinco escolas de pensamento



Fonte: Fecher e Friesike (2014, p. 19, tradução nossa).

De acordo com o que pôde ser sintetizado pelos autores por meio de uma extensa revisão bibliográfica:

2.3.1. Escola Pública

Segundo essa teoria a ciência precisa ser acessível a um público mais amplo, tanto para os cientistas, como para o coletivo de não-especialistas interessados. Pelo que expõe Fecher e Friesike (2014), é possível reconhecer a existência de duas correntes dentro dessa escola, a primeira se refere à acessibilidade do processo de produção da pesquisa. A segunda aborda a compreensibilidade do resultado da pesquisa, ou seja, o seu produto propriamente dito. Segundo os autores as tecnologias da web social e web 2.0 funcionam como instrumentos que permitem aos cientistas, não somente difundir os resultados de suas pesquisas mais facilmente, mas também estabelecer uma ponte entre a pesquisa e os cidadãos¹⁴ fazendo com que eles se integrem ao processo de produção.

Os blogs são uma das ferramentas mais recentes que os cientistas têm utilizado para comunicar as suas ideias com o público geral, apesar de os caracterizados como científicos ou tecnológicos serem apenas 4% do geral, de acordo com o que informa Bonetta (2017).

Em consonância com o que também foi apurado nessa corrente de Ciência Aberta, as ferramentas de Web 2.0 podem ser usadas para uma finalidade além da acessibilidade, como por exemplo, também podem ser utilizadas para aumentar a confiança pública.

Ao abordar a segunda corrente da Escola Pública, Fecher e Friesike (2014) destacam na literatura autores que declaram que a ciência também deve ser compreendida por um público mais amplo, o que demanda uma mudança adaptativa no estilo de escrita a tornando mais simples.

¹⁴ O princípio exposto está em conformidade com o conceito de Ciência Cidadã, que retrata a participação do público de não- cientistas, especialistas e amadores ativamente no processo de criação científica.

Também é argumentado que as ferramentas da Web 2.0 propiciam uma maior interação pública, onde o pesquisador pode se tornar uma figura pública e o corretor de suas próprias informações.

2.3.2. Escola de medição

Essa vertente da Ciência Aberta versa sobre a preocupação com os padrões indicadores do impacto científico, segundo explica Fecher e Friesike (2014). Para eles é indiscutível que o impacto de um artigo é medido pelo número médio de citações que esse recebe. Esse número de citações influencia outros aspectos da carreira científica do pesquisador que vão desde a sua reputação até o financiamento de outras pesquisas.

Os autores ainda afirmam que no contexto da Ciência Aberta é comum o questionamento sobre como o impacto dos artigos publicados serão mensurados, já que atualmente o modelo de revisão por pares se mostra demorado, sendo que o impacto científico está mais relacionado com o periódico do que com o artigo publicado e as novas formas de publicação como diários abertos e blogs normalmente não podem ser atrelados a um fator de impacto.

Ao analisar o uso microblog Twitter como uma estrutura de citações e referências, Weller e Puschmann (2011) explicam que por meio do uso de hashtags e retweets é possível identificar o comportamento de citações¹⁵ dentro da plataforma, que se caracteriza como bastante diferente do tradicional.

Para Priem et. al. (2012, p. 40):

“à medida que os fluxos de trabalho acadêmicos migram cada vez mais para a Web, antigos usos “clandestinos” como leitura, bookmarking, compartilhamento, discussão e classificação estão começando a deixar rastros on-line. Eles estão se tornando visíveis em páginas da Web, em blogs, em downloads em mídias sociais como Twitter. e em gerenciadores de referência social como

¹⁵Segundo Weller e Puschmann (2011) existem duas tipos fundamentais de citação ao se analisar o Twitter: as citações externas, que são todos os links incluídos nos tweets, e as internas, que são os retweets dentro da própria plataforma, ambos fazendo referência a artigos acadêmicos revisados por pares.

CiteULike, Mendeley e Zotero”. (PRIEM et. al., 2012, p. 40, tradução nossa).

Nas visões de Fecher e Friesike (2014), a opinião de Priem et. al. (2012) evidencia a necessidade de um novo parâmetro de medição de impacto científico que seja alternativo, rápido e inclua as novas formas de contribuição científica, já que a tradicional modelo de medição de citações reflete apenas um espectro dos usos acadêmicos que os cientistas fazem.

2.3.3. Escola Pragmática

Fecher e Friesike (2014) explicam que, para os defensores desta corrente, a Ciência Aberta “é um método para tornar a pesquisa e a disseminação do conhecimento mais eficientes” ao incluir um pensamento externo e permitir a colaboração científica por meio das ferramentas da Web 2.0.

Nessa corrente está muito presente a noção de “inovação aberta”, que segundo afirmam os autores, se relaciona diretamente aos princípios da Ciência Aberta. Tacke (2011) explica que desde o século XIX estão sendo criadas tentativas para fazer com que o conhecimento especializado se torne mais popular e de melhor compressão seja por meio de escritas mais simplificadas ou por meio de peças de teatro. Para o autor, a adoção de formas alternativas para tornar a ciência mais compreensível para as grandes massas pode ser compreendido como Ciência Aberta.

Mais além, Tacke (2011) elucida que os atuais usos da web 2.0 permitem que a ciência não seja apenas um produto entregue ao público já pronto, de forma que os processos que levaram aquele conhecimento permaneçam na obscuridade e omitidos. Para ele, a internet age como um canal em que os pesquisadores podem receber dúvidas e sugestões e ainda fornecer um retorno ao público interessado. Ademais, o autor enfatiza que:

Na Wikipédia, podemos não apenas ler textos, mas também escrever. No YouTube, podemos não apenas assistir a filmes, mas também definir e comentar por conta própria. A distinção entre produtores e consumidores de conteúdo se suaviza. (TACKE, 2011, tradução nossa).

Os autores Fecher e Friesike (2014) e Tacke (2011) concordam que esse contexto de troca de informações e experiências de forma aberta permite a elaboração de pesquisas coletivas que aumentam não só a produtividade científica, mas também respondem positivamente a crescente complexidade da produção do conhecimento.

Com o foco na colaboração científica, Fecher e Friesike (2014) fazem alusão ao trabalho de Neylon and Wu (2009), que afirmam que, em busca da construção de uma ciência coletiva, os pesquisadores buscam ferramentas eletrônicas de interação social que facilitam e aceleram a descoberta científica.

Sendo assim, essa corrente se destaca pelas perspectivas de colaboração que são facilitadas, não só pelas ferramentas da Web 2.0, mas também pela criação de novos instrumentos que facilitem a colaboração e troca de recursos. Hey e Trefethen (2005) notificam a comunidade científica sobre projetos que envolvem a criação de *middlewares* que permitirão aos cientistas terem um ambiente seguro e controlado para compartilhamento colaborativo de recursos de pesquisas. Os autores afirmam que os “conjuntos de serviços de *middleware* e as redes globais de pesquisa de alta velocidade constituirão a nova Cyberinfraestrutura para a pesquisa colaborativa”.

2.3.4. Escola de Infraestrutura

É perceptível que, a certo modo, a Escola de Infraestrutura se ligue à Pragmática, visto que, segundo os autores Fecher e Friesike (2014), a Escola de Infraestrutura tem como cerne a preocupação com “a infraestrutura técnica que permite práticas de pesquisa emergentes na Internet”, sendo, portanto, a Ciência Aberta, “um desafio tecnológico”. Infraestruturas essas, que permitem, entre outras ações, a colaboração para a criação da pesquisa científica.

Hey e Trefethen (2005) entendem que desde seu início, a internet serviu como uma ferramenta para o mundo científico, entretanto, agora ela representa um aparato que pode ir muito além da colaboração científica.

Além de estar apta a acessar informações de diferentes sites, eles desejam integrar, federar e analisar informações de diversas fontes de dados distribuídas e dispersas (incluindo arquivos de dados, bem como redes de sensores e tags de identificação) e acessar e controlar recursos computacionais e equipamentos experimentais em locais remotos. (HEY e TREFETHEN, 2015, p. 818, tradução nossa).

Hey e Trefethen (2015) dão continuidade ao seu raciocínio sobre as infraestruturas tecnológicas de colaboração científica citando a visão de John Taylor (Diretor geral do Research Councils at the Office of Science and Technology - OST) que reconhecia nas ferramentas de colaboração científica um serviço de tecnologia da informação que poderia ser fornecido por demanda, assim como a eletricidade, gás, água ou telefone ao invés de ser comprada.

2.3.5. Escola Democrática

Esta escola parte de um foco central na disponibilização do acesso ao conhecimento com foco principal nos produtos das pesquisas científicas como dados científicos, fontes de materiais, representações digitais de materiais pictóricos e gráficos ou material multimídia, de acordo com os ensinamentos de Fecher e Friesike (2014). Para os autores, nessa escola, em resumo: “qualquer produto de pesquisa deve estar disponível gratuitamente.”.

“A razão pela qual nos referimos ao discurso sobre o livre acesso a produtos de pesquisa como a escola democrática resulta de sua lógica inerente de que todos devem ter o mesmo direito de acessar o conhecimento, especialmente quando ele é financiado pelo Estado.” (FECHER e FRIESIKE, 2014, p.25, tradução nossa).

Ainda sobre as reflexões dos autores sobre o tema, as correntes centrais dessa escola de pensamento se concentram em duas correntes: o acesso aberto a publicações e aos dados abertos. Sobre a publicação aberta de dados, Molloy (2011) acredita que quanto mais eles forem disponibilizados abertamente, maior será a transparência, reprodutibilidade e, logo, por consequência, mais eficiente será o processo científico, visto que “ciência é construída sobre dados: sua coleta, análise, publicação, reanálise, crítica e reutilização”.

2.3.6. Dados Abertos: conceituação e classificação

2.3.6.1. Dados: conceituação

A medida que as comunidades de pesquisa empregam mais tecnologia na elaboração de suas atividades, mais dados serão produzidos. De acordo com os Kim e Stanton (2013), durante as últimas décadas, houve um enorme esforço para promover o compartilhamento de dados entre os cientistas. Os autores ainda informam que muitas revistas adotaram uma postura mandatória em suas políticas no que é referente aos pesquisadores compartilharem os dados basilares de suas pesquisas publicadas.

Anteriormente, foram citadas as considerações de Borgman (2008) referentes a falta de consenso do que seria o conceito de dados. Esse fenômeno pode ser atribuído em parte, pela grande diversificação dos tipos de dados existentes e as diferentes formas que os pesquisadores os empregam em seus campos científicos. Gómez, Méndez e Hernández-Pérez (2016, p. 547) possuem uma postura complementar a de Borgman (2008) ao relatarem que “para cada disciplina do campo científico existe uma interpretação do que são dados ou conjunto de dados de investigação, sua natureza e como são coletados”, sendo que mesmo dentro de um campo específico, como é o caso das Humanidades, o conceito do que seja dado de pesquisa pode variar, o que, na opinião dos pesquisadores, pode dificultar a sua recuperação.

A OCDE (2007, p. 13) definiu dados de pesquisa como sendo:

“Dados de pesquisa” são definidos como registros factuais (pontuações numéricas, registros textuais, imagens e sons) usados como fontes primárias para pesquisa científica, e que são comumente aceitos na comunidade científica como necessários para validar os resultados da pesquisa. Um conjunto de dados de pesquisa constitui uma representação parcial e sistemática do assunto investigado. (OCDE, 2007, p. 13, tradução nossa).

Molloy (2011) acredita que são os dados que fornecem evidências para o conhecimento científico, que é a própria base da ciência. Quanto a definição

do que sejam dados, Boulton (2012) os conceitua como sendo “números, caracteres ou imagens que designam um atributo ou fenômeno”. Eles subdividem dados em duas categorias: brutos ou primários e os derivados dos dados primários.

Ainda conceituando dados, a National Science Board (2005) se refere a dados como sendo:

[...] qualquer informação que possa ser armazenada em formato digital, incluindo texto, números, imagens, vídeo ou filmes, áudio, software, algoritmos, equações, animações, modelos, simulações, etc. Tais dados podem ser gerados por vários meios, incluindo observação, computação ou experimento.

Há nessa classificação, uma diferenciação das demais quando a National Science Board (2005) entende que os dados podem ser classificados em “texto, números, imagens, vídeo ou filmes, áudio, software, algoritmos, equações, animações, modelos, simulações, etc”, sendo que as categorias observacionais, computacionais ou experimentais, se referem a origem dos dados, não a sua natureza. Para a National Science Board (2005) estabelecer essa diferenciação se torna “crucial para escolhas feitas para arquivamento e preservação”.

A National Science Board (2005) argumenta que:

- Dados observacionais são observações diretas de fenômenos que se tornam registros históricos e por isso, “dados observacionais são geralmente arquivados indefinidamente.”.
- Dados computacionais são os resultados da execução de um modelo ou simulação computacional, por esse motivo, se houver a correta descrição dos objetos e condições dos experimentos, “a preservação em um repositório de longo prazo pode não ser necessária, pois os dados podem ser reproduzidos”.
- Dados experimentais aparentam ter uma complexidade mais elevada. Segundo os autores, aqueles dados experimentais que possam ser reproduzidos com precisão não carecem de armazenamento indefinidamente, no entanto, ainda é argumentado que nem sempre todas as condições do experimento são possíveis de serem reproduzidas, ou a sua

execução gera custos muito elevados. “Nestes casos, a preservação a longo prazo dos dados é garantida.”.

Em uma análise do livro de Viktor Mayer-Schönberger e Kenneth Cukier, Abreu (2014) entende que quando se tem como objetivo conhecer uma pessoa, basta que sejam observados e analisados os dados que permeiam o círculo desse indivíduo, os dados que, de certa forma passam fornecer indícios de quem esse indivíduo é e dos hábitos que possui: “os contatos e ligações estabelecidas, relacionando praticamente todos os aspectos da vida para, a partir das ferramentas e sistemas do Big Data¹⁶, identificar padrões de comportamento, fazer correlações e análises preditivas.”. Esses dados podem possuir origem e finalidades distintas apesar de remeterem à vida de um único indivíduo.

Então cabe lembrar e contextualizar a diferenciação que Hernández-Pérez e Garcia-Moreno (2013) fizeram a respeito das espécies de dados: eles os diferenciam em:

- Estatísticos;
- Geográficos;
- De transporte;
- Meteorológicos;
- Financeiros
- Ambientais;
- Governamentais;
- Científicos e
- Culturais.

Os autores ainda afirmam que as tecnologias da informação possibilitaram compilar as grandes quantidades de dados em:

- **Dados pessoais:** que abrangem os hábitos de compras, relações pessoais, o registro de preferências online registradas a partir do Facebook, Twitter, Youtube e outros, viagens, saúde e etc.

¹⁶ O termo “Big data” foi definido pela MGI (2013, p. 4) como sendo referente à “refere-se a conjuntos de dados que são volumosos, diversos e oportunos.”

- **Dados sobre o meio ambiente:** que contemplam os níveis de contaminação, tipos e quantidade de partículas estranhas no ar e etc.
- **Dados sobre objetos e produtos:** como edifícios, a qualidade de produtos alimentícios a partir de leituras de códigos de barras e;
- **Dados sobre os processos:** abrangendo dados sobre a rota de ônibus, ou a localização de uma carga, origem e destino de viagens.

A National Science Board (2005) argumenta que existem uma gama de características pelas quais os dados são diferenciados, sendo que essas diferenças incluem:

- A natureza dos dados
- Sua reprodutibilidade;
- O nível de processamento ao qual foram submetidos.

Entretanto, quando Costa (2017, p. 47) analisa o conceito de conceito de Big data, e-science, dados de governo e dados abertos a autora compreende uma relação existente entre essas áreas e seus diversos tipos de dados entendendo que há a possibilidade de combinação dessas áreas, gerando novas análises e produtos. A autora percebe então, quatro espécimes de dados que foram ilustrados na figura a seguir:

Figura 7: Aspectos conceituais do Big Data.



Fonte: Costa (2017, p. 48)

Ao avaliar com maior profundidade o esquema ilustrativo de Costa (2017), compreende-se que independentemente da finalidade para que tenham sido coletados, os dados podem ser utilizados para outra destinação a depender do contexto e da interpretação a qual possam ser atribuídos, ou seja, é possível acreditar que qualquer dado, independentemente de sua fonte, pode vir a ser utilizado como um dado de pesquisa.

Rodrigues e Saraiva (2010) afirmam que alguns dados não são produzidos com a finalidade de investigação, mas acabam por se tornar seu objeto, como é o caso dos dados de redes sociais que são utilizados para finalidades de estudo nos campos sociológicos.

A respeito dos dados de pesquisa em formato digital, a OECD (2007 p. 9) possui uma postura que vai ao encontro das afirmações feitas por Rodrigues e Saraiva (2010), e conclui que eles estão sendo cada vez mais utilizados para além do projeto que original para que fossem coletados, além disso, destacam que dados que originalmente não foram coletados para o propósito científico são utilizados como engrenagens desses campos:

Os dados administrativos das instituições dos países membros da OCDE, como informações sobre emprego, são agora amplamente utilizados nas ciências sociais, bem como na formulação de políticas. Dados de organizações de saúde pública desempenham um papel crescente no avanço das ciências da vida. Da mesma forma, dados geoespaciais coletados por diversas organizações governamentais são essenciais para pesquisas ambientais e de outros tipos. A lista continua. (OECD, 2007, p. 9).

O fator mutacional do propósito dos dados os atribui uma característica informacional, a multiforma, já que todos os tipos de dados podem vir a se transformar em uma fonte de informação a depender da pesquisa a qual se destinam o que ressalta a necessidade de gestão e preservação dessas fontes de informação compartilhadas. A exemplo, Pepe et al. (2014) citam que os dados primários são utilizados por astrônomos em todo mundo de várias maneiras distintas produzindo dados derivados.

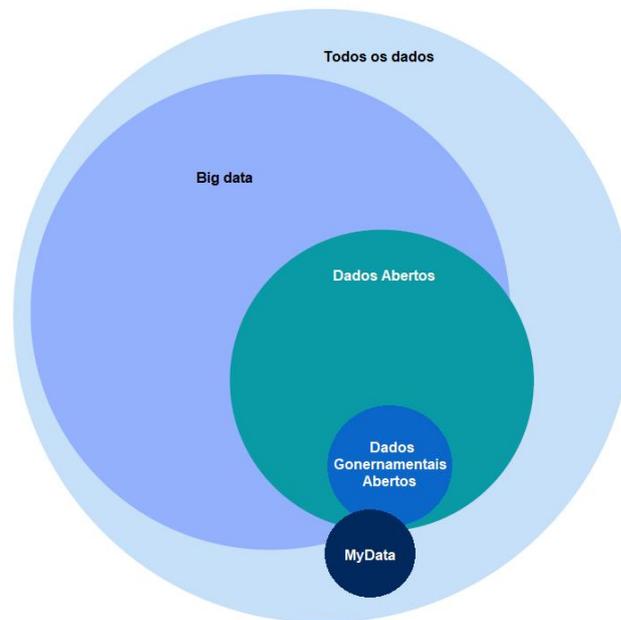
2.3.6.2. Dados abertos: conceituação

A prática da publicação em acesso aberto está cada vez mais popular entre os cientistas, de acordo com o que afirma o Institution of Environmental Sciences (IES) (2013). Segundo o IES, da prática de publicação científica em acesso aberto também emergiu o fortalecimento da das noções de compartilhamento dos dados científicos de forma aberta.

[...] o foco do debate mudou para incluir o fornecimento de conjuntos de dados originais juntamente com artigos científicos para permitir que outros pesquisadores / profissionais utilizem os dados para publicações subsequentes ou para verificar a confiabilidade dos resultados da pesquisa original através de um processo chamado 'data-mining'. (INSTITUTION OF ENVIRONMENTAL SCIENCES, 2013, p.1, tradução nossa).

A diferenciação sobre os tipos de dados em relação à classificação de dados abertos foi condensada pelo MGI (2013) na figura seguinte:

Figura 8: Como os dados abertos se relacionam com outros tipos de dados



Fonte: McKinsey Global Institute analysis

Fonte: McKinsey Global Institute analysis (2013, p. 4, tradução nossa).

O instituto explica que os dados classificados como “*MyData*” são aqueles que “envolvem o compartilhamento de informações coletadas sobre um indivíduo (ou organização) com esse indivíduo. ”, sendo que esses podem ou não ser classificados como abertos de acordo com a autorização dada pelo indivíduo. Os autores afirmam que dados governamentais são aqueles divulgados pelo governo, sem que necessariamente estejam abertos.

Por fim, eles concluem que os dados abertos geralmente fazem parte do Big Data, mas também podem ser um conjunto pequeno. Os dados abertos, de acordo com eles, “podem ser locais, nacionais ou globais e podem ser obtidos tanto de fontes governamentais quanto comerciais”.

O acesso aos dados científicos se tornou um requisito para o desenvolvimento da pesquisa moderna, segundo afirma González (2006). No entanto, o autor também externa a sua preocupação com uma grande tendência de comercialização desses dados, pois essa comercialização se traduziria como uma redução na produção científica geral.

A comercialização dos dados citada por González (2006) representa um estreitamento no número de pessoas que conseguiriam ter acesso a esses dados. Em um lado oposto, se encontram os dados que são disponibilizados de

forma aberta para que qualquer interessado possa utilizá-lo. O IES (2013) explica que os dados abertos são “informações que estão disponíveis para qualquer um usar, para qualquer finalidade, sem nenhum custo.”. O instituto elenca critérios para que dados sejam considerados como abertos:

- Acessível (preferencialmente pela internet) a não mais do que o custo de reprodução, sem limitações baseadas na identidade ou intenção do usuário;
- Em um formato digital, legível por máquina para interoperar com outros dados; e
- Livre de restrições de uso ou redistribuição em suas condições de licenciamento.

De acordo com o que ensina Spinak (2013), o fenômeno conhecido como “Dados Abertos” também é referenciado como “Informação líquida” fazendo referência aos dados que “são gravados em mídia digital, estão continuamente sendo atualizados, sem restrições de acesso, amplamente disponíveis e em formatos que podem ser compartilhados via protocolos abertos.”.

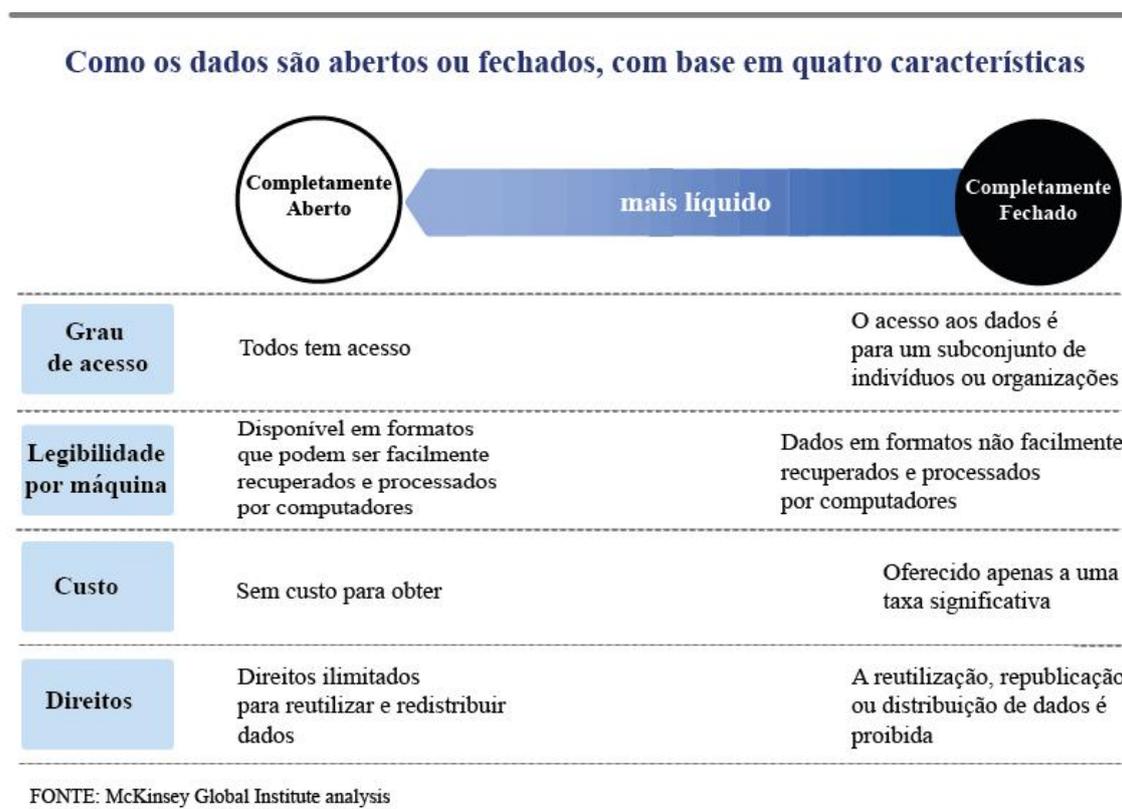
O McKinsey Global Institute (MGI) (2013) analisa que apesar de sempre ter sido possível compartilhar informações abertamente, quando essas informações passaram do meio físico para o digital, esse processo de disseminação foi facilitado abrindo caminho para o compartilhamento outros tipos de dados que não os de pesquisa. O mesmo instituto também considera que os conjuntos de dados abertos compartilham características básicas. Para o MGI (2013), são eles:

- Acessibilidade: os usuários possuem permissão para acessar os dados;
- Legibilidade por máquina: os dados podem ser processados automaticamente;
- Custo: os dados podem ser acessados gratuitamente ou a um custo insignificante.

- Direitos: as limitações no uso, transformação e distribuição de dados são mínimas

O MGI (2013) ainda considera que dentro das quatro dimensões mencionadas acima, os conjuntos de dados podem variar entre completamente aberto e completamente fechado, conforme é ilustrado na figura 13.

Figura 9: Como os dados são abertos ou fechados, com base em quatro características



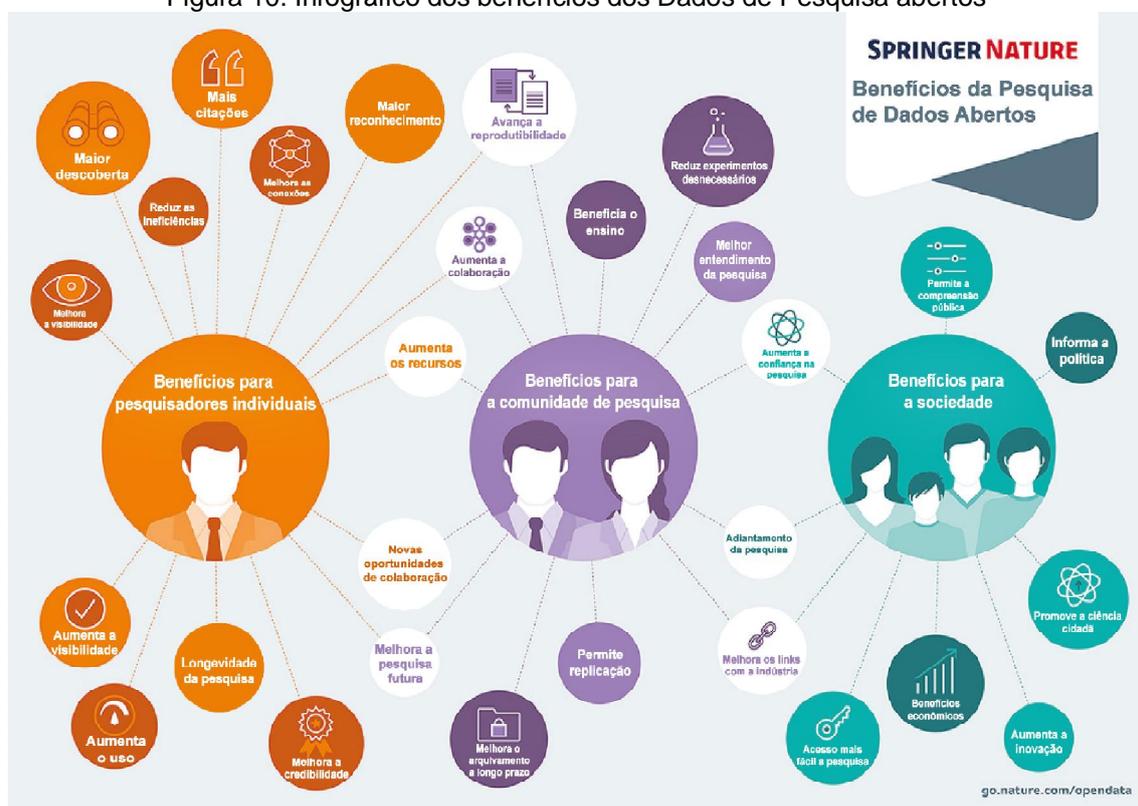
Fonte: McKinsey Global Institute (2013, p. 3)

O contexto de “dilúvio de dados” em que os pesquisadores estão inseridos os estimula a reproduzir não apenas as pesquisas já concluídas, mas também as perguntas que as originou, segundo afirma Borgman (2011, p. 2). Tenopir et al (2011, p. 1) atestam que o processo de compartilhamento de dados vai além da disseminação e preservação, mas também inclui o seu fornecimento para acesso, uso e reutilização. Os autores listam como vantagens para o compartilhamento de dados:

- A re-análise dos dados fornecendo uma espécie de verificação dos resultados obtidos;
- Fornecer diferentes abordagens e interpretações;
- Quando bem administrados, a preservação auxilia a garantir a integridade dos dados;
- Segurança contra a má conduta de fabricação e falsificação de resultados.

A Astell (2017) categorizou as vantagens do compartilhamento de dados em três grandes vertentes de beneficiados, conforme é ilustrado a seguir:

Figura 10: Infográfico dos benefícios dos Dados de Pesquisa abertos



Fonte: Astell (2017, tradução nossa)

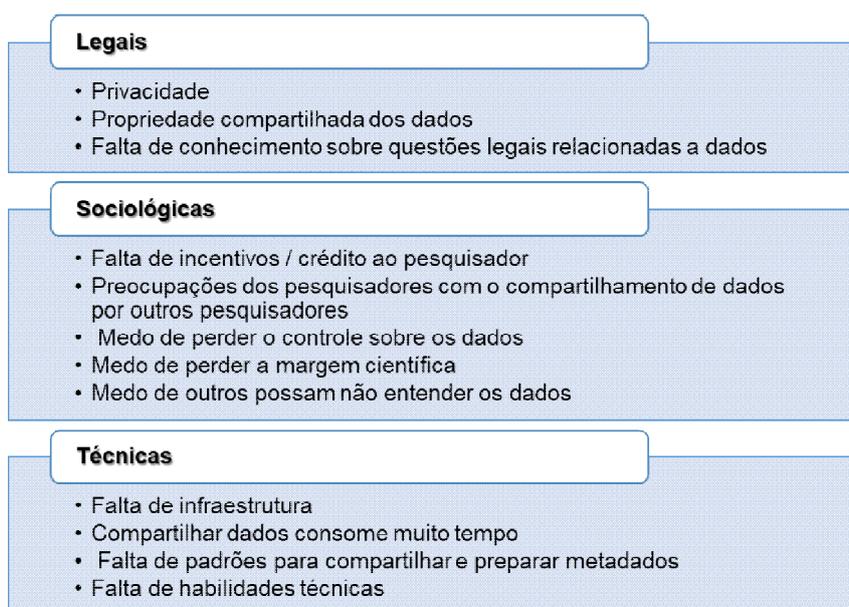
As grandes categorias foram: “Benefícios para o pesquisador individual”; “Benefícios para a comunidade de pesquisa” e “Benefícios para a sociedade”. Como exemplos de benefícios mútuos o autor cita: o avanço da reprodutibilidade, o aumento da colaboração, o aumento de recursos, o

aumento na confiança científica e a integração do cidadão com o processo de pesquisa científica.

Ao relatar os impactos do compartilhamento de dados abertos, o relatório *“Open data: Unlocking innovation and performance with liquid information”* do McKinsey Global Institute (MGI, 2013) conclui que “se os dados abertos estivessem disponíveis, somente nos EUA, o PIB poderia ser aumentado em US \$ 3 a US \$ 5 trilhões por ano.”, além de indicar uma mudança na maneira como professores ensinam, alunos aprendem, como decisões são tomadas e empregos são procurados. Segundo o autor, os dados abertos se tornaram “um instrumento para quebrar as lacunas de informações em todos os setores econômicos, de bens e serviços e acadêmicos.”.

Nesse contexto, Molloy (2011) afirma que a corrente do Conhecimento Aberto exerce uma importante influência sobre o compartilhamento de dados de pesquisa já que por meio de seus preceitos os cientistas especificam quais são as permissões e quais direitos que terceiros tem para reutilizar os dados. Entretanto, se torna importante salientar que, apesar das vantagens para o compartilhamento de dados de pesquisa, há também fatores que impedem o seu fluxo.

Figura 11: Barreiras para o compartilhamento de dados



Fonte: DAMVAD baseado Tenopir et al. (2011), Enke et al. (2012), EC (2012), Kvale (2012), PARSE.Insight (2010). 2014. Tradução nossa.

O Grupo DAMVAD (2014) aponta que os principais fatores que exercem uma influência negativa sobre o compartilhamento de dados científicos são as questões legais, as questões sociológicas e as questões técnicas.

A despeito das vantagens no compartilhamento de dados de pesquisa, por serem itens que possuem características específicas, é necessário observar como o seu ciclo de vida atua como componente indissociável do processo de pesquisa científica.

2.3.6.3. Ciclo de Vida dos Dados de Pesquisa

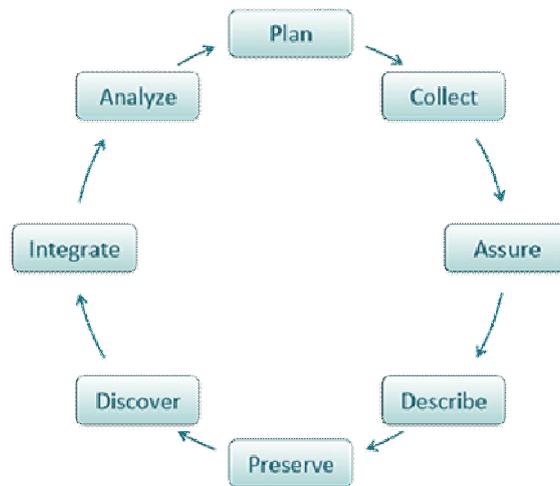
Sayão e Sales (2015) afirmam que “os dados e as coleções de dados de pesquisa possuem um tempo de vida maior que os projetos de pesquisa que os criaram. “. Isso possibilita que eles possam ser utilizados para outras pesquisas futuras além das que os originaram.

Biblioteca da Universidade de Queensland (Austrália) entende que “uma boa administração de dados de pesquisa é necessária para validar os resultados da pesquisa e manter a integridade dos resultados da pesquisa”. Em sua visão, o Gerenciamento de Dados de Pesquisa abrange as etapas de:

- Planejamento;
- Coleta
- Organização;
- Gerenciamento;
- Armazenamento;
- Segurança;
- Backup;
- Preservação e for fim
- O efetivo compartilhamento dos dados.

Em o DataONE forneceu graficamente um modelo de gerenciamento e preservação para dados de pesquisa, entretanto, alerta que a depender do campo científico, variações poderão ser observadas nas etapas descritas.

Figura 12: modelo de gerenciamento e preservação para dados de pesquisa



Fonte: DataOne

O ciclo descrito pela DataONE contempla oito etapas:

- Planejar: aborda a descrição dos dados e como estes serão compilados e como os dados serão gerenciados e disponibilizados ao longo de sua vida útil;
- Coleta: etapa onde observações são feitas e os dados são colocados em formato digital;
- Garantir: garante a qualidade dos dados por meio de verificações e inspeções
- Descrever: os dados são descritos com precisão e detalhadamente usando os padrões de metadados apropriados;
- Preservar: os dados são enviados para um arquivo de longo prazo apropriado;

Descobrir: dados potencialmente úteis são localizados e obtidos, juntamente com as informações relevantes sobre os dados (metadados);

Integrar: dados de fontes diferentes são combinados para formar um conjunto de dados homogêneo que pode ser prontamente analisado;

Análise: os dados são analisados.

Segundo o que afirmam os gestores da Biblioteca da Universidade de Sevilla, a correta gestão dos dados de pesquisa faz parte do processo de investigação no qual foram produzidos.

Figura 13: O ciclo dos dados científicos



Fonte: Biblioteca de la Universidad de Sevilla

Segundo o que é informado pelo ciclo dos dados científicos desenvolvido pela Biblioteca da Universidade de Sevilla, o ciclo de dados científicos se divide em quatro etapas básicas, que por sua vez se subdividem em outras ações:

1. Estabelecer um plano para gestão dos dados;
2. Atribuir metadados;
3. Padrões, armazenamento e cópias de segurança e
4. Acesso e uso, licenças e uso ético.

O ciclo descrito pela Biblioteca da Universidade de Sevilla e do DataONE se tornam importantes como elemento dessa dissertação pois eles elevam a discussão a importância da correta atribuição de metadados que descrevam o conteúdo dos dados que estão sendo disseminados. Este tópico voltará a ser objeto de observação quando o tema “repositório de dados” estiver em foco.

Por meio da percepção da importância da disseminação dos dados científicos para o avanço e validação científica, diversas áreas do

conhecimento demonstram adotar o compartilhamento de dados como forma de estimular a produção científica.

No entanto, mesmo com particularidades padronizadas, diferentes áreas do conhecimento já demonstram adotar certas particularidades no tocante ao compartilhamento de dados científicos. “As ciências físicas e astronômicas têm uma reputação bem estabelecida por serem disciplinas com uma forte cultura de compartilhamento de dados.”, conforme explica Pepe et al. (2014).

2.3.7. Contestações a respeito da Ciência Aberta

A despeito da corrente que acredita que a Ciência Aberta representa o futuro das publicações científicas, Mirowski (2018) aponta sérias falhas no movimento. O autor questiona em qual sentido a ciência estaria “fechada” e quais seriam as pessoas responsáveis por sua abertura. Ele acredita que o movimento de Ciência Aberta é um “artefato do atual regime neoliberal da ciência, que reconfigura tanto as instituições quanto a natureza do conhecimento, de modo a melhor se adequar aos imperativos do mercado.”.

