



UNIVERSIDADE DE BRASÍLIA  
FACULDADE DE CIÊNCIA DA INFORMAÇÃO  
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA INFORMAÇÃO

MILTON SHINTAKU

**Federação de Repositórios Científicos: identificação, análise e proposta  
de modelo baseado nas tendências tecnológicas e da Ciência**

Tese 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 à obtenção do grau de Doutor em Ciência da Informação.

Orientador: Emir José Suaiden

Co-orientador: Claudio Gottschalg Duque

Brasília, 2014

## FOLHA DE APROVAÇÃO

Nome: Milton Shintaku

Título: **Federação de Repositórios Científicos:  
Identificação, análise e proposição de modelo  
baseado nas tendências tecnológicas e da  
Ciência**

Tese apresentada na Faculdade de Ciência da  
Informação - FCI da Universidade de Brasília -  
UnB para qualificação do grau de doutor em  
Ciência da Informação.

Aprovado em:

### **Banca Examinadora**

---

Profa. Dra. Emir José Suaiden  
Presidente da Banca

---

Profa. Dra. Silvana Aparecida Borsetti Gregório Vidotti  
Membro Externo

---

Dra. Cecília Leite de Oliveira  
Membro Externo

---

Profa. Dra. Dulce Maria Baptista  
Membro Interno

---

Profa. Dra. Rosemeire Tavares  
Membro Interno

---

Prof. Dr. Rodrigo Rabello  
Membro Suplente

## DEDICATÓRIA

Dedico esse trabalho a minha esposa Noriko Lúcia Sabanai e a minha filha Lumie Sabanai Shintaku, companheiras de risos e lágrimas. Nem todas as palavras podem exprimir o meu amor.

Ao Saudoso Jaime Robredo, por ter acreditado em mim, encorajando-me a fazer o doutorado.

## AGRADECIMENTOS

Desde antes do início desse trabalho muitas pessoas ajudaram, apoiaram, contribuíram, encorajaram..... Entretanto, primeiramente, quero agradecer a Deus por ter me dado a vida e a oportunidade de estudar;

À minha família pelo apoio e amor incondicional;

À Universidade de Brasília, por meio do Programa de Pós-Graduação em Ciência da Informação da Faculdade de Ciência da Informação pela oportunidade;

Ao meu querido orientador Prof. Dr. Emir José Suaiden pelo apoio e compreensão, sempre estando à disposição. Pela sabedoria e paciência que me guiaram pela pesquisa. Pela confiança depositada, meus sinceros agradecimentos;

Ao meu co-orientador Prof. Dr. Claudio Gottschalg Duque pelas palavras de incentivo, pelo apoio, sugestões que tanto enriqueceram o estudo.

Aos professores da banca pela disposição em avaliar, Silvana Vidotti, Cecília Leite, Dulce Baptista, Rosemeire Tavares, Fernanda Monteiro e Rodrigo Rabello, meu muito obrigado;

Aos professores do Programa de Pós-Graduação da Faculdade de Ciência da Informação, especialmente a Georgete Medleg, Lillian Alvares e Rogério Henrique Araujo Junior. Aos funcionários da Secretaria de Pós-Graduação da Faculdade de Ciência da Informação, sempre dispostos a ajudar, Martha Araújo, Jacqueline Farrapeira e Vivian Miatelo;

Aos Colegas de Curso, que tanto me ajudaram;

A Diretora do Instituto Brasileiro de Informação em Ciência e Tecnologia, Dra. Cecília Leite de Oliveira pelo apoio e disposição em ajudar; Finalmente, mas não menos importante, aos amigos do Ibict, Andrea, Diego, Marcos, Priscila, Ronnie, Washington, Waltinho, Zezinha, Magda, Michelli, Mariana e tantos outros pelo apoio e palavras de carinho.

## SUMÁRIO

Folha de aprovação .....	II
Dedicatória .....	III
Agradecimentos.....	IV
Lista de Figuras .....	VIII
Lista de quadros.....	XI
Lista de Siglas.....	XII
Resumo .....	XVII
Abstract .....	XVIII
1 Introdução .....	1
1.1 Justificativa .....	2
1.1.1 Inovação .....	6
1.2 Problema .....	7
1.3 Objetivo .....	10
1.3.1 Objetivo Geral .....	10
1.3.2 Objetivos Específicos.....	10
1.4 Contribuições Esperadas da Pesquisa .....	11
1.5 Restrições .....	12
1.6 Premissas e Hipóteses .....	12
2 Contexto da pesquisa .....	15
2.1 Repositórios institucionais.....	16
2.2 Federações de repositórios .....	20
3. Revisão de Literatura .....	23
3.1. A informação na web.....	24
3.1.1. Tipos de informação na web.....	26
3.2. Bancos de dados .....	32
3.2.1. Metadados e esquemas de metadados.....	34
3.3. Tendências Tecnológicas .....	38
3.3.1. Big Data .....	39
3.3.2. Mineração de dados.....	41
3.3.3 Computação em nuvens .....	43

3.3.4 Mídias Sociais .....	45
3.3.5. Web-Science.....	47
3.4. Comunicação Científica na era digital .....	50
3.5. Tendências na Comunicação Científica.....	51
3.5.1. Acesso aberto.....	54
3.5.2. Arquivos abertos .....	56
3.5.3. Dados Abertos.....	61
3.5.4. Ciência Aberta .....	64
3.5.5 Curadoria Digital.....	66
3.5.6. E-Science .....	68
3.6. Repositórios, redes e federações .....	71
3.6.1. Repositório .....	72
3.6.2. Interoperabilidade .....	76
3.6.3. Redes.....	79
3.6.4. Federações .....	82
3.6. Conclusão do levantamento de literatura.....	86
4. Metodologia .....	90
4.1. Referencial teórico .....	91
4.2. Modelos.....	94
4.2.1 Modelo conceitual .....	97
4.2.2 Método de representação IDEF0 .....	99
4.3 Desenho da pesquisa.....	103
4.3.1 Abordagem.....	104
4.3.2 Técnicas de Pesquisa.....	106
4.3.3 Universo da Pesquisa e amostras .....	109
4.3.4 Objeto da Pesquisa.....	110
4.3.5. Variáveis.....	112
4.3.6. Métodos para análise e interpretação de dados.....	114
5. Resultados e discussões .....	116
5.1 Federações de repositórios: conceitos, políticas, características e tendências. 117	
5.1.1. Networked Digital Library of Theses and Dissertation .....	119

5.1.2 Biblioteca Digital Brasileira de Teses e Dissertações .....	122
5.1.3 Portal Brasileiro de Acesso Aberto à Informação Científica .....	124
5.1.4 Repositório Científico de Acesso Aberto de Portugal .....	126
5.1.5 LaReferencia .....	128
5.1.6 Sistema Nacional de Repositorios Digitales .....	130
5.1.7 Biblioteca Digital Colombiana .....	131
5.1.8 Red Mexicana de Repositorios Institucionales .....	133
5.1.9 Acceso Libre a Información Científica para la Innovación .....	135
5.1.10 Repositório da Produção Científica do Conselho de Reitores das Universidades Estaduais Paulistas .....	136
5.1.11 Sobre as federações .....	138
5.2 características dos repositórios que influenciam na federação de repositórios	146
5.2.1 Análise dos portais dos repositórios .....	147
5.2.2 Análise da adoção das tendências tecnológicas pelos repositórios .....	161
5.2.3 Análise da adoção das tendências da Ciência nos repositórios.....	167
5.2.4 Análise da oferta de meta-informação pelos repositórios .....	174
5.3. Tendências tecnológicas e da ciência no desenvolvimento de Federações .....	177
5.3.1 Desenvolvimento de Federações .....	178
5.3.2 Tendências tecnológicas e as federações .....	195
5.3.3 Tendências da ciência e as federações .....	207
6 Considerações finais .....	226
6.1 Modelos de federações .....	227
6.1.1 Considerações sobre as hipóteses .....	227
6.1.2 Contribuições da tese.....	230
6.2 Estudos futuros.....	238
Referências .....	240

## LISTA DE FIGURAS

Figura 1 - Gráfico da distribuição dos repositórios por tipo.....	17
Figura 2 - Gráfico da distribuição dos repositórios de acesso aberto por país.....	17
Figura 3 - Estrutura da Revisão de Literatura.....	23
Figura 4 - Modelo de comunicação científica de Costa.....	28
Figura 5 - Modelo de comunicação científica projetada para 2020 .....	30
Figura 6 - Diagrama do Processo de mineração .....	42
Figura 7 - Base conceitual da web-science .....	48
Figura 8 - Modelo de comunicação de Hurd para 2020 .....	52
Figura 9 - Representação dos conceitos e parâmetros dos Arquivos abertos.....	58
Figura 10 - Diagrama das relações entre os conceitos do OAI-ORE.....	61
Figura 11 - Esquema representativo da federação de repositórios científicos.....	87
Figura 12 - Relação entre as tendências tecnológicas e a federação de repositórios ...	87
Figura 13 - Relação entre as tendências da Ciência e a Federação.....	89
Figura 14 – Estrutura de uma federação de repositórios .....	92
Figura 15 – conceituação de federação de repositório.....	93
Figura 16 - Esquema de um modelo conceitual para federação de repositórios. ....	99
Figura 17 – representação de um diagrama IDEF0 padrão .....	102
Figura 18 – Representação do refinamento de processo do IDEF0 .....	102
Figura 19 - Modelo de estratégia aninhada concomitante .....	106
Figura 20 - Detalhe da página do oásisbr apresentando resultado de busca .....	125
Figura 21 - Mídias sociais nos repositórios.....	162
Figura 22 - Mobilidade nos repositórios.....	163
Figura 23 - Computação em nuvens e os repositórios .....	164



Figura 24 - Big Data e os repositórios.....	165
Figura 25 - Dificuldades na adesão das tendências tecnológicas.....	166
Figura 26 - Políticas de depósito dos resultados de pesquisa.....	169
Figura 27 - Ciência Aberta e os repositórios.....	172
Figura 28 - A geração de informação estratégica nos repositórios.....	175
Figura 29 - Diagrama O de uma federação de repositórios.....	178
Figura 30 - Diagrama IDEF0 de Federar Repositórios.....	180
Figura 31 - Diagrama Coletar Metadados.....	186
Figura 32 - Diagrama IDEF0 Gerenciar repositórios .....	187
Figura 33 - Diagrama Executar Coleta .....	189
Figura 34 - Diagrama de Mapear Metadados .....	190
Figura 35 - Diagrama Indexar Metadados .....	192
Figura 36 - Diagrama Apresentar Metadado.....	193
Figura 37 - Diagrama Apresentar Metadados com os Indicadores.....	194
Figura 38 - Diagrama Federalizar repositórios com Tendências Tecnológicas .....	195
Figura 39 - Federar repositórios com mídias sociais .....	196
Figura 40 - Interoperar com Mídias Sociais.....	197
Figura 41 - Diagrama Analisar Dados Mídias Sociais.....	198
Figura 42 - Federar Repositórios com Técnicas Responsivas .....	202
Figura 43 - Federar Repositórios com Big Data .....	204
Figura 44 - Federar Repositórios com Big Data para Indexação .....	205
Figura 45 - Federar Repositórios com Tendências da Ciência.....	208
Figura 46 - Federar Repositórios com Resposta à coleta de Metadados.....	212
Figura 47 - Responder à Coleta de Metadados .....	214
Figura 48 - Federar Repositório com Oferta de Perfil .....	220

Figura 49 - Federar Repositório com Curadoria Digital .....	223
Figura 50 - Executar Curadoria .....	224
Figura 51 – Relação entre as Federações e as Tendências da Ciência .....	238