Mirowski (2018) segue argumentando que uma das consequências do processo de abertura científico seria a geração de um ceticismo por parte da comunidade de religiosos em relação aos resultados obtidos. Para o autor, a comunidade de religiosos acredita que a ciência é irrelevante para as causas públicas e para os neoliberais que suspeitam que os cientistas atuam de acordo com as flutuações do mercado para produzir conhecimento.

Quanto a “Ciência Cidadã”, Mirowski (2018) afirma que seu conceito não remete ao que realmente seja convicto como democrático pois o processo de inserção científico é visto, pelo autor como uma contribuição de uma mão de obra não remunerada para gerar dados que serão processados pelos cientistas.

Pode-se observar a tristemente vazia noção de "democracia" que iguala uma quantidade maior de pessoas inscritas em papéis de apoio menores (e não remunerados) com um maior grau de participação democrática, quando, na verdade, eles servem principalmente como exército de reserva passiva de trabalho. (MIROWSKI, 2018, p. 177, tradução nossa).

Paralelo aos argumentos de Mirowski (2018) a maioria da literatura científica consultada para elaboração dessa dissertação adota uma postura a favor das práticas abertas de construção científicas que vigoram atualmente e se iniciaram a partir dos movimentos de Acesso Aberto.

A prática aberta, segundo Ortellado (2008, p. 187), possui tanto suas raízes nos antigos valores da ciência quanto nos recentes processos econômicos e tecnológicos que abalaram a estrutura da ciência, como a “crise dos periódicos”, que tomou as bibliotecas universitárias a partir do final dos anos 1980, e o advento da World Wide Web”.

O capítulo seguinte busca abordar quais foram as raízes do Acesso Aberto sobre dupla perspectiva: a internacional e a partir dos movimentos ocorridos no âmbito da América Latina.

2.4. MODELOS DE COMUNICAÇÃO CIENTÍFICA

Na perspectiva de revolução digital, as tecnologias de informação e comunicação desempenharam um papel fundamental no atual fluxo da comunicação científica. Conforme compreende Assante et al. (2015), os serviços das TIC's não são apenas destinados a apoiar a investigação científica, mas também a publicar e reutilizar os produtos resultantes de pesquisa. Os autores também destacam a importância dos pesquisadores poderem publicar resultados que possam ir além da literatura, como conjuntos de dados, experiências ou qualquer tipo de resultado de pesquisa que seja importante para a ciência.

No entanto, conforme afirma Russel (2001, p. 271), não se pode afirmar com precisão a qual categoria de comunicação a auto publicação corresponde, visto que ela possui características tanto dos meios formais e informais de comunicação. Sendo formal no que tange a universalidade do público que pode acessá-la e informal, já não faz parte de um formato estabelecido, como um periódico.

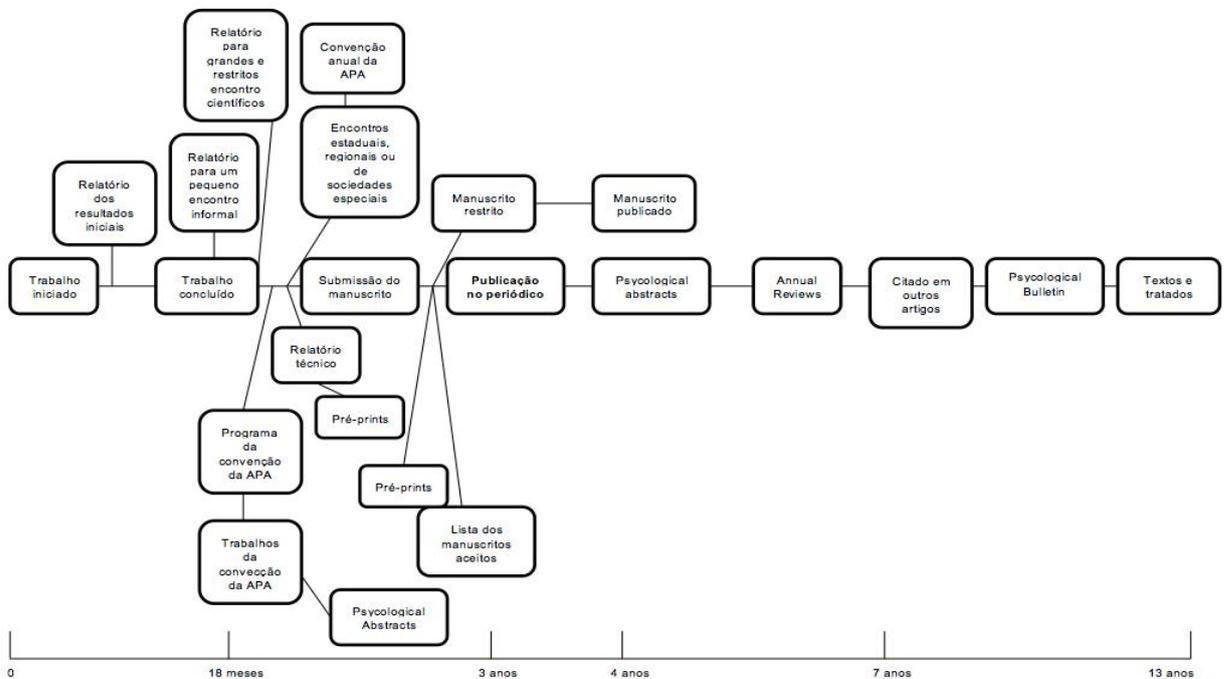
Como exemplo, Garvey e Griffith em 1979 já demonstravam o discernimento da necessidade retratar a desvinculação que ocorria no processo de comunicação científica dentro da área de Psicologia no que se referia ao canal formal de comunicação. Segundo os pesquisadores, os cientistas da área já reportavam os resultados iniciais de suas investigações de maneira informal em relatórios para a apreciação de pequenos simpatizantes para posteriormente, apresentar os resultados em colóquios.

Como resultado de suas intervenções, Garvey e Griffith (1979) ilustraram suas constatações em um modelo (Figura 3) que, segundo o raciocínio de Freitas (2015, p. 42), examina o processo tradicional de publicação científica, no qual o pesquisador submete a um periódico impresso um artigo com os resultados de suas pesquisas, que, se caso aprovado, será publicado transformando a ciência que ele continha em informação pública.

No entanto, na proposta de seu modelo INISIST, Garvey e Griffith (1979) não deixam de evidenciar os caminhos informais que as mesmas publicações

podem percorrer no que tange aos objetivos de serem comunicados e ao público atingido.

Figura 14: Modelo de comunicação científica de Garvey e Griffith



Fonte: Garvey, 1979 (com adaptações e traduzido, apud Costa 2017, p. 24)

A construção teórica de Garvey e Griffith em 1979 também foi analisada por Costa (2017, p. 25), entretanto a autora o fez de uma perspectiva temporal destacando que: “de acordo com os dados apresentados por Garvey e Griffith (1979), quando os artigos foram publicados em periódicos, 70% dos autores já haviam iniciado novas pesquisas e, desses, 60% já tinham até concluído um novo trabalho.”, indicando a lentidão no ciclo produção – compartilhamento – uso da informação científica.

A lentidão causada por esse sistema de disseminação do conhecimento atesta a real necessidade caminhos ou ações que o alterem fazendo com que ocorra a aceleração da transmissão da informação que deve sair de sua fonte e fluir até seus pares em enquanto há vida útil.

Como observa Hurd (2000, p. 1280), desde que o modelo proposto por Garvey e Griffith foi criado, as tecnologias de informação emergentes alteraram e aprimoraram as opções para comunicação inserindo o computador na dinâmica e por fim, resultando em sistemas de informação que não dependem

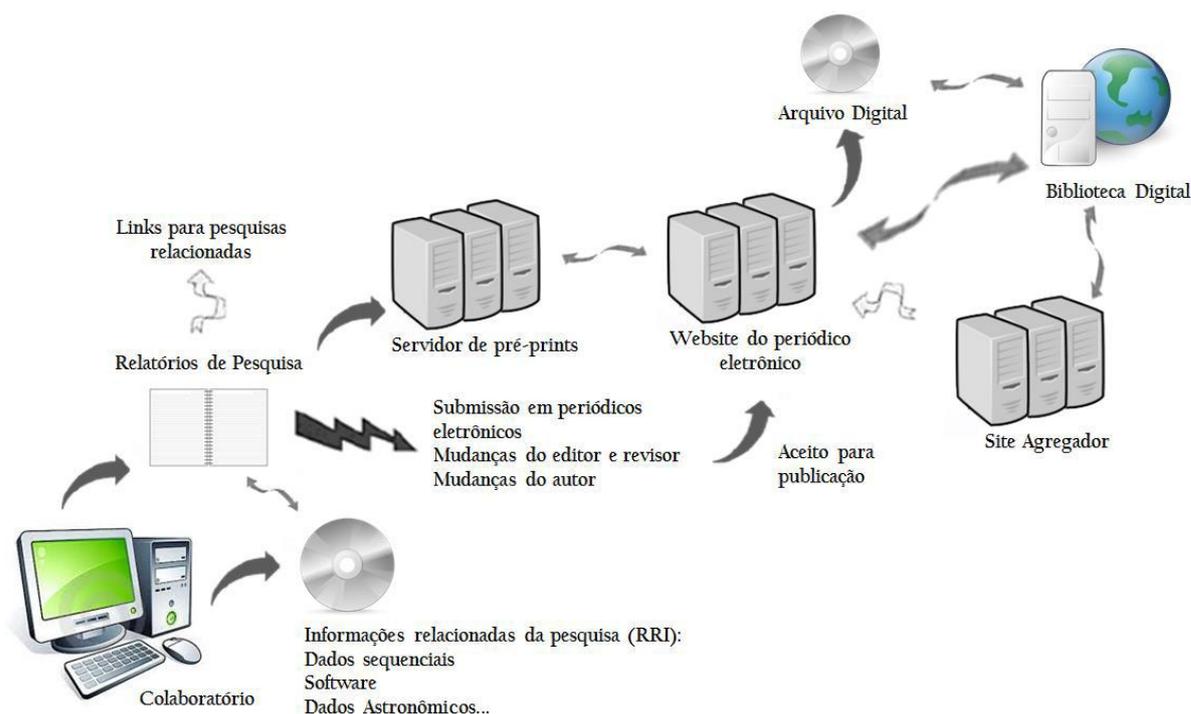
de papel para a sua existência, como é o caso dos periódicos eletrônicos. Um dos fatores que contribuíram para as mudanças nesse modelo clássico de comunicação foi o alto custo das assinaturas dos periódicos científicos aliado ao avanço das TIC's, conforme reafirmam Rosa e Gomes (2010, p. 21).

Historicamente o sistema tradicional de publicação científica tinha o papel como a base de seus fluxos, mas hoje, como fundamenta Hurd (2000, p. 1280) os computadores acoplados em rede permitem novas descobertas ao conectarem cientistas de diferentes partes do mundo, que trocam e analisam os resultados de suas pesquisas em tempo real. Segundo afirma a Organisation for Economic Co-Operation and Development – OECD – (2007, p. 9), os computadores e a internet criaram campos de atuação que vão além da aplicação dos resultados das pesquisas desenvolvidas, eles ampliaram sua atuação para as fontes de pesquisa do material base dos dados de pesquisa.

Apesar de destacar a profunda mudança causada pelo uso dos computadores na comunicação científica, ao desenvolver seu próprio arquétipo que retrataria como teria se desenvolvido o sistema de comunicação científica até o ano de 2020, Hurd (figura 2, 2000, apud Freitas 2015) pressupõe que as organizações tradicionais que participam do atual modelo como, editores e revisores, ainda estariam presentes, apesar de crer que novas organizações possam surgir durante o período de transição.

O modelo transformado do sistema de comunicação científica apontado por Hurd (2000) prevê um cenário onde o funcionamento da comunicação científica não está dissociado do processo de revisão por pares e as figuras dos colégios invisíveis continuam existindo. Seu conjunto de ideias representou a possibilidade do arquivamento de dados relacionados à pesquisa em servidores que poderiam ser acessados por terceiros que desejassem fazer uso dos dados, além de retratar uma cultura mais rápida de compartilhamento de informações científicas por meio dos servidores de pré-prints, publicação em periódicos eletrônicos e bibliotecas digitais.

Figura 15: Comunicação Científica - um modelo para 2020



Fonte: Hurd (2000, apud Freitas, 2015, p. 29)

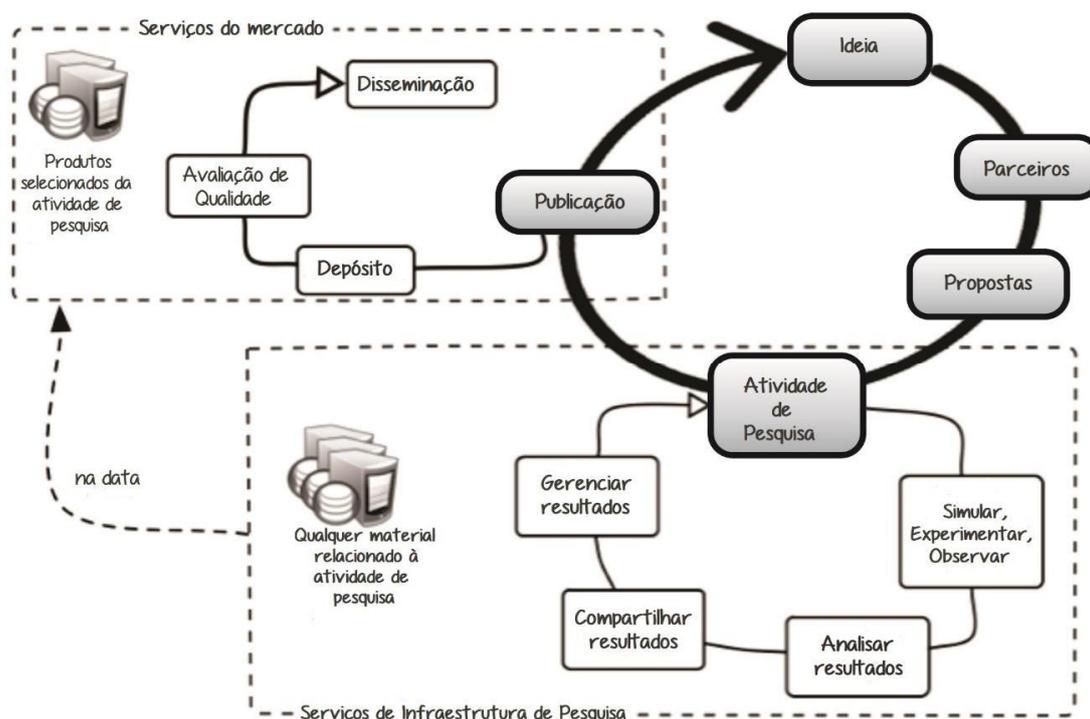
Quinze anos após Hurd, Assante et al. (2015) refletem que o ato da publicação científica possui uma nova interpretação holística. Neste novo ambiente descrito pelos autores os avanços das tecnologias de informação e comunicação possibilitaram que os pesquisadores pudessem escolher o tipo de publicação para os resultados de suas pesquisas que julgassem ser válidos, não somente os em forma de literatura, mas também conjunto de dados, experiências entre outros.

Para os autores os resultados positivos dessas práticas são evidentes:

1. Melhor interpretação dos resultados científicos.
2. Resultados de pesquisas mais rigorosos, possivelmente automatizados e evoluídos.
3. Práticas de recompensas científicas onipresentes
4. Maximização do reuso da pesquisa, desse modo reduzindo os custos da pesquisa. (ASSANTE et al. 2015, p. 2, tradução nossa).

Mediante as reflexões feitas por Assante (2015), a publicação do produto científico está inserida no ciclo de vida da pesquisa, conforme é ilustrado na figura 8.

Figura 16: Publicação no Ciclo de vida da Pesquisa



Fonte: Assante (2015, p. 9, tradução nossa).

Em consonância com o modelo desenvolvido por Assante (2015), o trabalho científico é realizado dentro das Infraestruturas de Pesquisa, que correspondem elementos relativos à:

- Organização, como papéis e procedimentos;
- Estrutura, como laboratórios e edifícios e a
- Tecnologia, como internet, telescópios e aplicativos.

Neste ambiente retratado, os pesquisadores, além de realizarem as suas pesquisas, fazem uso de produtos científicos existentes para produzirem seus trabalhos, conforme afirma Assante (20015). Segundo o mesmo autor, os serviços de TIC que permeiam o ambiente de infraestrutura de pesquisa se destinam não somente a investigação científica, mas também a publicação e reutilização dos produtos de investigação deles resultantes.

De acordo com a conclusão de Costa (2000, p. 101) a introdução de novas tecnologias de informação no ciclo de comunicação científica levou à proposição de um modelo híbrido onde os meios impressos e eletrônicos coexistem e variam de acordo com a área do conhecimento que se manifestem.

Em consonância com essa corrente de raciocínio, Russel (2001, p. 272) aponta que as publicações eletrônicas estão sendo vistas como possível alternativa para os autores que se sentem incapazes de satisfazer suas demandas de produção da informação por meio da publicação tradicional em periódicos. Ainda segundo Russel (2001, p. 273), as universidades poderiam desempenhar um papel vital na publicação dos resultados de pesquisas por meio de uma imprensa universitária, o que reduziria a quantidade de dinheiro gasto com a aquisição de periódicos.

Na visão de Russel (2001, p. 273):

Outros autores sugeriram que os acadêmicos de todas as disciplinas poderiam "publicar" seus artigos sobre os arquivos abertos de um "Sistema de Comunicação Acadêmica", uma rede eletrônica onde as universidades retomariam o controle sobre quem e o que é "publicado". Os conselhos de avaliação com membros nomeados e eleitos pelas instituições participantes seriam responsáveis por classificar cada artigo de acordo com as contribuições que ele faz no campo especializado antes de entrar no sistema. (RUSSEL, 2001, p. 273, tradução nossa).

2.4.1. O compartilhamento de dados e os novos modelos de Comunicação Científica

Conforme explica Klump et al., "o conhecimento científico é comunicado por meio de uma literatura científica", a sessão anterior dessa dissertação se dedicou a precisar os modelos de comunicação que poderiam abranger as formas pelas quais essa literatura é produzida e disseminada. No entanto, os mesmos autores ainda afirmam que "o conhecimento é fundamentalmente derivado de dados", sendo que as mesmas influências tecnológicas observadas nos modelos de comunicação científica anteriormente descrito também pôde ser observado nas formas de coleta e análise dos dados científicos.

Em paralelo ao pensamento de Hurd e Russel sobre a influência que as tecnologias podem ter sobre o sistema de comunicação científica, Assante (2015) contempla que o atual estado evolutivo das TIC's modernizou a forma de armazenamento, acesso e compartilhamento de dados, ultrapassando as

expectativas em termos velocidade e resiliência dos dados e mudando a forma como a informação é permutada e entregue a outros cientistas, sendo um exemplo, o uso de ferramentas que vão além da publicação de artigos em formato digital, possibilitando a coleta, curadoria e criação de dados de pesquisa.

A tarefa de definir o que são “dados” foi tomada por diversos autores, cada qual com sua diferente perspectiva. O Comitê Consultivo de Sistemas de Dados Espaciais (CCSDS) definiu “dados” como sendo:

Dados: uma representação reinterpretabil da informação em uma maneira, adequada para comunicação, interpretação ou processamento. Exemplos de dados incluem uma seqüência de bits, uma tabela de números, os caracteres em uma página, a gravação de sons feitos da fala de uma pessoa ou um espécime de rocha da lua. (COMITÊ CONSULTIVO DE SISTEMAS DE DADOS ESPACIAIS, 2012, p. 10, tradução nossa).

Entretanto, em uma perspectiva social, Borgman (2008, p. 30) afirma que a noção do que venha a ser dados é uma variável que depende do “quem”, “quando” e para que propósito, já que uma evidência informacional não se comprova útil para todas as áreas do conhecimento.

Em geral, os dados podem ser diferenciados em: “dados estatísticos, geográficos, de transporte, meteorológicos, financeiros, ambientais, governamentais, científicos e culturais”, conforme alegam Hernández-Pérez e García-Moreno (2013, p. 259). Os autores ainda afirmam que as tecnologias são responsáveis pela grande quantidade de dados coletados, sendo que a rápida coleta de dados é um dos fatores que estimulam a produção de novos conhecimentos científicos.

A Comissão Europeia de direções-gerais para investigação, inovação (RTD), redes de comunicações de conteúdo e tecnologia (CONNECT) (2014) observa que a transição do sistema de pesquisa científica é impulsionada pelos pesquisadores que integram o sistema digital, pela globalização da comunidade acadêmica, bem como pela crescente demanda social para enfrentar as demandas sociais dos nossos tempos, o que por sua vez se reflete em todo ciclo de pesquisa, reorganizando as relações existentes entre os atores envolvidos no sistema.

O modelo tradicional do ciclo de pesquisa deixa de reproduzir com exatidão as ações desenvolvidas nos processos de comunicação científica já que novas dinâmicas foram desenvolvidas visando à resolução das próprias necessidades dos atores da comunidade científica. As novas dinâmicas criadas entre os atores da comunidade científica, associadas ao crescente uso das tecnologias de comunicação, possibilitaram uma nova cultura de circulação de conhecimento onde se destaca a colaboração entre cientistas e o compartilhamento de dados científicos de forma aberta.

Sobre os novos espaços ocupados pela ciência e as novas dinâmicas do sistema científico Abdo (2014) reflete que:

[...] a produção de conhecimento está consistentemente ocupando espaços não institucionais, fora de universidades, dos institutos de pesquisa, e mesmo das empresas, num processo que desafia termos como “extensão”, “parceria” ou mesmo “colaboração”. (ABDO, 2014, p. 462)

Um dos motivos mais plausíveis para justificar a colaboração entre os cientistas não se encontra em uma atividade altruísta, mas sim na percepção do benefício mútuo, na qualidade da informação, na reputação do próprio pesquisador e até mesmo na arrecadação de financiamentos para pesquisas, conforme reflete Andrade (2015).

Costa (2017, p 37) analisa as iniciativas internacionais de colaborações científicas e chega à conclusão que a organização virtual dos cientistas em projetos que necessitavam de um comportamento colaborativo para seu desenvolvimento deu início a uma ciência colaborativa que envolviam diferentes recursos de diferentes países, o que contribuiu para a virtualização da ciência e a transformação do pensamento científico.

Andrade (2015) também conjecturou que a “comunicação e a colaboração estão diretamente ligadas à produção de conhecimento”, que por sua vez se liga, não somente aos seus resultados, mas também e principalmente a sua construção.

A coerência do pensamento de Andrade (2015) é reforçada quando interligado ao raciocínio de Borgman (2008, p. 30) que sustenta que “dados são a ‘cola’ que une a pesquisa colaborativa”. Segundo Borgman (2008, p. 30) os

pesquisadores trabalham em conjunto gerando dados, e esses dados, são o produto essencial desta colaboração.

O compartilhamento de dados científicos é facilitado por meio do depósito dos dados coletados em infraestruturas tecnológicas onde os cientistas disponibilizam os seus próprios dados e podem acessar os das pesquisas de seus pares, segundo afirma Kim (2013).

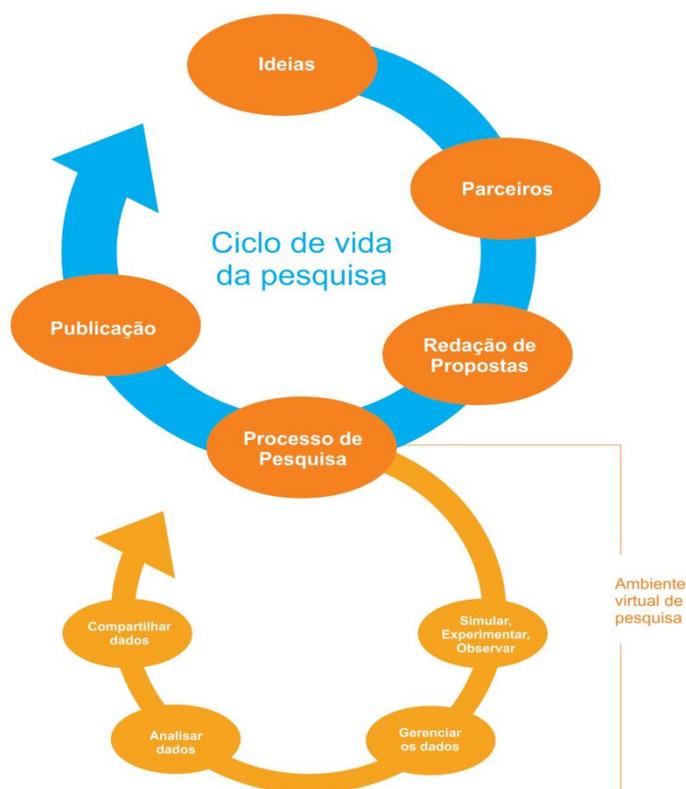
Os cientistas Gómez, Méndez e Hernández-Pérez (2016, p. 547) acreditam que cada disciplina das áreas do conhecimento interpretam o conceito de dados ou *datasets* de pesquisa, sua natureza e como se agrupam de uma forma singular. Entretanto, conforme acredita Murray-Rust (2008), os “dados em si são um recurso comunitário que ganha cada vez mais importância”. Aos olhos do autor, a ciência avança melhor quando um maior número de cientistas tem acesso aos dados de pesquisas anteriores já que isso evita a repetição no custeio da pesquisa e permite a reinterpretação dos mesmos dados para uma nova pesquisa.

Esta corrente de pensamento é reafirmada por Tenopir et al. (2011) quando constatam que os dados são a infraestrutura da ciência e a medida em que a ciência se torna mais intensa e colaborativa, o compartilhamento de dados de pesquisa se torna mais importante.

“O compartilhamento de dados é uma parte valiosa do método científico, permitindo a verificação dos resultados e ampliando a pesquisa de resultados”, segundo Tenopir et al (2011). O compartilhamento de dados científicos reestrutura o modo clássico do ciclo de comunicação científica “e abre novas formas de interlocução e de socialização no mundo científico, além de contribuir para a racionalização dos recursos financeiros públicos aplicados na pesquisa científica”, segundo a visão de Sayão e Sales (2014, p. 78).

O Joint Information Systems Committee (2010) projetou um diagrama para ilustrar o ciclo de vida de uma pesquisa científica (Figura 9), onde se é possível observar modo como o compartilhamento de dados está inserido no ciclo de vida da pesquisa, que se encontra, mais precisamente, durante o processo de pesquisa.

Figura 17: Diagrama do ciclo de vida da pesquisa



Fonte: Joint Information Systems Committee (JISC) (2010, tradução nossa).

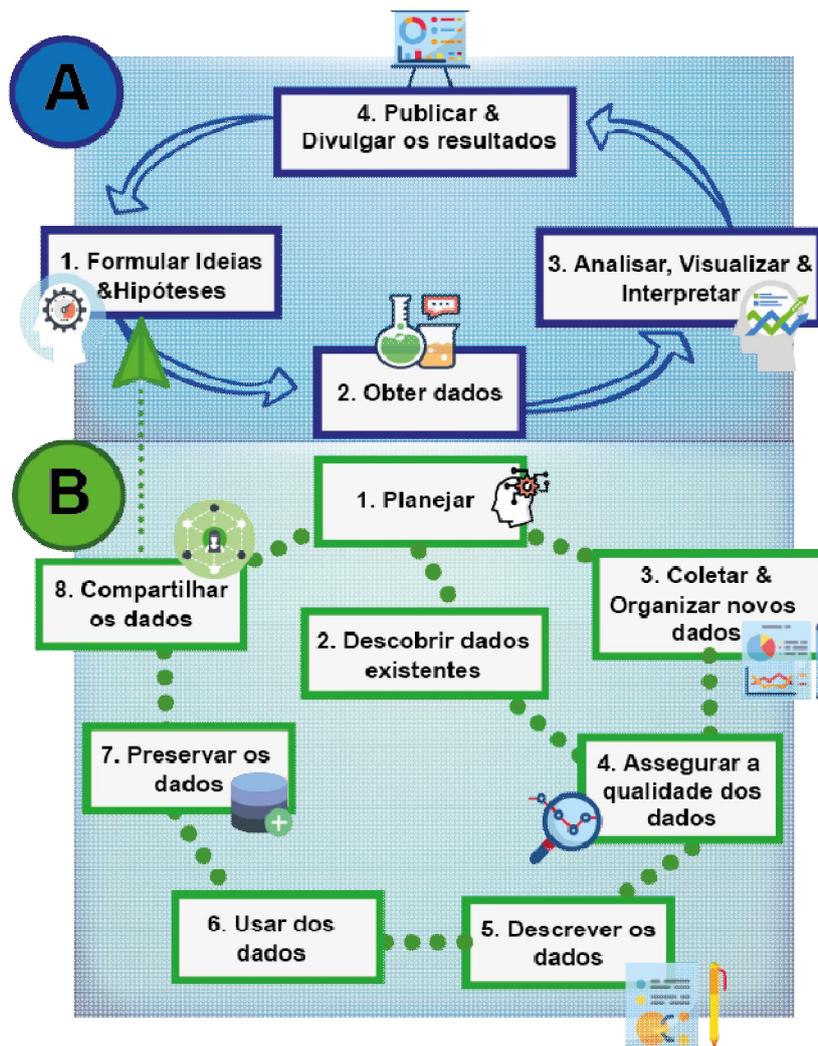
No ideal proposto, o ciclo de vida da pesquisa científica se inicia com uma ideia, passando pelas etapas de obtenção de parcerias e de redação de propostas, que segundo o JISC (2010), normalmente é uma exigência da maioria dos financiadores, chegando, enfim, no processo de execução da pesquisa onde os dados científicos são obtidos por meio de simulações, experiências e observações, para que possam ser gerenciados, analisados e compartilhados com outros membros da comunidade científica.

O resultado desse processo segue, então, para a publicação de material científico que servirá de matéria embasadora para o surgimento de novas ideias que darão início novamente ao mesmo ciclo.

Apesar do modelo proposto por JISC apresentar uma nova dimensão ao aprofundar o processo de execução da pesquisa, no que é tocante ao compartilhamento de dados de pesquisa, ele possui uma lacuna referente a orientações e formas de compartilhamento e preservação dos dados que foram compartilhados.

Essa lacuna pode ser preenchida por meio da compreensão de parte do modelo elaborado por Michener (2015). Nesse ideal, o ciclo de vida da pesquisa é associado ao ciclo de vida dos dados de pesquisa, que fornece informações relacionadas a afinidade inexoravelmente existente entre as duas etapas do trabalho científico:

Figura 18: Relacionamento do ciclo de vida da pesquisa (A) com o ciclo de vida do dado (B)



Fonte: Michener com adaptações (2015, p. 2, tradução nossa).

Para Michener (2015) os fomentadores da atividade científica, cada vez mais, reconhecem o valor dos dados de pesquisa e por essa razão, passam a exigir a apresentação de planos de gerenciamento de dados como parte da proposta a ser financiada, sendo, portanto, necessário que no transcorrer da pesquisa, que entre outros elementos:

- Se determine os requisitos e padrões do patrocinador da pesquisa, já que, seguindo o autor, diferentes comunidades de pesquisa normalmente possuem diferentes padrões para gerenciar e disseminar seus dados.
- Sejam identificadas quais serão as naturezas dos de dados a serem coletados, que pode variar segundo tipo, fonte, volume e formatos de arquivo.
- Definir como os dados serão coletados, organizados e gerenciados, que pode variar desde a criação de tabelas a demanda de um sistema de gerenciamento de banco de dados, como por exemplo um repositório específico para essa necessidade.
- Seja assegurada a qualidade do dado coletado;
- Descrever os dados;
- Utilizar os dados previamente coletados e
- Que se já assegurada a preservação dos dados para posterior compartilhamento tomando como pontos relevantes:
 - Por quanto tempo esses dados deverão ficar disponíveis
 - Como deverá acontecer o armazenamento dos dados e quais serão os parâmetros utilizados;
 - Como os dados armazenados serão disseminados no futuro.

Russel (2001, p. 271) considera que o intercâmbio de opiniões e dados se torna essencial na fase experimental da pesquisa. Por meio do modelo proposto pelo autor, compreende-se que o compartilhamento de dados de pesquisa é um elemento que suscita a formulação de novas hipóteses e ideias. Sayão e Sales (2014, p.80) expressaram o mesmo raciocínio a respeito do re- uso dos dados ao afirmar:

A ciência como um todo avança com maior qualidade, menor custo e mais eficiência quando abre a possibilidade para que o maior número possível de pesquisadores disponha de vias de acesso aos dados acumulados por seus antecessores e contemporâneos. Isso evita, objetivamente, o custo da duplicação de esforços e permite novas interpretações em diferentes contextos científicos para esses dados

e, além do mais, permite que eles sejam integrados e retrabalhados de forma mais criativa, descortinando horizontes para novas pesquisas. (SAYÃO E SALES 2014, p.80).

Por razão do compartilhamento de dados científicos se configurar com tanta importância para a manutenção do ciclo da pesquisa científica, as ações relacionadas ao planejamento, organização e preservação tornam-se cruciais no âmbito deste novo paradigma, pois elas proveem suporte para outros atos relacionados ao reuso dos dados de pesquisa pelos pesquisadores das mais diversas áreas do conhecimento.

O re-uso de dados já é observado em diversas áreas do conhecimento, entretanto, há variações que são observadas nos campos científicos quanto a intensidade, a finalidade do compartilhamento e re-uso dos dados. Sayão e Sales (2012, p. 179) afirmam que essa diversidade pode ser causada devido à noção de que sejam dados científicos varie consideravelmente em cada área do conhecimento e que dentro dessas áreas, os dados são gerados com propósitos diferenciados.

Murray-Rust (2008) também observaram a dissemelhança entre as áreas do conhecimento no que é referente a publicação e re-uso de dados entre as disciplinas. Com base em informações fornecidas pelo autor, algumas disciplinas têm, por tradição, o costume de exigir que pesquisas que foram financiadas com recursos públicos tenham seus dados publicados e posteriormente anexados em alguns bancos de dados, enquanto outras já disponibilizam seus dados universalmente. Essas diferenças podem ser observadas e classificadas dentro das "*large science*" e "*small science*":

Outras [disciplinas] na "*large science*" têm uma política de reutilização de dados bem desenvolvida e exigem que dados de telescópios, satélites, aceleradores de partículas, fontes de nêutrons, etc. sejam disponibilizados universalmente para reutilização. Nessas áreas, as normas são geralmente suficientes para a prática de Dados Abertos. Nas "*small sciences*", em contraste, a unidade de pesquisa é o laboratório ou indivíduo ("pequeno" não reflete a importância da disciplina que pode ser numericamente muito grande). Essas disciplinas tipicamente resultam em muitas publicações independentes que relatam experimentos individuais (eu cunhei o neologismo "hipopublicação" para expressar a natureza desarticulada dessas informações). (MURRAY-RUST, 2008, [s.p.], tradução nossa, grifo nosso).

Nota-se um fenômeno onde, não só as áreas da ciência começaram a trabalhar colaborativamente, mas com as particularidades relacionadas às diferenças que podem ser observadas em cada área do conhecimento. Segundo afirma Covi (2000, p. 1284) foi a habilidade de realizar trabalhos colaborativamente além das barreiras da distância e tempo que criou novas oportunidades para que os pesquisadores pudessem compartilhar recursos escassos como instrumentos e os conjuntos de dados resultantes das pesquisas.

As características de compartilhamento referentes aos resultados de pesquisa que cada área possui ampla complexidade e por essa razão será estudado com maior profundidade no capítulo a seguir.

2.5. AS DIFERENÇAS CULTURAIS CIENTÍFICAS NA TRANSMISSÃO DO CONHECIMENTO

Os diferentes objetos de estudo e métodos científicos forneceram diversas categorias que possibilitaram a diferenciação das disciplinas acadêmicas, no entanto, as diferenciações que podem ser estipuladas para essas áreas não se resumem a esses padrões, também foi observado que diferentes áreas do conhecimento também possuem condições específicas para o estudo e para a comunicação dos resultados de suas pesquisas científicas.

Alguns aspectos do campo científico se aplicam de forma geral para todos os pesquisadores, no entanto, outras características são bem delimitadas a alguns pesquisadores e suas respectivas áreas acadêmicas, como por exemplo, o estágio da carreira em que se encontram, os métodos de pesquisa utilizados e os dados gerais, conforme acreditam Eynden et al. (2016).

Mueller (2000, p. 23) informa que apesar da literatura produzida por diferentes áreas do conhecimento apresentar particularidades que as diferenciam de acordo com seus temas e abordagens, essa literatura também possui características comuns por sofrerem influências de um conjunto de fatores comuns ao grupo.

“Comunidades científicas podem então ser definidas como o agrupamento de pares que compartilham um tópico de estudo, desenvolvem pesquisas e dominam um campo de conhecimento específico, em nível internacional.”, conforme determina Costa (200, p. 88):

Dentre as várias dessemelhanças que podem ser observadas quanto a construção e transmissão do conhecimento no âmbito das comunidades científicas - que compreendem seus atores, hábitos e ambientes - estudos relacionados aos meios e práticas do compartilhamento da informação nessas diferentes áreas do conhecimento são objeto de estudo dos cientistas da informação. Essas pesquisas possuem como enfoque questões quanto a percepção dos atores, infraestrutura e meios de produção e disseminação do conhecimento produzido.

Pinheiro (2014) que reflete que as diferenças encontradas nas áreas do conhecimento são tidas como um fenômeno natural, mesmo se tratando da comunicação científica dentro dessas classificações. Costa (2000, p. 89) ainda ressalta que além das diferenças encontradas dentro das mesmas áreas de uma mesma divisão do conhecimento – também chamadas de subdisciplinas pela autora – existem aquelas que variam entre diferentes países e as diferenças intradisciplinares.

Snow (1905, p. 51) ensina que mesmo sendo admissível englobar cientistas puros e aplicados na mesma cultura científica, há um grande fosso e vasto entre eles, sendo que, muitas vezes, os cientistas puros e engenheiros divergem totalmente entre si, pois seus comportamentos tendem a ser muito diferentes.

Em concordância com o pensamento traçado por Meadows (1999, p. 61), se as diferenças entre matérias e disciplinas são significativas, é comum que tais diferenças também sejam observadas nos padrões de comunicação entre os acadêmicos das comunidades de tais áreas. Sendo também comum que áreas distintas, e mesmo dentro de uma área, apresentem diferentes perspectivas no tocante a abordagem a um conteúdo, o que faz com que formas diferentes de comunicação que se adaptem às suas necessidades sejam demandadas, conforme afirma Mueller (2005).