## LISTA DE QUADROS

Quadro 1 - Distribuição, por tipo, dos documentos científicos em repositórios brasileiros .....	19
Quadro 2 - Tipos de metadados .....	34
Quadro 3 - Tipos de mídias sociais e seus objetivos .....	46
Quadro 4 - Quadro sumarizado dos conceitos do arquivo aberto apresentado na convenção de Santa fé.....	57
Quadro 5 - Sumário do conceitos do oai-pmh .....	59
Quadro 6 - Comparação entre o E-Science e a pesquisa tradicional .....	70
Quadro 7 - sumário representativo dos objetivos específicos e técnicas de pesquisa	107
Quadro 8 - RElação entre as hipóteses e as variáveis.....	114
Quadro 9 - relação entre resultados de busca pelo mesmo termo em idiomas diferentes.....	121
Quadro 10 - Resumo da utilização da curadoria pelos repositórios brasileiros .....	173
Quadro 11 - Sumário das políticas de federação .....	233

## LISTA DE SIGLAS

ALICE	Repositório Acesso Livre à Informação Científica da Embrapa
ALICIA	Acceso Libre a Información Científica para la Innovación
API	acesso programado às aplicações
BDCOL	Biblioteca Digital Colombiana
BDM	A Biblioteca Digital de Monografias
BDS	Biblioteca Digital Sonora
BDTD	Biblioteca Digital Brasileira de Teses e Dissertações, como este termo pode apresentar ambiguidades nesse trabalho será adotado como padrão as seguintes formas: BDTD nacional para a BDTD mantida pelo Instituto Brasileiro de Informação em Ciência e Tecnologia e BDTD local para as BDTD mantidas pelas universidades.
BDTD da USP	Biblioteca Digital de Teses e Dissertações da Universidade de São Paulo
BID	Banco Interamericano de Desenvolvimento
CDWA	Categories for the Description of Works of Art
CICyT	Cosejo Interistitucional de Ciencia y Tencologia
CLARA	Cooperação Latino Americana de Redes Avançadas
CONACYT	Consejo Nacional de Ciencia y Tecnologia
CONCYTEC	Consejo Nacional de Ciencia, Tecnología e Innovación Tecnológica

CONFOA	Conferência Luso-brasileira de Acesso Aberto
CNPq	Conselho Nacional de Desenvolvimento em Ciência e Tecnologia
CSS	Cascading Style Sheets
CUDI	Corporación Universitaria para el Desarrollo de Internet
DC	Dublin Core
DRIVER	Digital Repository Infrastructure Vision for European Research
CRUESP	Conselho dos Reitores das Universidades Estaduais Paulistas
EAD	Encoded Archival Description
EdUFBA	Editora da Universidade Federal da Bahia
EMBRAPA	Empresa Brasileira de Pesquisa Agropecuária
ETD	Eletronic Theses and Dissertation
ETD-MS	Eletronic Theses and Dissertation - Metadata Schema
Fapesp	Fundação de Amparo à Pesquisa do Estado de São Paulo
FCCN	Fundação para a Computação Científica Nacional
FCT	Fundação para a Ciência e Tecnologia
FEB	Federação de Repositórios Educa Brasil
FINEP	Financiadora de Estudos e Projetos
Fiocruz	Fundação Oswaldo Cruz
GTI Unesp	Grupo de Tecnologia da Informação da Unesp
HTML	HyperText Markup Language

IaaS	Infrastructure as a Service
Ibict	Instituto Brasileiro de Informação em Ciência e Tecnologia
ICAM	<i>Integrated Computer-Aided Manufacturing</i>
Icict	Instituto de Comunicação e Informação Científica e Tecnológica em Saúde
IDEF	<i>Integration DEFinition</i>
INFOTECA	Informação Tecnológica em Agricultura
ISO-639	International Organization for Standardization - 639
LAMBDA	Laboratório de Automação de Museus, Bibliotecas Digitais e Arquivos
LE-UnB	Livros Eletrônicos da UnB
LOM	Learning Object Metadata
MARC	Machine-Readable Cataloging
METS	Metadata Encoding and Transmission Standard
MODS	Metadata Object Description Schema
MTD-BR	Metadata Theses and Dissertation - BRasil
NDLTD	Networked Digital Library of Theses and Dissertations
OA	<i>Open Access</i>
OA	<i>Open Archive</i>
oasisbr	Portal Brasileiro de Acesso Aberto à Informação Científica
OAI-ORE	Open Archives Initiative Object Reuse and Exchange

OAI-PMH	Open Archive Initiative - Protocol Metadata Harvest
OPAC	Catálogos Online de Acesso Público
OpenAIRE	Open Access Infrastructure for Research in Europe
ore	Object Reuse and Exchange
PUC-Rio	Pontifícia Universidade Católica do Rio de Janeiro
qdc	Qualified dublin core
RCAAP	Repositório Científico de Acesso Aberto de Portugal
RDF	Resource Description Framework
REMERI	Red Mexicana de Repositorios Institucionales
RENATA	Red Nacional Académica de Tecnología Avanzada
RIUFBA	Repositório Institucional da Universidade Federal da Bahia
RIUFSC	Repositório Institucional da Universidade Federal de Santa Catarina
RIUnB	Repositório da Universidade de Brasília
RNP	Rede Nacional de Pesquisa
ROAR	Registry of Open Access Repositories
RODA	Repositório de Objetos Digitais de Aprendizado
SciELO	<i>Scientific Eletronic Library Online</i>
SGBD	gerenciadores de banco de dados relacionais
SNRD	Sistema Nacional de Repositorios Digitales
TCC	Trabalhos de Conclusão de Curso

TEDE	Teses E Dissertações Eletrônicas
TIC	Tecnologia da Informação e Comunicação
UEL	Universidade Estadual de Londrina
UFBA	Universidade Federal da Bahia
UFPR	Universidade Federal do Paraná
UFRGS	Universidade Federal do Rio Grande do Sul
UFSC	Universidade de Santa Catarina
UMI	University Microfilms International
UMIC	Agência da Sociedade do Conhecimento
UnB	Universidade de Brasília
UNESP	Universidade Estadual Paulista “Júlio de Mesquita Filho”
UNICAMP	Universidade Estadual de Campinas
URI	Uniform Resource Identifier
USP	Universidade de São Paulo
VLTS	Visionary Technology in Libray Solutions
www	World Wide Web
XML	eXtensible Markup Language
XSLT	Extensible Stylesheet Language Transformations



## RESUMO

A consolidação dos repositórios institucionais das universidades e institutos de pesquisa tem se firmado como um marco na disseminação da informação científica de forma gratuita na Internet, no que permite o acesso ao texto integral dos documentos. Com isso, a integração dessas iniciativas torna-se, naturalmente, o próximo passo, possibilitando a oferta consolidada de informações oriundas de vários repositórios científicos. Entretanto, apresenta também desafios técnicos e tecnológicos, principalmente, em relação às tendências tecnológicas e da ciência. Neste sentido, o presente estudo identifica, analisa e propõe um modelo de federação de repositórios científicos, de forma a colaborar com um tema interdisciplinar no âmbito da Ciência da Informação. Um estudo de caráter exploratório, com abordagem mista, focado na pesquisa bibliográfica e *survey*, que visa obter uma base conceitual que atenda aos objetivos do estudo. Com isso, apoiar futuros trabalhos, visto a inovação do tema e os poucos trabalhos correlatos, apresentando um campo fértil a pesquisa. Por fim, revela-se a importância de estudos que se alinhem aos aspectos teóricos e práticos em iniciativas que facilitam o acesso à informação científica, como forma de apoio à própria ciência.

**Palavras-chave:** Federação de repositório; Tendências tecnológicas e da ciência, informação científica e estratégica, Repositório Institucional

## ABSTRACT

The consolidation of institutional repositories of universities and research institutes has established itself as a landmark in the dissemination of scientific information free of charge on the Internet, while allowing access to the full text of the documents. Thus, the integration of these initiatives becomes naturally the next step, enabling the provision of consolidated information from various scientific repositories. However, it also presents technical and technological challenges, particularly in relation to technological trends and science. In this sense, this study identifies, analyzes and proposes a federation model of scientific repositories in order to collaborate with an interdisciplinary theme within the Information Science. An exploratory study with a mixed approach, focused on literature and survey research that aims at a conceptual basis that meets the objectives of the study. Thus, support future work, since the theme of innovation and the few related works, presenting a fertile field research. Finally, reveals the importance of studies that align theoretical and practical aspects in initiatives that facilitate access to scientific information, in support to science itself.

**Keywords:** Repository Federation; Science and Technology trends; Scientific information , Institutional Repository

# 1 INTRODUÇÃO

---

A produção de informação científica aumenta numa razão significativa, criando um montante de informação cada vez maior. Após a Segunda Grande Guerra, Bush (1945) em reflexão sobre a informação científica existente a época, constatou o crescimento representativo e os problemas relacionados ao acesso deste montante de informação. Assim, desde a metade do século passado, evidencia-se a preocupação com a quantidade de informação e as suas implicações.

Levando-se em conta que informação existente alimenta a geração de novas informações, quanto maior a quantidade de informação disponível de fácil acesso, maior será a quantidade de informações produzidas. Atualmente, com tanta disponibilidade de informação, principalmente, em meio digital, a quantidade de informação disponível e produzida tem crescido em escala exponencial.

Com o advento da internet, se por um lado, algumas facilidades foram desenvolvidas para potencializar o acesso à informação existente na rede mundial, como os motores de busca, por outro lado, o montante de informação total dificulta sua utilização. Johnson (2003, p. 86) analisa que os motores de busca são mecanismos para solucionar a desordem natural da web, criando uma organização artificial ao caos informacional, que cresce à medida que a quantidade de informação é agregada à rede mundial.

Nesse sentido, a quantidade de informação disponível na web compõe um cenário cada vez mais complexo, os motores de busca indexam milhares de páginas e retornam a cada busca uma infinidade de resultados. Nesse contexto, em análise da sobrecarga informacional Bawden e Robinson (2009) revelam que a incapacidade de controlar a situação, sobre que informação é relevante ou potencialmente relevante, pode levar a uma frustração e a um sentimento de opressão.

Ruff (2002, p. 11-12), analisando causas, sintomas e soluções da sobrecarga informacional, aponta que as organizações podem contribuir para minimizar os problemas daí decorrentes com a criação de repositórios, entre outras ações. Com esse

procedimento, pode-se organizar a informação, estruturando-a de forma a facilitar o acesso e a avaliação de relevância.

Evidencia-se o papel dos repositórios no contexto da quantidade de informação disponível na web em armazenar e gerir documentos digitais, por longos períodos, fornecendo acesso adequado aos seus usuários. (VIANA; MÁRDERO ARELLANO; SHINTAKU, 2005). Portanto, os repositórios ofertam serviços de preservação e de acesso numa estrutura organizada para recuperação dos documentos que contêm.

Nesse ponto, devido a importância dos repositórios, autores como Bjork (2009) e Costa (2008) inserem os repositórios, como elementos, nos modelos de comunicação científica, com destacado papel de facilitar o acesso ao documento científico. Assim, repositórios auxiliam tanto no processo de levantamento de literatura, ao permitir acesso a um acervo organizado via web, quanto na disseminação de resultados de pesquisa, com o depósito das publicações.

Ampliando as possibilidades sobre o uso dos repositórios, Leite e Costa (2006, pp. 217-218) revelam a possibilidade do uso de repositórios como ferramenta para gestão do conhecimento em ambientes acadêmicos, já que aumenta a eficiência da comunicação e o compartilhamento, assim potencializando processos de internalização e uso da informação. Neste sentido, salienta-se a função dos repositórios como um agente catalizador da produção de informação.

Por fim, apresenta-se evidências de que os repositórios são sistemas de informação organizados num ambiente, de forma geral, desorganizado da web. Nesse sentido, ofertam funcionalidades que apoiam a geração de novos conhecimentos, não apenas como um depósito de documentos digitais, mas como um elemento ativo no processo de geração de conhecimento, que o alimenta e mantém os frutos, ofertando-os em um sistema cíclico.

## **1.1 JUSTIFICATIVA**

Os repositórios institucionais são iniciativas importantes para a visibilidade das instituições, possibilitando acesso aos documentos científicos via web e assegurando a preservação e autenticidade. Estão presentes principalmente nas universidades e

institutos de pesquisas, devido à produção acadêmica dessas instituições, pois tornaram-se canais de disseminação científica.

No entanto, torna-se um desafio atualmente integrar os repositórios, por meio de uma rede, formando uma federação que oferte informações e serviços consolidados, seguindo as tendências tecnológicas e científicas. As diferenças existentes entre os repositórios apresentam um cenário diversificado, impondo restrições e desafios à integração.

Apesar da federalização, os repositórios mantêm-se independentes, no que possuem uma gestão própria, não vinculada à federação. A integração das informações sobre os documentos mantidos pelos repositórios aumenta a possibilidade de acesso aos acervos. Essa ação potencializaria a visibilidade dos documentos, autores, repositório e instituição.

No que se refere à visibilidade ofertada pelos repositórios, Aguillo e colaboradores (2010) desenvolveram uma metodologia webométrica para classificar as universidades por meio de seus repositórios. Essa classificação baseou-se na premissa que os repositórios disponibilizam a produção acadêmica da instituição. As universidades recebem pontuação pela produção acadêmica disponível no repositório. Essa questão alinha-se às colocações de Costa e Leite (2006), sobre o impacto dos repositórios nas pesquisas e na visibilidade da instituição.

Em um modelo de comunicação científica mais completo, Bjork (2007) coloca os repositórios como fonte de informação para novas pesquisas e como facilitador para dar acesso aos resultados de pesquisas. Revela-se a importância dos repositórios na geração de novas informações científicas, num ciclo em que ao disponibilizar a informação, facilita o acesso e, por consequência, seu uso como fonte para novos conhecimentos.

A integração atual, em forma de federação de repositórios, baseia-se nos preceitos dos Arquivos Abertos, cuja interoperabilidade é vista como o seu cerne, como observam Triska e Café (2001). Esses preceitos, consolidados e difundidos em 1999 pela convenção de Santa Fé, estão firmados e conhecidos, mas ainda encontram alguns

desafios para o pleno estabelecimento. Portanto, para a interoperabilidade, estabelece os seguintes aspectos:

- uso de um protocolo de comunicação comum, que permita a comunicação automática entre máquinas e programas que permitam a oferta e coleta de metadados e/ou documentos completo;
- uso do padrão eXtensible Markup Language (XML), para a representação dos metadados a serem coletados, que facilita a troca de informações na internet;
- um conjunto mínimo de metadados, que descrevem os objetos digitais.

Nesse contexto, exceto o uso do XML, não há uma unanimidade de qual protocolo ou esquema de metadados utilizar, mesmo que, atualmente, haja certa predileção pelo protocolo Open Archive Initiative – Protocol for Metadata Harvesting (OAI-PMH) e o esquema de metadados Dublin Core (DC). Assim, requerem maiores estudos sobre essas questões e sobre os aspectos básicos da interoperabilidade.

O uso do protocolo Open Archives Initiative - Protocol Harvesting Metadado (OAI-PHM), torna as federações assíncronas, visto que o processo de coleta automática de metadados (harvesting) é um processo a ser executado em períodos de tempo regulares. Assim, um documento depositado em um repositório, só aparecerá na federação ou rede depois que for feita a coleta, que pode levar um tempo.

O esquema de metadados Dublin Core, se por um lado é simples e flexível, por outro não consegue descrever objetos digitais mais complexos, principalmente objetos compostos por vários arquivos, possuindo hierarquização, visto que não possibilita múltiplos níveis de descrição. No início os repositórios se propunham a manter, principalmente, artigos pós ou pré-print, atualmente podem gerenciar objetos digitais mais complexos, como bases de dados, objetos educacionais e outros, que trazem certa complexidade para a descrição com o DC.

Por esses motivos, a interoperabilidade como apresentada atualmente requer uma revisão, tanto na questão teórica, quanto na tecnológica. Nos últimos anos, houve mudanças significativas na Tecnologia da Informação e Comunicação (TIC),

principalmente na área de comunicação entre sistemas informatizados, que atualmente permitem ações que há algum tempo atrás não era possível. Entretanto, mudanças devem ser embasadas, pois trata-se de um padrão internacionalmente aceito.

Outro ponto que necessita de atenção refere-se aos serviços ofertados. Federações de Repositórios Científicos devem ofertar mais que uma ferramenta de busca como serviço, implementados com motores de busca genéricos. Por se tratar de um sistema de informação diferenciado, necessita de estudos que levantem os serviços requeridos pela comunidade e passíveis de implementação pelas limitações técnicas e tecnológicas.

Observando pelo viés dos participantes da rede ou federação, Tarapanoff, Araujo Júnior e Comier (2000) apresentam as redes e consórcios, como elementos condicionadores de retorno de investimento. Coloca-se, dessa maneira, os estudos referentes às federações de repositórios científicos tendo reflexos nas fontes, que em grande parte são vinculadas às universidades e instituições de pesquisa.

Do mesmo modo, critérios de avaliação são instrumentos importantes no gerenciamento de qualquer sistema informatizado ou não. Segundo Cohen (2008), os critérios de avaliação são orientadores que apoiam a realização dos objetivos. Assim, em um sistema tão complexo como as federações, estudos que definam critérios de avaliação são requeridos para apoiar desde a implantação até a manutenção.

A necessidade de atualização tecnológica é constante, em alguns casos, no entanto, exige estudos que a amparem. A conformidade com as tendências em desenvolvimento de sistemas informatizados deve ser ponto de estudo para a definição de modelos de federações de repositórios, visto que possui um componente informatizado importante. Assim, estar em consonância com as tendências tecnológicas é crucial para o sucesso das federações.

Verifica-se, portanto, que redes e federações de repositórios científicos ainda é um assunto pouco estudado na Ciência da Informação constituindo-se num campo a ser explorado. Ressalta-se a importância de estudos sobre redes e federações de repositórios científicos como forma de apoiar iniciativas, no que esses sistemas de informação

precisam estar alinhados às tendências da ciência, por serem canais atuais para a disseminação da informação científica.

### **1.1.1 INOVAÇÃO**

A comparação de uma tese com trabalhos correlatos não apenas assegura sua unicidade, mas, também apoia a contextualização do estudo, no âmbito das possibilidades de pesquisa sobre um tópico. Por esse motivo, revela-se um procedimento importante, pela quantidade de informação existente.

Nesse sentido, o presente estudo utiliza como fontes de informação as bases de dados on-line na web, tais como a Biblioteca Digital de Teses e Dissertações (BDTD) nacional e o Portal de Revistas Científicas da *Scientific Electronic Library Online* (SciELO), Portal de Periódicos Capes, além de portais universitários como o Portal da pós-graduação da Universidade Estadual Júlio de Mesquita (UNESP), campus Marília e o buscador *Google Acadêmico*. Assim, buscou-se assegurar um levantamento pontual, mas amplo.

O presente estudo tem como foco principal a questão das federações, que agrupam sistemas de informação autônomos procurando-se revelar os diversos aspectos que envolvem o tema, que podem apresentar-se de várias formas.

Nesse sentido, os repositórios são abordados frequentemente na literatura em Ciência da Informação, tanto que em uma busca na BDTD nacional, iniciativa mantida pelo Instituto Brasileiro de Informação em Ciência e Tecnologia (Ibict). Com um acervo de 97 instituições e mais de duzentas mil teses e dissertações em texto completo, foram encontradas 217 teses e dissertações com o termo repositório, todos os documentos ligados à Ciência da Informação.

Ao buscar, pelo termo exato, "federação de repositórios" ou "rede de repositórios", na BDTD nacional, nenhum registro é retornado. O mesmo ocorre no Portal da pós-graduação da Universidade Estadual Paulista (UNESP) e no Repositório Institucional da Universidade de Brasília (UnB), o que revela o ineditismo da abordagem do tema.



Focando em um aspecto das federações, como a interoperabilidade, tem-se 52 documentos na BDTD nacional. A análise dos metadados desses documentos revela estudos, ora sobre os sistemas de informação que implementam a interoperabilidade, ora sobre a infraestrutura da interoperabilidade. Assim, os estudos abordam os componentes da federação e não as federações propriamente.

No levantamento efetuado no Portal SciELO obteve-se os mesmos resultados do levantamento na BDTD nacional. Ratificam-se, assim, as evidências levantadas nas teses e dissertações. No entanto, percebeu-se que o tópico federação apresenta estudos em várias áreas do conhecimento, um termo polissemico, que apesar de acepção próxima, possui estudos distintos nas várias áreas, sem muita relação entre as visões das áreas.

Em uma busca mais ampla, o Google acadêmico retorna 36 resultados, sendo 23 de documentos brasileiros. Deste total, três registros são apenas referências à citação de outros documentos. Com isso, dos 20 registros restantes, 13 possuem relação com federação de objetos educacionais, projeto da Universidade Federal do Rio Grande do Sul (UFRGS) financiado pela Rede Nacional de Pesquisa (RNP). Os outros sete registros, três abordam bibliotecas digitais, mas apenas um está disponível, os outros apresentam *link* quebrado.

Os levantamentos apresentados acima revelam um tema pouco explorado, muito provavelmente, pelos estudos ainda estarem focados nos repositórios, visto que muitas universidades ainda não possuem repositórios institucionais ou estão em processo de implantação. Assim, apesar de poucos trabalhos correlatos, o tema apresentado está na vanguarda e pode ser considerado como o próximo passo após o estabelecimento dos repositórios.

## **1.2 PROBLEMA**

Os repositórios institucionais têm se tornado uma opção para facilitar o acesso aos documentos científicos produzidos por uma instituição, após terem sido publicados

nos canais tradicionais, ou seja, agregam artigos de eventos, artigos de periódicos, livros, teses, dissertações e outros. Assim, atendem a preferência, no que concerne a fontes de pesquisa, em todas as disciplinas, pois, geralmente, são multitemáticos.

Em um contexto em que os repositórios apresentam-se como parte da solução de preservar, gerir e dar acesso à informação disponível na web, revelam-se como um oásis de informações organizadas em um deserto desestruturado, como a internet se apresenta. Nesse sentido, formas de interligar os repositórios, para prover serviços, apresentam-se com a ampliação das facilidades ofertadas pelos repositórios.

Nesse sentido, as redes e federações de repositórios tornam-se fonte de estudos, não apenas por promover, levantar modelos ou cenários, compondo estudos históricos, mas, principalmente, por fornecer embasamento teórico para a criação, desenvolvimento e manutenção dessas iniciativas. Em análise da interoperabilidade de sistemas de informação, Marcondes e Sayão (2008, p. 1) consideram a pesquisa sobre integração de sistemas autônomos um grande desafio, por envolver quase todos os tópicos estudados pela Ciência da Informação.

Nesse contexto, o tema do presente estudo situa-se nesse escopo, das implementações de redes e federações de repositórios científicos, que possui questões técnicas, mas, também, muitos elementos teóricos, com intersecções com diversas disciplinas. Se a maior interface, primeiramente, apresenta-se com a Ciência da Computação, pelas questões, técnicas, outras relações com diversas disciplinas se apresentam, tais como Administração, Linguística e Direitos, como nos casos de estudos sobre direitos autorais de documentos interoperados, por exemplo.