Essa corrente de pensamento também foi seguida pelos autores Eason et al. (1997). Os autores acreditam que, como exemplo, o uso de periódicos

eletrônicos como meio de comunicação do conhecimento se destaca mais em algumas áreas do que em outras, mesmo estando presente em toda a estrutura de atividades acadêmicas.

Costa (2000, p. 85) analisa e identifica as mudanças que foram observadas nos sistemas de comunicação ressaltando as relações entre pesquisadores, pares e recursos de informação e comunicação. A pesquisadora resalta as novas interações dentro das comunidades científicas facilitadas pelas tecnologias de comunicação concluindo que há indicadores que apontam que quanto mais exata a disciplina, mais alto será o uso de tecnologia, lembrando, evidentemente que a proporção de recursos recebidos pelas áreas do conhecimento não é proporcional, sendo que as ciências naturais tradicionalmente tendem a receber mais recursos que as sociais e humanidades.

A exemplo da informação fornecida por Costa (2000), Covi (2000, p. 1284) cita dados que constata que jovens oceanógrafos que utilizaram uma nova forma de comunicação eletrônica (SCIENCEnet), mesmo estando geograficamente distantes conseguiram realizar a permuta de dados de pesquisa e obter o reconhecimento científico de seus pares.

Como pôde ser descrito ao longo dessa dissertação, a busca pelo reconhecimento dos demais cientistas e comunidades acadêmicas pode levar os cientistas a adotarem diversas estratégias para a disseminação do resultado das suas pesquisas, entre eles, os canais formais ou informais de comunicação.

Algumas dessas diferenças disciplinares foram elencadas por Eason et al. (1997) em suas pesquisas acerca das diferenças disciplinares quanto a adoção do periódico eletrônico em diferentes áreas do conhecimento:

Quadro 3: O uso da comunicação informal nas áreas do conhecimento

| | | |
|--|--|--|
| <p>Em áreas onde os profissionais necessitam se manter rapidamente atualizados.</p> | <p>A adoção de mecanismos informais é mais constante, sendo utilizados para divulgar resultados preliminares, sendo assim, o</p> | <p>A comunicação pode ser estabelecida por meio de conferências, mas, em algumas ciências, muitas vezes é feita através da</p> |
|--|--|--|

| | | |
|---|--|--|
| | tempo gasto com publicações revisadas por pares é mais longo. | distribuição de pre-prints. |
| Em áreas em que são mais comuns a realização de pesquisas individuais em assuntos especializados. | Normalmente, redes, reuniões nacionais e internacionais são consideradas vitais para manutenção do contato entre os acadêmicos, visto que poucas pessoas no mundo que têm interesses semelhantes. | |
| Nas áreas onde a velocidade do desenvolvimento alto que os cronogramas de publicações em geral se torna ineficaz para a divulgação dos resultados. | As publicação em periódicos funcionam principalmente para desempenhar o papel de arquivos históricos que garantem que o trabalho seja armazenado de forma adequada e permanente. | A real troca de informações acontece por meio de artigos de conferências que evitam grandes atrasos. |
| Acadêmicos se encontram sobrecarregados e não têm tempo para se envolver em buscas formais sobre assuntos significativos em seus campos. | Nesses casos os métodos de divulgação formais e informais são utilizados ao mesmo tempo para finalidades distintas. Os canais informais são usados para identificar o que pode ser encontrado com maior profundidade em periódicos e livros. | Essa prática possui dupla vantagem: economiza tempo e esforços enquanto e realizar uma verificação da real qualidade do material desconhecido, já que é recomendado por pares. |
| Em muitas das disciplinas aplicadas, busca por informações se dá na literatura | A literatura cinzenta pode ser considerada como fonte tão importante quanto os registros formais, por parte, | São considerados, nesses casos, estudos de casos, conclusões práticas, "dicas e sugestões" e etc. - que são |

| | | |
|-----------------|---|---|
| cizenta. | por ser mais atual. e isso tendia a ser a literatura cinzenta. A literatura formal era importante, mas não deveria receber "ênfase indevida". | úteis para os profissionais, mas não necessariamente aptos para um periódico acadêmico. |
|-----------------|---|---|

Fonte: Elaboração própria a partir das informações fornecidas por Eason et al. (1997).

Meadows (1999, p. 61) nos leva a crer que essas relações também possam ser medidas por meio da análise das referências que os trabalhos científicos apresentam em relação aos outros. O autor explica que essa interação pode ser medida pelo efeito de proximidade¹⁷ estudado por Prince.

Ao explorar as repercussões trazidas por essa corrente teórica Meadows (1999, p. 62) reflete tais resultados sobre a natureza cumulativa da informação concluindo na percepção de outra diferença observada entre os campos científicos:

Como resultado desse processo, os cientistas em geral precisam, ao realizar seus próprios estudos, estar cientes apenas dos trabalhos recentes. Ao contrário, as informações nas ciências sociais com frequência são menos facilmente codificadas, de modo que a literatura antiga continua sendo mencionada. As humanidades constituem em caso especial, pois a literatura antiga representa para elas muitas vezes a matéria-prima de suas investigações. (MEADOWS, 1999, p. 62).

Hurd (2000, p. 1282) também observa particularidades na transmissão e uso do sistema de comunicação científica nas áreas do conhecimento. Para a autora possivelmente algumas disciplinas específicas apresentarão variações quanto o uso de ferramentas inovadoras.

Já as pesquisas feitas por Russel (2001, p. 274) constataam que certas áreas do conhecimento apresentam facilidade para adotar tecnologias de informação e comunicação, sendo, portanto, outra variante entre as áreas

¹⁷ Meadows explica que o índice desse efeito é obtido por meio da proporção total das referências em um campo com a literatura publicada nos últimos cinco anos. Essa proporção também é conhecida como 'Índice Price'.

Meadows elucida mais detalhes sobre esse índice nas páginas 61 e 62 de seu livro.

científicas. O autor considera que é esperado que algumas práticas científicas sejam mais afetadas por novas tecnologias e outras em apenas certos campos.

Um possível fator que elucide a facilidade da adoção e utilização de novos mecanismos tecnológicos, que é um dos fatores marcantes dessa nova geração foi que, presumidamente, as gerações de jovens que nasceram entre os anos de 1970 e 1980 aprenderam a utilizar computadores ou softwares educacionais adquirindo habilidades nas tecnologias de comunicação eletrônicas que as anteriores não desenvolveram, e isso impacta diretamente nas práticas de construção do conhecimento científico colaborativamente por meio de instrumentos e laboratórios virtuais, conforme observa Covi (2000, p. 1285).

Essas habilidades eletrônicas adquiridas por essa geração também possuem impacto positivo quanto a aceitação de publicações eletrônicas como meio de divulgação dos resultados obtidos, bem como na maior aceitação do compartilhamento dos dados científicos provenientes de tais pesquisas.

Há certos campos onde o periódico científico não possui a mesma relevância científica outrora foi observada, essas áreas do conhecimento se aproveitam das facilidades oferecidas pelas práticas de comunicação inovadoras para mudar, não somente as práticas de comunicação científica, mas também a maneira como as pesquisas são realizadas, conforme observa Russel (2001).

Eason et al. (1997) afirmam que, assim como as práticas acadêmicas, o armazenamento, uso e disseminação da informação variam consideravelmente entre as áreas do conhecimento, também é possível inferir que a necessidade de revistas eletrônicas também é um ponto variante entre as áreas, sendo que esses hábitos podem ter diferentes efeitos na prática acadêmica.

Uma pesquisa realizada por Mueller em 2005 pesquisou quais são os canais que diferentes áreas do conhecimento adotam para realizar a comunicação de suas pesquisas para seus pares. Segundo a autora, “os dados foram tratados de modo a mostrar preferências por canais e frequências anuais de publicação”.

O período de abrangência do estudo está entre os anos de 1995 a 2002 e buscou investigar o comportamento de publicação dos resultados de pesquisa dos bolsistas dos programas de pós-graduação brasileiros que

possuísem bolsas de estudos pagas pelas agências Capes e CNPq, somando no total 226 pesquisadores na amostra analisada por Mueller (2005).

Quadro 4: Análise Mueller 2005

| Quanto as áreas do conhecimento analisadas | Quanto aos tipos de publicações analisados |
|---|---|
| <ul style="list-style-type: none"> Foi adotada neste trabalho a classificação estabelecida pela CAPES¹⁸, que agrupa as ciências e especialidades científicas e tecnológicas em oito grandes áreas do saber. | <ul style="list-style-type: none"> Artigos publicados em periódicos nacionais; Artigos publicados em periódicos estrangeiros; Livros e capítulos de livros; Trabalhos apresentados em congressos nacionais e; Trabalhos apresentados em congressos estrangeiros. |

Fonte: Adaptado do texto de Mueller 2005.

Em relação aos pesquisadores selecionados para o estudo, a autora afirma que sua pesquisa os distribuiu de acordo com as áreas do conhecimento delimitadas pela tabela da CAPES que se encontrava vigente durante o período da pesquisa. Nenhuma predileção por áreas foi feita durante a seleção. A relação dos pesquisadores e suas respectivas áreas do conhecimento foi detalhada na tabela que segue:

Quadro 5: Distribuição de pesquisadores por Área (n=226)

| Grandes áreas do saber | Quantidade de pesquisadores |
|-------------------------------------|-----------------------------|
| Ciências Exatas e da Terra | 64 |
| Ciências Biológicas | 16 |
| Engenharias | 28 |
| Ciências da Saúde | 21 |
| Ciências Agrárias | 15 |
| Ciências Sociais Aplicadas | 15 |
| Ciências Humanas | 53 |
| Linguística, Letras e Artes. | 14 |
| Total | 226 |

¹⁸ São consideradas pela Capes as grandes áreas: Ciências Exatas e da Terra; Ciências Biológicas, Engenharias; Ciências da Saúde; Ciências Agrárias; Ciências Sociais Aplicadas; Ciências Humanas e Linguística, Letras e Artes.

As subdivisões exatas de cada grande área pode ser visualizada e melhor analisada no documento fornecido pela Capes que pode ser acessado por meio do link: <<http://www.capes.gov.br/avaliacao/instrumentos-de-apoio/tabela-de-areas-do-conhecimento-avaliacao>>. Acesso em: 30 jan. 2018.

Fonte: Adaptado do texto de Mueller 2005.

Os resultados obtidos por Mueller (2005) foram reunidos e representados no quadro 5. Em uma abordagem não exaustiva sobre as peculiaridades dos domínios do conhecimento, apontou o seguinte estado da arte:

Quadro 6: Total de documentos por tipo e grande área do saber

| <i>Grande área</i> | Periódico Estrangeiro | Periódico Nacional | Anais de Congresso Estrangeiro | Anais de Congresso Nacional | Livros | Capítulos de Livros |
|--|-----------------------|--------------------|--------------------------------|-----------------------------|--------|---------------------|
| <i>Ciências Exatas e da Terra</i> | 516 | 282 | 133 | 238 | 13 | 20 |
| <i>Ciências Biológicas</i> | 152 | 71 | 28 | 31 | 2 | 22 |
| <i>Engenharias</i> | 109 | 111 | 409 | 589 | 12 | 45 |
| <i>Ciências da Saúde</i> | 198 | 199 | 14 | 71 | 20 | 92 |
| <i>Ciências Agrárias</i> | 49 | 330 | 33 | 190 | 27 | 48 |
| <i>Ciências Sociais e Aplicadas</i> | 46 | 104 | 21 | 76 | 36 | 76 |
| <i>Ciências Humanas</i> | 114 | 479 | 69 | 232 | 201 | 282 |
| <i>Linguística, Letras e Artes</i> | 49 | 185 | 32 | 30 | 52 | 101 |

Fonte: Mueller (2005)

Por meio dos resultados obtidos pela pesquisa de Mueller (2005), é possível comprovar a variante preferência das áreas acadêmicas por canais de publicação científica, como foi discorrido pela literatura. A preferência pela publicação dos resultados de pesquisa em periódicos pôde ser observada nas áreas de Ciências Exatas, Biológicas, e da Saúde. Demonstrando ser menos usado nos campos de Ciências Sociais Aplicadas, Linguística, Letras e Artes Ciências Agrárias e Humanas.

Em relação a extensão e intervalo de publicações em periódicos nas áreas do conhecimento, Eason et al. (1997) afirmam que dentro das “ciências naturais, na maior parte, os artigos são curtos e a publicação é frequente, nas

ciências humanas e sociais são mais longos e menos frequentes. ” Sendo para que para os autores se torna mais comum que as áreas de humanidades e ciências sociais sejam mais dedicadas e realizar as suas publicações livros e monografias.

As áreas de Ciências Humanas, Ciências Sociais Aplicadas e Lingüística, Letras e Artes, conforme os dados apresentados no quadro 6, comprovam seguir as indicações feitas pela literatura tendo o número de suas publicações mais significativas em veículos como livros e capítulos de livros. Eason et al. (1997) argumentam que em campos como os das ciências sociais e humanidades existem poucas pessoas trabalhando no mesmo assunto, o que exige dos cientistas uma apresentação do contexto da pesquisa, além de uma justificativa e suas conclusões, o que torna os livros e longos artigos veículos mais apropriados ao conteúdo das suas pesquisas.

Hurd (2000, p. 1282) acredita que até mesmo o uso de textos pré-prints pode ser considerado como fator para a diferenciação entre os campos científicos, já que seu uso como canal de comunicação científica também possui diferentes graus de adoção nas diferentes áreas do conhecimento.

Alguns observadores sugeriram que as culturas de pre-print não se espalham em especialidades em que patentes sobre produtos e processos são a norma. Em uma era digital, podemos esperar ver impressões eletrônicas, o equivalente eletrônico dos pre-prints, desempenharem um papel maior na alta Física da energia em Química ou em Fonocologia. (HURD, 2000, p. 1282). Tradução nossa.

Russel (2001, p. 275) informa que os “pre-prints se tornaram o principal método para relatar novas descobertas dentro das comunidades científicas para pesquisadores em campos como matemática, física, computação e linguística”.

No tocante a publicação em diferentes veículos de comunicação em cada área do conhecimento é possível identificar que Hurd (2000) e Eason et al. (1997) possuem correntes de raciocínio complementares. No entanto, Eason et al. (1997) ainda aprofunda seu conhecimento sobre o tema afirmando que essas diferenças ainda podem ser encontradas dentro de um mesmo campo científico. Segundo os autores, “em algumas disciplinas, há áreas que

publicam em livros e raramente publicam em periódicos e outras que publicam com frequência e se concentram em artigos curtos.”.

Costa (2000, p.87) afirma que “na medida em que limites para comunidades são vistos como independentes de proximidade física, o foco principal tende a se concentrar na questão das interações entre seus membros”. Russel (2001, p. 274) afirma que a crescente colaboração entre os cientistas faz com que os mesmos procurem por formas de comunicação mais dinâmicas e flexíveis, sendo que o advento das redes de telecomunicações especializadas e da internet deram “aos cientistas o tipo de capacitação de informação que eles estavam procurando”.

Em um cenário em que o avanço das tecnologias de informação e comunicação propiciaram um ambiente favorável para a troca de informações, é possível observar um grau de adoção de práticas relacionadas ao crescimento do compartilhamento de informações por meios digitais nas áreas do conhecimento.

A convergência de tais tecnologias de comunicação e computação estão modificando a forma como as comunidades científicas lidam com a investigação científica no limiar de uma nova ciência baseada em dados, conforme afirmam Hey e Hey (2006, p. 515). De acordo com os autores, está próxima a emergência de uma quarta nova metodologia de pesquisa chamada de “e-Science” ou ciência baseada em dados.

Frank, Yakel e Faniel (2015, p. 143) acreditam que as infraestruturas do conhecimento são construídas ao redor de normas disciplinares e fatores externos, visto que essas normas são definidas pelos autores como sendo “um comportamento voluntário que prevalece dentro de um determinado grupo”. Pesquisadores, ao realizarem suas atividades investigativas se envolvem em comportamentos que são predominantes as suas comunidades disciplinares, incluindo aquelas que se referem às práticas de coleta, gestão e registro dos dados, segundo pensamento expresso pelos autores. De acordo com os autores Hey (2006, p. 515), está próxima a emergência de uma quarta nova metodologia de pesquisa chamada de “e-Science” ou ciência baseada em dados, sendo que muitas áreas serão vastamente transformadas pela grande quantidade de dados científicos que disponibilizados em uma proporção nunca antes vista.

Sendo a forma de realizar pesquisa uma variante entre as disciplinas científicas, é percebido que até mesmo a forma de coletar os dados que embasam as suas descobertas salta como um aspecto que as diferenciam. As particularidades relacionadas ao compartilhamento de dados de pesquisa, bem como o aprofundamento sobre as suas características serão abordados na próxima sessão desta dissertação.

2.5.1. Diferenças disciplinares relacionadas a dados de pesquisa

O compartilhamento de dados de pesquisa vem se tornando essencial na maioria das atividades científicas desenvolvidas nas variadas áreas do conhecimento, sendo que, no âmbito dessa dissertação, se demonstra ser necessário obter maior aprofundamento no que se refere ao processo de compartilhamento de dados nos diferentes campos científicos.

Os pesquisadores, em decorrência das investigações dos novos métodos e instrumentos de investigação, estão produzindo um dilúvio dados que nunca havia sido percebido anteriormente. Entretanto, assim como as demais características científicas já citadas, o compartilhamento de dados científicos parece variar de acordo com o campo científico em que está inserido, conforme Borgman (2012) afirma.

Meadows (2014) relata que em uma pesquisa realizada com 2.886 pesquisadores de diferentes áreas do conhecimento pela Wiley, revelou que dos 52% que relatam compartilhar dados:

Figura 19: Meios de compartilhamento



Fonte: Meadows (2014, tradução nossa) com adaptações.

- Compartilham dados em materiais suplementares em periódicos - 67%.
- Usam um repositório específico de disciplina - 19%
- Usam um repositório geral - 6%;
- Além dos citados, muitos relataram realizar o compartilhamento de dados de maneiras informais e muitas vezes não-permanentes como:
 - Conferências (57%);
 - Compartilhamento por e-mail, contato direto etc. (42%).
 - Sites pessoais, institucionais ou de projeto - 37%

Kim e Stanton (2013) investigaram os fatores que influenciam individual e institucionalmente os pesquisadores a realizarem ou não o compartilhamento de dados em diferentes áreas acadêmicas. Os autores concluem que existem três fatores que exercem influências sobre os cientistas:

1. **Fatores institucionais:** política da agência de financiamento, exigências de periódicos e contrato com patrocinadores;
2. **Fatores tecnológicos:** metadados e repositórios de dados e
3. **Fatores individuais:** características pessoais, benefício percebido, esforço percebido e risco percebido

A percepção da diversidade dos dados que são gerados por diferentes propósitos em diferentes comunidades científicas e por meio de diferentes processos foi noticiada por Sales e Sayão (2015, p. 7), o que ressalta a compreensão das diferenças existentes nos campos científicos no que é referente aos dados coletados e utilizados pelos pesquisadores das diversas áreas.

Frank, Yakel e Faniel (2015, p. 144) argumentam que, como exemplo, a coleta de dados no campo da Arqueologia envolve práticas e instrumentos que, durante o processo de captura da informação pode até mesmo destruir os ambientes em que os artefatos se encontravam, o que para a área se constitui como prática habitual devido ao valor atribuído a informação retirada a partir da captura da evidência e sua transformação em informação.

Apesar do fato de o uso de dados ser uma prática fundamental em todas as disciplinas científicas, as práticas específicas que circundam a coleta, gestão e armazenamento são específicas de cada área do conhecimento e são baseadas em fatores históricos e lógicos, bem como em um entendimento do que se constitui uma evidência e cada campo específico. (FRANK, YAKEL e FANIEL, 2015, p. 144, tradução nossa).

Como exemplo, Gómez, Méndez e Hernández-Pérez (2016, p. 546) alegam que as áreas de Ciências Sociais e Humanidades não coletam todos os seus dados de investigação digitalmente, adotando formatos e formas variadas.

Por exemplo, na Sociologia os dados procedentes de pesquisas e entrevistas podem ser facilmente captados digitalmente, não obstante, na Arqueologia o resultado dos dados observacionais estará mais ligado ao objeto e a informação contextual sobre o dito objeto [coordenadas geográficas, amostras e desenhos do objeto sobre o papel, ou fotografias ou vídeos (digitais)]. (GÓMEZ, MÉNDEZ e HERNÁNDEZ-PÉREZ, 2016, p. 546, tradução nossa).

Outro aspecto que passou a caracterizar mais uma diferença cultural entre as disciplinas acadêmicas foi verificado a partir do diagnóstico das tendências de abertura do compartilhamento dos dados de pesquisa e investigações científicas.

Meadows (2014) cita tais diferenças e as classifica em diferentes categorias. Em relação ao compartilhamento de dados de pesquisa em diferentes países, a autora relata que:

Os alemães demonstraram ser os melhores compartilhadores de dados de pesquisa, totalizando 55% dos entrevistados. Segundo a pesquisa, eles relatam realizar o compartilhamento “para aumentar a visibilidade de seu trabalho e para garantir transparência e reutilização do público”.

Figura 20: Motivação dos pesquisadores para compartilhar dados



Fonte: Meadows (2014, tradução nossa) com adaptações.

Entretanto, segundo a mesma pesquisa, os chineses são aqueles que atestam serem os menos propensos ao compartilhamento de dados, somando um total de 36%, “com metade dos entrevistados afirmando que isso é porque não é um requisito de financiador.”. Quanto a comunidade de pesquisadores brasileiros, eles afirmam que “estariam dispostos a compartilhar seus dados se obtivessem crédito ou atribuição adequados para isso”.

Se tratando do não compartilhamento de dados, Meadows (2014) relata que:

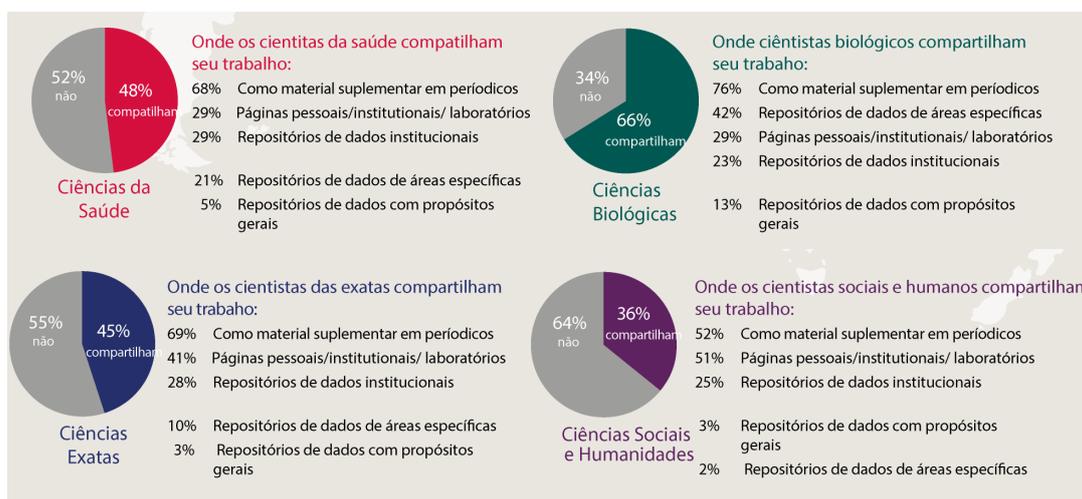
A principal razão citada pelos entrevistados para não compartilhar dados é sobre questões de confidencialidade de IP, especialmente em ciências da saúde, onde foi citada como uma razão por 68% dos entrevistados, em comparação com uma média de 42% em todas as disciplinas. Nenhuma exigência de financiador foi a segunda maior razão dada para não compartilhar dados (36%), enquanto as preocupações sobre a coleta e o possível uso indevido ou má interpretação dos dados ficaram em terceiro lugar pelos entrevistados (26% cada). (MEADOWS, 2014, tradução nossa).

Gómez, Méndez e Hernández-Pérez (2016, p. 546) corroboram a afirmam que diferentes disciplinas científicas estão adotando diferentes posturas quanto ao compartilhamento de dados e também demonstram como a disponibilização desses dados é feita para a comunidade científica.

As diferenças no compartilhamento de dados pelas Ciências Sociais e Humanidades começa desde a fonte de dados que essas áreas utilizam para as suas pesquisas, conforme afirmam Gómez, Méndez e Hernández-Pérez (2016, p. 547). Os autores informam que as áreas “costumam utilizar dados provenientes, em geral, de todos os tipos de fontes”, sendo que essa pode ser um dos fatores que fazem com que a maioria dos pesquisadores optem por não compartilhar os dados de suas pesquisas em repositórios de dados.

A respeito do compartilhamento de dados em diferentes áreas do conhecimento, Rodrigues e Saraiva (2010, p. 15) reconhecem a necessidade de preservação em todas as áreas do conhecimento, porém alerta que “o tipo de dados a preservar e o período de preservação é variável”. Meadows (2014) também expõe a mesma interpretação sobre as particularidades e variâncias no compartilhamento de dados, conforme demonstrado na figura abaixo:

Figura 21: Onde os cientistas compartilham dados



Fonte: Meadows (2014, tradução nossa) com adaptações.

A autora conclui que as Ciências Sociais e Humanidades, assim como as Ciências Exatas possuem uma motivação ligada ao aumento da visibilidade e impacto de seus trabalhos. Os das Ciências da Biológicas, no entanto, afirmam que compartilhariam mais se tivessem a garantia de obtenção de

crédito. Os pesquisadores das Ciências da Saúde expõem estarem “mais preocupados com questões éticas e de privacidade em torno do compartilhamento de dados.”.

Uma pesquisa com 1329 cientistas desenvolvida por Tenopir et al (2011) acerca dos hábitos de compartilhamento de dados científicos com pesquisadores de diversas áreas acadêmicas revela que:

Tabela 2: Compartilhamento de dados por disciplina

| Outros podem acessar meus dados facilmente | | |
|---|---------------------|-------------------|
| | Concordo fortemente | Concordo um pouco |
| Ciências Sociais | 11 (5,4%) | 36 (17,8%) |
| Ciência da Computação / Engenharias | 12 (10,3%) | 29 (24,8%) |
| Ciências Exatas | 17 (11,3%) | 41 (27,3%) |
| Ciências Ambientais & Ecologia | 56 (12,0%) | 124 (26,5%) |
| Ciência atmosférica | 12 (23,5%) | 13 (25,5%) |
| Biologia | 28 (15,6%) | 50 (27,9%) |
| Medicina | 2 (6,5%) | 2 (6,5%) |
| outros | 12 (13,0%) | 21 (22,8%) |

Fonte: Tenopir et al. (2011)

De acordo com os autores Hey e Hey (2006, p. 516), essa nova era quem que os dados exercem papel dominante trouxe novos desafios para os cientistas e por essa razão:

Eles precisarão das habilidades e tecnologias dos cientistas da computação e da comunidade de bibliotecas para gerenciar, pesquisar e selecionar esses novos recursos de dados. As bibliotecas não estarão imunes à mudança neste novo mundo de pesquisa. O advento da web está mudando a cara da publicação acadêmica e o papel dos editores e bibliotecas.

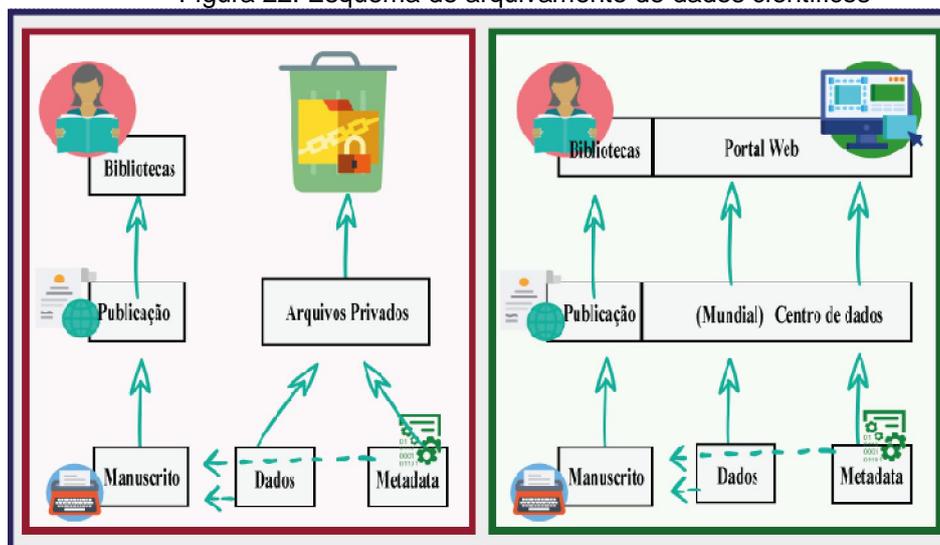
2.6. REPOSITÓRIOS DE DADOS DE PESQUISA

Como já foi percebido por meio das argumentações expostas ao longo dessa dissertação, a colaboração, independente da proximidade geográfica, foi facilitada pelos avanços infraestruturais dos meios de comunicação e também das ferramentas científicas e tecnológicas que se fazem necessárias para apoiar o desenvolvimento das pesquisas. Além desse fato, a percepção da importância do compartilhamento dos dados de pesquisa está sendo cada vez mais percebida por diferentes atores científicos.

Vários periódicos, como por exemplo a Nature e Science, passaram a exigir dos autores de artigos aceitos para publicação que eles compartilhassem seus conjuntos de dados brutos como uma condição para a publicação do trabalho, segundo o que informam Savage e Vickers (2009, p. 6).

Entretanto, conforme explica IES (2013, p.1), apesar de que cada vez um número maior de organizações exija a publicação do conjunto de dados junto com o artigo, a maioria dos dados ainda são armazenados em arquivos privados, sendo que normalmente esses dados são perdidos. Essa corrente de pensamento também é refletida nas afirmações de Klump (2006). O autor afirma que que a prática convencional atualmente é guardar os dados em arquivos privados e não em repositórios destinados a essa finalidade.

Figura 22: Esquema de arquivamento de dados científicos



Fonte: Klump (2006, p. 80, tradução nossa) com adaptações

O autor explica que o diagrama à esquerda representa o fluxo de informações que são atualmente praticados. Nessa estrutura, os dados e metadados que embasam a pesquisa científica são guardados em arquivos e catálogos pessoais que possivelmente serão excluídos ou perdidos com o tempo, mesmo com a publicação do produto da pesquisa sendo depositados em bibliotecas.

A direita, o autor ilustra uma abordagem baseada na publicação dos dados em centros e na organização do conteúdo para portais científicos ou bibliotecas. Para ele, as publicações interligadas com os dados criarão novos produtos científicos com valor agregado.

É importante observar que o autor cita basicamente duas formas de armazenamento de dados, o arquivo pessoal e os centros de dados. No entanto, a partir dos preceitos da Ciência aberta, esses dados podem ser compartilhados de várias formas. Conforme Whyte e Pryor (2011, p.6), existem seis estágios de compartilhamento de dados de forma aberta:

Quadro 7: Tipos de compartilhamento de dados

| Tipo de compartilhamento | Descrição do compartilhamento |
|--------------------------------|---|
| Gestão privada | O compartilhamento é feito dentro de um grupo de pesquisa, com o objetivo de facilitar o acesso e a reutilização pelos pesquisadores do grupo. |
| Compartilhamento colaborativo | O compartilhamento é feito entre os membros de um consórcio de projeto ou programa e somente eles podem acessar o conteúdo disponibilizado. |
| Intercâmbio entre pares | O compartilhamento é feito com condições estipuladas entre os membros de uma rede de pares por meio de uma plataforma web de rede social. |
| Governança transparente | O compartilhamento nesse estágio é feito com o objetivo de divulgação a uma parte externa, como por exemplo, para a avaliação de uma pesquisa. financiadores para avaliação de pesquisa. |
| Compartilhamento da comunidade | Nesses casos de compartilhamento, o acesso ou reutilização dos dados são limitados aos membros do grupo de pesquisa ou comunidade acadêmica. A troca de dados é facilitada por colaborações ou ambientes virtuais de pesquisa e recursos licenciados para acesso / uso educacional. |
| Compartilhamento público | Nesse grau de compartilhamento os recursos são disponibilizados para que qualquer pessoa possa ter |

| | |
|--|---|
| | acesso a eles por meio da descrição dos recursos por meio de metadados que reflitam as principais características das obras. Essas pesquisas possuem poucas ou nenhuma restrição de acesso. |
|--|---|

Fonte: Elaboração própria a partir das informações de Whyte e Pryor (2011).

Eynden et al. (2011) complementam os estágios de compartilhamento traçados por Whyte e Pryor (2011) elencando que os dados científicos podem ser compartilhados:

- Depositando-os em um data center especializado, arquivo de dados ou banco de dados;
- Submetê-los a uma revista para apoiar uma publicação;
- Depositá-los em um repositório institucional;
- Disponibilizá-los on-line através de um projeto ou site institucional;
- Disponibilizá-los informalmente entre pesquisadores em base peer-to-peer.

Entre as formas de compartilhamento de dados abertos se destacam os em repositórios científicos criados com essa finalidade. Um tipo de exemplo desses repositórios são os repositórios de dados.

“Os repositórios de dados de pesquisa são grandes infraestruturas de banco de dados configuradas para gerenciar, compartilhar, acessar e arquivar conjuntos de dados. ”, conforme define Uzwyszyn (2016). O autor ainda classifica os repositórios de dados em duas categorias: especializados ou gerais.

Os repositórios on-line também podem agregar dados de especialistas globalmente ou localmente, coletando dados de uma universidade ou consórcio de pesquisadores da universidade para benefício mútuo. A ideia simples é que compartilhar dados melhora os resultados e impulsiona a pesquisa e a descoberta. Um repositório permite exame, prova, revisão, transparência e validação dos resultados de um pesquisador por outros especialistas além do artigo acadêmico publicado. (UZWYSHYN, 2016, tradução nossa).

O site <<https://how-to.usopendata.org/en/latest/>> entende que os repositórios de dados agem como um “local para padronizar práticas e uma demonstração do uso desses dados. Em um sentido prático, um repositório

serve como um local central e pesquisável para que as pessoas (tanto fora quanto dentro do governo) encontrem dados.”.

Rodrigues e Saraiva (2010) acreditam que:

A forma e o volume desses registos ou dados científicos foram naturalmente evoluindo, crescendo em dimensão e complexidade, de acordo com a própria evolução da investigação científica, dos seus objectos, metodologias e instrumentos. De igual modo, foram-se registando alterações nas formas de armazenar, preservar, aceder e partilhar os dados produzidos no âmbito da actividade científica. (RODRIGUES e SARAIVA, 2010, p.10).

A ampla tendência de disseminação de dados científicos, gera demandas relacionadas a sua gestão e preservação a longo prazo. Para Sales e Sayão (2012) os pesquisadores têm apontado problemas significativos na estrutura de comunicação atual relacionados com “a gestão de dados de pesquisa, sua preservação, seu reuso e os processos de agregação de valor, cujas metodologias são coletivamente chamadas de curadoria digital”.

Maurer (2003, p. 6) relata que bons bancos de dados científicos possuem certas características específicas. Entre elas se destacam:

- **Cobertura:** para o autor o valor de um banco de dados deve aumentar com o seu tamanho. Isso inclui, além da organização dos dados, o rastreamento de resultados obscuros e/ ou não publicados.

- **Responsividade:** Para o autor, o valor de um banco de dados está relacionado com o seu uso.

- **Ciência Aberta:** de acordo com o autor, bancos de dados científicos comerciais costumam restringir o acesso e “capacidade dos usuários de copiar, redistribuir, extrair ou modificar dados”, o que, na visão dele, limita o número de usuários tornando o banco de dados menos valioso para a sociedade.

Segundo Sales e Sayão (2012), “mudança nos fluxos da comunicação científica e da rápida obsolescência tecnológica de seus suportes, a necessidade de criação de métodos para tratamento e recuperação de dados científicos toma uma proporção ainda maior”. Li e Banach (2011) afirmam que na era digital, a preservação da informação se tornou uma atividade de grande complexidade por ser uma informação frágil e repleta de ameaças “incluindo a obsolescência tecnológica e a deterioração da mídia de armazenamento

digital”. O autor também reafirma que devido à alta taxa de mudança nas tecnologias, a informação pode se tornar inacessível dentro de uma década.

Se o registro acadêmico digital é para ser preservado, as bibliotecas precisam estabelecer as novas melhores práticas para preservação. Por sua vez, os criadores precisam ser mais proativos quanto ao arquivamento de seu trabalho. O desenvolvimento relativamente recente de repositórios institucionais (RIs) oferece alguma promessa em garantir a preservação de longo prazo da produção acadêmica digital. (LI e BANACH, 2011, tradução nossa).

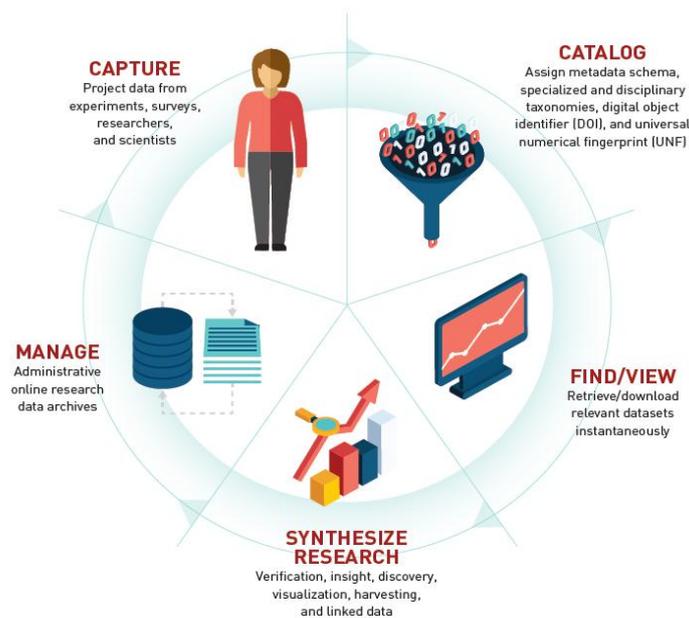
Lynch (2003) afirma que em sua opinião, Repositórios Institucionais são “um conjunto de serviços que uma universidade oferece aos membros de sua comunidade para o gerenciamento e disseminação de materiais digitais criados pela instituição e seus membros da comunidade.”.