A complexidade do tema reflete-se nos diversos aspectos passíveis de estudo, sendo a interoperabilidade apontada como aspecto principal, já que possui diversos aspectos em si própria. A interoperabilidade humana, com questões que envolvem comunidades, a interoperabilidade semântica, com os aspectos linguísticos, a interoperabilidade entre instituições e seus aspectos legais, entre outros, forma um cenário de muitas variáveis, que elevam o nível de complexidade do tema.

Outro aspecto a se considerar, com certa ênfase, é a oferta de serviços, que envolve as redes e federações de repositório. Dentre esses serviços, destacam-se as ferramentas de busca consolidada e a descoberta de recursos digitais na web, assim como possibilitar o acesso ao seu conteúdo integral. Em projeto apoiado pela RNP, Vicari (2008) considera que os serviços de busca e descoberta de conteúdo heterogêneo e distribuído na Web são um desafio para várias áreas de pesquisa.

Sobre as questões gerenciais, no ambiente das unidades de informação, Tarapanoff, Araujo Júnior e Commier (2000), propõem essas unidades, como detentoras da salvaguarda da informação organizacional, pois possuem características de prestação de serviços. Assim, a participação em redes e federações, com agregação dos serviços e facilidades dessas iniciativas, podem apresentar-se como mais uma prestação de serviço das unidades de informação.

Assim, temas desafiadores e interdisciplinares podem fornecer diversos problemas para serem estudados, dos mais práticos, para resolver uma questão pontual, até os totalmente teóricos. Dentre esses possíveis problemas, o presente estudo foca na proposição de modelos para implementação de uma federação de repositórios científicos.

Levantar requisitos, critérios de avaliação, necessidades e outros elementos que permitam criar um modelo de federação de repositórios que atenda as diversas características desses sistemas, requer um estudo interdisciplinar visto que tal instância envolve desde questões técnicas até questões políticas.

Por se tratar de uma iniciativa que congrega sistemas, com gerenciamento independente e com características heterogêneas, remete a certos problemas de integração, os quais precisam de estudos para resolução. Por exemplo, como aquelas não ofertadas por sistemas de motores de busca, que, até certo ponto integram sistemas de informações distintos.

A organização de registros de metadados oriunda, de forma automática, de repositórios ou bibliotecas digitais de várias instituições, por exemplo, revela-se um problema para os modelos responderem, seja por proposição tecnológica, ou por

geração ou uso de ferramentas semânticas que permitam organizar registros de metadados, não necessariamente com a mesma estruturação.

Outro ponto a ser destacado, refere-se às tendências dos sistemas informatizados na web. O *Gartner Group* apresentou em 2012 as forças que irão nortear o mercado da TIC e, por consequência, os sistemas informatizados de informação em um futuro próximo. Batizada de “nexus das forças” agrupam: computação em nuvens, mobilidade, mídias sociais e informação (*big data*).

Nesse sentido, um modelo deve alinhar-se às tendências tecnológicas e da ciência, verificando o impacto das novas tecnologias e ideias. Assim, aumenta a complexidade do estudo de um modelo que contemple as diversas facetas que envolvem as federações de repositórios, visto que exige a amplitude que envolva, pelo menos, as principais características da federação.

Nesse contexto, pode-se sintetizar o problema de pesquisa em uma pergunta: quais os requisitos necessários para o desenvolvimento de um modelo para federações de repositórios científicos, que atendam às tendências tecnológicas e da Ciência?

## **1.3 OBJETIVO**

### **1.3.1 OBJETIVO GERAL**

Identificar, analisar e propor um modelo para federação de repositórios científicos, que atenda as tendências tecnológicas e da Ciência.

### **1.3.2 OBJETIVOS ESPECÍFICOS**

- Levantar os conceitos, orientações, políticas, características, tendências e problemas que caracterizem as federações de repositórios;
- Mapear as principais características dos repositórios que influenciam na federação de repositórios;

- Verificar a adaptação das novas tendências tecnológicas e da ciência no desenvolvimento de federação;

## **1.4 CONTRIBUIÇÕES ESPERADAS DA PESQUISA**

A amplitude e a atualidade do tema estudado possibilitam várias contribuições que podem ser aplicadas em outros estudos ou de forma mais prática, visto que por apresentar um modelo, pode ser utilizado no apoio ao desenvolvimento de federações de repositórios. Nesse sentido pode-se esperar as seguintes contribuições:

- revelar um modelo para federação de repositórios, com embasamento teórico, que atenda as várias necessidades que esse tipo de sistema de informação requer;
- apresentar um cenário atual dos repositórios científicos, revelando o potencial desses sistemas para a instituição mantenedora e para os usuários;
- enriquecer o montante de informação científica sobre federação de repositórios, no viés da Ciência da Informação, perpassando linhas de pesquisa como a Comunicação Científica, Gestão da Informação e Organização da Informação, entre outros.

A interdisciplinaridade da Ciência da Informação, por sua vez, possibilita que seus estudos permeiem disciplinas correladas e apresentem contribuições. Nesse ponto o presente estudo busca amparo, conceitualmente, em outras disciplinas e, portanto, pode contribuir, visto a relação existente:

- no que concerne à Informática: apoiar o desenvolvimento de sistemas informatizados para implementação de federação de repositórios, ofertando base teórica que ampare o desenvolvimento de facilidades;
- no que concerne à Administração: os critérios de avaliação possibilitados pelos sistemas informatizados são entradas para sistemas de informações gerenciais, com alinhamento voltado a gestão. Nas federações de repositórios, que agregam informações de várias

instituições, pode-se ter um cenário, permitindo o monitoramento ambiental para tomada de decisões.

- no que concerne à Linguística: apoiar as questões terminológicas e lexicais desse tema, que ainda requer atenção, pois por ser relativamente novo apresenta-se em formação.

## **1.5 RESTRIÇÕES**

Pelas diversas possibilidades relacionadas aos repositórios que por definição é um sistema que oferta facilidades em depositar, preservar e dar acesso a objetos digitais, abrem-se várias formas e tipos de repositórios, que, por sua vez, possibilitam a criação de vários tipos de federações, dependendo da finalidade e tipos de repositórios agregados. Nesse ambiente composto por várias possibilidades, o presente estudo apresenta as seguintes restrições:

- o escopo restringe-se às federações de repositórios científicos;
- a pesquisa se limita à análise teórico-metodológica, com a proposição de um modelo factível, conforme momento tecnológico atual e conceitos vigentes.

## **1.6 PREMISSAS E HIPÓTESES**

O presente estudo baseia-se na proposição de modelos, com aspectos relevantes, para contribuir na melhoria dos sistemas informatizados. Ao reproduzir, de forma controlada, um recorte da realidade, permitindo uma análise mais detalhada dos aspectos que compõem o todo. Assim, contribuir para obter mais conhecimento sobre federações de repositórios e possibilitar uma análise mais profunda do tema.

Nesse sentido, amplia-se a discussão sobre a federação de repositórios científicos, visto que o desenvolvimento de modelos pode, também, propor metas, como objetivo a uma situação ideal. Assim, é necessário desenvolver critérios de avaliação,

que orientem a evolução dessas iniciativas, com o propósito de atender melhor todos os envolvidos nas federações.

Outro elemento norteador do presente estudo firma-se na necessidade de embasamento teórico para o desenvolvimento e avaliação dos sistemas informatizados. Esse elemento torna a ferramenta mais abrangente e universal, diminuindo muito sua obsolescência. Visto que, com base teórica, seguem-se padrões aceitos internacionalmente, alterando processos e metadados, de forma a estarem em conformidade com regras e orientações seguidas por outras iniciativas.

Também, alinha com Marcondes e Sayão (2008), afirmando que a federação de repositórios é o mais alto nível de interoperabilidade, promovendo maior robustez e complexidade. Ressalta-se, nesse ponto, a heterogeneidade dos repositórios que fazem parte da federação, criando um ambiente mais complexo para a integração pela interoperabilidade.

Nesse mesmo sentido, acorda com as proposições levantadas por Tarantino e Cernauskas (2009) relatando que as federações de repositórios podem ofertar serviços mais complexos, além de acesso ao conteúdo de vários repositórios. Nesse ponto ao atribuir maior complexidade às federações, requer-se estudos mais profundos sobre as possibilidades de serviços, que podem ser ofertados pela federação.

Por fim, alinha-se com as premissas relacionadas ao movimento de acesso aberto e com os arquivos abertos, visto que os repositórios, constituintes das federações, na grande maioria, inserem-se nesse contexto. Mesmo que esses dois movimentos, muitas vezes, sejam confundidos até na sigla em inglês, *Open Access* (OA) e *Open Archive* (OA), são movimentos distintos, que possuem intercalações, mas definições e conceitos distintos.

Nesse contexto podem-se apresentar as seguintes hipóteses:

- considerando o quantitativo de repositórios científicos brasileiros é possível a construção de um modelo de federação que atenda a várias finalidades;

- considerando o quantitativo de tipos de repositórios existente no país é fortemente viável que o modelo possa ser utilizado para criação de federações de repositórios não científicos;
- com base nos resultados da pesquisa, o modelo apresentado pode amparar o desenvolvimento de sistemas sustentáveis e estáveis, se utilizado para implementação de ferramentas informatizadas para a criação de federação de repositórios;
- considerando a federação de repositórios como um sistema que organiza um cenário heterogêneo, é possível a aplicação de formas diferenciadas de representação, que contextualizem a informação apresentada;
- considerando que o estudo inclui várias linhas de pesquisa da Ciência da Informação, assim como disciplinas correlatas, pode-se presumir que a pesquisa contribuirá para a Ciência da Informação, em suas questões epistemológicas no que diz respeito às relações entre tópicos de estudo e interdisciplinaridade.



## 2 CONTEXTO DA PESQUISA

A informação científica, entre tantos outros tipos de informação, possui um status diferenciado, por ter sido avaliado pelos pares e ser resultado de um processo em que o método possibilita a reprodução. Tradicionalmente é publicada em trabalhos monográficos como livros, teses e dissertações; em artigos de eventos e em artigos de periódicos. Essas formas de publicação, no entanto, apresentam-se em discussão em face às mudanças, principalmente, reveladas pela era digital e pelos avanços da tecnologia da informação e comunicação.

Um claro exemplo de documento não avaliado pelos pares, mas com forte apelo científico, são os artigos resultantes de palestras, que geralmente são produzidos por estudiosos reconhecidos de uma determinada disciplina. Nas publicações, resenhas e editoriais tampouco são avaliados, mas margeiam a ciência. A divulgação científica em jornais e revistas baseia-se em informações científicas e cada vez tem se tornado mais eficaz. Lievrouwn (1992) considera a divulgação científica como parte do modelo de comunicação da ciência, podendo inclusive definir os caminhos da pesquisa.

Disciplinas próximas às artes como Música, Teatro, Artes Plásticas, Teoria Literária, entre outras, apresentam características muito distintas de outras disciplinas no que diz respeito às formas de produção e publicação de documentos científicos, além de possuírem produções científicas diferenciadas. Acentuam-se as diferenças disciplinares no que diz respeito aos canais preferenciais para publicação de resultados de pesquisas e como fonte de informação para uso.

Nesse contexto, discute-se a questão da produção científica, produção acadêmica e a produção intelectual de uma universidade ou instituto de pesquisa. Estas questões afetam as políticas de formação de acervo, definem a tipologia documental aceita pelo repositório, se aceitarão apenas documentos avaliados pelos pares ou se outros tipos de documentos também farão parte do acervo.

## 2.1 REPOSITÓRIOS INSTITUCIONAIS

Dentre as características de repositórios institucionais, destaca-se obviamente a instituição responsável que mantém o repositório, assegurando a infraestrutura necessária para manter o acervo por longo tempo, ou seja, a preservação do acesso. A instituição, por isso repositório institucional, fornece ao repositório, além da credibilidade, as políticas quanto à tipologia de documentos que irão compor o acervo.

Dessa forma, no que concerne a formação do acervo destaca-se a questão da multitematicidade. Para Café e colaboradores (2003) o repositório institucional é uma reunião de repositórios temáticos em uma única instalação, no que se refere aos tópicos dos documentos que fazem parte do acervo. Revela-se, por esse motivo, a necessidade da organização do repositório institucional, assim como uma biblioteca, possuir no acervo uma grande quantidade de tipos de documentos, pertencentes a várias disciplinas, organizado conforme critérios próprios ou aceitos pela comunidade.

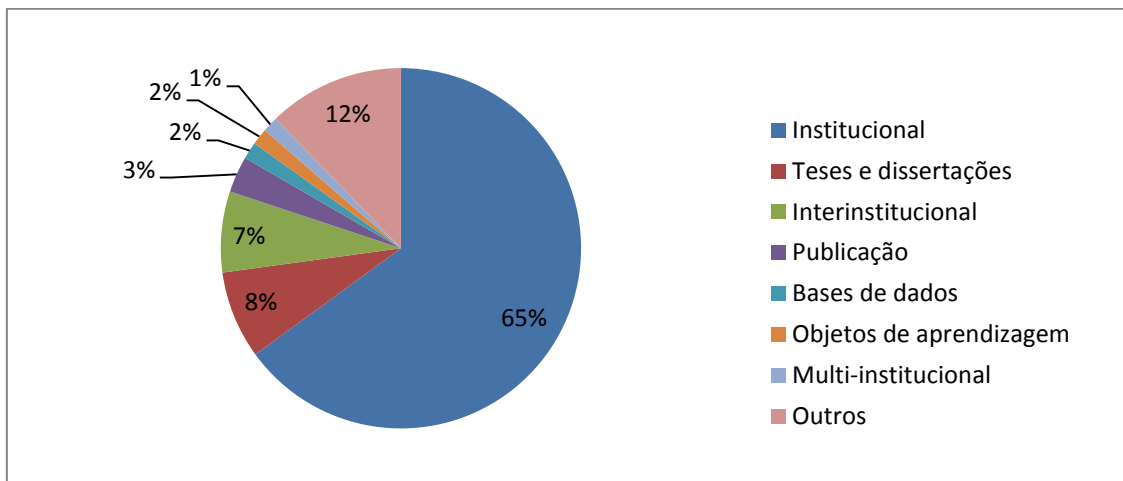
Outro ponto importante que caracteriza o repositório institucional é a vinculação de pelo menos um dos autores à instituição responsável pelo repositório. Assim, o acervo contido em um repositório institucional é representativo da produção da instituição, levando em conta as questões legais relacionadas, em alguns casos, aos direitos de publicação, geralmente cedidos pelos editores de revistas, livros etc.

Na criação dos repositórios, destaca-se o ArXiv, um repositório acadêmico temático na área das Ciências Exatas criado em 1991 por Paul Ginsparg no Los Alamos National Laboratory, atualmente mantido pela Universidade de Cornell. O grande destaque desse repositório era a inovação, um repositório de *preprints* totalmente de acesso aberto, no qual um artigo ainda não publicado poderia ser acessado, lido e comentado. Este tipo de repositório, ainda hoje apresenta uma inovação na comunicação científica, visto que amplia o acesso a um trabalho não totalmente finalizado.

Nesse sentido, o Movimento de acesso aberto à informação científica teve forte influência nos repositórios, denominado de Via Verde por Hanard e colaboradores (2004). Segundo o Registry of Open Access Repositories (ROAR), em fevereiro de

2014, havia 3.478 repositórios registrados, sendo 2.335 repositórios institucionais científicos (Figura 1). Entretanto, possivelmente, esse quantitativo deve ser maior, visto que o cadastro no ROAR é voluntário e dinâmico.

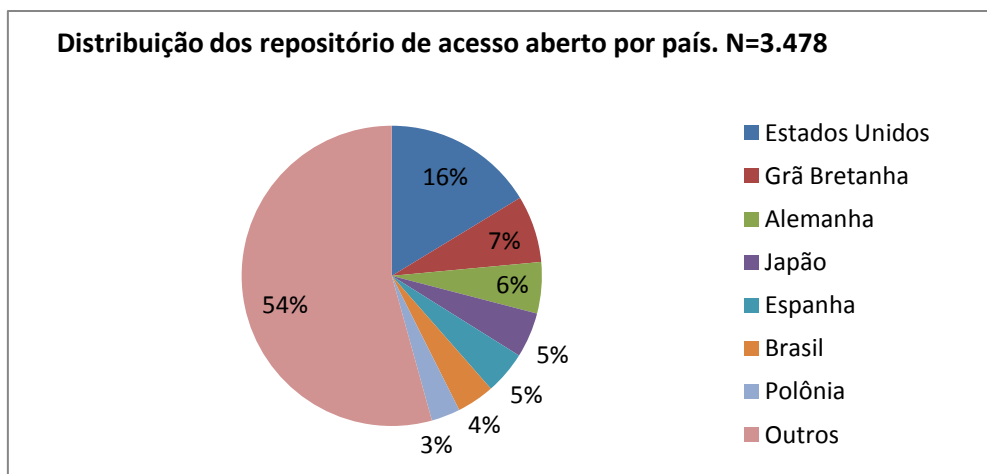
FIGURA 1 - GRÁFICO DA DISTRIBUIÇÃO DOS REPOSITÓRIOS POR TIPO



Fonte: ROAR 2014.

Quanto à distribuição geográfica dos repositórios (Figura 2), o Brasil se apresenta em destaque, sendo o sexto país com mais repositórios cadastrados, com pouco mais de quatro por cento do total, ou seja, 141 repositórios. Desse total de repositórios brasileiros apresentados no ROAR, há alguns registros duplicados e agrega todos os tipos de repositórios. Mesmo com a pouca precisão dos dados disponíveis pelo ROAR, apresenta indícios sobre o interesse brasileiro pelos repositórios.

FIGURA 2 - GRÁFICO DA DISTRIBUIÇÃO DOS REPOSITÓRIOS DE ACESSO ABERTO POR PAÍS



Fonte: ROAR 2014.

Historicamente no Brasil, os repositórios desenvolvidos com sistema informatizado para disseminação de documentação científica tem início com o projeto da Biblioteca Digital Brasileira de Teses e Dissertações (BDTD) nacional, desenvolvido pelo Ibict e financiado pela Financiadora de Estudos e Projetos (Finep). A BDTD nacional tem como objetivo criar repositórios que disponibilizam as teses e dissertações das instituições, de forma a dar maior visibilidade a essa literatura, em muitos casos restrita.

O projeto da BDTD nacional, coordenado pelo Ibict, teve início em 2001 com três instituições piloto: Pontifícia Universidade Católica do Rio de Janeiro (PUC-Rio), Universidade de São Paulo (USP) e Universidade Federal de Santa Catarina (UFSC), conforme relata Southwick (2006). Em 2013, 97 instituições possuíam repositórios de teses e dissertações e disponibilizavam mais de 200 mil teses e dissertações em texto completo.

Para a concretização do projeto da BDTD nacional, o Ibict desenvolveu e distribuiu um pacote de *software* denominado de TEDE (Teses E Dissertações Eletrônicas) para a criação de repositórios destinados à disponibilização de teses e dissertações eletrônicas com camada e protocolo OAI-PMH, muitas vezes denominadas de BDTD local. Outras instituições desenvolveram sistemas próprios, com destaque para a PUC-Rio e Universidade Estadual de Campinas (Unicamp), que desenvolveram sistemas mais alinhados com o conceito de repositórios institucionais.

O TEDE foi distribuído para mais de 100 instituições, das quais algumas estão mudando para repositórios desenvolvidos com o *software* livre DSpace, visto que o TEDE restringe-se à apenas teses e dissertações e muitas instituições necessitam divulgar outros tipos de documentos em formato. Mesmo assim, segundo Macedo e colaboradores (2014), o TEDE representa 45% dos repositórios brasileiros.

A PUC-Rio desenvolveu o Sistema Maxwell, que como discorre Lopes (2001, p. 63) é "[...] a ferramenta de integração do ambiente de ensino e biblioteca digital, assistido com a tecnologia de informação baseada na Web". O Sistema Maxwell se aproxima tanto de uma biblioteca digital, quanto de um repositório institucional, mas ainda apresenta certa predileção às engenharias.

No caso da Unicamp foi desenvolvido o Nou-Rau, que foi distribuído para outras instituições de forma gratuita, como a Universidade Estadual de Londrina (UEL). O Nou-Rau teve início em 2002 e pode ser considerado um dos primeiros *softwares* a implementar conceitos de repositórios institucionais, conforme análise de Marcondes e Sayão (2009).

Atualmente há certa tendência na criação de repositórios institucionais científicos com a utilização do DSpace. Isso se deve, possivelmente, ao apoio do IbiCT a esse *software* livre. O Ranking Web of Repositories apresenta 40 repositórios científicos brasileiros, sendo 37 desenvolvidos com o DSpace, que conforme Macedo e colaboradores totalizam cerca de 48% dos repositórios brasileiros.

Nesse contexto, em janeiro de 2014, os repositórios brasileiros de universidades e institutos de pesquisa disponibilizavam mais de 190 mil documentos científicos em texto integral (Quadro 1). Assim, o Brasil destaca-se na América Latina em relação ao número de repositórios, visto que a Colômbia possui 22 repositórios, Argentina 19, México 14 e Chile 9.

**QUADRO 1 - DISTRIBUIÇÃO, POR TIPO, DOS DOCUMENTOS CIENTÍFICOS EM REPOSITÓRIOS BRASILEIROS**

Quantitativo de documentos em repositórios brasileiros por tipo								
Dissertação	Artigo	Tese	Artigo de evento	Outro	Livro	Relatório	Capítulo de livro	Total
176.922	115.423	63.648	30.751	23.189	1.324	839	692	<b>1942.257</b>

Fonte: o autor baseado no oasisbr: 2014.

A visibilidade dos repositórios brasileiros pode ser avaliada pelo Ranking Web of Repositories, entre outros indicadores, no qual dez repositórios brasileiros figuram entre os 25 maiores repositórios da América Latina. Conforme o mesmo site, em avaliação mundial, o Lume, repositório da Universidade Federal do Rio Grande do Sul aparece na 17<sup>o</sup> posição, sendo o repositório de maior visibilidade na América Latina.

Nesse contexto, pelo quantitativo de repositórios e pela representatividade dos acervos mantidos, apresenta-se um cenário propício à criação de federações com o propósito de fortalecer os repositórios, potencializar a sua visibilidade e ofertar serviços consolidados.

## 2.2 FEDERAÇÕES DE REPOSITÓRIOS

Etimologicamente a palavra federação tem origem latina da palavra *foederatio*, um substantivo de processo do verbo *foedero*, que significa unir-se por aliança. Da mesma forma, em português o termo federação é um substantivo de processo do verbo federar, ou seja, o processo ou resultado do processo de unir-se por meio de aliança, contrato ou adesão.

Teixeira (2001), num enfoque mais pedagógico, revela que as federações se embasam em dois conceitos complementares, a independência dos elementos federados e a união e o fortalecimento da integração. Assim, tornam-se uniões de elementos independentes, mantidos por regras e relacionamentos, com determinado propósito claro.

Baseado nessa premissa, uma federação de repositórios restringe o termo a ser uma reunião de repositórios independentes, firmados por acordos ou adesões, com o propósito de ofertar serviços consolidados. Com essa premissa levanta-se um cenário sobre as federações de repositórios existentes, de forma a contextualizar a pesquisa.

Nesse contexto é possível afirmar que a primeira federação de repositórios foi a Networked Digital Library of Theses and Dissertations (NDLTD), que se restringia a agrupar apenas repositórios de teses e dissertações eletrônicas. Segundo Suleman e colaboradores (2001) a NDLTD teve início em 1996 nos laboratórios da universidade Virginia Tech, tornando-se posteriormente uma federação mundial de repositórios de teses e dissertações eletrônicas.