Nota-se, na presente conjuntura da pesquisa que, dentre os meios de compartilhamento de dados de pesquisa, os repositórios institucionais já se consolidaram como canal para vários campos acadêmicos, sendo que, alguns repositórios institucionais, como o repositório institucional da Universitat Pompeu Fabra já possuem coleções destinadas ao compartilhamento de dados científicos.

Entretanto, Arano et al. (2011) alertam que a publicação de dados de pesquisa “requerem uma descrição mais detalhada, a fim de facilitar sua busca e recuperação por outros pesquisadores da área”, por essa razão, a publicação de dados em repositórios institucionais representa um ponto de dificuldade, visto que essas são estruturas com configurações destinadas a publicação e preservação de diferentes tipos de objetos digitais.

Segundo afirma Uzwyshyn (2016), há uma sequência de ações que os repositórios de dados devem seguir para consolidar uma infraestrutura para a gestão dos dados científicos:

Figura 23: O ciclo de vida dos repositórios de dados de pesquisa



Fonte: Uzwyszyn (2016)

Essas ações formam desenvolvidas para garantir a preservação de diferentes tipos de dados por um longo prazo e incluem, segundo Uzwyszyn (2016), começa com um projeto de pesquisa e a “captura de dados iniciais e progride para upload, catalogação, adição de esquema de metadados disciplinares e designação de DOI e / ou UNF”. Segue para encontrar conjuntos de dados relevantes e a administração desses recursos.

Arano et al. (2011) argumentam que a implementação de infraestruturas científicas que atuem como ferramenta para o compartilhamento de dados em que os pesquisadores encontrem e colham dados necessários para a elaboração de suas pesquisas requer um planejamento onde muitos recursos, humanos e econômicos serão empregados.

Os autores descrevem que essas infraestruturas são “é um conjunto de instalações, meios técnicos e serviços necessários, tanto para permitir o acesso, uso e reutilização dos dados, de modo a garantir a autoria e integridade dos mesmos.”, o que amplia a possibilidade das infraestruturas poderem se manifestar em “redes, redes, aplicações e recursos de computação, workbenches experimentais, centros de supercomputação, repositórios e comunidades virtuais”. No entanto, somente os repositórios agem

como instalações que conseguem preservar e divulgar dados e informações em acesso aberto com as descrições necessárias para a sua recuperação.

Por essa razão, essa dissertação possui como foco o sexto estágio de acesso aos dados de pesquisa descrito por Whyte e Pryor (2011) e no terceiro elencado por Eynden et al. (2011), ou seja, os repositórios de dados que compartilham e disseminam dados de pesquisa de forma aberta e livres de barreiras.

Rodrigues e Saraiva (2010) concebem o conceito de repositório como: “um sistema informático em que existe uma plataforma de armazenamento de objectos representados em ficheiros, capaz de incorporar novos objectos à medida que são produzidos ou submetidos”. Entretanto, para a equipe, o termo “Repositório de dados” designa algo que vai muito além da complexidade envolvida em repositórios comuns, “objectos, uma vez que cada conjunto de dados tem características próprias e por isso pode requerer um tratamento diferenciado”.

Esforços estão sendo percebidos na direção do desenvolvimento de repositórios de dados e outras ferramentas que facilitam o compartilhamento de dados pelos pesquisadores, conforme Kim e Stanton (2013) afirmam. Sendo assim os repositórios de dados se configuram como uma importante ferramenta na gestão de dados de pesquisa.

O Rodrigues e Saraiva (2010) afirmam que:

Dado o número crescente de repositórios institucionais, tem sido sugerido que eles possam ser a resposta, ou pelo menos parte dela, à necessidade de curadoria dos dados produzidos na investigação. De facto, existem já várias centenas de repositórios institucionais em estágio de produção, mas a sua utilização para albergar, preservar e dar acesso a conjuntos de dados científicos é ainda muito reduzida. (RODRIGUES e SARAIVA, 2010, p. 17)

Hernández-Pérez e García-Moreno (2013) informam que a maioria das universidades e centros de pesquisas no mundo já gerenciam repositórios institucionais para realizarem o armazenamento dos resultados de pesquisa das pesquisas científicas de seus membros.

Para Kim e Stanton (2013) os repositórios de dados se incorporam rapidamente à infraestrutura mundial de informação científica e, dessa forma, as coleções de dados podem ser usadas, reusadas e compartilhadas. No

entanto, Hernández-Pérez e García-Moreno (2013) informam que além da infraestrutura tecnológica, os planos de gestão de dados devem contemplar:

- A descrição dos dados;
- A definição de padrões de qualidade;
- As propriedade intelectual e direitos;
- As licenças e
- As políticas de arquivos e preservação.

Se tratando da descrição dos dados, esta é feita por meio da utilização de metadados. A National Science Board (2005) acredita que:

Para tornar os dados utilizáveis, é necessário preservar a documentação adequada relacionada ao conteúdo, estrutura, contexto e fonte (por exemplo, parâmetros experimentais e condições ambientais) da coleta de dados - chamados coletivamente de metadados. Idealmente, os metadados são um registro de tudo o que pode interessar a outro pesquisador.

O Relatório MGI (2013) afirma que o compartilhamento de dados abertos segue quatro características básicas:

- Acessibilidade: uma ampla gama de usuários deve ter permissão para acessar os dados.
- Legibilidade da máquina: os dados podem ser processados automaticamente.
- Custo: os dados podem ser acessados gratuitamente ou a um custo insignificante.
- Direitos: limitações no uso, transformação e distribuição de dados são mínimas.

Olijhoek, Mitchell e Bjørnshauge (2015) alertam para a existência de gestores que adotam uma postura híbrida no tocante a disponibilização dos conteúdos disponibilizados com o signo de “aberto”, mas, no entanto, adotam práticas restritivas de acesso. Conforme afirmam: “Os editores argumentam que o modelo híbrido é apenas um estado de transição no caminho para acesso totalmente aberto, mas disso somos duvidosos.”.

Molloy (2011) elucida que as atuais práticas de compartilhamento científico não estão em consonância com as reais intenções da abertura da ciência. O autor ainda destaca, com a finalidade de reduzir as confusões relacionadas ao compartilhamento de dados abertos, que “houve a necessidade de estender o OKD com um novo conjunto de princípios específicos para o campo científico.”, os chamados “Princípios de Phanton”.

Murray-Rust et al. (2010) afirmam que, por meio dos Princípios de Phanton, foram estabelecidas quatro recomendações para a publicação aberta de dados e pesquisa. Esses pontos devem ser observados para a averiguação do compartilhamento de dados de pesquisa.

1. Publicar dados com uma declaração explícita dos desejos e expectativas dos editores em relação à reutilização e redefinição dos dados. “Esta declaração deve ser precisa, irrevogável e baseada em uma declaração legal apropriada e reconhecida na forma de uma renúncia ou licença.”.
2. Utilizar uma licença já reconhecida e apropriada ao compartilhamento de dados, como a licença Creative Commons.
3. Encorajar o uso de licenças que limitem a produção de obras derivadas utilizando o princípio de acesso aberto.
4. É altamente recomendável que os dados, especialmente quando financiados publicamente, sejam explicitamente colocados no domínio público.

O grupo Force 11 acredita que a descoberta dos dados disponibilizados está atrelada às características de legibilidade por máquina atribuída a esses dados, o que se torna um dos grandes desafios da e-Science. Eles consideram que um conjunto de princípios e práticas devem ser seguidos para que a comunidade, os provedores e consumidores de dados possam descobrir, acessar e reutilizar esses dados com maior facilidade.

Reconhecendo os grandes desafios da ciência contemporânea o Force 11 reuniu quatro diretrizes para atender, apoiar e direcionar o melhor compartilhamento de dados, os Princípios FAIR (***F**indable, **A**ccessible,*

Interoperable, Reusable), em português as diretivas indicam que os dados científicos devem ser localizáveis, acessíveis, interoperáveis e reusáveis.

Em relação ao critério de “Localizável”, a Superintendência de Tecnologia da Informação da Universidade de São Paulo explica que os dados devem ser:

1. (meta) dados ¹⁹são relacionados a um identificador global único e permanente;
2. dados são descritos por metadados ricos;
3. Metadado inclui claramente e explicitamente o identificador dos dados que ele descreve e
4. (meta)dados são registrados ou indexados em um recurso passível de pesquisa.

Sobre o critério de “Acessibilidade”, Aventurier (2017) cita que:

1. Os dados são acessíveis na medida em que podem ser sempre obtidos por máquinas e seres humanos;
2. Mediante autorização adequada;
3. Através de um protocolo bem definido e
4. Assim, máquinas e seres humanos serão capazes de julgar a acessibilidade atual de cada objeto de dados.

Sobre o critério de “Interoperabilidade”, a Superintendência de Tecnologia da Informação da Universidade de São Paulo explica que:

1. (meta)dados utilizam uma linguagem formal, acessível, compartilhada e amplamente aplicável para a representação do conhecimento;
2. (meta)dados utilizam vocabulários que seguem os princípios FAIR e
3. (meta)dados incluem referências qualificadas a outros (meta)dados.

¹⁹ Os autores utilizam o termo (meta) dados “indicar que o princípio é verdadeiro para os metadados, bem como para os elementos de dados coletados no objeto de dados, entretanto esse princípio em questão pode ser implementado independentemente para cada um deles.

A respeito dos critérios para a “Reutilização” dos dados científicos, o Force 11 elenca que:

1. Objetos de dados²⁰ devem estar em conformidade com os princípios de localização e interoperabilidade;

2. (Meta) dados devem ser suficientemente bem descritos e ricos para que possam ser automaticamente (ou com o mínimo esforço humano) ligados ou integrados, como com outras fontes de dados.

3. Os objetos de dados publicados devem se referir a suas fontes com metadados e proveniência suficientemente ricos para permitirem a citação apropriada.

Rodrigues e Saraiva (2010) concordam com os Princípios FAIR e acreditam que para que o acesso aos dados preservados seja garantido é necessário que “os conjuntos de dados sejam acompanhados de informação que descreva a sua origem (tempo ou espaço, métodos e instrumentos de recolha), âmbito, autoria, propriedade e condições de reutilização, ou seja, de metadados”.

Conforme explicam Klump et al (2006), um dos incentivos que fariam com que os pesquisadores se dedicassem mais atividade de compartilhamento de dados seria se essas publicações pudessem ser citáveis, o que aumentaria, segundo os autores, a reputação dos pesquisadores em meio aos seus pares. Os autores citam dois principais critérios para que os dados sejam aceitos como uma publicação: a persistência e a qualidade. Eles explicam que:

- a) Para que os dados sejam citáveis eles precisam ser referenciados de maneira persistente já que, uma URL pode mudar com facilidade, o que faz com que os dados sejam perdidos. Sendo assim, a utilização de identificadores como o DOI ou a URN atribuem aos conjuntos de dados endereços confiáveis.
- b) Os autores explicam que para que a qualidade dos dados seja salvaguardada é necessário que esses tenham: credibilidade, usabilidade e interpretabilidade.

²⁰ Os autores explicam que o termo 'Objeto de Dados' se refere “à combinação de elementos de dados + seus metadados + um identificador único.”

2.6.1. A UTILIZAÇÃO DE METADADOS PARA A DESCRIÇÃO EM REPOSITÓRIOS DE DADOS

Boulton (2012) conceitua metadados como sendo “dados sobre os dados”. Ainda afirma que os metadados contém “informações sobre um conjunto de dados” que podem ser informações sobre como o dado foi gerado, quem os criaram e também como. “Também pode ser técnico, descrevendo sua estrutura, termos de licenciamento e padrões aos quais se conforma.”.

De acordo com a Biblioteca da Universidade de Queensland (2018), metadados são “informações estruturadas usadas para descrever seus dados de pesquisa, facilitando a descoberta, recuperação e reutilização.”. Ainda informa que o uso de metadados apropriados e armazenados junto com os dados de pesquisa facilita a sua recuperação, por isso são exigidos pelos gestores da biblioteca.

A Australian National Data Service (ANDS) (2016) resumem o que “metadados” realmente significam em seu conjunto de obras como:

- Metadados são informações sobre um objeto ou recurso que descreve as características desse objeto, como conteúdo, qualidade, formato, localização e direitos de acesso;
- Os metadados podem ser usados para descrever objetos físicos (fragmentos de pote, espécimes, etc.), bem como objetos digitais (documentos, imagens, conjuntos de dados, software, etc.);
- Os metadados podem assumir muitas formas diferentes, desde texto livre (por exemplo, arquivo leia-me) até texto padronizado, estruturado, conteúdo extensível legível por máquina para descoberta de recursos, fornecendo informações pesquisáveis.

Eyden et al. (2011) explicam que se tratando de gerenciamento de dados, metadados são:

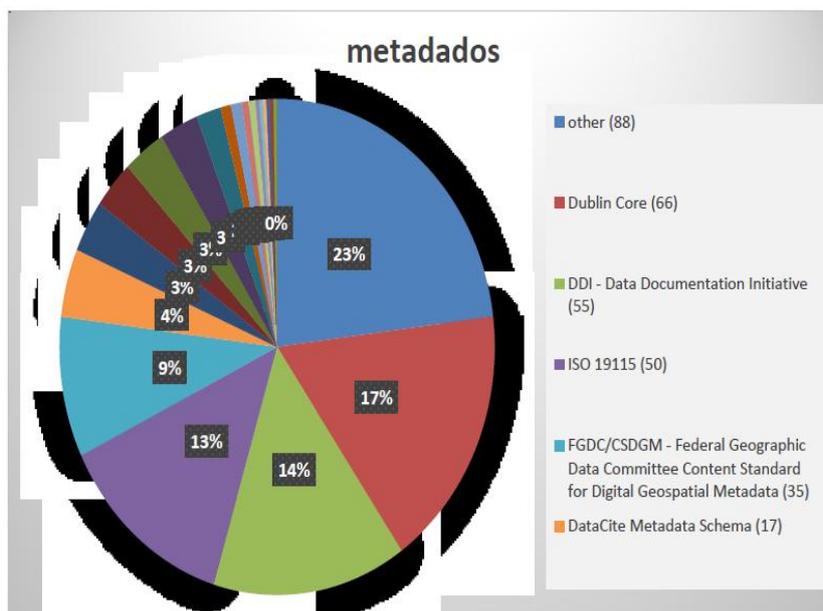
[...] um subconjunto da documentação central padronizada e estruturada de dados que explicam a **origem, finalidade, referência de tempo, localização geográfica, criador, condições de acesso e termos de uso** de uma coleta de dados. (EYDEN et al., 2011, p. 10, tradução nossa, grifo nosso).

Além desses critérios, Green, Macdonald e Rice (2009) elencam que ainda restam algumas considerações a respeito do esquema de metadados que os gestores dos repositórios devem considerar, a exemplo:

- Quem poderá acessar os metadados e se será de forma gratuita.
- Se o acesso aos metadados será controlado.
- Se será autorizado o re-uso dos metadados;
- Se o sistema de repositório permitirá a coleta de metadados do conjunto de dados conforme as orientações das diretrizes OAI-PMH.

Costa e Braga (2016) ao analisar as características dos repositórios cadastrados no Re3data quanto ao padrão de metadados utilizados, contataram que existem três padrões de metadados de descrição amplamente utilizados pelos repositórios de dados:

Figura 24: Padrões de metadados dos repositórios



Fonte: Costa e Braga (2016)

Segundo o que foi constatado na pesquisa realizada pelos autores, padrão Dublin Core de metadados está presente na maioria relativa dos repositórios

que declararam as informações relativas aos seus metadados, totalizando 17% dos repositórios.

O padrão Dublin Core de metadados foi desenvolvido pela Dublin Core Metadata Initiative (DCMI). O DCMI (2018) informa que as especificações adotadas visam o apoio a descrição dos recursos em quinze elementos.

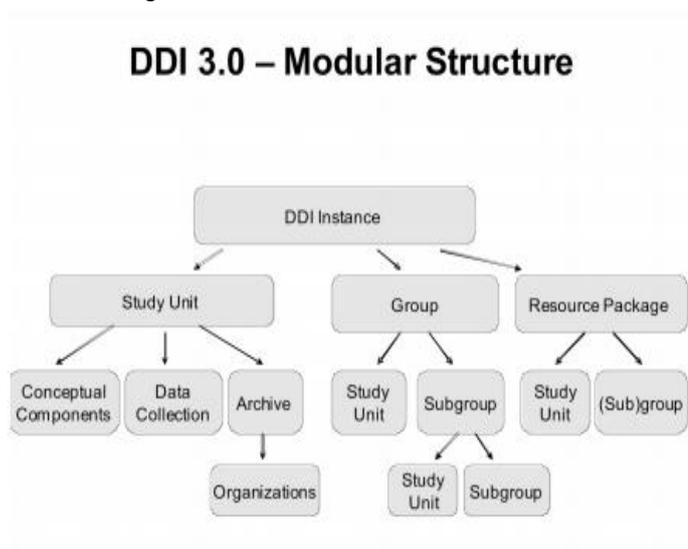
Grácio (2002) entende que o padrão do DCMI se caracteriza por:

1. Simplicidade: como um entendimento semântico simples que dispensa, extensos treinamentos;
2. Interoperabilidade semântica;
3. Consenso internacional: sendo que mais de vinte países adotam o esquema de descrição.
4. Extensibilidade: o padrão DC é um modelo simplificado de descrição, que permite a criação de novos elementos para atender uma necessidade de descrição de um determinado recurso.
5. Flexibilidade: seus elementos são opcionais, podem ser repetidos se necessário, e modificáveis utilizando-se de qualificadores, que mostraremos nesse capítulo.

O segundo padrão mais utilizado, estando presente em 14% dos repositórios declarados segundo a pesquisa de Costa e Braga, é o Data Documentation Initiative (DDI). Conforme afirma Martinez Uribe (2008), este é um padrão criado em 2003 com bases para o arquivamento dos conjuntos de dados das ciências sociais.

O autor declara que o esquema de descrição do DDI possui a seguinte estrutura desenvolvida para cobrir todas as etapas do ciclo de vida dos dados de pesquisa.

Figura 25: Estrutura modular do DDI



Sobre o uso desse esquema de descrição em repositórios, Martinez Uribe (2008), informa que:

Repositórios digitais podem e devem tirar proveito do DDI para classificar e organizar conjuntos de dados. Isto, acreditamos, é possível nos três sistemas de repositórios digitais mais comuns, o DSpace, o ePrints e o Fedora. Embora o DSpace e o e-Prints usem diferentes esquemas de metadados, ambos podem ser modificados para aceitar e usar o DDI. O próprio Fedora não apresenta problemas, pois pode usar muitos formatos de metadados, conforme necessário.

Em terceira colocação do uso de metadados, se encontra o ISSO 19115, estando presente em 13% dos repositórios. Segundo Costa (2017) este padrão foi elaborado com foco para a descrição de informações geográficas.

Embora a catalogação dos metadados seja uma parte fundamental da recuperação dos dados de pesquisa, as estratégias de preservação devem contar com infraestruturas de softwares adequados a sua preservação.

2.6.2. A UTILIZAÇÃO DE SOFTWARES PARA A CRIAÇÃO DE REPOSITÓRIOS DE DADOS

A respeito da gestão dos dados científicos, Hernández-Pérez e Garcia-Moreno (2013) alertam que os repositórios de dados devem ser separados dos repositórios institucionais já que ambos possuem características que os diferenciam substancialmente.

Com essa premissa, alguns softwares foram desenvolvidos com vistas a solução de necessidades específicas para o compartilhamento de dados. Essas necessidades básicas foram descritas por <https://howto.usopendata.org/en/latest/>. O site chama a atenção que alguns softwares para a gestão de dados de pesquisa devem:

- Converter automaticamente os formatos de dados mesmo que os gestores disponibilizem em apenas um formato. “ (por exemplo, CSV), ele gerará XML, JSON, Excel etc.”.
- Permitir a visualização dos conjuntos de dados diretamente no navegador “permitindo que as pessoas mapeiem, classifiquem, pesquisem e combinem conjuntos de dados, sem precisar de nenhum conhecimento de como programar.”.
- Permitir a organização, possibilitando que outras entidades complementem os dados com outros dados que possuem, “ (por exemplo, uma agência de transporte estadual pode reunir todos os dados de transporte das localidades e republicá-los)”.
- Suportar o carregamento de arquivos para serem armazenados no repositório ou apontar para o repositório em um endereço de site existente onde o arquivo reside.
- Fazer uma divisão ampla nos três tipos de hospedagem disponíveis: hospedagem comercial, hospedagem própria e hospedagem gratuita.

Rodrigues e Saraiva (2010) afirmam que já existem diversas plataformas de repositórios. Segundo os mesmos autores essas plataformas “foram desenvolvidas para recolher, preservar e divulgar literatura científica, mas

presentemente podem ser usadas para agregar outros tipos de conteúdos digitais.”.

Há os seguintes softwares para a construção dos repositórios de dados citados e analisados pelo site < <https://how-to.usopendata.org/en/latest/>>:

Quadro 8: Softwares de repositórios de dados

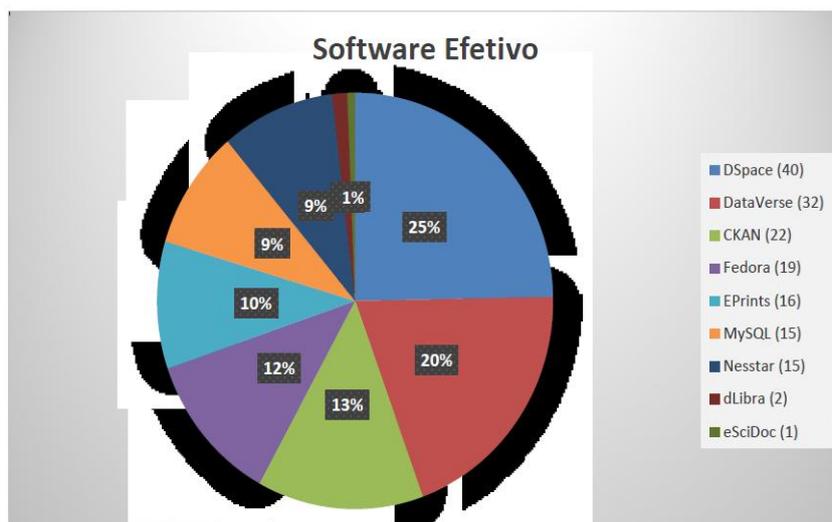
| Nome | Tipo |
|-------------------|----------------------|
| ArcGIS Open Data | Hospedado |
| CKAN | Fonte aberta |
| DataHub | Grátis, hospedado |
| DKAN | Fonte aberta |
| GitHub | Grátis, hospedado |
| JKAN | Fonte aberta |
| Junar | Hospedado |
| NuData | Hospedado |
| OpenData.city | gratuito e hospedado |
| OpenDataSoft | Hospedado |
| Open Data Catalog | Fonte aberta |
| Socrata | Hospedado |

Fonte: <https://how-to.usopendata.org/en/latest/The-Basics-of-Open-Data/Data-Repositories/> com adaptações. Tradução nossa.

Uzwysbyn (2016) aponta que ainda existem outras possibilidades relação ao software de repositório de dados que também foram criadas especialmente para realizar a gestão de dados científicos como o Dataverse, HUBzero e Chronopolis. Apesar da ampla variedade de softwares dedicados exclusivamente a gestão de dados de pesquisa, ainda existem gestores que optam por utilizar outras fontes de repositórios/ bibliotecas digitais, como é o caso do DSpace, Fedora, e Hydra.).

Em sua pesquisa sobre o compartilhamento de dados em repositórios, Costa e Braga (2016) apontam que:

Figura 26: Software utilizados pelos repositórios de dados de pesquisa

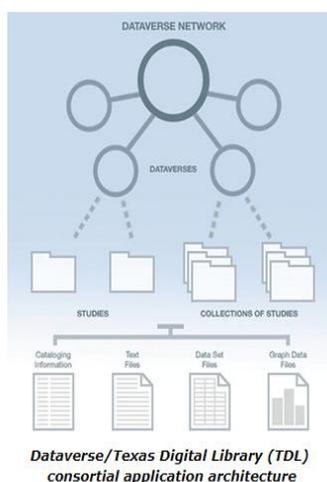


Fonte: Costa e Braga (2016)

Segundo o que pode ser constatado pelos autores, “dentre os repositórios que indicaram o software utilizado, a maioria (25%) está relacionada ao uso do DSpace”, um software gratuito mantido pelo DuraSpace que se adapta facilmente as necessidades dos repositórios, segundo Costa e Braga.

Em segunda colocação se encontra o DataVerse em 20% dos repositórios que declararam o software utilizado. Conforme afirma o Dataverse.org, o software possui um código aberto “para compartilhar, preservar, citar, explorar e analisar dados de pesquisa” através da web. De acordo com o Dataverse. Org, o software funciona conforme ilustração abaixo:

Figura 27: Dataverse / Biblioteca Digital do Texas (TDL)



Fonte: Uzwyshyn (2016)

[...] arquivos virtuais chamados Dataverses. Cada dataverse contém conjuntos de dados e cada conjunto de dados contém metadados descritivos e arquivos de dados (incluindo documentação e código que acompanham os dados). Como método de organização, os dataverses também podem conter outros dataverses. (DATAVER.ORG, 2018).

“O software CKAN foi identificado como o terceiro mais utilizado entre os sistemas analisados, estando presente em 13% dos repositórios de dados de pesquisa.”. Segundo o site < <https://ckan.org/> >. O "CKAN" é o acrônimo para "Comprehensive Knowledge Archive Network," e foi criado pela UK-based Open Knowledge.

3. PRINCÍPIOS E RECOMENDAÇÕES SOBRE O ACESSO ABERTO EM REPOSITÓRIOS DE DADOS DE PESQUISA

Este tópico da dissertação tem a intenção de funcionar como elemento compilador de todas as estratégias e recomendações para a gestão de repositórios de dados de pesquisas em diferentes áreas acadêmicas expostas ao longo dessa dissertação. O objetivo central dessa ação é providenciar uma matriz de elementos e dimensões que combinadas possam indicar, com base na literatura levantada nessa dissertação, quais são os principais pontos a serem observados em repositórios de dados para que sejam considerados como repositórios pertencentes ao movimento da Ciência Aberta, bem como mapear suas principais características infraestruturais.

Com uma finalidade de expandir a capacidade didática da organização dos princípios e entendimentos expostos ao longo dessa dissertação, optou-se pela segmentação dos princípios estabelecidos em blocos.

Princípios e Entendimentos



Características básicas de RD's

A literatura da área mostra que os repositórios de dados foram desenvolvidos a partir de uma necessidade específica do compartilhamento de dados.

RD'S possuem características específicas quanto a:

- Seu tipo, podendo ou não ser especializado em uma natureza específica.
- A área do conhecimento a que se destinam.
- Sua origem geográfica ou quanto a uma conexão a uma determinada instituição.
- Sua finalidade acadêmica ou científica.

Ciência Aberta

A partir da revisão de literatura feita, tem-se que Ciência Aberta faz referência a uma revolução científica que remete à:

A disponibilização online dos produtos e processos da pesquisa científica onde basicamente:

- O acesso ao conteúdo deve ser facilitado de modo que eles possam ser reutilizados;
- O conteúdo deve ser disponibilizado em íntegra, sem restrições de propriedade intelectual;
- A colaboração científica deve ser estimulada. Esse estímulo pode ocorrer por meio da infraestrutura onde os dados são disponibilizados.



Gestão e funcionamento

O levantamento bibliográfico dessa dissertação indicou que a gestão dos dados de pesquisa é feito por meio de uma infraestrutura tecnológica que disponha de:

Software e metadados:

Convém que tenham destinação específica a gestão dos dados que disseminam.

Critérios de persistência:

Com a finalidade de assegurar a melhor recuperação dos dados, é aconselhável que os dados sejam disponibilizados com identificadores persistentes de seus conteúdos.

Política de gestão:

Para que erros sejam minimizados durante a curadoria dos dados, é conveniente que haja um plano de gestão e que este tenha fácil acesso ao público do banco de dados.

Sendo assim, partindo dos princípios estabelecidos e todos os pontos elencados ao longo da dissertação, considera-se que os Repositórios de dados analisados obedeçam às características básicas mencionadas para que sejam avaliados como parte integrante e contribuinte do movimento em prol da Ciência Aberta.

4. METODOLOGIA

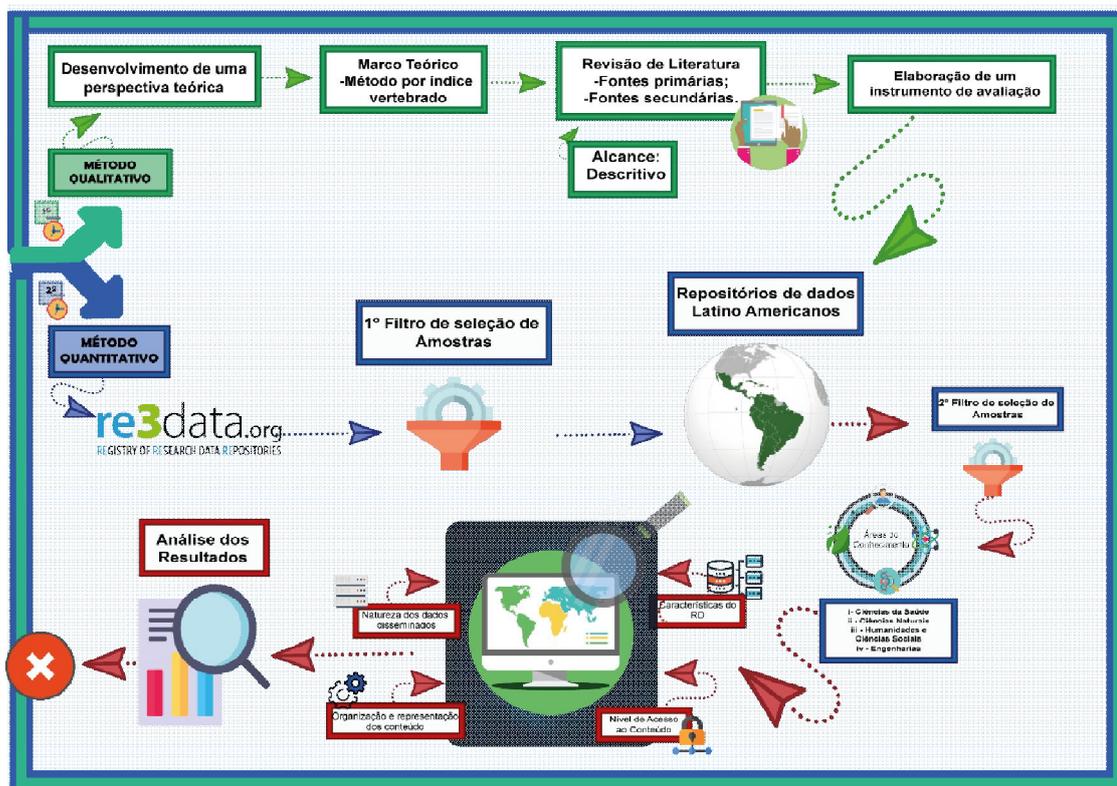
Este estudo possui como objetivo a identificação das características e contribuições dos repositórios de dados de pesquisas científicas da América Latina para o desenvolvimento da Ciência Aberta. Para o alcance dessa meta foi adotado um conjunto de procedimentos que caracterizam a metodologia científica da pesquisa. Segundo Kauark, Magalhães e Medeiros (2010, p. 24) “não se deve dizer que se faz ciência, mas que se produz ciência através de uma pesquisa. Pesquisa é, portanto, o caminho para se chegar ao conhecimento”. De acordo com os mesmos autores, a pesquisa necessita de instrumentos que assegurem a precisão das respostas encontradas. Tais instrumentos metodológicos serão apresentados neste capítulo.

4.1. PROCEDIMENTOS

Para a mensuração das características e contribuições dos repositórios de dados científicos da América Latina para o desenvolvimento da Ciência Aberta foram consideradas as principais orientações da literatura científica em relação aos princípios e pilares fundamentais sobre o tema. Os pontos essenciais foram identificados e elencados em uma lista de verificação que será aplicada na apuração das características dos repositórios de dados apurados no âmbito geográfico da América Latina.

A ilustração a seguir foi elaborada com o intuito de demonstrar graficamente os procedimentos adotados para a obtenção dos dados que irão consolidar o objetivo geral traçado.

Figura 22: Plano de pesquisa



Fonte: Elaboração própria

A metodologia da pesquisa retratada foi elaborada com base no questionamento geral que orienta essa pesquisa: “Quais as características e contribuições dos repositórios de dados de pesquisa científicas da América Latina para o desenvolvimento da Ciência Aberta?”. A partir desse ponto, desenhou-se um cenário onde esse questionamento conduz a pesquisa a buscar uma metodologia mista que pudesse responder esse questionamento.

Hernández Sampieri, Fernández Collado e Baptista Lucio (2013) expõem uma dupla significação do enfoque misto. Primeiramente, os autores argumentam que o método misto representa um conjunto de processos de pesquisa que “implicam a coleta e a análise de dados quantitativos e qualitativos, assim como sua integração e discussão conjunta, para realizar inferências e como produto de toda a informação coletada (metainferências) [...]”. Posteriormente, eles argumentam que o método misto reúne em uma única pesquisa os métodos quantitativo e qualitativo, de tal forma que ambos conservem as suas estruturas originais, com a finalidade de obter uma “fotografia” mais completa do fenômeno estudado. Por meio da integração as

abordagens qualitativa e quantitativa procurou-se obter um instrumento melhor, mais complexo e profundo para o alcance do objetivo proposto por esse estudo.

Dentro da abordagem qualitativa deste método, buscou-se inicialmente o desenvolvimento de uma perspectiva teórica que demonstrasse o atual cenário do conhecimento científico a respeito da Ciência Aberta, Dados abertos e os seus desdobramentos conceituais. Entre as funções do desenvolvimento de uma perspectiva teórica para um estudo, Hernández Sampieri, Fernández Collado e Baptista Lucio (2013) citam três pontos fundamentais que interagem diretamente com essa dissertação. São eles:

- Amplia o horizonte do estudo ou guia o pesquisador para que se centre em seu problema;
- Leva a formulação de hipóteses ou afirmações que mais tarde deverão ser submetidas à prova na realidade ou ajuda a não formulá-las por razões bem fundamentadas e
- Proporciona uma estrutura de referência para interpretar os resultados do estudo.

Os três pontos foram essenciais para a formulação da do marco teórico adotado na dissertação, como também para a construção do índice geral inicial para a elaboração da revisão de literatura presente nesse trabalho.

Sobre a ação de se criar um índice geral para a construção o marco teórico:

A experiência mostra que outra maneira rápida e eficaz de construir um marco teórico é desenvolver, primeiro, seu índice experimental, global ou geral, e aprimorá-lo até que se torne bem específico, para depois colocar a informação (referências) em seu lugar correspondente dentro do esquema. A essa operação podemos dar o nome de “vertebrar” o marco ou perspectiva teórica (criar a coluna vertebral desta). (HERNÁNDEZ SAMPIERI, FERNÁNDEZ COLLADO e BAPTISTA LUCIO, 2013, p. 91).

Inicialmente o índice geral era composto por três pilares centrais sobre o tema: “Comunicação científica”; “Dados de pesquisa” e “Organização da informação”, como pode ser observado na figura a seguir. Esse índice geral sofreu um processo de “vertebramento” em subtemas de acordo com a necessidade de cada abordagem.

Figura 28: Índice vertebrado inicial

| | |
|------------------------------|--|
| Fundamentação Teórica | |
| 1. | Comunicação Científica |
| 1.1 | Diferenças Disciplinares |
| 2. | Dados de Pesquisa |
| 2.1 | Compartilhamento de dados de pesquisa |
| 2.1.1 | Padrões de compartilhamento de dados de pesquisa em diferentes áreas do conhecimento |
| 2.2 | Dados científicos abertos |
| 3. | Organização da Informação |
| 3.1 | Curadoria de Dados Científicos |
| 3.2 | Repositório de dados científicos |
| 3.2.1 | Organização dos Repositórios |
| 3.2.2 | Metadados |

Fonte: Arquivos pessoais (ago. 2017)

O índice vertebrado inicialmente proposto sofreu diversas alterações ao longo da elaboração da dissertação, bem como devido as orientações fornecidas pelos orientadores dessa pesquisa.

A partir do índice vertebrado, construiu-se uma revisão de literatura a partir de fontes primárias e secundárias de informação tendo um alcance descritivo. O que Hernández Sampieri, Fernández Collado e Baptista Lucio (2013) chamam de “alcance da pesquisa”, Gil (2001) denomina de “nível de pesquisa” e explica que pesquisas de nível descritivo buscam “a descrição das características de determinada população ou fenômeno ou o estabelecimento de relação entre as variáveis”, sendo comum, entre as pesquisas descritivas aquelas que têm por objetivo estudar as características de um determinado grupo, o que nesse estudo se aplica as características da Ciência Aberta e do compartilhamento de dados de pesquisa.

Com base na literatura levantada, foi elaborado um instrumento para avaliação dos repositórios de dados de pesquisa pré-selecionados a partir a abordagem quantitativa desse estudo.

Em um momento posterior a abordagem qualitativa, foi aplicada da abordagem quantitativa. Sobre o método quantitativo Creswell (2010) afirma que esta abordagem origina dados objetivos e medidas que conduzem a interpretações significativas dos dados.

Da abordagem quantitativa dada a essa pesquisa, buscou-se no Re3data.org (Registry of Research Data Repositories) a fonte do universo dos dados a serem coletados e analisados. Tornando-o então, parte essencial da abordagem quantitativa que foi adotada.

A partir dos repositórios cadastrados no diretório, a pesquisa adotou a América Latina como a região geográfica para a seleção dos repositórios, por possuir uma trajetória particular no compartilhamento dos resultados da investigação científica, conforme foi demonstrado ao longo da revisão de literatura.

O prosseguimento do plano de pesquisa traçado classifica os repositórios averiguados em quatro grandes áreas do conhecimento: Ciências da Saúde; Ciências Naturais; Humanidades e Ciências Sociais e Engenharias. Áreas essas que mais se destacam no compartilhamento de dados de científicos em repositórios, segundo informações do próprio diretório de repositórios de dados.