A NDLTD, em fevereiro de 2014, disponibilizava acesso ao texto integral a mais de 3 milhões teses e dissertações, oriundas de 142 instituições. O Brasil aparece em destaque nessa federação, tendo o Ibict como provedor, disponibilizando mais de 250 mil teses e dissertações, colocando a língua portuguesa como a terceira maior presença, logo depois do inglês e do alemão.

Nesse sentido, em 2002 nasce a BDTD nacional como uma federação de repositórios de teses e dissertações nos moldes da NDLTD, oferecendo, entre outros serviços, um portal de busca consolidada, sendo um dos principais veículos do acesso aberto brasileiro, juntamente com o portal SCiELO. Em 2013, inicia-se um processo de modernização tecnológica para torna-la uma federação mais flexível e ofertar serviços mais apropriados à comunidade acadêmica brasileira.

Num aspecto mais abrangente, no que se refere à tipologia de documentos científicos, mas focado no acesso aberto, em 2008 foi criado o Repositório Científico de Acesso Aberto de Portugal (RCAAP). Paralelamente, no Brasil o desenvolvimento do oasis.br transcendeu o conceito de federação de repositórios, visto que agrega, além de repositórios, periódicos científicos e outras fontes da mesma forma que o RCAAP.

Em 2009, um memorando de entendimento entre Brasil e Portugal, assinado pelo Ministério da Ciência, Tecnologia e Ensino Superior da República Portuguesa e o Ministério da Ciência e Tecnologia da República Federativa do Brasil, possibilitou a integração do RCAAP com o oasis.br. Assim, tanto o RCAAP, quanto o oasis.br compartilham o acervo, que é composto por mais de 500 mil documentos, dando visibilidade a produção científica em língua portuguesa de acesso aberto.

Em âmbito regional o LaReferencia agrupa iniciativas nacionais de nove países da América Latina, nomeadamente, Argentina, Brasil, Chile, Colômbia, Equador, El Salvador, México, Peru e Venezuela. Iniciados os trabalhos em 2010 com a reunião técnica em Lima (Peru), em 2013 apresentou o portal piloto do projeto com mais de 550 mil documentos, sendo que o Brasil participa com o oasis.br, com mais de 400 mil documentos, sendo o destaque dessa federação.

O cenário apresentado revela um ambiente ainda em formação, com a necessidade de estudos que forneçam subsídios para melhoria dos sistemas, pois os sistemas mais antigos como a NDLTD e BDTD nacional se restringem a teses e dissertações eletrônicas. RCAAP e oasis.br, por sua vez, que podem ser considerados espelhos, agregam outros os tipos de sistemas que disponibilizam documentos científicos. Por fim, o LaReferencia ainda se apresenta incipiente, com poucos membros

e muito díspares na produção, em que o Brasil (418.891 documentos) possui cerca de 500 vezes a produção do Equador (707).

No que se refere à infraestrutura para federações pode-se destacar dois projetos, o Digital Repository Infrastructure Vision for European Research (DRIVER) e o Open Access Infrastructure for Research in Europe (OpenAIRE). Esses dois projetos são europeus, entretanto, muitas das orientações podem ser implementadas e adaptadas para federações em outras regiões.

O DRIVER se autodenomina confederação de repositórios digitais europeus e apresenta um conjunto de pacote de *software* e orientações para a criação de federações, com o propósito do compartilhamento e reutilização da informação científica e material educacional, por meio da aplicação de uma camada de serviços. Para o estabelecimento da rede DRIVER foram desenvolvidas as diretrizes para apoiar todos os envolvidos na sua implementação, tanto nas questões técnicas, quanto nas tecnológicas.

O OpenAIRE, por sua vez, tem características mais políticas para a implementação dos preceitos do movimento de acesso aberto (*Open Access*). Com esse projeto procura-se fornecer apoio para o estabelecimento do acesso aberto nas universidades e instituições de pesquisa europeias, visto que a adoção do acesso aberto altera o negócio científico vigente de publicação.

Revela-se que o DRIVER e o OpenAIRE são complementares, com forte apelo para a construção de uma infraestrutura voltada a adoção do acesso aberto à informação científica na Europa. Essa integração de projetos apresenta-se eficaz em um ambiente multicultural e com diferenças marcantes como o europeu.

Nesse contexto, revela-se que as federações de repositórios apresentam várias facetas, como as técnicas, tecnológicas e políticas para a sua implementação, além das questões de infraestrutura física. Da mesma forma que ao integrar acervos de repositórios distintos, que podem ter políticas, também, distintas, apresenta desafios para adaptar-se a todas as diferenças e permitir a integração.

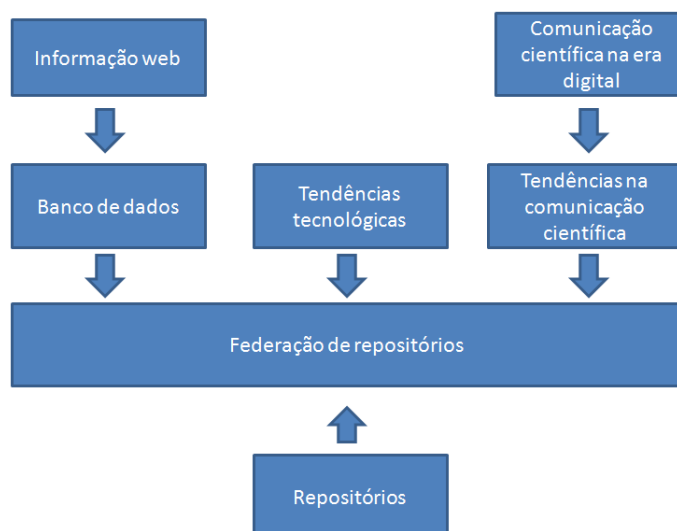


### 3. REVISÃO DE LITERATURA

---

Com o intuito de embasar a presente pesquisa, a Revisão de Literatura apresenta os tópicos que apoiam a proposta de um modelo de federação de repositórios, numa estrutura que converge à formulação de um referencial teórico que atenda as tendências vigentes (Figura 3). Com isso cobre-se os aspectos que envolvem o estudo, de forma a ofertar a conceituação necessária ao estudo.

**FIGURA 3 - ESTRUTURA DA REVISÃO DE LITERATURA**



Fonte: o autor: 2014.

Para tanto, inicia-se com a discussão sobre a informação na internet, a questão mais básica, de forma a contextualizar a pesquisa, que se apresenta no ambiente virtual. Destaca-se a heterogeneidade da web, um ecossistema artificial, que abriga uma variedade de tipo de informações, num cenário complexo e pouco organizado. Para o presente estudo, o tipo de informação focada é a informação científica e estratégica.

Segue pela discussão sobre os bancos de dados, visto que informações podem ficar hospedadas nesses tipos de sistema. Nesse caso, informações e dados científicos podem ser depositados em banco de dados. Da mesma forma que informações

estratégicas podem ser extraídas de bancos de dados, pois são locais próprios para depósito de informação institucional. Discorre sobre os bancos de dados e a forma de armazenamento de informações.

Como todos os sistemas informatizados, os bancos de dados seguem tendências, que na maioria das vezes, tem relação íntima com as necessidades da comunidade de usuários e a evolução tecnológica. Assim, destacam-se as tendências tecnológicas que são passíveis de implementação em bancos de dados acadêmicos, apresentando um cenário atual, importante para uma área dinâmica como a informática.

Na parte relativa à Comunicação científica, apresenta-a no momento atual em que a informatização se torna cada vez mais frequente. Assim, revela-se o cenário atual, pela ótica dos estudiosos, revelando como a tecnologia da informação e comunicação tem alterado o processo de geração de novos conhecimentos. Complementando a Comunicação Científica na era digital, Verificou-se as tendências que tem se apresentado na Ciência. Como tudo, a Ciência está em constante mutação e estudar as tendências ajuda na definição de modelos, tornando-os mais duradouros.

Por fim, na seção de Repositórios, Redes e Federações de Repositórios conceitua-se essas bases de dados que ofertam serviços específicos na web. Assim, completa-se a revisão de literatura, abrangendo todos os aspectos necessários ao estudo.

### **3.1. A INFORMAÇÃO NA WEB**

A World Wide Web (www), a grande rede de alcance mundial ou simplesmente web teve início no início da década de 90 do século XX, com as páginas hipertextuais, mesmo que em muitas vezes o termo seja utilizado como sinônimo de internet, apesar de serem coisas distintas. A internet é uma rede, mais relacionada com a estrutura física, enquanto a web é composta de múltiplos serviços que utiliza a internet como infraestrutura.

Liu (2011, p. 4-5), em análise sobre as características da web, destaca o rápido crescimento das informações disponíveis na plataforma, que a torna uma fonte

importante de informações, apresentando-se, também, como desafio e oportunidade para obter as meta-informações. Entre outras características pode-se destacar:

- **o enorme montante de informação e seu contínuo crescimento.** A gama de informação disponível é tão abrangente que se pode encontrar informações sobre quase qualquer coisa na web;
- **a vasta quantidade de tipos de dados, podendo ser estruturado, semi-estruturado ou não-estruturado.** Da mesma forma que pode ser apresentado em vários formatos, como: texto, imagem, áudio, vídeo e outros;
- **A informação na web se apresenta de forma heterogênea.** A mesma informação pode ser representada de várias formas, palavras ou formatos;
- **Uma quantidade significativa de informação possui *links* com outras informações.** Os *hiperlinks* entre as páginas de um mesmo site são utilizados como forma de organização, em sites distintos podem representar relações, em que páginas muito apontadas podem apresentar maior qualidade ou confiabilidade;
- **A informação na web se apresenta com ruído.** Apenas uma pequena parte das informações apresentadas em uma página é útil, visto que tipicamente uma página possui informações de organização, como os menus, propagandas, informações de direitos autorais, avisos e outros;
- **A web possibilita a oferta de serviços.** Principalmente nos sites de comércio e negócios há uma variedade de serviços que requerem a oferta e execução de operações úteis aos usuários;
- **A web é dinâmica.** As informações disponíveis na web são atualizadas constantemente, mudam a cada instante;
- **A web é uma sociedade virtual.** Possibilita a interação entre pessoas, organização e sistemas.

Sob esse aspecto, apresentam-se três elementos que influenciam o uso da informação na web: o montante de informação, a relevância da informação e sua credibilidade. Numa visão mais contemporânea sobre os problemas relativos à

sobrecarga informacional, Bawden e Robinson (2009, p. 3-4) consideram que a incapacidade de controlar a situação, sobre que informação é relevante ou potencialmente relevante, pode levar a uma frustração e um sentimento de opressão.

No que envolve a credibilidade da informação na web, revela-se que não é exclusivo ou recente, mas toma uma dimensão maior pela facilidade de acesso, replicação da informação, possibilidade de discussão e outros. Serra (2006) aponta a falta de um *gatekeeper* e a quebra o monopólio da informação como fatores que interferem na questão da credibilidade das informações postadas na web.

Yamamoto e Tanaka (2009) corroboram a premissa de que a pouca credibilidade recai por falta de um responsável, como nos casos dos canais de comunicação de massa. Os referidos autores citam que grande parte da informação postada na web é anônima ou não autorizada, o que permite que opiniões sem fundamento possam ser publicadas, gerando informações sem credibilidade.

Outro ponto a destacar é que a informação na web é destinada a humanos. Segundo Devedžic (2006) a informação na web é predominantemente exposta em linguagem natural, facilmente entendida por humanos mesmo estando em múltiplos idiomas, mas que dificulta o processamento automatizado, visto que máquinas não conseguem interpretar o significado da informação. A informação na Web é *machine readable*, que possibilita que programas acessem as informações e a leiam, mas a compreensão da informação ainda é exclusividade humana.

Em um contexto informacionalmente complexo, constata-se que a informação na web compõe-se de um montante humanamente impossível de gerenciar, mas que por ser voltada a compreensão humana, programas não conseguem manipular corretamente. Recai sobre os usuários a árdua tarefa de julgar a credibilidade e veracidade, além do que é ou não pertinente as suas necessidades. Então, torna-se a tarefa primordial dos estudiosos classificarem as informações constantes na web, como suporte ao uso.

### **3.1.1. TIPOS DE INFORMAÇÃO NA WEB**

A informação na web pode ser categorizada por vários critérios devido à quantidade e variedade de tipos, formatos e conteúdos disponíveis. Em termos estruturais as informações na web estão contidas em páginas, sites e portais, possibilitando uma primeira categorização da informação, sob a questão da hospedagem. Da mesma forma que se pode categorizar as páginas web por serem multimídia, hipermídia e hipertexto.

Outro ponto que apoia a categorização se refere à questão do acesso, visto que parte da informação contida na web é invisível. As informações contidas em bancos de dados não podem ser indexadas por motores de busca, tornando-as menos acessíveis, como relatam Sherman e Price (2001). Com isso, se por um lado oferta certa privacidade a informações sensíveis, por outro cria restrição ao acesso a uma parcela importante de informações.

Nesse contexto, em relação aos aspectos mais técnicos tem-se que a informação na web pode ser classificada quanto:

- Estrutura - página, site ou portal;
- Tipo de mídia - discreta (texto e imagem), contínuo (áudio, vídeo e animação);
- Organização - base de dados estruturado ou desestruturado;
- Visibilidade - visível ou invisível.

A categorização, também, pode ser feita em relação ao conteúdo, em relação aos tópicos, finalidade ou o tipo de informação. No que se trata da classificação por assunto, pouco difere das formas de classificação tradicional. Pelo tipo de informação, por sua vez, apresenta uma adaptação do que existe para um cenário digital, compreendendo: informação científica, informação tecnológica, informação religiosa, informação jurídica entre tantas outras, cada qual com suas características.

### **3.1.1.1. A INFORMAÇÃO CIENTÍFICA NA WEB**

Em contrapartida à falta de credibilidade na internet, as informações científicas se apresentam confiáveis, com acurácia e credibilidade. Essas informações são

asseguradas pelo uso de uma metodologia amplamente aceita pela comunidade científica, além da validação pelos pares. Assim, a informação científica compartilha a web com outros tipos de informação, mas possuem um *status* diferenciado.

A geração de informação científica ocorre na comunicação científica, que para Meadows (1984) é o cerne da ciência, visto que, para a informação se tornar científica, deve ser comunicada para ser posta a validação. Ziman (1984) considera a comunicação científica como a instituição social principal da ciência, atuando na geração do conhecimento científico, de forma que transcende o pesquisador, tornando-se o processo pelo qual preserva e alimenta a própria ciência.

Nesse sentido, os modelos de comunicação científica representam seus processos e elementos, de forma a permitir a análise para levantar características e possibilidades. Costa (2008), atualizando o modelo de Garvey e Griffith (1979), apresenta a utilização do formato digital nos processos de pesquisa, revelando a tendência para o periódico e anais de conferência no formato eletrônico, além do depósito em repositório (figura 4). Todas essas iniciativas pode estar na Web.

FIGURA 4 - MODELO DE COMUNICAÇÃO CIENTÍFICA DE COSTA



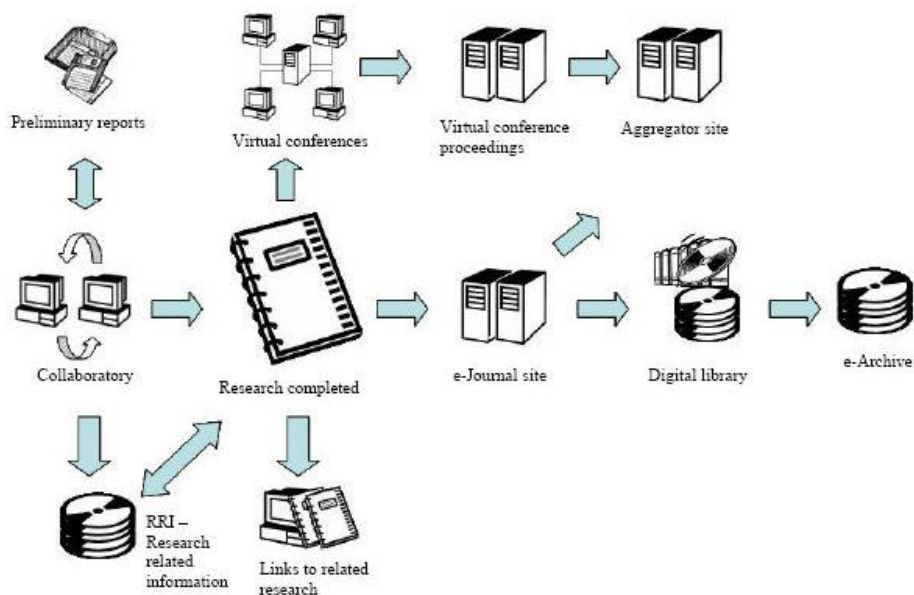
Fonte: Costa (2008)

A informação científica ofertada pela Web se apresenta em várias atividades da pesquisa, revelando a importância da internet nos processos da comunicação científica. O modelo de Bjork (2007) apresenta a comunicação científica, detalhado em 113 processos, sendo que a web é mencionada nas seguintes atividades:

- Obter fundos: Os órgãos de fomento, em muitos casos, fazem uso de informações na internet para avaliar candidatos que pleiteiam fundos para pesquisa, tais como bases de citações, google acadêmico e outros;
- Executar a pesquisa: No processo de obter os conhecimentos existentes para embasar a pesquisa, atualmente, a web é uma grande fonte de informação. No processo de coleta, também, há repositórios de dados abertos e outras bases de dados disponíveis na web;
- Comunicar resultados: A etapa em que a web se apresenta mais presente. Na comunicação formal pelas publicações a informação da web está presente em quase todas as atividades, visto a existência de sistemas na web que publicam resultados de pesquisas, tais como periódicos eletrônicos e anais de eventos. No que concerne a facilitar a recuperação dos resultados, os sistemas de buscas, os indexadores e os repositórios de acesso aberto são algumas das iniciativas que estão disponíveis na web.

Hurd (2000), em uma projeção da comunicação científica em 2020, apresenta um cenário totalmente permeado pela Web (figura 5). Destaca-se nesse modelo a questão da produção do conhecimento colaborativo, mediada por computadores, que atualmente é proporcionada pela web. Da mesma forma que aposta nos eventos virtuais com publicação de anais eletrônicos, periódicos eletrônicos, bibliotecas digitais e agregadores, todos presentes na internet.

FIGURA 5 - MODELO DE COMUNICAÇÃO CIENTÍFICA PROJETADA PARA 2020



Fonte: Hurd (2000)

Revela-se que a informação científica na web se faz presente pelas iniciativas disponíveis na internet, tais como os diretórios, sistemas de buscas, repositórios, bibliotecas digitais, periódicos científicos, anais de eventos e tantos outros. Desse modo, as informações se apresentam como documentos científicos, principalmente os artigos de revistas e eventos, livros, dados de pesquisa, registros de pesquisas e outros.

### 3.1.1.2. A INFORMAÇÃO ESTRATÉGICA NA WEB

A informação estratégica tem por objetivo apoiar a gestão, diminuindo as incertezas e dar vantagens a uma organização, tanto que, geralmente, são obtidas por sistemas de informação. Borges (1995) afirma que uma empresa se torna mais competitiva proporcionalmente a sua capacidade de obter, processar e disponibilizar informações de forma rápida e segura.

Destaca questões de sustentabilidade e sobrevivência das organizações pela capacidade de gerir as informações, tanto internas quanto externas. Monitorar o ambiente a qual a organização está inserida e as suas mudanças, verificar as tendências, entre tantas outras atividades que envolvem a informação se torna ponto crucial nas empresas (BORGES; 1995).



Balarine (2002) discute a questão das tecnologias da informação no âmbito estratégico, em que revela a importância dos sistemas de informação na web para as organizações. Baseado em Porter (1982) e Bickerton, Bickerton & Simpson-Holley (1999) o referido autor advoga a importância da internet nos negócios e a integração entre o sistema de informação e as atividades dos negócios, com forma de obter vantagens no gerenciamento da organização.

Sistemas de informação requerem informações de entrada para que possa executar um processamento e fornecer informações de saída. Carvalho (2001) considera que as informações obtidas na web são formais, por estarem registradas e estruturadas, no que os emails são informais, mesmo que tenham origem em um mesmo meio. Informações disponíveis na web podem tornar-se fontes para sistemas de informações. Da mesma forma, os resultados do processamento podem ser disponibilizados na web. Assim, a informação na web pode apresentar características estratégicas.

Corroborando essa premissa, Miranda (1999) afirma que a informação estratégica na web pode ser classificada sob os vieses do cliente, concorrente, cultural, demográfico, ecológico e outros, que revelam tendências em sob aspectos distintos. Independente da classificação, essas informações podem ser extraídas pela web, nos diversos tipos de sites e portais. Por exemplo:

- Informações sobre clientes são corriqueiramente ofertados e utilizados nos sistemas de venda online, nas opções que ofertam produtos aos clientes de forma seletiva. Os produtos são ofertados conforme o perfil do cliente, resultado de um processamento estratégico;
- Informações sobre o ambiente estão disponíveis nos sites, nas propagandas e outros locais. Com isso pode-se monitorar o ambiente e os concorrentes;
- Indicadores e resultados de pesquisa estão disponíveis na web apresentando informações demográficas e culturais;
- Listas de discussões e redes sociais revelam tendências em comportamento de determinados grupos e outros.

A capacidade de coletar informações de forma automatizada, além das formas de alertas automáticos existentes, transforma a web em uma fonte inesgotável de informações estratégicas, em que o desafio é a seleção e o processamento. Colocam-se as tecnologias da informação e o planejamento de sistemas de informação no cerne das discussões sobre informação estratégica na web.

## **3.2. BANCOS DE DADOS**

Da forma mais simplista pode-se dizer que banco de dados é uma coleção organizada de dados. Então, por essa definição, banco de dados não tem relação direta com a informática, mas com um conjunto de informações em formato físico ou digital. Um fichário contendo registros bibliográficos, um conjunto de fitas de músicas, arquivos de fotos, pela definição são bancos de dados físicos. Da mesma forma, um conjunto de fotos, textos, registros digitais também formam bancos de dados.

No âmbito da Tecnologia da Informação e Comunicação (TIC), Becker (2000) define banco de dados como um conjunto auto-descritivo de dados que representam um domínio de informação. Essa definição restringe o banco de dados a ser digital, em que seus itens de informação tenham alguma relação. Reduz-se o banco de dados a questão operacional de armazenamento, organização e recuperação dos dados.

Em alguns casos, na Língua Portuguesa faz-se a diferenciação de dois termos que em Inglês é único, "*database*". Banco de dados e Base de dados na TIC são entidades distintas, mas que em outras áreas representam a mesma coisa. Para os informáticos banco de dados é uma abstração mais ampla, que em muitos casos compõe-se de várias bases de dados. Nessa relação de pertinência, um banco de dados sem bases de dados não tem função, enquanto não há base de dados gerenciada sem que componha um banco.

Especificamente para a web de uma forma geral, pode-se considerar todo portal que oferta serviços informacionais como um banco, em uma visão mais voltada à Ciência da Informação, que difere da visão da Ciência da Computação, que se restringe às ferramentas e técnicas. Essa visão mais ampla está de acordo com o entendimento voltado às fontes de informação, em que cada base de dados compreende uma fonte.