Segue-se então para a execução das variáveis adotadas à investigação por meio da lista de verificação previamente elaborada com base na revisão de literatura feita no corpo da dissertação. Esta estratégia de investigação está associada a técnica quantitativa do método de pesquisa e se materializa por meio de um levantamento.

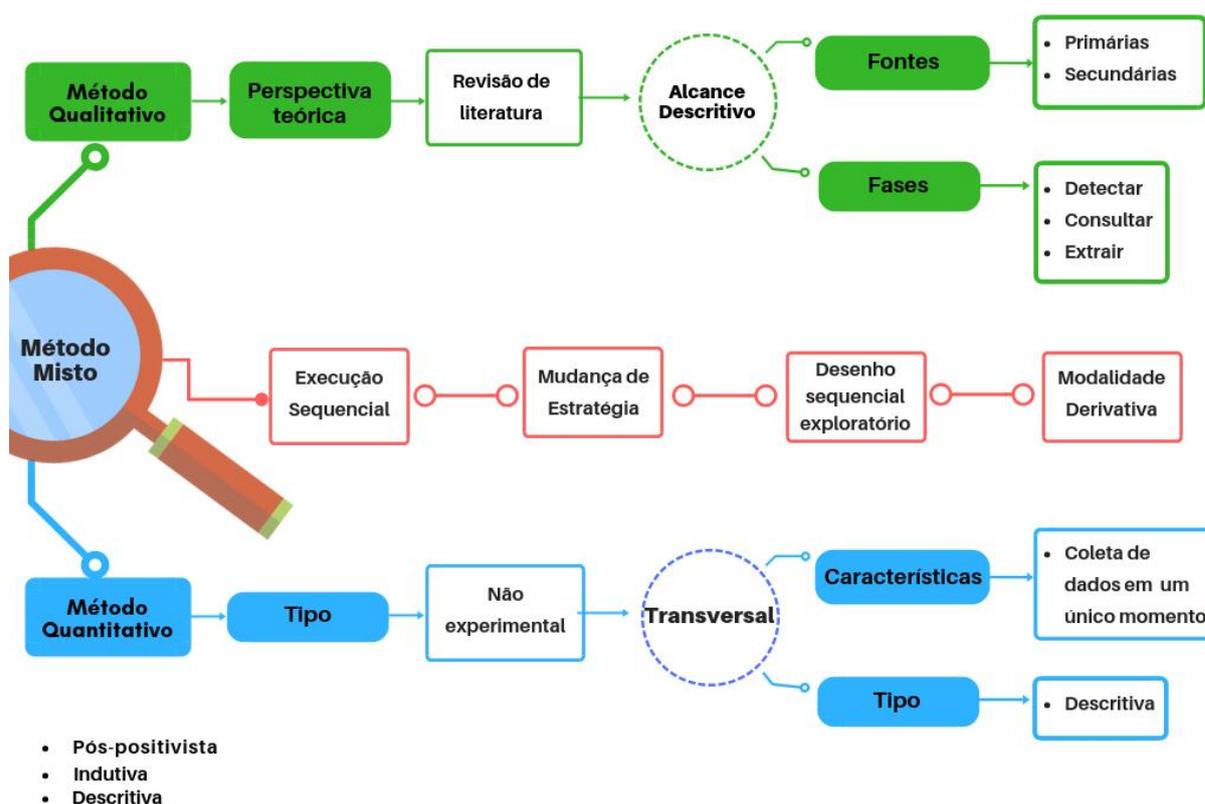
De acordo com Creswell (2007) o levantamento “dá uma descrição quantitativa ou numérica de tendências, atitudes, ou opiniões de uma população ao estudar a amostra dela”.

4.1.1. Desenho da Pesquisa

Entende-se que a formulação de um problema exija um plano de ação que demonstre de maneira prática as etapas fixadas para o desenvolvimento da pesquisa. Quanto a esse plano de ação, Hernández Sampieri, Fernández Collado e Baptista Lucio (2013) entendem que “o termo desenho se refere ao plano de ação ou estratégia criado para obter a informação que deseja”.

A figura feita com base nos ensinamentos fornecidos pelos autores busca sintetizar de forma gráfica as principais características do plano de ação adotado nessa pesquisa quantitativa:

Figura 29: Desenho da Pesquisa



Fonte: Elaboração própria

O plano de pesquisa desenvolvido possui os elementos da concepção pós-positivista, que segundo Creswell (2010, p. 29) se baseia em uma filosofia na qual “as causas provavelmente determinam os efeitos ou resultados”. Seguindo as premissas do autor, esta pesquisa de mestrado acredita que a trajetória de

publicação de documentos resultantes da investigação científica influência como essas disciplinas se comportam na atualidade, inclusive no que é referente ao compartilhamento de dados em repositórios, sendo, portanto, possível mensurar e estruturar esse comportamento em números.

Sendo assim, essa pesquisa de cunho pós-positivista foi construída em três pilares que sustentaram a sua base lógica e determinarão os rumos da sua investigação: o método lógico, o seu objetivo e a sua natureza.

O método lógico adotado para essa pesquisa foi o indutivo, método pelo qual, segundo Gil (2011), se inferem princípios gerais para um fenômeno a partir de observações específicas. Segundo o autor, neste método, a observação revela as conclusões prováveis dos fenômenos estudados. No caso dessa dissertação, a observação sob a lente indutiva tem o objetivo de inferir características que possam ser generalizáveis aos demais repositórios que possuem a mesma finalidade na mesma área de conhecimento.

Essa pesquisa pode ser classificada com base no seu objetivo, segundo Gil (2012, p. 26-9), Kauark, Magalhães e Medeiros (2010, p. 28) Ferrari (1982, p. 240) e Babbie (1990, p. 52) como descritiva, pois ela busca a identificação e enumeração das características e fenômenos que podem ser observadas nos repositórios de dados científicos latino americanos.

A abordagem adotada é do tipo mista. Sobre o método misto, Onwuegbuzie e Johnson (2006) afirmam que esta abordagem “envolve a combinação de forças complementares e fraquezas não sobrepostas da pesquisa quantitativa e qualitativa”.

Dentro da abordagem mista, adotou uma execução sequencial. Segundo Hernández Sampieri, Fernández Collado e Baptista Lucio (2003), quando se realiza um estudo com proposta mista o pesquisador deve permanecer atento aos tempos do método do estudo, tanto no que se refere as etapas de coleta e análise dos dados.

Os autores explicam que na execução sequencial do método misto primeiramente se coleta e analisa os dados qualitativos ou qualitativos e, em um momento posterior se trabalha com o outro método. Eles fundamentam sua argumentação nos ensinamentos do autor Chen (2006), que considera que esse desenho pode ser aplicado a duas estratégias, a “Mudança de estratégia” e a

“Estratégia contextual “dissimulada””. A primeira “aplica os métodos qualitativos para “clarear” e produzir teoria fundamentada e depois utilizar métodos quantitativos para “aquilatá-la””. A segunda atua na utilização da “abordagem qualitativa para coletar informação do contexto com a finalidade de facilitar a interpretação de dados quantitativos ou “reconciliar” descobertas”. Por essa razão a “mudança de estratégia” foi adotada nessa etapa da dissertação.

Quanto à tipologia da estratégia, a que mais se adequava aos objetivos do estudo foi a Sequencial exploratória. Segundo Creswel (2010, p. 248), esta tipologia concentra o seu peso maior na coleta de dados qualitativos, já que eles serão combinados para a análise dos dados quantitativos em uma segunda etapa da pesquisa. Ainda de acordo com o autor, “o objetivo dessa estratégia é utilizar os dados e resultados quantitativos para auxiliar na interpretação dos resultados qualitativos”.

Dentro da abordagem sequencial exploratória, para Hernández Sampieri, Fernández Collado e Baptista Lucio (2003, p. 566) DEXPLOS, a finalidade desse estudo se encontra na área “derivativa”, pois “a coleta e análise dos dados quantitativos são elaborados com base nos resultados qualitativos”, sendo a sequência composta por três etapas:

1. Coleta e análise dos dados qualitativos;
2. Utilização dos resultados para a construção de um instrumento quantitativo e
3. Administração do instrumento para uma amostra probabilística de uma população para validá-lo.

As três etapas descritas pelos autores serão seguidas nessa dissertação para melhor busca e análise dos resultados para a problemática exposta. A via qualitativa do método adotado busca estabelecer uma perspectiva teórica sobre o tema Ciência aberta a partir de uma revisão de literatura com alcance descritivo, como já mencionado na sessão anterior.

Dentro do método quantitativo adotado, a pesquisa segue pela corrente de estudos com desenhos não-experimentais. Segundo Hernández Sampieri, Fernández Collado e Baptista Lucio (2013) orientam, em pesquisas não-

experimentais, os fenômenos são observados como ocorrem em seu contexto natural, sem a necessidade da manipulação intencional das variáveis.

Tendo sido determinado que o objetivo central da pesquisa é fazer uma análise das características atuais dos repositórios de dados latino americanos, entende-se que o estudo adota uma postura com corte transversal em um esquema único de coleta de dados.

Sobre desenhos transversais de pesquisa, Hernández Sampieri, Fernández Collado e Baptista Lucio (2013) explicam que:

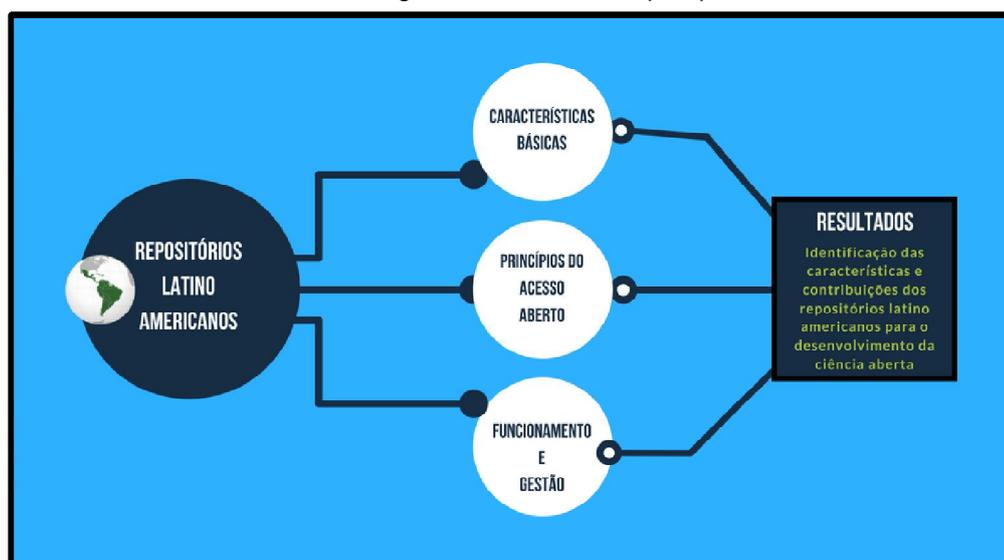
Os desenhos de pesquisa transversal coletam dados em um só momento, em um tempo único. Seu propósito é descrever variáveis em analisar sua incidência e inter-relação em um momento determinado. É como tirar uma fotografia de algo que acontece. (HERNÁNDEZ SAMPIERI, FERNÁNDEZ COLLADO e BAPTISTA LUCIO, 2013, p. 170)

O corte transversal para coleta de dados adotado se caracteriza como descritiva em seu tipo, pois possui o intuito de indicar incidências de fenômenos ou características dos repositórios de dados em modalidades específicas do universo estipulado.

O instrumento que mais se adéqua ao plano de pesquisa proposto é o levantamento. De acordo com Creswell (2007) o “levantamento dá uma descrição quantitativa ou numérica de tendências, atitudes ou opiniões de uma população”.

A elaboração da lista de verificação, que fornece a base para o levantamento quantitativo, evidencia a preocupação da dissertação com o alcance de um entendimento sobre quais são os pontos de equilíbrio para a caracterização de um repositório de dados de pesquisa de acordo com os reais objetivos e delineamentos da Ciência Aberta, fato que independe da área do conhecimento a qual se destinem. Esses pontos de caracterização foram divididos em categorias para melhor descrição das variáveis adotadas:

Figura 30: Resumo da pesquisa



Fonte: Elaboração própria

4.1.2. Delimitações e limites da pesquisa

O problema identificado por essa pesquisa gerou quatro objetivos específicos que orientaram o desenho metodológico adotado para a coleta dos dados. Suas especificidades serão abordadas a seguir.

4.1.2.1. Universos e fontes de dados

Barbie (2014) reflete que como parte do desenvolvimento de uma pesquisa científica, além de conceitos e dimensões, deve-se definir o objeto do estudo, que normalmente se caracteriza por ser um grupo. Ele conceitua que a “população” do estudo é esse grupo de quem se deseja obter informações. Selltiz et al. (1980 apud Sampieri, Collado e Lucio, 2013) afirmam que “universo” e “população” são termos sinônimos que descrevem o “conjunto de todos os casos que preenchem determinadas especificações”. Para o alcance dos objetivos propostos essa dissertação adotou dois universos diferentes, são eles:

- a) Literatura científica sobre a ciência aberta, repositórios de dados e seus desdobramentos;

b) Repositórios de dados latino americanos cadastrados no Re3data.

4.1.2.1.1. *Universo “a”*

O universo “a” foi estruturado e concebido por meio da busca de conceitos chave em documentos científicos que abordassem os temas:

- Comunicação científica;
- Ciência aberta e Movimento de Acesso Aberto;
- Diferenças disciplinares e
- Repositório de dados.

As fontes de informação selecionadas para a coleta dos documentos foram:

1. Google e Google acadêmico;
2. Portal de Periódicos da CAPES e
3. Citação em outros documentos relevantes para o tema.

Em ambos os motores de busca foram utilizadas palavras-chave em três idiomas: português, inglês e espanhol, sendo que em relação a:

- Portal de Periódicos da CAPES

As buscas foram realizadas dentro do campus Darcy Ribeiro- UnB - que permite que os pesquisadores possam ter acesso ao conteúdo das bases de dados de acesso livre e acesso restrito.

De acordo com a Capes o Portal de Periódicos:

O Portal de Periódicos, da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Capes), é uma biblioteca virtual que reúne e disponibiliza a instituições de ensino e pesquisa no Brasil o melhor da produção científica internacional. Ele conta com um acervo de mais de 53 mil títulos com texto completo, 129 bases referenciais, 11 bases dedicadas exclusivamente a patentes, além de livros, enciclopédias e obras de referência, normas técnicas, estatísticas e conteúdo audiovisual. (CAPES, 2018).

Dentro do Portal, os documentos foram buscados dentro do campo “assunto”. Os critérios para a seleção dos trabalhos recuperados foram:

1. Estar disponível em texto completo;

- a. Os resultados obtidos não sofreram refinamento por data, ou tipo de recurso, tampouco software necessário para a sua visualização.
2. A relevância considerada para seleção foram “os mais acessados”.
 - Google e Google acadêmico²¹

O Google acadêmico é um motor de busca de trabalhos acadêmicos em formatos variados fornecido gratuitamente pelo Google. Nesse motor de busca, os critérios de seleção foram os mesmos utilizados nas buscas feitas no Portal de Periódicos da Capes.

No entanto, como essa pesquisa buscava fontes além das acadêmicas tradicionais, pois, como exposto ao longo da revisão de literatura, a internet possibilitou que informações de relevância científica fluíssem para além dos periódicos e livros, o portal do Google foi adicionado à lista de buscadores apesar de recuperar uma ampla variedade de documentos que não correspondiam ao escopo da pesquisa, o que gerou a necessidade de maior refinamento quanto aos critérios de relevância. Foram excluídas:

1. Fontes de informação que não eram relevantes para a ciência;
2. Fontes de informação que não possuam relevância científica para o estudo em questão;
3. Fontes de informação que não apresentavam o documento na íntegra e
4. Fontes de informação que traziam informações em idiomas diferentes dos selecionados previamente.

No final dos refinamentos, a dissertação considerou como documentos válidos para a construção da fundamentação teórica:

- Artigos;
- Teses;
- Dissertações;
- Relatórios de eventos e
- Textos de blogs científicos.

²¹ <https://scholar.google.com.br>

4.1.2.1.2. Universo “b”

O universo “b”, ou seja, os repositórios de dados de pesquisa latino americanos, foi abordado em sua totalidade, logo dentro deste aspecto, não houve a necessidade de seleção de amostras em nenhum de seus tipos.

Atualmente a América Latina é uma região composta por países que possuem como língua oficial o espanhol, o português e o francês, totalizando 20 países: Argentina, Bolívia, Brasil, Chile, Colômbia, Costa Rica, Cuba, Equador, El Salvador, Guatemala, Haiti, Honduras, México, Nicarágua, Panamá, Paraguai, Peru, República Dominicana, Uruguai e Venezuela, conforme indica o mapa a seguir:

Figura 31: Mapa da América Latina



Fonte: Elaboração própria

Segundo o que pôde ser apurado, no âmbito geográfico da América Latina, nem todos os países possuem repositórios de dados cadastrados no Re3data, logo essa pesquisa adotou a autodeterminação dos gestores desses repositórios como caráter para validar a origem de cada repositório identificado.

4.1.3. Método de coleta e Análise dos dados

Para o alcance dos objetivos traçados inicialmente para essa pesquisa foi necessário que fossem utilizadas vertentes de investigação que atuaram de forma complementar. A primeira, de natureza teórica, foi a revisão da literatura científica exposta na com base nos documentos científicos da área, a segunda, de natureza prática, foi a elaboração da lista de verificação que atuou de como instrumento para a coleta de dados.

A vertente de natureza teórica foi a base para a criação do instrumento de verificação. A consolidação da perspectiva teórica também possibilitou o alcance do primeiro objetivo traçado: “Identificar, com base na literatura, princípios e entendimentos básicos da ciência aberta e repositório de dados.”.

A vertente prática se consolidou com a ficha de análise elaborada com vistas a destacar três características dos repositórios de dados analisados:

- Características básicas do repositório;
- Características relacionadas aos princípios norteadores da Ciência Aberta destacados pela literatura científica e
- Características relacionadas ao esquema de funcionamento e gestão.

Mais precisamente, a pesquisa investiga pontos centrais que foram destacados na literatura científica, como:

O grau de abertura ou nível de acesso aos seus conteúdos, que podem ser;

- i. Aberto ou;
- ii. Acesso embargado ou;
- iii. Acesso restrito ou;
- iv. Acesso fechado.

A natureza e nível de tratamento dos dados disseminados, que podem abranger;

- v. Observacionais e/ ou;
- vi. Computacionais e/ ou;
 - i. Experimentais e/ou;

- ii. Brutos e/ou
- iii. Derivados.

As licenças para atribuição de autoria ou reutilização, que podem ser manifestadas em:

- i. Possuir declaração quanto a propriedade intelectual ou;
- ii. NÃO possuir declaração quanto a propriedade intelectual;
- iii. Possuir declaração quanto a reutilização de dados ou;
- iv. NÃO possuir declaração quanto a reutilização de dados;

O esquema de metadados adotados para a descrição dos dados, que podem ser:

- i. Dublin Core ou;
- ii. DDI ou
- iii. Outros.

O software adotado para a gestão dos dados, que podem estar entre:

- i. Dspace ou;
- ii. DataVerse ou;
- iii. CKAN ou;
- iv. Outros.

O esquema de funcionamento e gestão²²

- i. Se busca avaliar se o repositório possui um plano de gestão de dados disponível.
- ii. Critérios de persistência ao conteúdo disponibilizado.

²² Caso o plano não esteja disponível no próprio repositório, será analisado a presença em um diretório específico para a disponibilização de políticas, o Melibea.

Sendo assim, resumidamente, o esquema de análise das características dos RD's latino americanos pode ser esquematizado da seguinte forma:

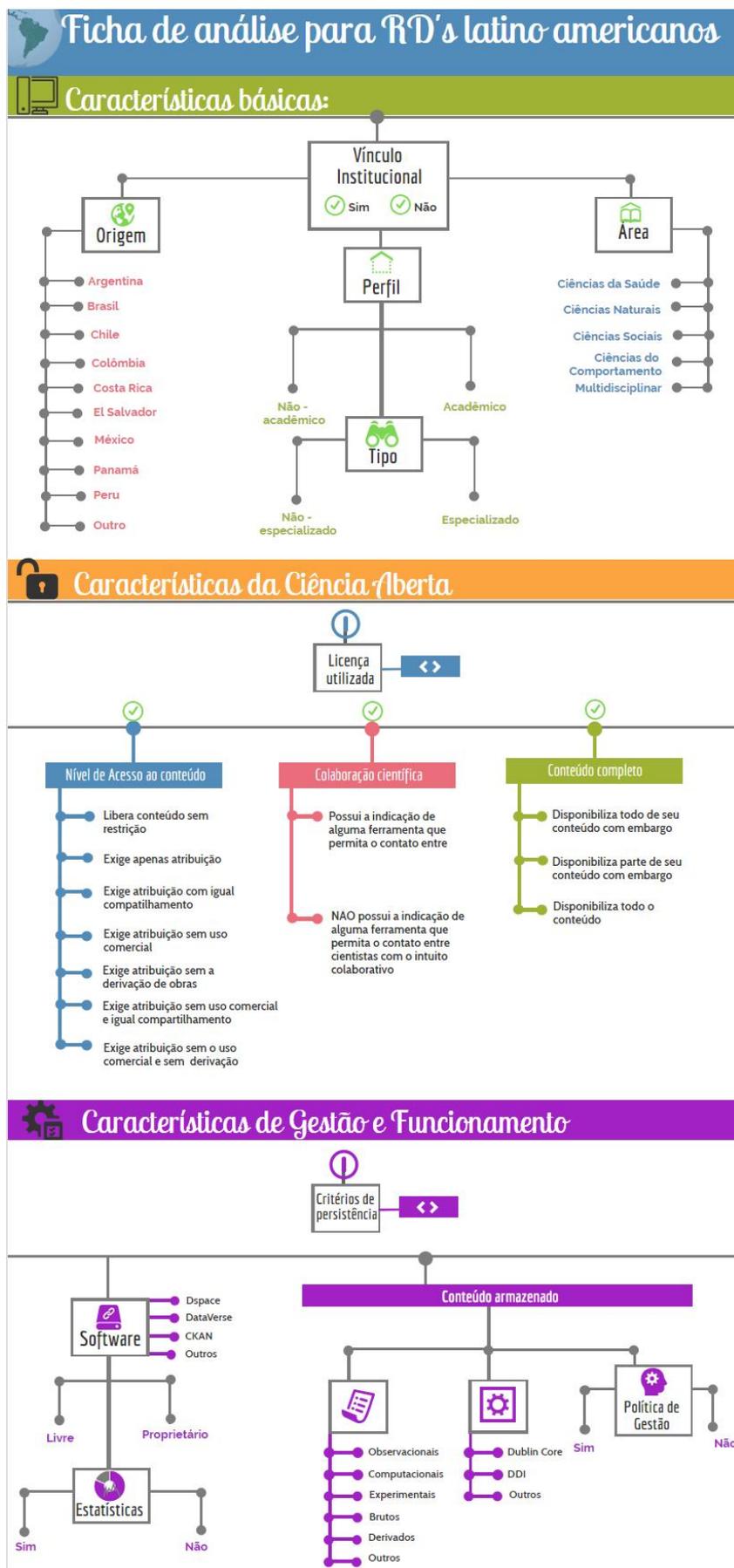
Figura 32: Esquema de análise das características dos RD's.



Fonte: Elaboração própria

Traçados os pontos chave para análise, foi construído um instrumento para a coleta dos dados. A ficha para a análise vertebrou os pontos principais a serem investigados com a finalidade de dar maior profundidade as questões investigadas. A seguir se encontra o “vertebramento” da ficha de análise construída:

Figura 33: Ficha para análise dos RDs



Fonte: Elaboração própria

Os pontos elencados no vertebramento da ficha de análise possuíam os seguintes objetivos de pesquisa:

- **Vínculo institucional:**

De acordo com o que foi apontado pela literatura, é comum a existência de repositórios do tipo institucional no que se refere a disseminação dos produtos das pesquisas científicas, ou seja, de teses, dissertações, artigos e etc. Se tratando de repositórios de dados, o mesmo poderia ser afirmado, já que, conforme foi destacado por Eason et al. (1997), os cientistas utilizam meios informais para a troca e informações?

Essa questão se desdobra em outras frentes de investigação, que são os pontos se esses repositórios possuem um perfil acadêmico, o que pode ser um indicativo quantitativo da quantidade de repositórios de dados de pesquisa presentes na América Latina. E também ao tipo de compartilhamento realizado, podendo ser geral ou específico, conforme a orientaram de Uzwyszyn (2016).

- **Origem:**

Este critério foi definido para que fossem identificadas as origens geográficas dos repositórios de dados. Com esse critério será possível segmentar todas as características específicas dos repositórios de acordo com a sua região.

- **Área do conhecimento:**

Este ponde de análise visa classificar os repositórios de dados latino americanos conforme a área do conhecimento a quais os dados compartilhados pertencem. Isso indicará possíveis diferenças disciplinares, assim como as que foram apontadas por Mueller (2015) quando a autora tratou das diferenças disciplinares em relação aos canais preferenciais que os pesquisadores adotam para disseminar o produto das suas pesquisas.

- **Nível de acesso ao conteúdo:**

Esse item visa a investigação de um dos pontos essenciais ligados as práticas de Ciência Aberta que foram expostos pela literatura. Autores como Borgman (2012), Boulton (2013) e Friesike (2014) citam que é

fundamental que os produtos disseminados sob signo da Ciência Aberta tenham um nível de acesso ao com o mínimo de restrição possível.

- **Conteúdo completo:**

A literatura levantada para a composição dessa dissertação também ressaltou que outro ponto essencial para a composição da Ciência Aberta é que as publicações sejam feitas de forma rápida e completa, significando que o conteúdo disseminado deve estar em sua íntegra, e não fragmentado em certas partes.

- **Colaboração científica:**

Conforme foi destacado pela literatura, os avanços tecnológicos possibilitaram que as práticas científicas se tornassem mais colaborativas. O que representa um nível de interação entre pesquisadores que possuem o mesmo objeto de estudo e podem estar separados geograficamente. Esse item de investigação buscou averiguar se de alguma forma, os usuários poderiam interagir tanto com o autor dos dados compartilhados, como com o público geral.

- **Crítérios de persistência:**

Conforme já elencado na revisão de literatura, a adoção de critérios de persistência como o DOI ou o Handle fazem com que os dados compartilhados possuam vantagens de preservação, que é um dos objetivos principais do seu depósito em repositórios, e da ampliação da visibilidade dos dados. Por essa razão, essa pesquisa adotou como ponto de investigação a quantidade de RD's que já utilizam critérios de persistência em suas gestões.

- **Software:**

De acordo com o que foi destacado por Hernández Sampieri, Fernández Collado e Baptista Lucio (2003), os softwares que são destinados a gestão de dados de pesquisa devem possuir uma série de características específicas que permitem a melhor gestão e acesso a esses dados. No entanto, essa dissertação buscou investigar quais são as plataformas mais utilizadas e se essas também seguem a filosofia do livre acesso a qualquer tipo de informação.

- **Conteúdo armazenado:**

Também se tornou objeto da investigação o conteúdo que esses repositórios armazenam e a forma como são descritos para a sua recuperação.

A partir da obtenção dos dados coletados, a pesquisa pretende analisar os resultados obtidos à luz da literatura científica levantada sobre tema das diferenças disciplinares quanto ao uso, compartilhamento e reuso de dados científicos.

Consolidadas as etapas metodológicas traçadas foi elaborado um quadro sistemático que resume e detalha os procedimentos de investigação adotados nessa dissertação.

Quadro 9: Sistematização da metodologia da pesquisa

| Objetivo específico | Universo | Amostra | Fonte | Método de Coleta |
|--|--|---|---|-------------------------------------|
| Identificar, com base na literatura, princípios e entendimentos básicos da ciência aberta e repositórios de dados. | Literatura científica sobre Ciência Aberta; Movimento de Acesso aberto e Repositório de dados. | Documentos científicos com relevância direta sobre os temas | Publicações indexadas no Google, Google acadêmico e Portal de Periódicos da Capes | Pesquisa documental |
| Mapear os Repositórios de Dados Científicos da América Latina. | América Latina | Países indicaram no Re3data possuir repositórios de dados ativos. | Registry of Research Data Repositories (re3data.org) | Levantamento no diretório de dados. |
| Caracterizar os Repositórios de dados de Científicos da América Latina de acordo com as dimensões apontadas pela literatura da área. | Repositórios de dados da América Latina. | | Lista de verificação elaborada com base na revisão de literatura. | Levantamento. |
| Identificar a organização e representação dos dados de pesquisa nos repositórios digitais de dados científicos. | Repositórios de dados da América Latina. | | Lista de verificação elaborada com base na revisão de literatura. | Levantamento. |

Fonte: Elaboração própria

5. ANÁLISE E DISCUSSÃO DOS DADOS

Este capítulo apresenta os resultados obtidos na realização da pesquisa. O universo que a pesquisa buscou investigar foi o dos repositórios latino americanos. Para delimitação desse universo usou o diretório de repositório de dados de pesquisa Re3data.org adotando como critério a autodeterminação dos gestores desses repositórios quanto a origem geográfica dos mesmos e a sua classificação como repositório disseminador de dados.

Segundo Pampel et al. (2013), o projeto do diretório Re3da começou a indexar os repositórios de dados de pesquisa em 2012 oferecendo aos pesquisadores e outras instituições envolvidas no campo científico uma visão geral do panorama dos repositórios de dados de pesquisa. Costa (2017) analisou o Re3data e conclui que “o diretório se apresenta como um registro global de repositórios de dados de pesquisa e cobre diversas áreas do conhecimento.”.

Ainda sobre o assunto, Vierkant et al. (2013, p. 3) afirmam que o Re3data.org promove a cultura do compartilhamento em diferentes disciplinas científicas dando maior visibilidade aos dados de pesquisa.

Este ambiente virtual, por melhor se adequar as necessidades desse estudo científico quanto ao fornecimento de informações sobre os repositórios de dados de pesquisa, foi selecionado como fonte de informações para a análise dos repositórios latino americanos.

Apesar da América Latina possuir 20 países em sua extensão, foi verificado, a partir dos registros indicados pelo diretório Re3data.org, que nem todos os países considerados latino americanos possuem repositórios de dados cadastrados como executantes da atividade de compartilhamento de dado, o que justifica o número de países verificados ser menor do que a do apurado geograficamente.

A tabela a seguir indica, segundo o que foi averiguado no diretório, a quantidade de repositórios por região.

Tabela 3: Quantidade de RD's Latino Americanos

| Rótulos de Linha | Contagem de Área do Conhecimento |
|------------------|----------------------------------|
| Argentina | 4 |
| Brasil | 8 |
| Chile | 1 |
| Colômbia | 2 |
| Costa Rica | 1 |
| México | 9 |

| | |
|-------------|----|
| Panamá | 2 |
| Peru | 2 |
| Total Geral | 29 |

Fonte: Elaboração própria

Os países que, por meio da autodeterminação, mais possuem repositórios de dados, em âmbito latino americano, são o Brasil e o México. No entanto, todos os 8 países que possuem repositórios farão parte da análise desta pesquisa, totalizando 29 repositórios a serem analisados a partir da lista de verificação elaborada.

Ressaltou-se na durante a coleta de dados, que alguns desses repositórios atuavam em conjunto com outras localidades geográficas. Esse fato destaca o princípio da colaboração entre pares para o melhor desenvolvimento científico que foi estipulado na Escola Pragmática da Ciência Aberta e consolida, na prática, um dos pilares apontados por Gezalter (2009) para a construção da Ciência Aberta: “3.

Uso de ferramentas baseadas na web para facilitar a colaboração científica”. Andrade (2015) também fez referência a essa modalidade de pesquisa, ao explicar que uma das recomendações feitas pela Ciência Aberta é que a atividade científica seja mais colaborativa.

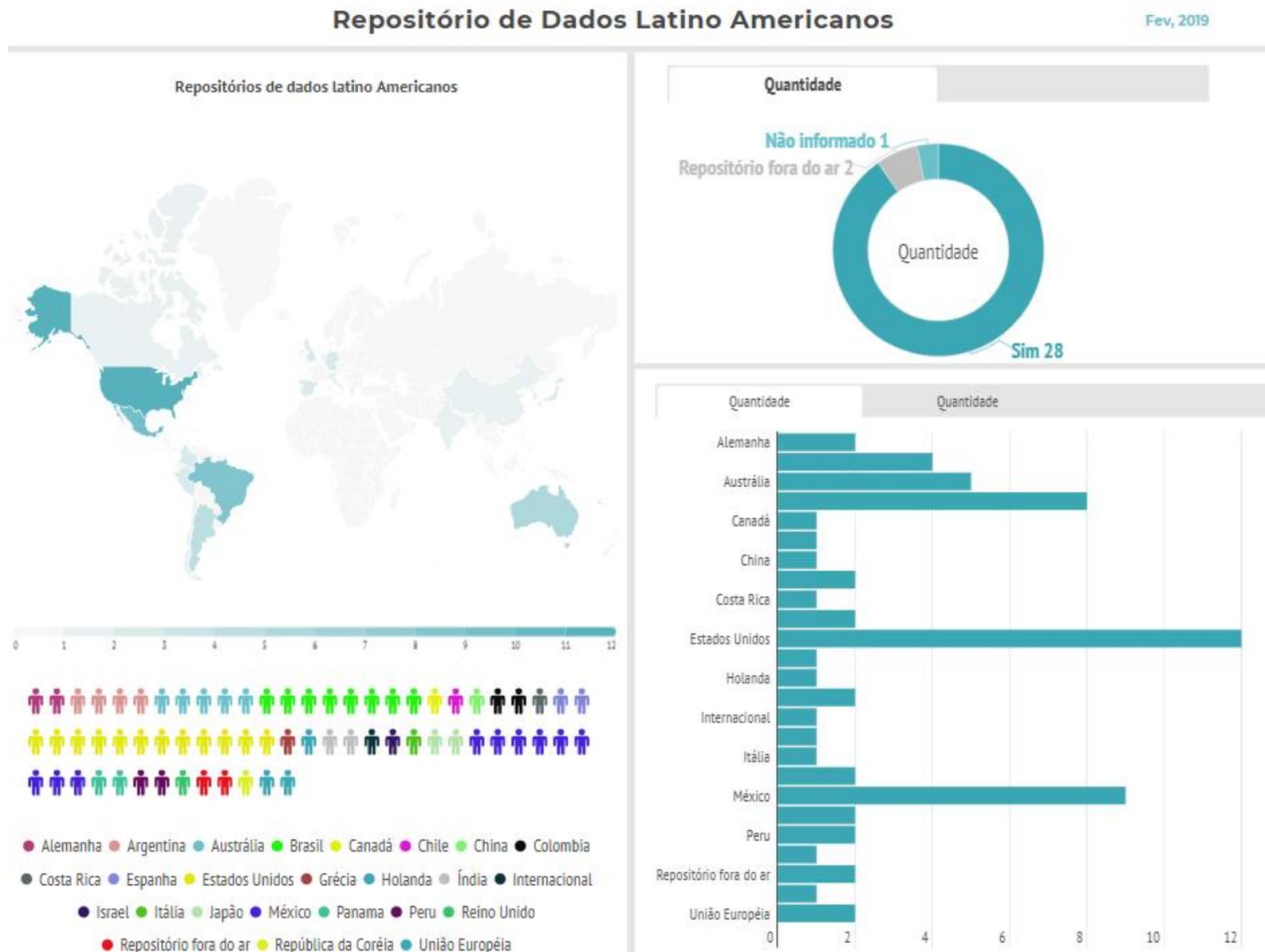
Além das tabelas²³ de coleta de dados anexada a essa dissertação, os países da América Latina e suas relações com outras localidades podem ser visualizadas graficamente na figura 32.

A coleta de dados também apontou que nem todos os repositórios autodeclarados como repositório de dados dentro do cadastro do RE3da, realmente se encaixam nessa classificação por:

- Apenas direcionar a outros sites, que realmente se caracterizam como um disseminador de dados.
- Não compartilharem dados;
- Estiverem fora do ar no momento da recolha de dados.

²³ As tabelas contendo todos os dados que foram coletados durante a fase de investigação quantitativa foram anexados a essa dissertação.

Figura 34: Quantidade de RD's Latino Americanos



Fonte: Elaboração própria

Excluídos aqueles repositórios da listagem de análise geral, a amostra analisada nessa pesquisa foi:

Quadro 10: Amostra de RDs analisados

| Repositório | URL | Área Geográfica |
|--|---|-----------------|
| The Global Agricultural Trial Repository and Database | http://www.agtrials.org/ | Colômbia |
| CIAT Dataverse | https://dataverse.harvard.edu/dataverse/CIAT | Colômbia |
| WorldClim - Global Climate Data | http://worldclim.org/ | Brasil |
| GLOBE | http://globe.umbc.edu/ | Brasil |
| International Ocean Discovery Program | http://publications.iodp.org/ | Brasil |
| PPBio Data Repository | https://ppbiodata.inpa.gov.br/metacatui/ | Brasil |
| IBICT Dataverse Network | https://repositoriopesquisas.ibict.br | Brasil |
| Base de Dados Científicos da Universidade Federal do Paraná | https://bdc.c3sl.ufpr.br/ | Brasil |
| CEDAP Research Data Repository - research data | https://cedap.ufrgs.br/jspui/handle/2050011959/90 | Brasil |
| Exploration and Production Data Bank | http://rodadas.anp.gov.br/en/concession-of-exploratory-blocks/brazil-round-3/bdep-round-3 | Brasil |
| eFish Genomic Database Repository | http://www.molecularfisherieslaboratory.com.au/efish/ | Chile |
| EcoCyc Database | https://ecocyc.org/ | México |
| ComBase | http://www.combase.cc/index.php/en/ | México |
| Gran Telescopio CANARIAS Public Archive | http://gtc.sdc.cab.inta-csic.es/gtc/index.jsp | México |
| Mexican Health and Aging Study | http://www.mhasweb.org/ | México |
| Constrained Local UniversE Simulations | https://www.clues-project.org/cms/ | México |
| Banff International Research Station for Mathematical Innovation and Discovery | https://open.library.ubc.ca/cIRcle/collections/birs | México |
| Integrated Taxonomic Information System | https://www.itis.gov/ | México |
| CIMMYT Research Data & Software Repository Network | https://data.cimmyt.org/dataverse/root | México |
| Center for Tropical Forest Science and Forest Global Earth Observatory | http://www.forestgeo.si.edu/ | Panamá |
| Center for Tropical Forest Science - Panama | http://ctfs.si.edu/webatlas/ | Panamá |
| Repositorio Institucional USIL | http://repositorio.usil.edu.pe/handle/123456789/11 | Peru |
| Repositorio de datos del Ministerio de Educación del Perú | http://datos.minedu.gob.pe/ | Peru |

Fonte: Elaboração própria

5.1. Análise dos critérios de Características Básicas dos RDS

5.1.1. Relação Tipo de dados - Área Geográfica – Área do Conhecimento

Como já descrito ao longo da dissertação, existem características variáveis entre as áreas conhecimento. Segundo o que foi estudado a partir das pesquisas de Mueller (2015) e Meadows (1999), essas diferenças disciplinares se manifestam desde a adoção de um canal preferencial para a publicação científica quanto no compartilhamento de dados.

A variável adotada nessa sessão buscou estabelecer uma conexão entre as variáveis “Área geográfica” “Área do conhecimento” e “Conteúdo armazenado” com a finalidade de mapear o atual estado da arte nesse requisito.

5.1.1.1. Brasil

As variáveis se relacionam de modo a indicar que o compartilhamento de dados em repositórios brasileiros abrange:

Figura 35: Relação Tipo de dados - Área Geográfica – Área do Conhecimento – Conteúdo armazenado - Brasil

| Área Geográfica | Área do conhecimento | Conteúdo armazenado |
|-----------------|-------------------------------------|---|
| Brasil | Climatologia | Conjunto de dados; Planilhas; Dados espectrais; Document... |
| Chile | Ecologia | Dados estatísticos; Textos simples; gráficos; Imagens |
| Colômbia | Geociências | Documentos padrão |
| México | Geologia; Geofísica | Documentos padrão; Imagens; Dados estatísticos; Dados b... |
| Panamá | Multidisciplinar | Gráficos; Estatísticos; Dados Brutos |
| Peru | Oceanografia; Geociências | Não especificado |
| | Agricultura | Qualquer formato de dados |
| | Agricultura; Climatologia | Documentos padrão; Bases de dados; Imagens; Dados esta... |
| | Astronomia | Audiovisual |
| | Ciências da Saúde; Ciências Sociais | Dados brutos |
| | Educação | Dados brutos; Dados estatísticos |
| | Genética | Dados brutos; Imagens |
| | | Dados estatísticos; Dados brutos; Softwares; Textos |

Fonte: Elaboração própria

Climatologia; Ecologia, Geologia, Geofísica, Oceanografia e particularmente a área Multidisciplinar.

O Brasil, a partir do levantamento realizado, possui repositórios de dados concentrados na grande área de Ciências Exatas e da Terra. É possível encontrar nas palavras de Eason et al. (1997) uma possível explicação para a concentração dos repositórios de dados nessas áreas, afinal, conforme eles afirmaram, as ciências mais duras tendem a ter um desenvolvimento cumulativo de conhecimento e trabalham adicionando novos conhecimentos aos trabalhos de seus pares.

Quanto a natureza dos dados compartilhados em repositórios de dados brasileiros, estes possuem natureza estatística, imagética, bruta e o IBICT Dataverse Network afirma fazer a preservação de todos os tipos de dados, o que não pode ser comprovado já que o repositório não abre todas as suas coleções para o público, o que expressamente contraria os princípios de acessibilidade estruturado no relatório da MGI (2013) e Declaração de Budapeste.

5.1.1.2. Chile

Em relação ao país Chile, foi identificado que único repositório de dados da região, o eFish Genomic Database Repository, compartilha dados de natureza especializada nas áreas de Genética e Zoologia.

Figura 36: Relação Tipo de dados - Área Geográfica – Área do Conhecimento – Conteúdo armazenado - Chile

| Área Geográfica | Área do conhecimento | Conteúdo armazenado |
|-----------------|---|---|
| Brasil | Genética; Zoologia | Dados brutos; Imagens |
| Chile | Agricultura | Documentos padrão; Bases de dados; Imagens; Dados esta... |
| Colômbia | Agricultura; Climatologia | Audiovisual |
| México | Astronomia | Conjunto de dados; Planilhas; Dados espectrais; Document... |
| Panamá | Ciências da Saúde; Ciências Sociais | Dados brutos |
| Peru | Climatologia | Dados brutos; Dados estatísticos |
| | Ecologia | Dados estatísticos; Dados brutos; Softwares; Textos |
| | Educação | Dados estatísticos; Textos simples; gráficos; Imagens |
| | Genética | Documentos padrão |
| | Geociências | Documentos padrão; Dados estatísticos; Imagens; Dados au... |
| | Geologia; Geofísica | Documentos padrão; Código fonte; Softwares |
| | Matemática; Biologia; Ciências Naturais | Documentos padrão; Dados estatísticos; dados brutos |
| | | Documentos padrão; Imagens; Dados estatísticos; Dados b... |

Fonte: Elaboração própria

Este repositório dissemina dados de natureza imagética e bruta.

5.1.1.3. Colômbia

A relação Área Geográfica-Área do Conhecimento na Colômbia mostra que os Repositórios de dados abrangem as áreas de:

- Agricultura;
- Climatologia

Figura 37: Relação Tipo de dados - Área Geográfica – Área do Conhecimento – Conteúdo armazenado - Colômbia

| Área Geográfica | Área do conhecimento | Conteúdo armazenado |
|-----------------|---|---|
| Brasil | Agricultura | Documentos padrão; Dados estatísticos;Imagens;Dados au... |
| Chile | Agricultura; Climatologia | Planilhas;Documentos padrão;Imagens;Dados estatísticos;... |
| Colômbia | Astronomia | Documentos padrão; Bases de dados; Imagens; Dados esta... |
| México | Ciências da Saúde; Ciências Sociais | Audiovisual |
| Panamá | Climatologia | Conjunto de dados; Planilhas; Dados espectrais; Document... |
| Peru | Ecologia | Dados brutos |
| | Educação | Dados brutos; Dados estatísticos |
| | Genética | Dados brutos; Imagens |
| | Genética; Zoologia | Dados estatísticos; Dados brutos; Softwares; Textos |
| | Geociências | Dados estatísticos; Textos simples; gráficos; Imagens |
| | Geologia; Geofísica | Documentos padrão |
| | Matemática; Biologia; Ciências Naturais | Documentos padrão; Código fonte; Softwares |
| | | Documentos padrão; Dados estatísticos; dados brutos |

Fonte: Elaboração própria

Quanto ao conteúdo armazenado, os repositórios de dados colombianos abrigam:

- Documentos padrão;
- Dados estatísticos;
- Imagens;
- Dados audiovisuais;
- Dados brutos e
- Planilhas.

5.1.1.4. México

O levantamento realizado indica que o México, mesmo tendo o mesmo número de repositórios de dados ativos que o Brasil, tem uma cobertura superior no que se refere as áreas do conhecimento abrangidas.

Figura 38: Relação Tipo de dados - Área Geográfica – Área do Conhecimento – Conteúdo armazenado - México

| Área Geográfica | Área do conhecimento | Conteúdo armazenado |
|-----------------|---|---|
| Brasil | Agricultura | Audiovisual |
| Chile | Astronomia | Dados brutos |
| Colômbia | Ciências da Saúde; Ciências Sociais | Dados brutos; Dados estatísticos |
| México | Genética | Dados estatísticos; Dados brutos; Softwares; Textos |
| Panamá | Matemática; Biologia; Ciências Naturais | Documentos padrão; Código fonte; Softwares |
| Peru | Microbiologia alimentar | Documentos padrão; Dados estatísticos; dados brutos |
| | Taxonomia; Biologia; Ciências da Saúde | Simulações; Artigos; Imagens; Vídeos; Observações |
| | Agricultura; Climatologia | Taxonomias |
| | Climatologia | Documentos padrão; Bases de dados; Imagens; Dados esta... |
| | Ecologia | Conjunto de dados; Planilhas; Dados espectrais; Document... |
| | Educação | Dados brutos; Imagens |
| | Genética; Zoolonia | Dados estatísticos; Textos simples; gráficos; Imagens |
| | | Documentos padrão |

Fonte: Elaboração própria

- Agricultura;
- Astronomia;
- Ciências da Saúde;
- Ciências Sociais;
- Genética;
- Matemática;
- Biologia;
- Ciências Naturais;
- Taxonomias;
- Microbiologia alimentar e
- Biologia.

Igualmente diferentemente do Brasil, os repositórios de dados mexicanos se cobrem uma maior quantidade de grandes áreas do conhecimento, mas também se destaca na grande área de Ciências Exatas e da Terra.

Em relação aos dados disseminados, os RD's mexicanos compartilham:

- Documentos padrão;
- Dados estatísticos;
- Imagens;
- Dados audiovisuais;
- Dados brutos;
- Simulações;
- Vídeos;
- Observações;
- Taxonomias;
- Código fonte e
- Softwares.

Uma possível consequência da variedade superior que o México possui em relação aos tipos de dados pode ser baseada nos estudos feitos por Mueller (2015) que lista as diferenças disciplinares entre as áreas do conhecimento em relação aos canais preferenciais de publicação científica.

Meadows (1999) também possui a crença de que existem diferenças significativas entre as áreas do conhecimento e essas diferenças se refletem nos padrões de comunicação, conforme pode ser observado.

5.1.1.5. Panamá

Os dois repositórios panamenhos analisados apenas compartilham dados em repositórios na área de:

- Ecologia

Figura 39: Relação Tipo de dados - Área Geográfica – Área do Conhecimento – Conteúdo armazenado - Panamá

| Área Geográfica | Área do conhecimento | Conteúdo armazenado |
|-----------------|---|---|
| Brasil | Ecologia | Não especificado |
| Chile | Agricultura | Documentos padrão; Bases de dados; Imagens; Dados esta... |
| Colômbia | Agricultura; Climatologia | Audiovisual |
| México | Astronomia | Conjunto de dados; Planilhas; Dados espectrais; Document... |
| Panamá | Ciências da Saúde; Ciências Sociais | Dados brutos |
| Peru | Climatologia | Dados brutos; Dados estatísticos |
| | Educação | Dados brutos; Imagens |
| | Genética | Dados estatísticos; Dados brutos; Softwares; Textos |
| | Genética; Zoologia | Dados estatísticos; Textos simples; gráficos; Imagens |
| | Geociências | Documentos padrão |
| | Geologia; Geofísica | Documentos padrão; Dados estatísticos; Imagens; Dados au... |
| | Matemática; Biologia; Ciências Naturais | Documentos padrão; Código fonte; Softwares |
| | | Documentos padrão; Dados estatísticos; dados brutos |

Fonte: Elaboração própria

Não foi possível mensurar quais os tipos de dados compartilhados nesses repositórios. Ambos os repositórios exigem que os usuários preencham um formulário em que sejam listados os dados de interesse e as intenções de uso desses dados.

O condicionamento de acesso aos dados científicos dos repositórios panamenhos impede a total transparência do processo científico que é defendido pela Ciência Aberta e aproxima esses repositórios mais de um esquema de funcionamento de um arquivo privado.

5.1.1.6. Peru

O levantamento realizado indica que o Peru possui repositórios que compartilham dados nas áreas de:

- Ecologia e
- Educação

Figura 40: Relação Tipo de dados - Área Geográfica – Área do Conhecimento – Conteúdo armazenado - Peru

| Área Geográfica | Área do conhecimento | Conteúdo armazenado |
|-----------------|---|---|
| Brasil | Ecologia | Documentos padrão; Bases de dados; Imagens; Dados esta... |
| Chile | Educação | Planilhas; Documentos padrão |
| Colômbia | Agricultura | Audiovisual |
| México | Agricultura; Climatologia | Conjunto de dados; Planilhas; Dados espectrais; Document... |
| Panamá | Astronomia | Dados brutos |
| Peru | Ciências da Saúde; Ciências Sociais | Dados brutos; Dados estatísticos |
| | Climatologia | Dados brutos; Imagens |
| | Genética | Dados estatísticos; Dados brutos; Softwares; Textos |
| | Genética; Zoologia | Dados estatísticos; Textos simples; gráficos; Imagens |
| | Geociências | Documentos padrão |
| | Geologia; Geofísica | Documentos padrão; Dados estatísticos; Imagens; Dados au... |
| | Matemática; Biologia; Ciências Naturais | Documentos padrão; Código fonte; Softwares |
| | | Documentos padrão; Dados estatísticos; dados brutos |

Fonte: Elaboração própria

Quanto aos dados compartilhados em repositórios no Peru, eles se manifestam nas seguintes categorias:

- Documentos padrão;
- Bases de dados;
- Dados brutos;
- Qualquer formato de dados;
- Simulações;
- Observações e
- Planilhas

Para a pouca quantidade de repositórios, o Peru se destaca na variedade dos tipos de dados compartilhados.

Em resumo, o quadro que pode ser observado na América Latina em relação ao compartilhamento de dados de acordo com a sua natureza gera o seguinte estado da arte:

Figura 41: Segmentação do tipo de dados por região geográfica



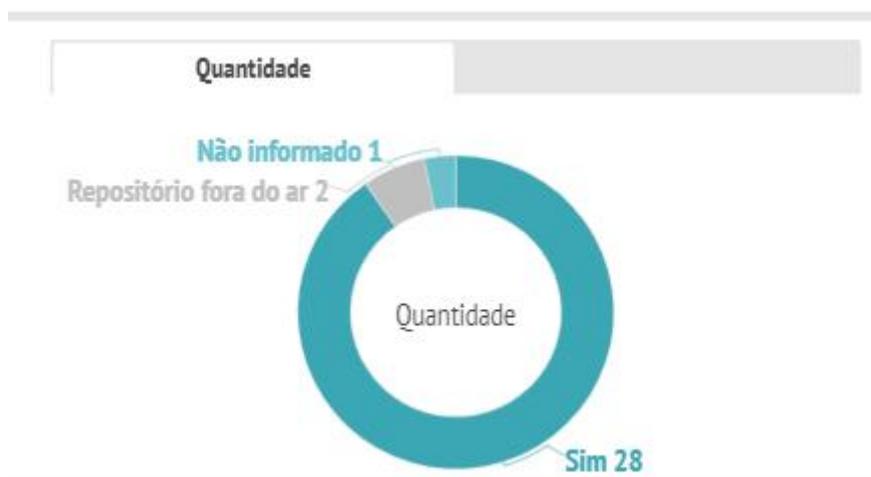
Fonte: Elaboração própria

5.2. Relação Área Geográfica- Vínculo institucional

Dos repositórios que se auto identificaram como repositórios que compartilham dados, 28 indicaram ter vínculo institucional, podendo ser esse vínculo estabelecido por meio de uma instituição formal ou de um grupo de pesquisa oficializado.

Do total de repositórios analisados, dois se encontravam fora do ar no momento da coleta de dados e um não informou ter qualquer tipo de vínculo com alguma instituição ou grupo.

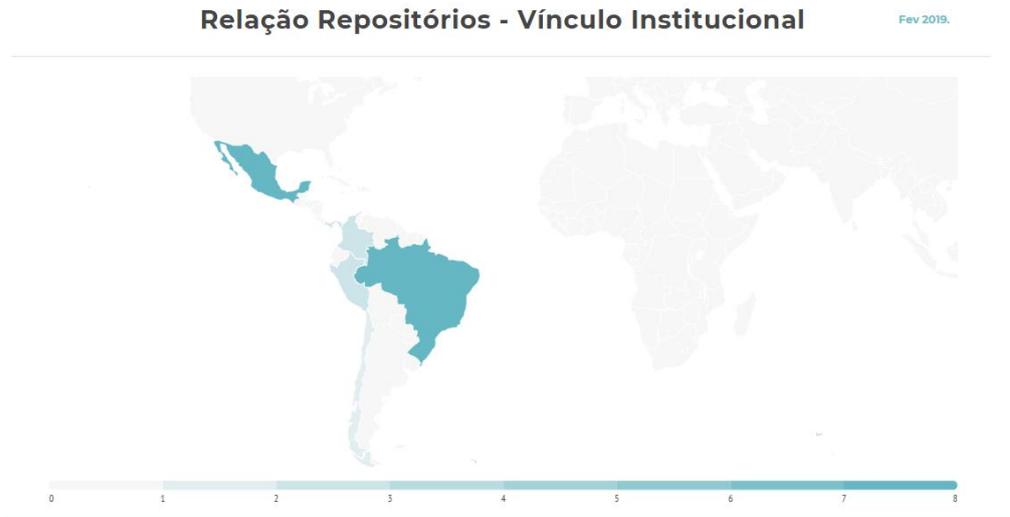
Figura 42: Vínculo institucional dos RD's



Fonte: Elaboração própria

No entanto, dos repositórios que caracterizavam como disseminadores de dados, averiguou-se que todos os 23 repositórios possuíam vínculo institucional, sendo que:

Figura 43: Relação Área Geográfica- Vínculo institucional



Fonte: Elaboração própria

Dos repositórios detectados como possuindo vínculo institucional:

- 8 são brasileiros;
- 1 é chileno;
- 2 são colombianos;
- 8 são mexicanos;
- 2 são panamenhos e
- 2 são peruanos.

A pesquisa mostrou que existe uma grande tendência de vínculo institucional nos repositórios de dados latino americanos. Foi percebido que esse vínculo tanto se estabelece entre a relação repositório-instituição quanto na relação repositório-grupo de pesquisa.

A literatura científica corrobora esses vínculos quando afirma que cada vez mais as práticas de investigação fogem do ambiente tradicional das instituições de ensino – pesquisa – extensão das grandes universidades, conforme Abdo (2014) indicou.

5.3. Relação Área Geográfica- Perfil acadêmico

O perfil acadêmico de um repositório de dados indica uma característica importante sobre a natureza de seus dados, pois eles já são coletados com uma finalidade científica declarada. Em repositórios não acadêmicos, há a possibilidade de disseminação de dados científicos ou que eles sejam utilizados por pesquisadores e transformados em dados científicos. No entanto, foi constatado que dentre os repositórios:

Figura 44: Relação Área Geográfica- Perfil acadêmico



Fonte: Elaboração própria

Oito já possuem uma finalidade acadêmica na curadoria de seus dados. Desses oito repositórios:

- 3 são brasileiros;
- 1 é chileno e
- 4 são mexicanos.

Há nesse dado, uma comprovação das premissas do ciclo de vida de uma pesquisa descrito por Assante (2015), onde foi destacado a atuação dos

próprios ambientes de pesquisas como publicadores e disseminadores de seus próprios produtos e subprodutos de pesquisa.

5.4. Análise dos critérios de Características de Acesso Aberto dos RD's

As variáveis estipuladas para a avaliação dos repositórios de dados quanto às diretrizes do Movimento de Ciência Aberta foram estipuladas com base nos pilares para essa corrente que foram descritos na revisão de literatura traçada.

5.4.1. Relações Área Geográfica – Tipo de licença adotada – Conteúdo completo

O tipo de licença adotada pelos gestores do repositório de dados se relaciona diretamente com o Movimento de Ciência Aberta, segundo preceitos de Andrade (2015) já descritos durante a revisão de literatura. A autora afirma que a abertura de dados deve ser feita de forma que os dados possam ser acessados e reutilizados. O que por sua vez, remete ao questionamento se o conteúdo que está sendo disponibilizado, mesmo que de forma aberta, é um conteúdo que se encontra na íntegra ou fracionado.

Como já defendido na revisão literária, Maurer (2013) indica três pilares fundamentais para a Ciência Aberta, sendo que o primeiro seria que as publicações que estivessem dentro do movimento deveriam ser “a) uma publicação completa, franca e rápida resultados;”.

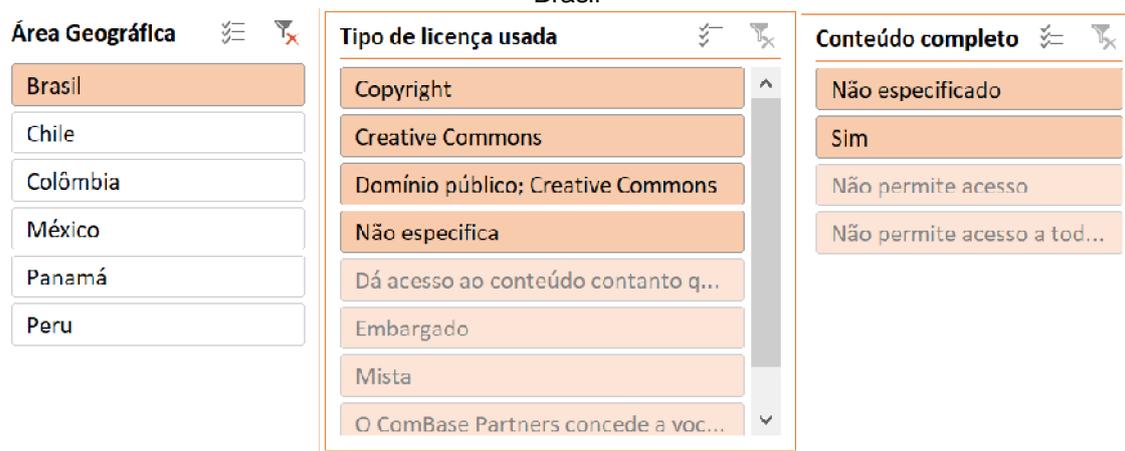
O resultado que a coleta de dados retornou sobre os assuntos foi:

5.4.1.1. Brasil

Os tipos de licença adotada pelos gestores do repositório de dados brasileiros são:

- Copyright;
- Creative Commons;
- Domínio público e
- Não especificada pelos gestores.

Figura 45: Relações Área Geográfica – Tipo de licença adotada – Conteúdo completo – Brasil



Fonte: Elaboração própria

O país segue uma tendência de disponibilização completa do conteúdo, sendo que, dos oito repositórios identificados na região, seis disponibilizam o conteúdo completo de suas publicações.

Os resultados obtidos indicam que, nesse requisito, o Brasil segue os princípios estabelecidos na revisão de literatura por Molloy (2011), Frecher e Friesike (2014) e Maurer (2013) que indicam como elemento basilar que os dados compartilhados estejam em publicação aberta, completa e sem restrições de propriedade intelectual.

5.4.1.2. Chile

O perfil do repositório chileno quanto à adoção de licenças se mostrou estar de acordo com o que a literatura afirma. A licença adotada foi Creative Commons.

Figura 46: Área Geográfica – Tipo de licença adotada – Conteúdo completo – Chile

| Área Geográfica | Tipo de licença usada | Conteúdo completo |
|-----------------|-------------------------------------|-----------------------------|
| Brasil | Creative Commons | Sim |
| Chile | Copyright | Não especificado |
| Colômbia | Dá acesso ao conteúdo contanto q... | Não permite acesso |
| México | Domínio público; Creative Commons | Não permite acesso a tod... |
| Panamá | Embargado | |
| Peru | Mista | |
| | Não especifica | |
| | O ComBase Partners concede a voc... | |

Fonte: Elaboração própria

Sendo que o conteúdo de seus repositórios é disponibilizado na íntegra.

5.4.1.3. México

Quanto as licenças adotadas, o México demonstra ter uma particularidade em relação aos outros países, visto que, além de utilizar as licenças:

- Creative Commons;
- Ter acesso embargado ao seu conteúdo;
- Um dos repositórios não cita uma licença específica, mas declara que os usuários podem ter acesso ao conteúdo, contando que cite a fonte, o que é típico de uma licença Creative Commos.
- O repositório ComBase declara ter uma licença de uso próprio que permite ao usuário “reproduzir ou distribuir qualquer parte do conteúdo para uso pessoal, pesquisa acadêmica ou científica ou estudo privado e para fins comerciais internos. ”, entretanto possui um conteúdo fechado com base em cadastro.

Figura 47: Área Geográfica – Tipo de licença adotada – Conteúdo completo – México

| Área Geográfica | Tipo de licença usada | Conteúdo completo |
|-----------------|-------------------------------------|-----------------------------|
| Brasil | Creative Commons | Não especificado |
| Chile | Dá acesso ao conteúdo contanto q... | Sim |
| Colômbia | Embargado | Não permite acesso |
| México | Mista | Não permite acesso a tod... |
| Panamá | Não especifica | |
| Peru | O ComBase Partners concede a voc... | |
| | Copyright | |
| | Domínio público; Creative Commons | |

Fonte: Elaboração própria

5.4.1.4. Colômbia

A Colômbia demonstra utilizar tanto as licenças Creative Commons, que são licenças próprias de quem dissemina o conteúdo de uma forma aberta, como o Copyright, que por sua vez é uma licença própria de conteúdos que são restritos.

Essa ação está de acordo com a forma de disponibilização do conteúdo, que possui uma abordagem mista. Um dos repositórios analisados não permite o acesso ao seu conteúdo, enquanto o outro, disponibiliza parcialmente.

Figura 48: Área Geográfica – Tipo de licença adotada – Conteúdo completo – Colômbia

| Área Geográfica | Tipo de licença usada | Conteúdo completo |
|-----------------|-------------------------------------|-----------------------------|
| Brasil | Copyright | Não permite acesso |
| Chile | Creative Commons | Não permite acesso a tod... |
| Colômbia | Dá acesso ao conteúdo contanto q... | Não especificado |
| México | Domínio público; Creative Commons | Sim |
| Panamá | Embargado | |
| Peru | Mista | |
| | Não especifica | |
| | O ComBase Partners concede a voc... | |

Fonte: Elaboração própria

5.4.1.5. Panamá

O Panamá não especifica ambos os aspectos analisados: O tipo de licença e nem se disponibilizam o conteúdo na íntegra.

Figura 49: Área Geográfica – Tipo de licença adotada – Conteúdo completo – Panamá

The screenshot displays a web interface with three filter panels. The first panel, 'Área Geográfica', lists countries: Brasil, Chile, Colômbia, México, Panamá (highlighted), and Peru. The second panel, 'Tipo de licença usada', lists license types: Não especifica, Copyright, Creative Commons, Dá acesso ao conteúdo contanto q..., Domínio público; Creative Commons, Embargado, Mista, and O ComBase Partners concede a voc... The third panel, 'Conteúdo completo', lists content availability options: Não especificado, Não permite acesso, Não permite acesso a tod..., and Sim.

Fonte: Elaboração própria

No entanto, cabe ressaltar que foi apurado durante o levantamento dos dados que, um dos repositórios do Panamá exige que os usuários preencham um formulário em que sejam especificadas quais são os conteúdos de interesse e qual será a destinação do uso, para que então, os dados sejam fornecidos.

Neste ponto, essa prática se configura como fora dos padrões apontados pela literatura, pois a revisão de literatura indica e recomenda a livre circulação do conhecimento, de acordo com o que foi afirmado por Gonzales (2006).

5.4.1.6. Peru

Como já citado anteriormente, o Peru possui dois repositórios de dados cadastrados no RE3data.org. Quanto ao tipo de licença adotada, um segue a licença copyright, enquanto o outro não especifica o tipo de licença adotada.

Figura 50: Área Geográfica – Tipo de licença adotada – Conteúdo completo – Peru

| Área Geográfica | Tipo de licença usada | Conteúdo completo |
|-----------------|-------------------------------------|-----------------------------|
| Brasil | Copyright | Não especificado |
| Chile | Não especifica | Sim |
| Colômbia | Creative Commons | Não permite acesso |
| México | Dá acesso ao conteúdo contanto q... | Não permite acesso a tod... |
| Panamá | Domínio público; Creative Commons | |
| Peru | Embargado | |
| | Mista | |
| | O ComBase Partners concede a voc... | |

Fonte: Elaboração própria

O mesmo parecer sobre a necessidade de disponibilização completa do conteúdo foi defendido pela Declaração de Bethesda (2003) onde foi declarado que mesmo os materiais suplementares à pesquisa deveriam estar em uma versão completa incluindo também uma cópia da permissão para o uso livre dos materiais disponibilizados.

5.4.2. Colaboração científica

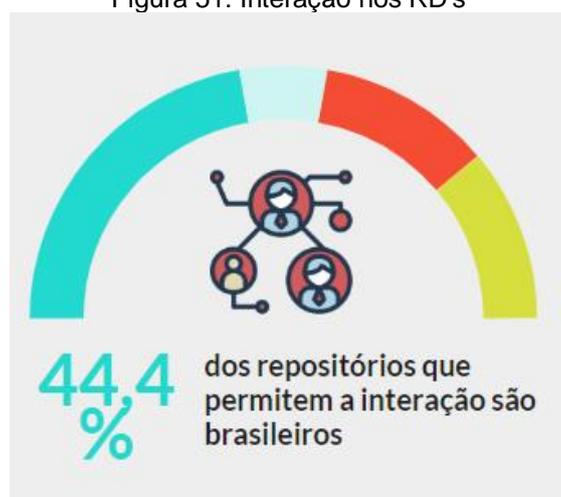
A colaboração científica foi um ponto altamente levantado na literatura a respeito da Ciência Aberta. Durante a apuração da bibliografia sobre o tema, autores como Tenopir et al. (2011), Merton (1974) e Canessa e Zennaro (2008) destacaram a importância da colaboração científica para o avanço da ciência.

A colaboração científica também foi abordada pelos autores Fecher e Friesike (2014) que declararam que “em busca da construção de uma ciência coletiva, os pesquisadores buscam ferramentas eletrônicas de interação social que facilitam e aceleram a descoberta científica.”.

Tendo essa questão em destaque, a pesquisa já havia analisado a presença da colaboração científica entre os países latino americanos com outros países, no entanto, a pesquisa também buscou investigar se as ferramentas tecnológicas utilizadas para a disponibilização dos dados também permitiam que os usuários, além de ter acesso aos dados disseminados pudessem interagir de alguma forma com os produtores ou compiladores dos dados disponibilizados ou ainda fossem capazes de deixar contribuições públicas para o conjunto de dados.

Sendo assim, foi averiguado que nove repositórios permitiam alguma forma de interação entre a comunidade científica, sendo que destes:

Figura 51: Interação nos RD's



Fonte: Elaboração própria

- 4 eram brasileiros;
- 1 era chileno;
- 2 colombianos e
- 2 eram mexicanos.

Esse resultado converge com a literatura científica onde foi destacado pelos autores Hey e Trefethen (2005) uma corrente de colaboração que está sendo facilitada tanto pelas ferramentas da Web 2.0, como pela criação e adoção de novas ferramentas para a troca de informações, opiniões e troca de recursos.

Os nove repositórios indicados com tendências na colaboração científica convergiam para a construção de uma Ciência Aberta. Segundo o levantamento bibliográfico dessa dissertação, a colaboração científica não só aumenta a qualidade das informações coletadas, mas também a reputação dos pesquisadores envolvidos e a arrecadação de recursos, segundo o que foi destacado por Andrade (2015).

5.5. Análise dos critérios de Características de Gestão e Funcionamento dos RD's

5.5.1. Relações Área Geográfica – Critérios de Persistência – Estatística

Conforme foi exposto na revisão de literatura, um dos critérios de avaliação da qualidade de um repositório é a adoção de critérios de persistência que, não só auxiliam na visibilidade do autor dos dados, de acordo com o que afirmaram Klump et al (2006), mas também auxiliam no arquivamento e curadoria dos dados.

Outra faceta que auxilia a mensurar a visibilidade de um repositório se dá por meio da utilização de estatísticas que permitem avaliar o desempenho que o repositório tem dentro da comunidade.

A respeito da adoção de critérios de persistência e a adoção de estatísticas, foi observado que:

5.5.1.1. Brasil

Os repositórios brasileiros apresentaram a utilização de dois critérios de persistência, o DOI e o Handle. Dois dos repositórios analisados que eram pertencentes ao Brasil não adotaram o uso de critérios de persistência.

Figura 52: Relações Área Geográfica – Critérios de Persistência – Estatística – Brasil

| Área Geográfica | Critérios de persistência | Estatísticas |
|-----------------|---------------------------|--------------|
| Brasil | DOI | Não |
| Chile | DOI; Handle | Sim |
| Colômbia | Não | |
| México | Handle | |
| Panamá | Não especificado | |
| Peru | Sim | |

Fonte: Elaboração própria

Quanto ao uso de ferramentas de estatística, três repositórios disponibilizam estatísticas e um não.

5.5.1.2. Chile

Quanto ao uso de critérios de persistência e adoção de estatísticas, o repositório chileno “eFish Genomic Database Repository” apontou o uso tanto do DOI como a presença de estatísticas para os documentos depositados.

Figura 53: Relações Área Geográfica – Critérios de Persistência – Estatística – Chile

| Área Geográfica | Critérios de persistência | Estatísticas |
|-----------------|---------------------------|--------------|
| Brasil | DOI | Não |
| Chile | DOI; Handle | Sim |
| Colômbia | Handle | |
| México | Não | |
| Panamá | Não especificado | |
| Peru | Sim | |

Fonte: Elaboração própria

5.5.1.3. Colômbia

Quando a esses critérios de análise, o repositório de dados colombiano “The Global Agricultural Trial Repository and Database” não adota nenhum critério de persistência, entretanto, faz o uso de estatísticas em seus itens.

Figura 54: Relações Área Geográfica – Critérios de Persistência – Estatística – Colômbia

| Área Geográfica | Critérios de persistência | Estatísticas |
|-----------------|---------------------------|--------------|
| Brasil | DOI | Sim |
| Chile | Não | Não |
| Colômbia | DOI; Handle | |
| México | Handle | |
| Panamá | Não especificado | |
| Peru | Sim | |

Fonte: Elaboração própria

Já o repositório “CIAT Dataverse” utiliza ambas as ferramentas, sendo que o critério de persistência utilizado foi o DOI.

5.5.1.4. México

Dos oito repositórios mexicanos analisados nesses requisitos:

- Seis não utilizam critérios de persistência em seus arquivos;
- Um utiliza ambos os critérios: handle e DOI e
- Um utiliza apenas o handle.

Figura 55: Relações Área Geográfica – Critérios de Persistência – Estatística – México

| Área Geográfica | Critérios de persistência | Estatísticas |
|-----------------|---------------------------|--------------|
| Brasil | DOI; Handle | Não |
| Chile | Handle | Sim |
| Colômbia | Não | |
| México | DOI | |
| Panamá | Não especificado | |
| Peru | Sim | |

Fonte: Elaboração própria

Dos mesmos oito repositórios:

- Cinco não utilizam estatísticas;
- Três a utilizam.

5.5.1.5. Panamá

O Panamá não adota ambos os itens analisados, tanto não adotam critérios de persistência tampouco usam estatísticas.

Figura 56: Relações Área Geográfica – Critérios de Persistência – Estatística – Panamá

| Área Geográfica | Critérios de persistência | Estatísticas |
|-----------------|---------------------------|--------------|
| Brasil | Não | Não |
| Chile | Não especificado | Sim |
| Colômbia | DOI | |
| México | DOI; Handle | |
| Panamá | Handle | |
| Peru | Sim | |

Fonte: Elaboração própria

5.5.1.6. Peru

Dos repositórios de dados mexicanos, o “Repositorio Institucional USIL” adota ambos os itens analisados.

Enquanto o repositório “Repositorio de datos del Ministerio de Educación del Perú” não adota nem critérios de persistência, tampouco o uso de estatísticas.

Figura 57: Relações Área Geográfica – Critérios de Persistência – Estatística – Peru

| Área Geográfica | Critérios de persistência | Estatísticas |
|-----------------|---------------------------|--------------|
| Brasil | Não | Não |
| Chile | Sim | Sim |
| Colômbia | DOI | |
| México | DOI; Handle | |
| Panamá | Handle | |
| Peru | Não especificado | |

Fonte: Elaboração própria

5.5.1. Relações Área Geográfica – Esquema de Software e metadados adotados

Seguindo o esquema de análise de características relacionadas a gestão e funcionamento dos repositórios de dados, duas particularidades se tornam importantes na análise proposta por essa dissertação: o software utilizado e o esquema de metadados utilizado.

Durante a revisão de literatura, Costa e Braga (2016) elucidaram que três softwares se destacam em âmbito mundial na criação de repositórios de dados de pesquisa: o DSpace; o Dataverse e o CKAN. Enquanto “Outros”; o Dublin Core e o DDI se destacam se tratando de esquemas de metadados.

Em um âmbito latino americano, foi averiguado que:

5.5.1.1. Brasil

No Brasil, apenas o repositório de dados do IBICT, o “IBICT Dataverse Network” utiliza o Dataverse como software para a gestão de seu conteúdo. O mesmo repositório utiliza o esquema descrição DDI para de seu conteúdo.

Figura 58: Relações Área Geográfica – Esquema de Software e metadados adotados - Brasil

The image shows a screenshot of a web-based interface with three columns: 'Área Geográfica', 'Software', and 'Metadados'. Each column has a list of items and a filter icon. The 'Área Geográfica' column lists: Brasil, Chile, Colômbia, México, Panamá, and Peru. The 'Software' column lists: Dataverse, DSpace software, Metacat, Não especificado, DSpace Software, DSpace, public domain software devel..., and WordPress. The 'Metadados' column lists: DDI, Dublin core, EML - Ecological Metad..., Não especificado, and ITIS Data Model.

| Área Geográfica | Software | Metadados |
|-----------------|---------------------------------|---------------------------|
| Brasil | Dataverse | DDI |
| Chile | DSpace software | Dublin core |
| Colômbia | Metacat | EML - Ecological Metad... |
| México | Não especificado | Não especificado |
| Panamá | DSpace Software | ITIS Data Model |
| Peru | DSpace | |
| | public domain software devel... | |
| | WordPress | |

Fonte: Elaboração própria

Se tratando da mesma região, dois repositórios indicam usar o DSpace e o esquema de Dublin Core para a gestão de conteúdo enquanto o “PPBio Data Repository” utiliza o software Metacat e o esquema “EML - Ecological Metadata Language” para descrição dos dados. Entretanto, quatro repositórios não especificaram nem o software utilizado, tampouco o esquema de metadados.

O que foi apurado nessa pesquisa sobre o uso do software DSpace e o Dataverse para a criação e manutenção de repositórios de dados, corrobora os resultados encontrados por Costa e Braga que apontaram os dois softwares como os mais utilizados em nível mundial.

5.5.1.2. Chile

O repositório chileno “eFish Genomic Database Repository” não especificou nem o software nem o esquema de metadados utilizado.

Figura 59: Relações Área Geográfica – Esquema de Software e metadados adotados - Chile

| Área Geográfica | Software | Metadados |
|-----------------|---------------------------------|---------------------------|
| Brasil | Não especificado | Não especificado |
| Chile | DSpace Software | DDI |
| Colômbia | Dataverse | Dublin core |
| México | DSpace | EML - Ecological Metad... |
| Panamá | DSpace software | ITIS Data Model |
| Peru | Metacat | |
| | public-domain software devel... | |
| | WordPress | |

Fonte: Elaboração própria

5.5.1.3. Colômbia

Em um âmbito geográfico colombiano, o repositório “The Global Agricultural Trial Repository and Database” não especificou o software utilizado, tampouco o esquema de metadados adotado. Entretanto o “CIAT

Dataverse” foi construído dentro do software livre Datavere utilizando o esquema de metadados DDI.

Conforme foi apontado na revisão de literatura por Costa e Braga (2016) é comum que sejam encontrados resultados que combinem o uso do Dataverse como software de gestão com o esquema DDI para descrição já que ambos são adotados em por grande parte dos gestores de repositórios em nível mundial.

Figura 60: Relações Área Geográfica – Esquema de Software e metadados adotados – Colômbia

| Área Geográfica | Software | Metadados |
|-----------------|---------------------------------|---------------------------|
| Brasil | Dataverse | DDI |
| Chile | Não especificado | Não especificado |
| Colômbia | DSpace Software | Dublin core |
| México | DSpace | EML - Ecological Metad... |
| Panamá | DSpace software | ITIS Data Model |
| Peru | Metacat | |
| | public domain software devel... | |
| | WordPress | |

Fonte: Elaboração própria

5.5.1.4. México

Dos oito repositórios mexicanos, quatro não especificaram quais os softwares utilizados. Idem o esquema de metadados empregado.

Figura 61: Relações Área Geográfica – Esquema de Software e metadados adotados - México

| Área Geográfica | Software | Metadados |
|-----------------|---------------------------------|---------------------------|
| Brasil | Dataverse | ITIS Data Model |
| Chile | DSpace | Não especificado |
| Colômbia | Não especificado | DDI |
| México | public-domain software devel... | Dublin core |
| Panamá | WordPress | EML - Ecological Metad... |
| Peru | DSpace Software | |
| | DSpace software | |
| | Metacat | |

Fonte: Elaboração própria

Dos quatro restantes:

- Constrained Local UniversE Simulations foi construído em WordPress;
- Banff International Research Station for Mathematical Innovation and Discovery utiliza o DSpace;
- Integrated Taxonomic Information System utiliza um software próprio de domínio público chamado “ITIS”.
- CIMMYT Research Data & Software Repository Network foi construído com o Dataverse.

5.5.1.5. Panamá

Os repositórios do Panamá não especificam o software e o esquema de metadados.

Figura 62: Relações Área Geográfica – Esquema de Software e metadados adotados - Panamá

| Área Geográfica | Software | Metadados |
|-----------------|---------------------------------|---------------------------|
| Brasil | Não especificado | Não especificado |
| Chile | DSpace Software | DDI |
| Colômbia | Dataverse | Dublin core |
| México | DSpace | EML - Ecological Metad... |
| Panamá | DSpace software | ITIS Data Model |
| Peru | Metacat | |
| | public domain software devel... | |
| | WordPress | |

Fonte: Elaboração própria

5.5.1.6. Peru

O repositório peruano “Repositorio Institucional USIL” utiliza o DSpace para a gestão de seu conteúdo e o Dublin como esquema de metadados.

Figura 63: Relações Área Geográfica – Esquema de Software e metadados adotados - Peru

| Área Geográfica | Software | Metadados |
|-----------------|---------------------------------|---------------------------|
| Brasil | DSpace Software | Dublin core |
| Chile | Não especificado | Não especificado |
| Colômbia | Dataverse | DDI |
| México | DSpace | EML - Ecological Metad... |
| Panamá | DSpace software | ITIS Data Model |
| Peru | Metacat | |
| | public-domain software devel... | |
| | WordPress | |

Fonte: Elaboração própria

Entretanto, o “Repositorio de datos del Ministerio de Educación del Perú” não especifica nem software idem o esquema de metadados.

5.5.2. Relação Área Geográfica – Política de Gestão e Política de Gestão presente no Melibea

No decorrer da dissertação, ressaltou-se a necessidade da formulação de estratégias pelos gestores de repositórios para a melhor curadoria e disseminação de seus conteúdos. Essa prática gera maior visibilidade e consolida as ações da preservação dos itens.

Se tratando da consolidação das práticas gerenciais adotadas, é fundamental que essas ações estejam descritas em um documento que elucide e direcione as ações dos gestores, autores e usuários.

Por ser um documento fundamental para a gestão do repositório, essa dissertação também buscou investigar quais repositórios latino americanos possuam e disponibilizavam documentos sobre a política de gestão dos repositórios. Essa investigação observou a presença do documento no próprio site do repositório. Complementarmente a essa avaliação, também se buscou no diretório de políticas abertas, o Melibea²⁴, a discriminação das políticas de gestão deles.

Os resultados obtidos foram:

5.5.2.1. Brasil

Se tratando da presença de políticas de gestão, dos oito repositórios brasileiros, apenas três possuem uma política de gestão de conteúdo, entretendo, nenhuma dessas políticas está cadastrada no diretório de políticas Melibea.

²⁴ <https://www.accesoabierto.net/politicas/>

Figura 64: Relação Área Geográfica – Política de Gestão e Política de Gestão presente no Melibea - Brasil

| Área Geográfica | Política de Gestão | Melibea |
|-----------------|--------------------|---------|
| Brasil | Não | Não |
| Chile | Sim | |
| Colômbia | | |
| México | | |
| Panamá | | |
| Peru | | |

Fonte: Elaboração própria

5.5.2.2. Chile

Na pesquisa realizada, nenhum repositório de dados chileno apresentou uma política de gestão. Tampouco foi encontrado no Melibea algum documento sobre a gestão.

Figura 65: Relação Área Geográfica – Política de Gestão e Política de Gestão presente no Melibea -Chile

| Área Geográfica | Política de Gestão | Melibea |
|-----------------|--------------------|---------|
| Brasil | Não | Não |
| Chile | Sim | |
| Colômbia | | |
| México | | |
| Panamá | | |
| Peru | | |

Fonte: Elaboração própria

Essa apuração vai de encontro as recomendações traçadas na revisão de literatura. Na revisão feita neste trabalho foi destacado a importância da presença de uma política para a eficiente gestão de um ambiente de compartilhamento. Kim e Stanton (2013) chegaram até a relacionar a presença de uma política das entidades com o próprio fator motivacional para o

compartilhamento de dados, visto que, se o compartilhamento consta como uma das exigências das instituições, os autores estariam condicionados a disponibilizar os dados de suas investigações.

5.5.2.3. Colômbia

No levantamento realizado sobre esse aspecto, foi revelado que:

- O repositório colombiano “The Global Agricultural Trial Repository and Database” possui uma política de gestão, todavia, ela não está cadastrada no diretório Melibea.

Figura 66: Relação Área Geográfica – Política de Gestão e Política de Gestão presente no Melibea - Colômbial

| Área Geográfica | Política de Gestão | Melibea |
|-----------------|--------------------|---------|
| Brasil | Não | Não |
| Chile | Sim | |
| Colômbia | | |
| México | | |
| Panamá | | |
| Peru | | |

Fonte: Elaboração própria

Paralelamente, o repositório “CIAT Dataverse” não demonstrou ter uma política, tampouco foi encontrado no Melibea um documento que fizesse menção a este documento no Melibea.

5.5.2.4. México

No âmbito geográfico mexicano apenas os repositórios:

- EcoCyc Database e
- Gran Telescopio CANARIAS Public Archive

Demonstraram não ter uma política de gestão do repositório. Também não foi encontrado no diretório Melibea documentos referentes a gestão desses repositórios.

Figura 67: Relação Área Geográfica – Política de Gestão e Política de Gestão presente no Melibea - México

| Área Geográfica | Política de Gestão | Melibea |
|-----------------|--------------------|---------|
| Brasil | Não | Não |
| Chile | Sim | |
| Colômbia | | |
| México | | |
| Panamá | | |
| Peru | | |

Fonte: Elaboração própria

Mesmo entre aqueles repositórios mexicanos que possuíam uma política de gestão, nenhum documento referente ao tema foi depositado no Melibea.

5.5.2.5. Panamá

Nenhum dos repositórios panamenses cadastrados no RE3data. org apresentou uma política de gestão. Tampouco foi localizado no diretório Melibea documentos referentes as políticas desses repositórios.

Figura 68: Relação Área Geográfica – Política de Gestão e Política de Gestão presente no Melibea - Panamá

| Área Geográfica | Política de Gestão | Melibea |
|-----------------|--------------------|---------|
| Brasil | Não | Não |
| Chile | Sim | |
| Colômbia | | |
| México | | |
| Panamá | | |
| Peru | | |

Fonte: Elaboração própria

5.5.2.6. Peru

Assim como os repositórios panamenses, os repositórios peruanos analisados não apresentaram uma política de gestão. Tampouco foi localizado no diretório Melibea documentos referentes as políticas desses repositórios.

Figura 69: Relação Área Geográfica – Política de Gestão e Política de Gestão presente no Melibea - Peru

| Área Geográfica | Política de Gestão | Melibea |
|-----------------|--------------------|---------|
| Brasil | Não | Não |
| Chile | Sim | |
| Colômbia | | |
| México | | |
| Panamá | | |
| Peru | | |

Fonte: Elaboração própria

5.6. Caracterização dos RD's de acordo com as dimensões apontadas pela literatura

Após o levantamento a fase de levantamento dos dados de acordo com o que foi estipulado pelo instrumento de investigação, que foi construído com bases em critérios expostos pela literatura científica, pode ser estipulado oito critérios básicos que os repositórios de dados deveriam seguir para fornecerem características que contribuíssem para a Ciência Aberta.

Esses oito pontos essenciais foram:

- 1 Se o repositório possuía uma **Licença intelectual** que orientasse os seus usuários quanto aos critérios estabelecidos para a utilização do conteúdo do RD;
- 2 Se o tipo da **Licença de propriedade intelectual** adotada pelo RD fosse do tipo aberta, não necessariamente do tipo Creative Commons, que é a mais recomendada pela literatura, mas que estabelecesse o acesso ao seu conteúdo e permitisse a reutilização de dados;
- 3 Se o RD de alguma forma incentivava ou possibilitava a **Colaboração científica** entre os pesquisadores e o público interessado no tema;
- 4 Se o conteúdo disseminado pelo repositório era um **Conteúdo completo**;
- 5 Se o repositório de dado analisado adotava **Critérios de persistência** para a identificação dos itens depositados;
- 6 Se o Software utilizado para a gestão do repositório também seguia os princípios da Ciência Aberta podendo ser identificado como livre;
- 7 Se o **Conteúdo armazenado** no banco de dados era realmente de natureza de dados ou se na realidade o que era disseminado era o produto científico já consolidado e
- 8 Se havia a presença de **Estatísticas** que servissem de indicadores do uso e alcance que a disseminação feita proporcionava.

A cada ponto de análise foi atribuída uma importância de 12,5% para a contribuição para o Movimento de Ciência Aberta, sendo que, aqueles em que o item de critério não pode ser observado, ou não foi declarado, receberam a pontuação "0". A soma total da satisfação de todos os critérios estipulados caracterizava o repositório de dados como contribuindo em 100% para o desenvolvimento da Ciência Aberta na América Latina.

Cabe ressaltar que para essa categorização, também foram tomados como evidência as observações que puderam ser apreciadas nos repositórios analisados.

Sendo assim, a avaliação dos critérios estipulou que:

Tabela 4: Caracterização²⁵ dos RD's de acordo com as dimensões apontadas pela literatura



| Coluna1 | Repositório | URL | Área Geográfica | Observações de análise | Análise |
|---------|------------------------------------|---|-----------------|--|---|
| 1 | Portal de Datos Genómicos del SNDG | http://datos.sndg.mincyt.gob.ar/ | Argentina | Não armazena conteúdo, logo não pode ser considerado um RD | Não contribui para a C.A. por não ser um Repositório. |
| 2 | Portal de Datos del Mar - SNDM | http://www.datosdelmar.mincyt.gob.ar/ | Argentina | Não armazena conteúdo, logo não pode ser considerado um RD | Não contribui para a C.A. por não ser um Repositório. |
| 3 | Portal de datos de Biodiversidad | http://datos.sndb.mincyt.gob.ar/ | Argentina | Não armazena conteúdo, logo não pode ser considerado um RD | Não contribui para a C.A. por não ser um Repositório. |
| 4 | Repositorio Institucional UCASAL | http://bibliotecas.ucasal.edu.ar/opac_css/index.php?lvl=cmspage&pageid=16 | Argentina | O repositório é um repositório universitário que armazena a produção científica, mas não os dados, logo não pode ser considerado um RD | Não contribui para a C.A. por não ser um Repositório. |

²⁵ Esta tabela é apenas um resumo da análise de todos os critérios. A tabela detalhada se encontra em Anexo.

| | | | | | |
|----|---|---|----------|--|------|
| 5 | The Global Agricultural Trial Repository and Database | http://www.agtrials.org/ | Colômbia | | 50% |
| 6 | CIAT Dataverse | https://dataverse.harvard.edu/dataverse/CIAI | Colômbia | | 88% |
| 7 | WorldClim - Global Climate Data | http://worldclim.org/ | Brasil | | 25% |
| 8 | GLOBE | http://globe.umbc.edu/ | Brasil | | 63% |
| 9 | International Ocean Discovery Program | http://publications.iodp.org/ | Brasil | | 88% |
| 10 | PPBio Data Repository | https://ppbiodata.inpa.gov.br/metacatui/ | Brasil | | 100% |
| 11 | IBICT Dataverse Network | https://repositoriopesquisas.ibict.br | Brasil | | 63% |
| 12 | Base de Dados Científicos da Universidade Federal do Paraná | https://bdc.c3sl.ufpr.br/ | Brasil | | 63% |
| 13 | CEDAP Research Data Repository - research data | https://cedap.ufrgs.br/jspui/handle/2050011959/90 | Brasil | | 50% |

| | | | | | |
|----|---|---|------------|--|-----|
| 14 | Exploration and Production Data Bank | http://rodadas.anp.gov.br/en/concession-of-exploratory-blocks/brazil-round-3/bdep-round-3 | Brasil | | 13% |
| 15 | eFish Genomic Database Repository | http://www.molecularfisherieslaboratory.com.au/efish/ | Chile | | 75% |
| 16 | International Long Term Ecological Research | https://www.ilter.net/work/ | Costa Rica | | 63% |
| 17 | EcoCyc Database | https://ecocyc.org/ | México | | 25% |
| 18 | The Taenia solium Genome Project | http://www.taeniasoilum.unam.mx/taenia/ | México | | 0% |
| 19 | California Coastal Atlas | https://californiacoastatlas.net/ | México | | 38% |
| 20 | Banco de Información para la Investigación Aplicada en Ciencias Sociales, Repositorio Institucional | https://biacs-dspace.cide.edu/ | México | | 0% |
| 21 | ComBase | http://www.combase.cc/index.php/en/ | México | | 38% |
| 22 | Gran Telescopio CANARIAS Public Archive | http://gtc.sdc.cab.inta-csic.es/gtc/index.jsp | México | | 25% |

| | | | | | |
|----|--|---|--------|--|------|
| 23 | Mexican Health and Aging Study | http://www.mhasweb.org/ | México | O repositório solicita ficha de cadastramento com intenções de uso para acesso aos dados | 25% |
| 24 | Constrained Local UniversE Simulations | https://www.clues-project.org/cms/ | México | | 63% |
| 25 | Banff International Research Station for Mathematical Innovation and Discovery | https://open.library.ubc.ca/cIRcle/collections/birs | México | | 100% |
| 26 | Integrated Taxonomic Information System | https://www.itis.gov/ | México | A base de dados permite o seu próprio download | 63% |
| 27 | CIMMYT Research Data & Software Repository Network | https://data.cimmyt.org/dataverse/root | México | | 69% |
| 28 | Center for Tropical Forest Science and Forest Global Earth Observatory | http://www.forestgeo.si.edu/ | Panamá | Exige que se preencha uma formulário que informe a finalidade de acesso e uso. | 0% |
| 29 | Center for Tropical Forest Science - Panama | http://ctfs.si.edu/webatlas/ | Panamá | Exige que se preencha uma formulário que informe a finalidade de acesso e uso. | 0% |
| 30 | Repositorio Institucional USIL | http://repositorio.usil.edu.pe/handle/123456789/11 | Peru | No que se refere a alguns dados, ele remete a outros repositórios. | 50% |

| | | | | | |
|----|---|---|------|--|-----|
| 31 | Repositorio de datos del Ministerio de Educación del Perú | http://datos.minedu.gob.pe/ | Peru | | 25% |
|----|---|---|------|--|-----|

Fonte: Elaboração própria

Dentro do âmbito geográfico argentino nenhum dos repositórios cadastrados no Re3data poderiam ser caracterizados como repositórios de dados. Ou por ser apenas redirecionadores de links, ou por não armazenar dados.

Entre os repositórios colombianos o que mais se destacou dentro do movimento de Ciência Aberta foi o “CIAT Dataverse”, com uma contribuição de 88% para o desenvolvimento do movimento.

O Brasil teve um repositório que alcançou a marca de 100% de contribuição, o “PPBio Data Repository”, um repositório especializado na área de Ecologia que dissemina todos os tipos de dados de pesquisa e desenvolvido no software Metacat.

Dos repositórios mexicanos, dois foram caracterizados com 0% por estarem indisponíveis no momento de coleta dos dados. O mesmo índice de contribuição foi dado aos repositórios panamenhos por condicionar a disponibilização e visualização do seu conteúdo ao preenchimento de formulários.

6. CONCLUSÕES

Esta pesquisa partiu da problemática da pergunta: “Quais as características e contribuições dos repositórios de dados de pesquisa científicas da América Latina para o desenvolvimento da Ciência Aberta? ”. Com a finalidade de solucionar essa questão, foram traçados quatro objetivos específicos:

1. Identificar, com base na literatura, princípios e entendimentos básicos da Ciência Aberta e repositórios de dados.
2. Mapear os Repositórios de Dados Científicos da América Latina.
3. Caracterizar os Repositórios de dados de Científicos da América Latina de acordo com as dimensões apontadas pela literatura da área.
4. Identificar a organização e representação dos dados de pesquisa nos repositórios digitais de dados científicos.

Para que esses objetivos fossem alcançados, foi adotada uma abordagem metodológica do tipo mista que foi esquematizada brevemente na figura a seguir:

Figura 70: Resumo do método



Fonte: Elaboração própria

A abordagem seguiu uma estratégia exploratória sequencial onde houve a coleta dos dados quali e quantitativos para a elaboração de um instrumento de pesquisa.

A partir da abordagem qualitativa desenvolveu-se uma perspectiva teórica que serviu de base para a criação de um instrumento de coleta de dados quantitativos que foi aplicado aos repositórios de dados cadastrados no Re3data.org no âmbito geográfico latino americano. Por meio da aplicação do instrumento de dados sobre os principais pontos abordados na revisão de literatura foi possível colher e analisar uma ampla gama de dados. Essa coleta serviu como elemento solucionador do problema inicialmente proposto e alcance dos objetivos traçados.

Sendo assim, para cada variável adotada é possível concluir que:

Dentre os 20 países que compõe a América Latina, apenas o Brasil, Chile, Colômbia, México, Panamá e Peru de fato possuem repositórios de dados de pesquisa, o que representa um percentual de 30% dos países latino americanos se engajando no compartilhamento de dados em repositórios.

Esses países juntos, agruparam um total de 23 repositórios de dados,

onde apenas oito possuíam um perfil notadamente acadêmico, o que não significa que os demais não contribuem efetivamente para o desenvolvimento da Ciência Aberta na América Latina, pois como a literatura aponta, os dados possuem um caráter multiforme podendo ser usados para um propósito científico, mesmo que esses não tenham sido colhidos com essa finalidade.

O que realmente serviu como indicativo das contribuições oferecidas para o Movimento da Ciência Aberta foram coletados em um segundo e terceiro momentos, que são variáveis ligadas diretamente aos grandes pontos levantados na literatura sobre a Ciência Aberta e o Acesso Aberto: as variáveis que indicam:

- A adoção de uma licença aberta para os conteúdos disseminados;
- A presença e estimulação da colaboração científica na coleta, análise e na elaboração de produtos científicos;
- A acessibilidade ao conteúdo completo dos dados disseminados;
- A adoção de critérios de persistência que garantam a preservação e a acessibilidade aos dados;
- A adoção de softwares de código aberto e esquemas de metadados como ferramentas que possibilitem a maior circulação da informação e
- A utilização de estatísticas que informem o alcance que o compartilhamento de dados está tendo dentro da comunidade científica.

Em relação aos pontos elencados, foi possível observar que:

- Esperava-se que adoção de licenças que permitem a circulação dos dados estivesse presente dentro dos repositórios, no entanto ela foi adotada por apenas 8 dos 23 repositórios analisados, o que representa um percentual de apenas 34% de contribuição dos repositórios nesse requisito, um valor que indica uma contribuição abaixo do mínimo de 50% que poderia ser esperado.
- A colaboração científica foi observada em mais de um aspecto, sendo que além de disponibilizarem meios para que os usuários possam estabelecer contato com os seus pares, a pesquisa realizada

também indicou que o compartilhamento de dados está sendo feita de forma colaborativa entre os países da América Latina e outros países.

- O nível de acesso aos conteúdos se mostrou ser satisfatório como elemento de contribuição para a Ciência Aberta estando presente em 10 repositórios, entretendo, cabe ressaltar que existem repositórios que mesmo declarando disponibilizar os seus documentos, exigem um cadastro prévio ou o preenchimento de formulários em que sejam expostas razões para a solicitação de alguns dados.
- A adoção de critérios de persistência foi observada em 9 dos 23 repositórios analisados, o que representa um percentual de 39% dos repositórios analisados. Apesar de ainda ser uma média inferior ao desejado, isso significa que está assegurada a preservação persistente de 39% do conteúdo disseminado em repositórios na América Latina.
- Quanto a adoção de softwares específicos para a criação de repositórios mencionados na literatura, a pesquisa demonstrou que 14 dos 23 repositórios analisados foram construídos com ferramentas diferentes daquelas apontadas pela literatura, o que representa uma perda nos requisitos da Ciência Aberta já que, softwares que foram desenvolvidos propriamente para o compartilhamento de dados possuem funcionalidades que os que foram desenvolvidos fora desse padrão não poderão atuar.
- A adoção de estatísticas nesses repositórios também foi pouco observada, sendo um total de 12 repositórios que não fornecem estatísticas dos seus serviços.

Quanto ao compartilhamento de dados, em uma esfera geral, até mesmo incluindo os repositórios que se auto declararam como repositórios de dados, mas apenas direcionam a links externos, possuem uma tendência de compartilhar dados nas seguintes áreas do conhecimento.

Tabela 5: Distribuição geral das áreas do conhecimento

| Área do conhecimento | Quantidade |
|-------------------------|------------|
| Ecologia | 5 |
| Biologia | 4 |
| Multidisciplinar | 4 |
| Agricultura | 3 |
| Genética | 3 |
| Astronomia | 2 |
| Ciências da Saúde | 2 |
| Climatologia | 2 |
| Oceanografia | 2 |
| Ciências Naturais | 1 |
| Ciências Sociais | 1 |
| Educação | 1 |
| Geociências | 1 |
| Geofísica | 1 |
| Geologia | 1 |
| Matemática | 1 |
| Microbiologia alimentar | 1 |
| Não informado | 1 |
| Zoologia | 1 |

Fonte: Elaboração própria

Os resultados obtidos quanto ao compartilhamento de dados categorizados pela área do conhecimento demonstram uma tendência de compartilhamento mais forte dentro das Ciências Exatas e da Terra.

Ao se analisar os pontos básicos de contribuição para o Movimento da Ciência Aberta, é possível concluir que apesar das contribuições feitas para a Ciência Aberta dentro da América Latina, pontos-chaves ainda permanecem pouco explorados.

7. REFERÊNCIAS BIBLIOGRÁFICAS

ABBOTT, Mark R. A New Path for Science. In: HEY, Tony; TANSLEY, Stewart; TOLLE, Kristin. **The Fourth Paradigm: data- intensive Science discovery**. Washington: Microsoft Research, 2009, p. 111-116. Disponível em: <<https://www.immagic.com/eLibrary/ARCHIVES/EBOOKS/M091000H.pdf>>. Acesso em 15 out 2018.

ABDO, Alexandre Hannud. Ciência Aberta, da ciência para todos à ciência com todos. **Liinc em Revista**, v. 10, n. 2, 2014. Disponível em: <<http://dx.doi.org/10.18617/liinc.v10i2.747>>. Acesso em: 1 mar. 2017.

ALBAGLI, Sarita. Ciência aberta em questão. In: ALBAGLI, Sarita; MACIEL, Maria Lucia; ABDO, Alexandre Hannud (Org.). **Ciência aberta**, questões abertas. Brasília: Ibict; Rio de Janeiro: Unirio.

ALBERTO, Pepe Goodman, et al. How do astronomers share data? Plos One, v. 9, n. 8, p. mar. 2014. Disponível em: <<http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0104798>>. Acesso em jun. 2018.

ANDRADE, Viviane Toraci Alonso de. Comunicação científica e ciência aberta: produção e circulação de conhecimento em redes digitais. **Comunicação e Sociedade**, São Bernardo do Campo, v. 37, n. 1, p. 259-287, jan./abr. 2015. Disponível em: <<https://www.metodista.br/revistas/revistas-ims/index.php/CSO/article/view/4817/4623>>. Acesso em: 26 ago. 2017.

ANDRADE, Viviane Toraci Alonso de. Comunicação científica na sociedade em rede: uma plataforma de ciência aberta para o Brasil. 2014. 227 f. **Tese** (Doutorado em Comunicação)—Universidade Federal de Pernambuco, Recife, 2014. Disponível em: <<https://repositorio.ufpe.br/handle/123456789/13158>>. Acesso em: 22 set. 2018.

APPEL, Andre Luiz. A e-science e as atuais práticas de pesquisa científica. 2014. 88 f., il. **Dissertação** (Mestrado em Ciência da Informação)—Instituto Brasileiro de Informação em Ciência e Tecnologia e a Universidade Federal do Rio de Janeiro/Escola de Comunicação, 2014. Disponível em: <<http://ridi.ibict.br/handle/123456789/854>>. Acesso em: 18 maio. 2018.

ARANO, Silvia; et al. La comunidad 'Recursos y datos primarios' de la Universitat Pompeu Fabra: los repositorios institucionales como infraestructuras científicas: estudio de caso. **Revista española de documentación científica**, v. 34, n. 3, p. 385-407, 2011. Disponível em: <<http://dx.doi.org/10.3989/redc.2011.3.834>>. Acesso em: jun. 2018.

ASSANTE, Massimiliano. Science 2.0 Repositories. 2015. 133 f., il. **Tesi** (Dottorato di Ricerca in Ingegneria dell'informazione)— Università di Pisa, Pisa,

2015. Disponível em: <<https://etd.adm.unipi.it/t/etd-04272015-145002/>>. Acesso em: 10 set 2017.

ASSANTE, Massimiliano; CANDELA, Leonardo; CASTELLI, Donatella; MANGHI, Paolo; PAGANO, Pasquale. Science 2.0 Repositories: Time for a Change in Scholarly Communication. **D-Lib Magazine**, v. 21, n. 1/2, 12 p. 2015. Disponível em: <<http://www.dlib.org/dlib/january15/assante/01assante.html>>. Acesso em: 8 set 2017.

AUSTRALIAN NATIONAL DATA SERVICE. **ANDS Guide: Metadata**. Dez. 2016. Disponível em: <http://www.ands.org.au/__data/assets/pdf_file/0004/728041/Metadata-Workinglevel.pdf>. Acesso em: jun. 2018.

AVENTURIER, Pascal. Princípios FAIR: critérios de qualidade para dados de pesquisa. **A publicação científica**: Blog do Pascal Aventurier sobre as publicações científicas e os dados de pesquisa, mar. 2017. Disponível em: <<https://publicient.hypotheses.org/1456>>. Acesso em: fev. 2019.

BABBIE, Earl. **Survey Research Methods**. 2.th ed. Belmont: Wadsworth, 1990. xx, 395 p.

BARBALHO, Célia Regina Simonetti. Periódicos científicos em formato eletrônico: elementos para sua avaliação. In: CONGRESSO BRASILEIRO DE CIÊNCIAS DA COMUNICAÇÃO, 28., 2005, Rio de Janeiro. **Anais...** Rio de Janeiro: Intercom: Sociedade Brasileira de Estudos Interdisciplinares da Comunicação, 2005. Disponível em: <<http://www.portcom.intercom.org.br/pdfs/144299010318091916156725937486754826190.pdf>>. Acesso em mar. 2018.

BAPTISTA, Ana Alice et al. Comunicação científica: o papel da Open Archives Initiative no contexto do Acesso Livre. **Encontros Bibli**, Florianópolis, n. esp., 1º sem. 2007. 17 p. Disponível em: <<https://doi.org/10.5007/1518-2924.2007v12nesp1p1>>. Acesso em: jun. 2018.

BARTLING, Sönke; FRIESIKE, Sascha. Towards Another Scientific Revolution. In: BARTLING, Sönke; FRIESIKE, Sascha. **Opening Science: The Evolving Guide on How the Internet is Changing Research, Collaboration and Scholarly Publishing**. Alemanha: Springer, 2014, p. 3-15. Disponível em: <<https://link.springer.com/content/pdf/10.1007%2F978-3-319-00026-8.pdf>>. Acesso em 16 out 2018.

BARROS, Moreno Albuquerque de. A Primavera Acadêmica e o custo do conhecimento. **Liinc em Revista**, v.8, n.2, set, 2012, Rio de Janeiro, p. 365-377. Disponível em: <<http://www.brapci.inf.br/index.php/article/download/41123>>. Acesso em jun. 2018.

BIBLIOTECA DA UNIVERSIDADE DE QUEEENSLAND. Biblioteca da Universidade de Queensland. 2018. Disponível em: <<https://www.library.uq.edu.au/>>. Acesso em: jun. 2018.

BIBLIOTECA DE LA UNIVERSIDAD DE SEVILLA. **El ciclo de los datos científicos**. 2018. Disponível em:< <http://guiasbus.us.es/accesoabierto/datos> >. Acesso em: jun. 2018.

BONETTA, Laura. Scientists Enter the Blogosphere. **Cell**, Washington, v.129, n. 3, jun. p. 443-445, 2017. Disponível em: <<https://doi.org/10.1016/j.cell.2007.04.032>>. Acesso em jan. 2019.

BORGMAN, Christine L. The Conundrum of Sharing Research Data. **Journal of the American Society for Information Science and Technology**, n.8, p. 1-41, 2012. Disponível em: <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1869155>. Acesso em: 11 mar. 2017.

BOULTON, Geoffrey. Reinventing Open Science for the 21st Century. In: **Uma década de acesso aberto na UMinho e no Mundo**. Lisboa: Universidade do Minho, 2013. p.239–250. Disponível em: <<http://hdl.handle.net/1822/26144>>. Acesso em 12 jan. 2019.

BOULTON, Geoffrey et al. **Science as an open enterprise**. Londres: The Royal Society, 2012. Disponível em: <https://royalsociety.org/~media/Royal_Society_Content/policy/projects/sape/2012-06-20-SAOE.pdf>. Acesso em jun. 2018.

BOURDIEU, Pierre. **Questões de sociologia**. Rio de Janeiro: Marco Zero, 1983. 208 p.

BUCKLAND, M. Information as thing. **JASIS**, n. 42, 351–36,1991. Disponível em: <<http://people.ischool.berkeley.edu/~buckland/thing.html>>. Acesso em: Acesso em: 28 jun. 2017.

CANESSA, Enrique; ZENNARO, Marco. **Difusión científica y las iniciativas de Acceso Abierto**: Recopilación de publicaciones seleccionadas sobre el Acceso Abierto al conocimiento. CeCalCULA – Centro de Cálculo Científico de la Universidad de Los Andes: Mérida, 2008. 204. Disponível em: <<http://eprints.rclis.org/13661/1/AcessoAbiertoConocimientoP.pdf>>. Acesso em: jan. 2019.

CARDOSO, Gustavo; JACOBETTY, Pedro; DUARTE, Alexandra. **Para Uma Ciência Aberta**. Lisboa: Mundos Sociais, 2012. 126 p. Disponível em: < https://www.researchgate.net/publication/235994691_Para_Uma_Ciencia_Aberta >. Acesso em: dez. 2018.

CASTELLS, Manuel. **A sociedade em rede**. 6. ed. São Paulo, SP: Paz e Terra, 2010. 698 p.

CHAN, Leslie; OKUNE, Angela; SAMBULI, Nanjira. O que é ciência aberta e colaborativa, e que papéis ela poderia desempenhar no desenvolvimento? In: ALBAGLI, Sarita; MACIEL, Maria Lucia; ABDO, Alexandre Hannud (Org.). **Ciência aberta, questões abertas**. Brasília: Ibict; Rio de Janeiro: Unirio. 306 p. 2015. Disponível em: <<http://livroaberto.ibict.br/handle/1/1060>>. Acesso em: 30 ago. 2017.

CONVENÇÃO DE SANTA FE PARA INICIATIVA DE ACESSO ABERTO. Santa Fé, 2000. Disponível em: <http://www.openarchives.org/sfc/sfc_entry.htm>. Acesso em jun 2018.

COSTA, Maíra Murrieta. Diretrizes para uma política de gestão de dados científicos no Brasil. 2017. 288 f., il. **Tese** (Doutorado em Ciência da Informação)—Universidade de Brasília, Brasília, 2017. Disponível em: <<http://repositorio.unb.br/handle/10482/24895>>. Acesso em: 01 abr. 2018.

COSTA, Michelli Pereira da. Características e contribuições da via verde para o acesso aberto à informação científica na América Latina. 2014. 226 f., il. **Dissertação** (Mestrado em Ciência da Informação)—Universidade de Brasília, Brasília, 2014.

COSTA, Michelli Pereira da. Fatores que influenciam a comunicação de dados de pesquisa sobre o vírus da zika, na perspectiva de pesquisadores. 2017. 269 f., il. **Tese** (Doutorado em Ciência da Informação)—Universidade de Brasília, Brasília, 2017.

COSTA, Michelli; BRAGA, Tiago. Repositórios de dados de pesquisa no mundo. **Cadernos BAD**, n. 2, jul-dez, p. 80-95. 2016. Disponível em: <<https://www.bad.pt/publicacoes/index.php/cadernos/article/view/1585>>. Acesso em: jun. 2018.

COSTA, Michelli Pereira da; LEITE, Fernando César Lima. Acesso Aberto no mundo e na América Latina: uma revisão a partir da BOAI. **Transinformação**, 2016, v. 28, n.1, pp. 33-46. Disponível em: <http://repositorio.unb.br/bitstream/10482/19433/2/AcessoAbertoMundo_%202016.pdf>. Acesso em: jun. 2018.

COSTA, Michelli Pereira da; LEITE, Fernando César Lima. **Repositórios institucionais da América Latina e o acesso aberto à informação científica**. Brasília: IBICT, 2017. 178 p.

COSTA, Sely M.S. Mudanças no processo de comunicação científica: o impacto do uso de novas tecnologias. In: MUELLER, Suzana Pinheiro Machado; PASSOS, Edilenice (Org.). **Comunicação científica**. Brasília: Editora Universidade de Brasília, 2000. 144 p. (Estudos avançados em Ciência da Informação).

COVI, Lisa M. Debunking the Myth of the Nintendo Generation: How Doctoral Students Introduce New Electronic Communication Practices into University Research. **Journal Of The American Society For Information Science**, New

Jersey, v. 51, n.14, 2000, p. 1284–1294. Disponível em: <[https://doi.org/10.1002/1097-4571\(2000\)9999:9999::AID-ASI1045>3.0.CO;2-Z](https://doi.org/10.1002/1097-4571(2000)9999:9999::AID-ASI1045>3.0.CO;2-Z)>. Acesso em 29 mar. 2018.

CRESWELL, John W. **Projeto de pesquisa: métodos qualitativo, quantitativo e misto**. 3. ed. Porto Alegre, RS: Artmed, 2010. 296 p.

DAMVAD. Sharing and archiving of publicly funded research data: Report to the Research Council of Norway. 2014, 74 p. Disponível em:<<https://www.innovationpolicyplatform.org/system/files/Research-Council-Norway-Survey-DAMVAD%284%29.pdf>>. Acesso em jun. 2018.

DATAVERSE. 2018. Disponível em: <<https://dataverse.org/>>. Acesso em: jun. 2018.

DECLARAÇÃO DE BERLIM SOBRE ACESSO AO CONHECIMENTO NAS CIÊNCIAS E HUMANIDADES, Berlim, 2003. Disponível em: <<https://repositorium.sdum.uminho.pt/about/DeclaracaoBerlim.htm>>. Acesso em: jun. 2018.

DECLARAÇÃO DE BETHESDA. 2003. Disponível em: <<http://legacy.earlham.edu/~peters/fos/bethesda.htm>>. Acesso em: jun. 2018.

DECLARAÇÃO DE BUDAPESTE DE ACESSO ABERTO. 2002. Disponível em: <<http://www.budapestopenaccessinitiative.org/read>>. Acesso em: jun. 2018.

DECLARAÇÃO DE SANTO DOMINGO. A CIÊNCIA PARA O SÉCULO XXI: uma visão nova e uma base de ação Conferência Mundial sobre Ciência, Santo Domingo, 10-12 mar. 1999. Disponível em: <<http://www.dominiopublico.gov.br/download/texto/ue000112.pdf>>. Acesso em: jun. 2018.

EASON, Ken et al. **A Comparative Analysis of the Role of MultiMedia Electronic Journals in Scholarly Disciplines**: final report. Loughborough: HUSAT Reseadrch Institute; Departamento f Human Sciences. Loughborough University, 1997. 81 p. Disponível em: <<http://opus.bath.ac.uk/35421/1/eason.pdf>>. Acesso em 28 dez. 2017.

EUROPEAN UNIVERSITY ASSOCIATION. Towards Open Access to Research Data: Aims and recommendations for university leaders and National Rectors' Conferences on Research Data Management and Text and Data Mining. Disponível em: <<http://eua.be/Libraries/publications-homepage-list/towards-open-access-to-research-data>>. Acesso em: 01 abr. 2018.

EYNDEN, Veerle Van den et al. **Managing and Sharing Data**. Disponível em: <<http://www.data-archive.ac.uk/media/2894/managingsharing.pdf>>. Acesso em jun. 2018.

FACILITATE OPEN SCIENCE TRAINING FOR EUROPEAN RESEARCH. **Open Science Definition**. Disponível em:

< <https://www.fosteropenscience.eu/taxonomy/term/100>> . Acesso em: 23 mar. 2019.

FAUSTO, Sibeles. Evolução do Acesso Aberto: breve histórico. **SciELO em Perspectiva**. Disponível em: < <https://blog.scielo.org/blog/2013/10/21/evolucao-do-acesso-aberto-breve-historico/#.Wxrpwfkvz4Z>>. Acesso em jun. 2018.

FECHER, Benedikt; FRIESIKE, Sascha. Open Science: One Term, Five Schools of Thought. In: BARTLING, Sönke; FRIESIKE, Sascha. **Opening Science: The Evolving Guide on How the Internet is Changing Research, Collaboration and Scholarly Publishing**. Alemanha: Springer, 2014, p. 3-15. Disponível em: <<https://link.springer.com/content/pdf/10.1007%2F978-3-319-00026-8.pdf>> . Acesso em 16 out 2018.

FISCHMAN, Josh. As Journal Boycott Grows, Elsevier Defends Its Practices. **The Chronicle of Higher Education**. Jan. 2012. Disponível em: <<https://www.chronicle.com/article/As-Journal-Boycott-Grows/130600/>>. Acesso em: fev. 2019.

FORCE11. Guiding Principles for Findable, Accessible, Interoperable and Reusable Data Publishing version b1.0. Disponível em: <<https://www.force11.org/fairprinciples>>. Acesso em fev. 2019.

FUNDAÇÃO OSWALDO CRUZ. Formação modular sobre Ciência Aberta. O que é Ciência Aberta. **CURSO 1**. Disponível em: < <https://cursos.campusvirtual.fiocruz.br/course/view.php?id=93>>. Acesso em: mar. 2019.

FRANK, Rebecca D.; YAKEL, Elizabeth; FANIEL, Ixchel M. Destruction /reconstruction: preservation of archaeological and zoological research data. **Archival science**, v.15, n.2, p 141-167. Disponível em: <<https://link.springer.com/content/pdf/10.1007%2Fs10502-014-9238-9.pdf>>. Acesso em: 18 jan. 2018.

FREITAS, Marília Augusta de. Diretrizes para o depósito da produção científica em repositórios institucionais. 2015. xv, 199 f., il. **Tese** (Doutorado em Ciência da Informação)—Universidade de Brasília, Brasília, 2015. Disponível em: <<http://repositorio.unb.br/handle/10482/19189>> . Acesso em: fev. 2018.

GARCIA, Maria Manuela Alves. O campo das produções simbólicas e o campo científico em Bourdieu. **Cadernos de Pesquisa**, São Paulo, n. 97, p. 64-72, maio 1996. Disponível em: <<http://publicacoes.fcc.org.br/ojs/index.php/cp/article/view/804/815>>. Acesso em: 20 mar 2018.

GEZELTER, D. **What, exactly, is Open Science?** The Open Science Project, jul. 2009. Disponível em:<<http://openscience.org/what-exactly-is-openscience/>>. Acesso em:12 jan. 2019.

GIL, Antonio Carlos. **Métodos e técnicas de pesquisa social**. 6. ed. São Paulo: Atlas, 2012. 200 p.

GÓMEZ, Nancy-diana; MÉNDEZ, Eva; HERNÁNDEZ-PÉREZ, Tony. Datos e Metadatos de investigación em Ciencias sociales e Humanidades: uma aproximación desde lós repositórios temáticos de datos. <<https://e-archivo.uc3m.es/handle/10016/24939#preview>>.

GONZALEZ, Andres Guadamuz. Open Science: Open Source Licenses in Scientific Research. North Carolina Journal Of Law & Technology, v. 7, n. 2, p. 321-366, 2006. Disponível em: <<https://scholarship.law.unc.edu/cgi/viewcontent.cgi?article=1076&context=ncjo>>. Acesso em: jan. 2019.

GRÁCIO, José Carlos Abbud. Metadatos para a descrição de recursos da Internet: o padrão Dublin Core, aplicações e a questão da interoperabilidade. 2002. 127 f., il. **Dissertação** (Mestrado em Ciência da Informação)— Universidade Estadual Paulista “Julio de Mesquita Filho” -UNESP, Marília, 2002. Disponível em: < https://www.marilia.unesp.br/Home/Pos-Graduacao/CienciadaInformacao/Dissertacoes/gracio_jca_dr_mar.pdf >. Acesso em jun. 2018.

GREEN, Ann; Macdonald, Stuart; RICE, Robin. **Policy-making for Research Data in Repositories: a Guide**. Data Information Specialists Committee –UK: Reino Unido, 29 p. 2009. Disponível em: <<https://www.coar-repositories.org/files/guide.pdf>>. Acesso em: jun. 2018.

GREENE, Lewis Joel. Mais visibilidade para a ciência brasileira: Biblioteca eletrônica já reúne mais de 50 revistas científicas. **Pesquisa FAPESP**, São Paulo, n. 57, 1 p., set. 2000. Disponível em: <http://revistapesquisa.fapesp.br/wp-content/uploads/2000/09/07_opini%C3%A3o.pdf >. Acesso em: 20 mar 2018.

HARNAD, Stevan, et al. The Access/Impact Problem and the Green and Gold Roads to Open Access. **Serials Review**. v.30, n. 4. 2004. Disponível em: <http://www.chss.uqam.ca/Portals/0/docs/articles/serial_review.pdf>. Acesso em: 28 ago. 2017.

HERNÁNDEZ-PÉREZ, Tony; GARCÍA-MORENO, María-Antonia. Datos abiertos y repositorios de datos: nuevo reto para los bibliotecários. **El profesional de la información**, v. 22, n. 3, maio/jun. 2013. Disponível em: <https://eprints.ucm.es/22025/1/datos_abiertos_EPI2013.pdf >. Acesso em 27 mar. 2018.

HERNÁNDEZ SAMPIERI, Roberto; FERNÁNDEZ COLLADO, Carlos; BAPTISTA LUCIO, Maria del Pilar. **Metodologia de pesquisa**. 3. ed. São Paulo: McGraw Hill, 2006. xxiv, 583 p.

HEY, Tony; HEY, Jessie. E-Science and its implications for the library community. **Library Hi Tech**, v. 24, n. 4, pp.515-528, 2006. Disponível em <<https://doi.org/10.1108/07378830610715383>>. Acesso em 9 maio.2018.

HEY, Tony; TREFETHEN, Anne E. Cyberinfrastructure for e-Science. **Science**, v. 308, n. 5723, pp.817-821, maio 2005. Disponível em <<http://science.sciencemag.org/content/308/5723/817/tab-pdf>>. Acesso em fev. 2019.

HJØRLAND, B. Domain analysis in information science: eleven approachestransitional as well as innovative. **Journal of Documentation**, v. 58, n. 4, p. 422-462, 2002.

HJØRLAND, Birger; ALBRECHTSEN, Hanne. Toward a new horizon in information science: domain-analysis. **Journal Of The American Society For Information Science**, 46, n. 6, p. 400-425, 1995. Disponível em: <[https://doi.org/10.1002/\(SICI\)1097-4571\(199507\)46:6<400::AID-ASI2>3.0.CO;2-Y](https://doi.org/10.1002/(SICI)1097-4571(199507)46:6<400::AID-ASI2>3.0.CO;2-Y)>. Acesso em: 30 jan. 2018.

HURD, Julie M.The Transformation of Scientific Communication: a model for 2020. **Journal Of The American Society For Information Science**, v. 51, n. 14, p.1279–1283, 2000. Disponível em: <[http://onlinelibrary.wiley.com/doi/10.1002/1097-4571\(2000\)9999:9999%3C::AID-ASI1044%3E3.0.CO;2-1/epdf](http://onlinelibrary.wiley.com/doi/10.1002/1097-4571(2000)9999:9999%3C::AID-ASI1044%3E3.0.CO;2-1/epdf)>. Acesso em: 11 set 2017.

INSTITUTION OF ENVIRONMENTAL SCIENCES. An Introduction To Open Access Data Sharing, **Institution of Environmental Sciences: IES**. 2013, 4 p. Disponível em: <https://www.the-ies.org/sites/default/files/documents/data_sharing_final.pdf>. Acesso em: 30 jan. 2019.

JOINT INFORMATION SYSTEMS COMMITTEE. How JISC is helping researchers. 2010. Disponível em: <<http://www.jisc.ac.uk/whatwedo/campaigns/res3/jischelp.aspx>>. Acesso em: jun. 2018.

KAUARK, Fabiana da Silva; MANHÃES, Fernanda Castro; MEDEIROS, Carlos Henrique. **Metodologia da pesquisa**: um guia prático. Itabuna: Via Litterarum, 2010.

KERLINGER, Frederick Nichols. **Metodologia da pesquisa em ciências sociais**: um tratamento conceitual. São Paulo, SP: Editora Pedagógica e Universitária, 1979. 378 p.

KIM, Youngseek; STANTON, Jeffrey M. Institutional and individual influences on scientists' data sharing behaviors: a multilevel analysis. In: **Annual Meeting of The Association For Information Science and Technology**: ASIST, 13., 2013, Montreal, Quebec. Proceedings... Maryland: American Society for Information Science Silver Springs, 2013. 14 p. Disponível em:

<<http://onlinelibrary.wiley.com/doi/10.1002/meet14505001093/epdf>>. Acesso em: 10 junho 2017.

KLUMP, Jens, et al. Data Publication In The Open Access Initiative. **Data Science Journal**, v. 5, n. 15, p. 79–83, jun. 2006. Disponível em: <<https://datascience.codata.org/articles/abstract/10.2481/dsj.5.79/>>. Acesso em: jan. 2019.

KURAMOTO, Hélio. Informação científica: proposta de um novo modelo para o Brasil. **Ciência da Informação**, Brasília, v. 35, n. 2, p. 91-102, maio/ago. 2006. P. 91-102. Disponível em: <<http://www.scielo.br/pdf/%0D/ci/v35n2/a10v35n2.pdf>>. Acesso em:

KURTZ, M. J., et al. The effect of use and access on citations. **Information Processing and Management**, v. 41, p. 1395–1402, 2005. Disponível em: <<http://arxiv.org/abs/cs.DL/0503029>>. Acesso em: jun. 2018.

LE COADIC, Yves-François. **A ciência da informação**. 2. ed. Brasília : Briquet de Lemos/Livros, 2004. 124 p.

LEITE, Fernando César Lima. **Como gerenciar e ampliar a visibilidade da informação científica brasileira**: repositórios institucionais de acesso aberto. Brasília: IBICT, 2009. 124 p. Disponível em: <<http://repositorio.unb.br/handle/10482/4841>>. Acesso em: 27 ago. 2017.

LI, Yuan; BANACH, Meghan. Institutional Repositories and Digital Preservation: Assessing Current Practices at Research Libraries. **D-Lib Magazine**, v. 17, n. 1/2, Jan. / Feb. 2011. Disponível em: <<http://www.dlib.org/dlib/may11/yuanli/05yuanli.html> >. Acesso em: jun. 2018.

LÓPEZ YEPES, J. La perspectiva informativa de la Documentación: la Documentación como ciencia de la información documental: el concepto de Information Science. In:_____. **La documentación como disciplina: teoría e história**. 2. ed. actual. y ampl. Pamplona: EUNSA, 1995. p. 153-196.

LUCE, Richard E. E-prints Intersect the Digital Library: Inside the Los Alamos arXiv. **International Spring School on the Digital Library and E-publishing for Science and Technology**, CERN, Geneva, Switzerland, 3 – 8 March 2002. Disponível em: <<https://indico.cern.ch/event/410468/sessions/165590/attachments/831393/1149886/12Luce.pdf>>

MARTINEZ, Luis, et al. The Data Documentation Initiative (DDI) and Institutional Repositories. 21 p. 2009. Disponível em: < http://www.disc-uk.org/docs/DDI_and_IRs.pdf >. Acesso em: jun. 2018.

MAURER, Stephen. New Institutions for Doing Science: From Databases to Open Source Biology. In: EUROPEAN POLICY FOR INTELLECTUAL PROPERTY CONFERENCE ON COPYRIGHT AND DATABASE

PROTECTION, PATENTS AND RESEARCH TOOLS, AND OTHER CHALLENGES TO THE INTELLECTUAL PROPERTY SYSTEM, 2003, The Netherlands, 2013. **Proceedings...** 2013, 16 p. Disponível em: <<https://www.researchgate.net/publication/237412391>>. Acesso em: fev. 2019.

MCKINSEY GLOBAL INSTITUTE (MGI). **Open data**: Unlocking innovation and performance with liquid information. McKinsey & Company: 2013. Disponível em: <https://www.mckinsey.com/~/_/media/McKinsey/Business%20Functions/McKinsey%20Digital/Our%20Insights/Open%20data%20Unlocking%20innovation%20and%20performance%20with%20liquid%20information/MGI_Open_data_FullReport_Oct2013.ashx>. Acesso em fev. 2019.

MEADOWS, Alice. "To share or not to share? That is the (research data) question...". The scholarly kitchen, 2014. Disponível em: <<http://scholarlykitchen.sspnet.org/2014/11/11/to-share-or-not-to-share-that-is-the-research-data-question>>. Acesso em: jun. 2018.

MELTON, R.K. Os imperativos institucionais da ciência. In: DEUS, Jorge Dias de; MERTON, Robert King. **A crítica da ciência: sociologia e ideologia da ciência**. Rio de Janeiro: Zahar, 1974. 240 p.

MEADOES, A.J. Avaliando o desenvolvimento da comunicação eletrônica. In: MUELLER, Suzana Pinheiro Machado; PASSOS, Edilenice (Org.). **Comunicação científica**. Brasília : Editora Universidade de Brasília, 2000. 144 p. (Estudos avançados em Ciência da Informação ; 1).

MEADOWS, A. J. A **comunicação científica**. Brasília : Briquet De Lemos, 1999. 268 p.

MICHENER, William K. Ten simple rules for creating a good data management plan. **PLoS: Computational Biology**, v. 11, n. 10, e1004525. Disponível em: <<http://journals.plos.org/ploscompbiol/article/file?id=10.1371/journal.pcbi.1004525&type=printable>>. Acesso em: 20 out. 2017.

MIROWSKI, Philip. The future(s) of open science. **Social Studies of Science**, v. 48, n. 2, p. 171-203, 2018. Disponível em: <<https://doi.org/10.1177/0306312718772086>>. Acesso em: jan. 2019.

MOLLOY, Jennifer C. The open knowledge foundation: open data means better science. **PLoS Biology**, v. 9, n. 12, e1001195, 4 p., 2011. Disponível em: <[doi:10.1371/journal.pbio.1001195](https://doi.org/10.1371/journal.pbio.1001195)>. Acesso em 25 jan. 2019.

MORRISON, Heather. **Scholarly Communication for Librarians**. Chandos Publishing, 2009 , 264 p. Disponível em: < <http://eprints.rclis.org/13177/> > Cap 6.

MUELLER, Suzana Pinheiro Machado. A publicação da ciência: áreas científicas e seus canais preferenciais. **DataGramZero**: Revista de Ciência da Informação, v. 6, n. 1, fev. 2005.

MUELLER, Suzana Pinheiro Machado. A ciência, o sistema de comunicação científica e a literatura científica. In: CAMPELLO, Bernadete Santos; CENDÓN, Beatriz Valadares; KREMER, Jeannette M. **Fontes de informação para pesquisadores e profissionais**. Belo Horizonte: Editora UFMG, 2007.

MURRAY-RUST, Peter. Open data in science. **Serials Review**, v. 34, n. 1, p. 52-64. Mar. 2008. Disponível em: <<http://precedings.nature.com/documents/1526/version/1>>. Acesso em: 19 out. 2017.

MURRAY-RUST, Peter, et al. Panton Principles, Principles for open data in science. Fev. 2010. Disponível em: <<https://pantonprinciples.org/>>. Acesso em dez. 2018.

OLIJHOEK, Tom; MITCHELL, Dominic; BJØRNSHAUGE, Lars. Criteria for open access and publishing. **ScienceOpen Research**. 2005. Disponível em: <10.14293/S2199-1006.1.SOR-EDU.AMHUHV.v1 >. Acesso em: 18 jun. 2018.

ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT. OECD Principles and Guidelines for Access to Research Data from Public Funding. França: OECD, 2007, 22 p. Disponível em: <<https://www.oecd.org/sti/sci-tech/38500813.pdf>>. Acesso em: jun. 2018.

ORTELLADO, Pablo. As políticas nacionais de acesso à informação científica. **Liinc em Revista**, v.4, n.2, set 2008, Rio de Janeiro, p. 186-195. Disponível em: <<http://revista.ibict.br/liinc/article/view/3164/2830> >.

PACKER, Abel L.; MENEGHINI, Rogério. O SciELO aos 15 anos: raison d'être, avanços e desafios para o futuro. In: PACKER, Abel L. et al. (Orgs.). **SciELO - 15 Anos de Acesso Aberto**: Um estudo analítico sobre Acesso Aberto e comunicação científica. Paris: UNESCO, 2014, 188 p. Disponível em: <<http://www.scielo.org/local/File/livro.pdf>>. Acesso em: .jun. 2018.

PAMPEL, Heinz, et al. Making research data repositories visible: the re3data.org registry. **PLoS ONE**, v. 8, n. 11, 2013. Disponível em: <<http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0078080>>. Acesso em: 27 jun. 2017. DOI: <https://doi.org/10.1371/journal.pone.0078080>.

PINHEIRO, Lena Vania Ribeiro. Do acesso livre à ciência aberta: conceitos e implicações na comunicação científica. **Revista Eletrônica de Comunicação, Informação & Inovação em Saúde**: RECIIS, v. 8, n. 2, jun. 2014 jun, p.153-165. Disponível em: <<https://www.reciis.icict.fiocruz.br/index.php/reciis/article/view/629/1269>>. Acesso em: 30 ago. 2017.

PONTIKA, Nancy, et. al. Fostering Open Science to Research using a Taxonomy and an eLearning Portal. In: IKNOW: 15th International Conference on Knowledge Technologies and Data Driven Business, Graz, Austria. Out.

2015. **Proceedings...** 2015, 9 p. Disponível em: <<http://dx.doi.org/doi:10.1145/2809563.2809571>>. Acesso em: 23 mar. 2019.

PRIEM, Jason, et al. Uncovering impacts: CitedIn and total-impact, two new tools for gathering altmetrics. In: WORKSHOP ON SEMANTIC PUBLISHING; 2012, Bremen. 2012, Bremen. **Proceedings...** 2012, p. 40-44. Disponível em: <https://www.researchgate.net/publication/260097152_Uncovering_impacts_CitedIn_and_total-impact_two_new_tools_for_gathering_altmetrics>. Acesso em: 25 jan. 2019.

PUSCHMANN, Cornelius. (Micro)Blogging Science?: Notes on Potentials and Constraints of New Forms of Scholarly Communication. In: BARTLING, Sönke; FRIESIKE, Sascha (Ed.). **Opening Science: The Evolving Guide on How the Internet is Changing Research, Collaboration and Scholarly Publishing**. Mannheim: Springer, 2014. p. 89-1016. Disponível em: <10.1007/978-3-319-00026-8_6>. Acesso em: dez. 2018.

RABELO, André. A primavera acadêmica: O livre acesso ao conhecimento científico. **SocialMente**, ScienceBlogs. 2012. Disponível em: <<http://scienceblogs.com.br/socialmente/2012/02/a-primavera-academica-o-livre-acesso-ao-conhecimento-cientifico/>>. Acesso em: fev. 2019.

REGISTRY OF RESEARCH DATA REPOSITORIES. Registry of Research Data Repositories: Re3data. Disponível em: <<https://www.re3data.org/>>. Acesso em: jun. 2018.

RODRIGUES, Eloy; SARAIVA, Ricardo. Os Repositórios de Dados Científicos: Estado da Arte. RCAAP, 2010. 54 p. Disponível em: <<https://repositorio-aberto.up.pt/bitstream/10216/23806/2/44632.pdf>>. Acesso em: jun. 2018.

ROSA, Flávia; GOMES, Maria João. Comunicação Científica: das restrições ao acesso livre. In: GOMES, Maria João; ROSA, Flávia Goulart Mota Garcia. **Repositórios institucionais: democratizando o acesso ao conhecimento**. Salvador: EDUFBA, 2010, p. 11-32. Disponível em: <<https://repositorio.ufba.br/ri/bitstream/ri/616/3/Repositorios%20institucionais.pdf>>. Acesso em: 21 ago. 2017.

RUSSELL, Jane M. Scientific Communication at the beginning of the twenty-first century. Blackwell Publishers Ltd., n. 28, p. 271-282, jun. 2001. Disponível em: <<http://onlinelibrary.wiley.com/doi/10.1111/1468-2451.00314/pdf>>. Acesso em: 09 maio 2018.

SALES, Luana Farias. Integração semântica de publicações científicas e dados de pesquisa: proposta de modelo de publicação ampliada para a área de ciências nucleares. 2014. **Tese** (Doutorado em Ciência da Informação)—Escola de Comunicação, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2014.

SANTOS, Boaventura de Sousa. **Um discurso sobre as ciências**. 7. ed. São Paulo: Cortez, 2010. 92 p.

SARMENTO, Fernanda, et al. Algumas considerações sobre as principais declarações que suportam o movimento Acesso Livre. In: WORLD CONGRESS ON HEALTH INFORMATION AND LIBRARIES, 2005, Salvador. **Anais...** Salvador, 2005. Disponível em: < <http://eprints.rclis.org/8512/> >. Acesso em mar. 2018.

SAVAGE, Caroline J.; VICKERS, Andrew J. Empirical Study of Data Sharing by Authors Publishing in PLoS Journals. **PLoS ONE**, vol. 4, n. 9, 2009. Disponível em: <<http://journals.plos.org/plosone/article/file?id=10.1371/journal.pone.0007078&type=printable>>. Acesso em: 11 mar. 2017.

SAYÃO, Luís Fernando; SALES, Luana Farias. Algumas considerações sobre os repositórios digitais de dados de pesquisa. **Revista Informação & Informação**, v. 21, n. 2, p. 90-115, 2016. Disponível em: <<http://dx.doi.org/10.5433/1981-8920.2016v21n2p90>>. Acesso em: 23 mar. 2019.

SAYÃO, Luís Fernando; SALES, Luana Farias. Dados abertos de pesquisa: ampliando o conceito de acesso livre. **Revista Eletrônica de Comunicação, Informação & Inovação em Saúde: Reciiis**, v. 8, n. 2, p. 76-92, 2014. Disponível em: <<https://www.reciis.icict.fiocruz.br/index.php/reciis/article/view/611/1252>>. Acesso em: 19 out. 2017.

SAYÃO, Luís Fernando; SALES, Luana Farias. Curadoria digital: um novo patamar para preservação de dados digitais de pesquisa. **Informação & Sociedade: Estudos**, v. 22, n. 3, p. 179-191, 2012. Disponível em: <<http://www.ies.ufpb.br/ojs/index.php/ies/article/view/12224/8586>>. Acesso em: 19 out. 2017.

SAYÃO, Luís Fernando; SALES, Luana Farias. **Guia de gestão de dados de pesquisa para bibliotecários e pesquisadores**. Rio de Janeiro: CNEN, 2015. 90 p.

SCIENCE INTERNATIONAL. **Open Data in a Big Data World**. Paris: International Council for Science (ICSU); International Social Science Council (ISSC); The World Academy of Sciences (TWAS); InterAcademy Partnership (IAP). 2015. 17 p. Disponível em: <https://council.science/cms/2017/04/open-data-in-big-data-world_long.pdf>. Acesso em fev. 2019.

SITEK, Dagmar; BERTELMANN, Roland. Open Access: A State of the Art. In: BARTLING, Sönke; FRIESIKE, Sascha. **Opening Science: The Evolving Guide on How the Internet is Changing Research, Collaboration and Scholarly Publishing**. Alemanha: Springer, 2014, p. 3-15. Disponível em: <<https://link.springer.com/content/pdf/10.1007%2F978-3-319-00026-8.pdf> >. Acesso em 16 out 2018.

SOUZA, Rosali Fernandez de. Áreas do Conhecimento, **Revista DataGramZero**: Revista de Ciência da Informação, v.5, n.2, abr. 2004. Disponível em: <<http://repositorio.ibict.br/bitstream/123456789/109/1/RosaliDatagramazero2004.pdf>>. Acesso em: 27 mar. 2018.

SOUZA, Rosali Fernandez de. Organização e representação de áreas do conhecimento em ciência e tecnologia: princípios de agregação em grandes áreas segundo diferentes contextos de produção e uso de informação. **Encontros Bibli**: Revista Eletrônica de Biblioteconomia e Ciência da Informação, Florianópolis, n. esp., 1º sem. 2006, p. 27-41. Disponível em: <<http://www.redalyc.org/html/147/14720365005/>>. Acesso em: 29 mar. 2018.

SOUZA, Sebastião de. **CDU**: como entender e utilizar a 2ª edição-padrão internacional em língua portuguesa. Brasília: Thesaurus, 2009. 163 p.

SPINAK, Ernesto. Open-Data: liquid information, democracy, innovation... the times they are a-changin'. **SciELO in Perspective**, 2013. Disponível em: <<https://blog.scielo.org/en/2013/11/18/open-data-liquid-information-democracy-innovation-the-times-they-are-a-changin/#.XFckRlxKiM8>>. Acesso em: jan. 2019.

SUBER, P. **Open Access Overview: focusing on open access to peer-reviewed research articles and their preprints**. In: **Open access**. Cambridge: MIT Press, 2012. (CHECAR NO ARTIGO FÍSICO)

SUPERINTENDÊNCIA DE TECNOLOGIA DA INFORMAÇÃO DA UNIVERSIDADE DE SÃO PAULO. Gestão de dados científicos. Disponível em: <<https://www.sti.usp.br/gestao-de-dados-cientificos/>>. Acesso em: fev. 2019.

TACKE, Oliver. **Raus aus dem Elfenbeinturm**: Open Science. olivertacke.de. 2011. Disponível em: <<https://www.olivertacke.de/2011/10/23/raus-aus-dem-elfenbeinturm-open-science/>>. Acesso em fev. 2019.

TENOPIR, Carol, et al. **Data Sharing by Scientists**: Practices and Perceptions. PLoS ONE, v. 6, n.6, 2011. Disponível em: <[doi:10.1371/journal.pone.0021101](https://doi.org/10.1371/journal.pone.0021101)>. Acesso em: 11 mar. 2017.

TERRY, Robert. Funding the Way to Open Access. **PLoS Biology**, Califórnia, v. 3, n. 3, p. 0364- 0366, mar. 2005. Disponível em: <<https://doi.org/10.1371/journal.pbio.0030097>>. Acesso em: 08 out. 2018.

TERRY, Robert; KILEY, Robert. Open access to the research literature: a funders perspective. In: JACOBS, Neal (Ed.). **Open Access**: Key strategic, technical and economic aspects. Chandos Publishing, 2006. 7 p. Disponível em: <http://eprints.rclis.org/7531/1/Chapter_10_final.pdf>. Acesso em: 27 ago. 2017.

TRISKA, Ricardo; CAFÉ, Lígia. Arquivos abertos: subprojeto da Biblioteca Digital Brasileira. **Ciência da Informação: Ci. Inf.**, Brasília, v. 30, n. 3, p. 92-96, set/dez. 2001. Disponível em: <<http://www.scielo.br/pdf/ci/v30n3/7291.pdf>>. Acesso em: jun. 2018.

TRUJILLO FERRARI, Alfonso. **Metodologia da pesquisa científica**. São Paulo, SP: McGraw-Hill do Brasil, 1982.T 318 p.

UZWYSHYN, Ray. Research Data Repositories: The What, When, Why, and How. **Computers in Libraries**, v. 36, n.3, abr. 2016. Disponível em: <<http://www.infotoday.com/cilmag/apr16/Uzwysyn--Research-Data-Repositories.shtml>>. Acesso em jun. 2018.

VILARINHO, Tatiane Ferreira. Padrões de comunicação científica nas ciências sociais: a interdisciplinaridade entre economia, ciência política, sociologia e antropologia. 2015. 221 f., il. **Tese** (Doutorado em Ciência da Informação)—Universidade de Brasília, Brasília, 2015.

VIERKANT, Paul; et al. **Schema for the description of research data repositories**: version 2.1, 24 p. 2013. Disponível em: <<http://doi.org/10.2312/re3.004> >. Acesso em: 10 maio 2016.

WELLER, Katrin; PUSCHMANN, Cornelius. Twitter for scientific communication: how can citations/ references be identified and measured? In *Proceedings of the ACM WebSci'11* (pp. 1–4). Koblenz: ACM Publishers. Disponível em: <https://www.researchgate.net/publication/228405782_Twitter_for_scientific_communication_How_can_citationsreferences_be_identified_and_measured>. Acesso em 24 jan. 2019.

ZIMAN, John M. **O conhecimento confiável**: uma exploração dos fundamentos para a crença na ciência. Campinas, SP: Papirus, 1996. 252 p.

Data, Scientific; Astell, Mathias (2017): Benefits of Open Research Data Infographic. figshare. Figure. <https://doi.org/10.6084/m9.figshare.5179006.v3>

8. ANEXOS

Manual de utilização das tabelas e segmentação dos dados

As tabelas anexadas a essa dissertação foram feitas no software Excel, no entanto sofreram um tratamento de desing para tornar a navegação e análise dos dados mais interativa com o leitor.

Essa tabela possui seis planilhas:

1. A Listagem geral dos repositórios cadastradas no Re3data.
2. As Características básicas dos repositórios analisados;
3. As Características relacionadas a Ciência Aberta dos repositórios analisados;
4. As Características de gestão e funcionamento dos repositórios analisados;
5. A análise dos dados e
6. A caracterização dos repositórios analisados em relação aos critérios essenciais da Ciência Aberta.

Para navegar pelas planilhas, basta clicar nos ícones correspondentes, como é demonstrado na figura:

| Coluna 1 | Repositório | URL |
|----------|---|---|
| 1 | Portal de Datos Genómicos del SNDG | http://datos.sndg.mincyt.gob.ar/ |
| 2 | Portal de Datos del Mar - SNDM | http://www.datosdelmar.mincyt.gob.ar/ |
| 3 | Portal de datos de Biodiversidad | http://datos.sndb.mincyt.gob.ar/ |
| 4 | Repositorio Institucional UCASAL | http://bibliotecas.ucasal.edu.ar/opac_css/index.php?lvl=cmspage&pageid=16 |
| 5 | The Global Agricultural Trial Repository and Database | http://www.agtrials.org/ |
| 6 | CIAT Dataverse | https://dataverse.harvard.edu/dataverse/CIAT |
| 7 | WorldClim - Global Climate Data | http://worldclim.org/ |
| 8 | GLOBE | http://globe.umbc.edu/ |
| 9 | International Ocean Discovery Program | http://publications.iodp.org/ |
| 10 | PPBio Data Repository | https://ppbiodata.inpa.gov.br/metacatu/ |
| 11 | IBICT Dataverse Network | https://repositoriopesquisas.ibict.br |

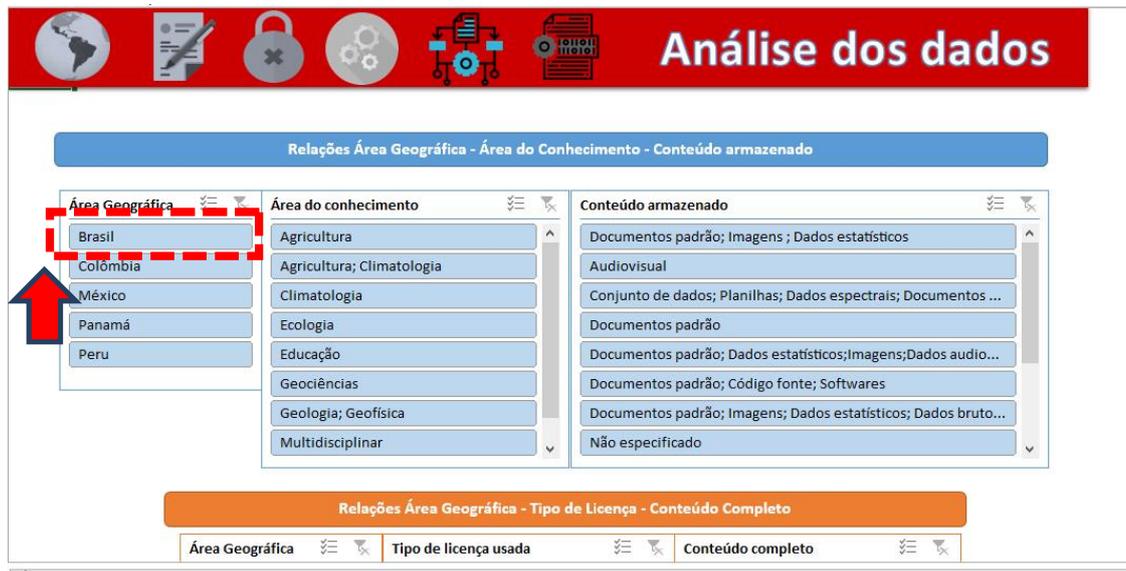
Base de Dados Científicos da

Para alterar as planilhas, basta que se clique na imagem do quesito correspondente, como ilustra a figura

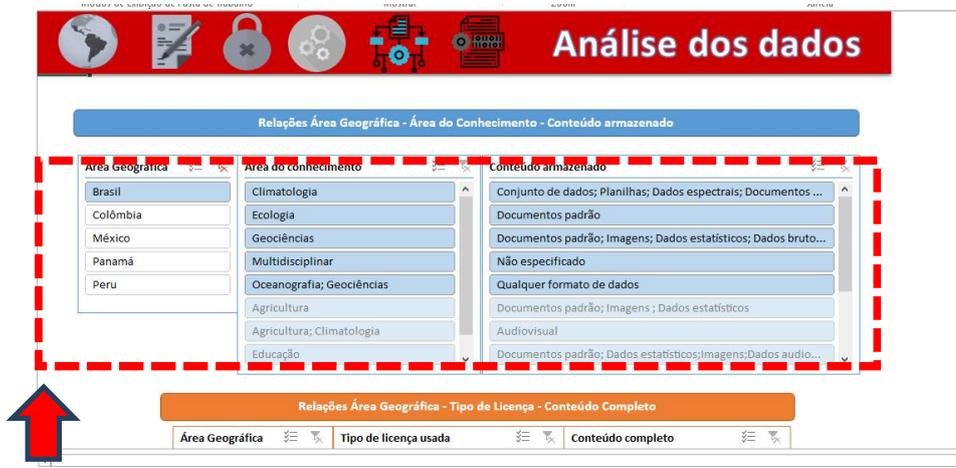
| Coluna 1 | Repositório | URL | Vínculo Institucional | Vínculo Institucional | Área Geográfica | Área do conhecimento | Perfil acadêmico | Tipo | Observações |
|----------|---|---|-----------------------|--|-----------------|---------------------------|------------------|---------------|-------------|
| 5 | The Global Agricultural Trial Repository and Database | http://www.agtrials.org/ | Sim | CGIAR Research Program on Climate Change, Agriculture and Food Security (CCAFS) | Colômbia | Agricultura; Climatologia | Não | Especializado | |
| 6 | CIAT Dataverse | https://dataverse.harvard.edu/dataverse/CIAT | Sim | The International Center for Tropical Agriculture (CIAT), a member of the CGIAR Consortium | Colômbia | Agricultura | Não | Especializado | |
| 7 | WorldClim - Global Climate Data | http://worldclim.org/ | Sim | Reference Center on Environmental Information | Brasil | Climatologia | Não | Especializado | |
| 8 | GLOBE | http://globe.umbc.edu/ | Sim | GLOBE (Global Collaboration Engine) | Brasil | Geociências | Não | Especializado | |
| 9 | International Ocean Discovery Program | http://publications.iodp.org/ | Sim | International Ocean Discovery Program (IODP) | Brasil | Oceanografia; Geociências | Não | Especializado | |
| 10 | PPBio Data Repository | https://ppbiodata.inpa.gov.br/metacatu/ | Sim | Data Observation Network for Earth | Brasil | Ecologia | Não | Especializado | |
| 11 | IBICT Dataverse Network | https://repositoriopesquisas.ibict.br | Sim | Instituto Brasileiro de Ciência e Tecnologia | Brasil | Multidisciplinar | Não | Geral | |
| 12 | Base de Dados Científicos da Universidade Federal do Paraná | https://bdc.c3sl.ufpr.br/ | Sim | Universidade Federal do Paraná | Brasil | Multidisciplinar | Sim | Geral | |

Para realizar a análise das relações que foram averiguadas na dissertação, basta clicar na classificação desejada. No exemplo, deseja-se estabelecer a relação: Quais são as áreas do conhecimento e o conteúdo armazenado nos repositórios de dados brasileiros?

Para realizar a análise, basta clicar em “Brasil” no campo “Área geográfica”.



A relação é estabelecida automaticamente:



Para estabelecer uma nova relação, é preciso limpar o filtro da pesquisa anterior com antecedência:

Relações Área Geográfica - Área do Conhecimento - Conteúdo armazenado

| Área Geográfica | Área do conhecimento | Conteúdo armazenado |
|-----------------|---------------------------|--|
| Brasil | Climatologia | Conjunto de dados; Planilhas; Dados espectrais; Documentos ... |
| Colômbia | Ecologia | Documentos padrão |
| México | Geociências | Documentos padrão; Imagens; Dados estatísticos; Dados bruto... |
| Panamá | Multidisciplinar | Não especificado |
| Peru | Oceanografia; Geociências | Qualquer formato de dados |
| | Agricultura | Documentos padrão; Imagens ; Dados estatísticos |
| | Agricultura; Climatologia | Audiovisual |
| | Educação | Documentos padrão; Dados estatísticos; Imagens; Dados audio... |

Relações Área Geográfica - Tipo de Licença - Conteúdo Completo

| Área Geográfica | Tipo de licença usada | Conteúdo completo |
|-----------------|-----------------------|-------------------|
|-----------------|-----------------------|-------------------|