É inegável a relação entre banco de dados e os serviços de recuperação da informação mantida pelas bases, que pode se apresentar de variadas formas. As ferramentas de busca, listagens e navegações são funcionalidades que fazem uso das facilidades ofertadas pelos bancos de dados, para facilitar a recuperar informações mantidas no banco, atendendo às características e necessidades dos usuários.

Nesse sentido, a evolução das ferramentas ofertadas pela TIC proporciona bancos de dados com organização dinâmica e compostos por itens de dados de diferentes tipos. Da mesma forma que apresenta banco de dados compostos por bases de dados heterogêneos, em um ambiente complexo, atendendo a necessidade de gerenciar dados de diferentes formatos e tipos.

Das possíveis formas de classificar os bancos de dados, o presente estudo classifica nos seguintes tipos de bancos de dados:

- **Banco de dados estruturados:** banco de dados cujas bases de dados apresentam um padrão nos registros, geralmente, formando tabelas em que cada registro é uma linha, podendo apresentar relações entre registros. Listas telefônicas são exemplos de banco de dados estruturados físicos, enquanto diretórios são bancos de dados estruturados na web.
- **Banco de dados desestruturados:** banco de dados cujas bases de dados não possui uma estrutura ou padrão entre seus registros, formado por itens de dados independentes, em que a relação entre os elementos não são de dependência. Bancos textuais e de fotos são exemplos de bancos de dados desestruturados, podendo ser físicos ou digitais.
- **Banco de dados semi-estruturados:** bancos de dados cujas bases apresentam características estruturadas e desestruturadas. Possuem itens de dados e descritores, ou seja, a parte estruturada descreve os itens de dados desestruturados. Repositórios e bibliotecas digitais são exemplos dos bancos de dados semi-estruturados.

Para cada tipo de banco de dados há ferramentas informatizadas que possibilitam o processamento dos dados. Os tradicionais sistemas gerenciadores de banco de dados relacionais (SGBD), para os bancos de dados estruturados, convivem com indexadores

de texto completo e mineradores de dados nos modernos serviços que recuperam e processam informações, principalmente nos bancos de dados semi-estruturados. Para tanto, o processamento nos bancos de dados semi-estruturados requer que os itens de informações possuam alguma descrição, os chamados metadados.

### 3.2.1. METADADOS E ESQUEMAS DE METADADOS

Etimologicamente metadado é um híbrido do prefixo grego meta (μετά) com o substantivo de origem latina dado (datum), que literalmente significa dados sobre dado, que não apresenta toda a amplitude do conceito. Ikematu (2001) apresenta algumas considerações sobre metadados, no que afirma que os metadados descrevem itens para possibilitar serviços, oferta um contexto temporal ao item, que possibilita caracterizá-lo no tempo e espaço e apoia o processo no qual dados se transformam em conhecimento.

O referido autor entende que: "A finalidade principal dos metadados é documentar e organizar de forma estruturada os dados das organizações, com o objetivo de minimizar duplicação de esforços e facilitar a manutenção dos dados" (IKEMATU; 2001, p. 1). Contextualiza, desta forma, os metadados na etapa de organização dos dados, na descrição de dados brutos para um melhor processamento e uso.

Ao se reconhecer o conceito dos metadados possibilita-se o levantamento da sua tipologia. Com isso, pode-se proporcionar uma melhor utilização desse recurso indispensável nos sistemas informatizados. Classificar os metadados conforme suas características permite um melhor entendimento e facilita o uso. Assim, pode-se caracterizar os metadados conforme quadro abaixo (quadro 2):

QUADRO 2 - TIPOS DE METADADOS

Tipos de Metadados	
Tipo	Descrição
<b>Descritivo</b>	Descreve e identifica o item de informação, de forma a facilitar a recuperação do item.
<b>Administrativo</b>	Facilita o gerenciamento e processamento do item, incluindo metadados: <ul style="list-style-type: none"> <li>• <b>Técnicos:</b> Inclui dados técnicos sobre a criação do item, que pode apoiar o processamento e uso por ferramentas;</li> <li>• <b>Preservação:</b> inclui dados que apoiam a preservação e arquivamento do item;</li> <li>• <b>Direitos:</b> relata os dados sobre a propriedade intelectual do Item</li> <li>• <b>Acesso:</b> relata sobre os direitos de acesso</li> </ul>
<b>Estrutural</b>	Apresenta a estrutura interna do item de informação e as relações com outros itens, forma a facilitar a apresentação automática do item.

Fonte: compilação do autor

Para Caplan (2003), metadados fornecem informações estruturadas para recursos de informações apresentadas em algum formato ou mídia, tanto para leitura humana ou por máquina. Entretanto, a autora reforça que os metadados devem seguir as orientações de algum esquema de metadados, independente da tipologia dos metadados.

Como ressaltado, a possibilidade de leitura por máquina só é possível pela padronização dos elementos de metadados por um esquema conhecido e aceito para determinado propósito. Caplan (2003) define esquemas de metadados como um conjunto de elementos de metadados com regras para o seu uso.

Assim, um esquema de metadados define os elementos, nomeando-os, no que é relacionado com a semântica do esquema de metadado. As regras para o uso dos elementos tem relação com os valores, ou seja, os conteúdos de cada elemento, que podem ser livres com apenas a orientação de preenchimento, ou valores fechados com o uso de vocabulários controlados.

Como o esquema é definido para um determinado fim, têm-se vários tipos de esquema de metadados, visto a diversidade de finalidades que as informações e sistemas possuem. Em muitos casos, adaptam-se esquemas de metadados mais genéricos para outros fins, pois uma das características dos esquemas é a flexibilidade nas regras, semânticas e valores. Abaixo alguns dos principais esquemas de metadados

### **Machine-Readable Cataloging - MARC**

Esquema para descrição de registros bibliográficos desenvolvidos pelo Library of Congress (<http://www.loc.gov/marc/marc.html>). Sendo um esquema de metadados detalhado e complexo possibilita que informações possam ser intercambiadas por computadores, principalmente utilizando o protocolo Z39.50. Este esquema de metadados é amplamente utilizado nos Catálogos Online de Acesso Público (OPAC) das bibliotecas. No formato que possibilita a troca de informação entre computadores apresenta certa dificuldade para leitura por humanos. A semântica dos elementos (rótulos dos campos) dá-se por código numérico, por exemplo: 245a é o rótulo para colocar o título. Uma variação é apresentada no formato XML, denominada de MarcXML, em que as tags são representadas pelo nomes dos campos

### **Metadata Object Description Schema - MODS**

Esquema de metadados para descrição bibliográfica e de objeto digital derivado do Marc e desenvolvido pelo Library of Congress (<http://www.loc.gov/standards/mods/>). Apresenta-se mais simplificado que o Marc, mas possui todos os principais campos necessários para descrever um item bibliográfico, com a vantagem que os rótulos são palavras, enquanto no Marc é um código numérico. Em alguns casos, agrupam-se informações de forma hierárquica, ou seja, uma tag agrupa subtags. Ex: a tag <titleInfo> pode ter até cinco subtags, em que somente a tag <title> título é obrigatório. O subtítulo é colocado em subtag separada <subTitle>.

### **Metadata Encoding and Transmission Standard - METS**

Esquema de metadados para objetos digitais desenvolvido pelo Digital Library Federation e mantido pela Library of Congress (<http://www.loc.gov/standards/mets/>). Estruturado em seções, apresenta-se simplificado e organizado, fácil de compreensão humana e por máquina por ser implementado com a tecnologia XML. Cada seção agrupa elementos para uma determinada finalidade, destacam-se as seções administrativa <admSec> , sobre o arquivo digital <fileSec> e sobre a estrutura <structLink>, que permitem descrever melhor objetos digitais complexos, compostos por vários arquivos.

### **Encoded Archival Description - EAD**

Esquema de metadados para descrição de coleção e inventário mantido pela Society of American Archivists e Library of Congress (<http://www.loc.gov/ead/>). Um esquema de metadados de cunho arquivístico que pode ser utilizado por bibliotecas, arquivos e museus para descrever seus acervos. Desenvolvido com a tecnologia XML é organizado em duas grandes seções, o cabeçalho (eadheader) e a descrição do acervo (archdesc). Esse esquema possibilita que instituições intercambiem informações sobre os seus acervos.

### **Learning Object Metadata - LOM**

Esquema de metadados para descrição de objeto educacional mantido pelo Institute of Electrical and Electronics Engineers (<http://ltsc.ieee.org/wg12/>). Estruturado em dez seções, apresenta detalhes necessários para descrever os objetos educacionais ou objetos de aprendizagem digitais, para que possam ser reutilizados. O propósito desse esquema é possibilitar a interoperabilidade de informações sobre objetos educacionais entre sistemas com o objetivo de facilitar a disseminação e reuso.

### **Dublin Core - DC**

Esquema de metadados para descrição objeto digital de forma geral mantido pelo Dublin Core Initiative, organização mantida por várias universidades do mundo (<http://dublincore.org/>). Simples e flexível tem se tornado o principal esquema de metadados utilizado na interoperabilidade, nos moldes do Open Archives. Composto por apenas quinze elementos, pode ser mais especificado se utilizar o sistema qualificado (Qualifier Dublin Core - QDC), em que cada elemento pode receber qualificações, por exemplo, o elemento relation (que indica as relações entre os conteúdos) pode receber qualificadores como: hasVersion (indica que o conteúdo descrito possui versão) ou isPartOf (indica que o conteúdo é parte de um conjunto).

### **Categories for the Description of Works of Art - CDWA**

Esquema de metadados para descrição registros de trabalhos artísticos e arquitetônicos mantido pelo J. Paul Getty Trust e College Art Association ([http://www.getty.edu/research/publications/electronic\\_publications/cdwa/index.html](http://www.getty.edu/research/publications/electronic_publications/cdwa/index.html)). Completo e complexo, apresenta 22 categorias de descritores, sendo que do primeiro ao décimo para descrever o item e do vigésimo ao vigésimo segundo para metadados administrativos, cobrindo todos os aspectos de um registro de uma obra de arte ou arquitetura. Desenvolvido utilizando tecnologia se preza a interoperar repositórios de obras de arte.

Os esquemas de metadados padronizam os bancos de dados semi-estruturados no que ofertam a parcela estruturada dos itens de informação. Nesse caso, os itens se compõem dos registros estruturados definidos pelo esquema e pelo objeto digital, numa

composição de complementaridade. Assim, podem-se utilizar ferramentas que atuam em bancos estruturados e desestruturados, aumentando a oferta de serviços.

Nesse contexto, bancos de dados compreendem sistemas que fazem uso de esquemas de metadados, tanto para a oferta de serviço, quanto para possibilidade de troca de informação com outros bancos. Tornando-se, assim, mais atual aos serviços integrados da web, em que o compartilhamento de informações facilita o acesso, objetivo de quase todos os bancos de dados.

### **3.3. TENDÊNCIAS TECNOLÓGICAS**

Como estudado anteriormente, os bancos de dados atendem as necessidades dos usuários e da instituição mantenedora, ofertam serviços a uma determinada comunidade que faz uso do banco. Assim, atendem, até certo ponto, os anseios por informações de uma comunidade, ou pelo menos parte da comunidade, que está em constante transformação, levando os bancos de dados a estarem em constante reavaliação e ajustes de forma a se manterem atualizados.

Neste ponto, a projeção comportamental da comunidade científica, em relação às necessidades sobre os bancos de dados, pode resultar nas tendências dos bancos de dados científicos. Essa questão alia-se ao que Kikuta e Ribeiro (2011, p. 1) discutem como: "Tendência é todo movimento social, espontâneo ou induzido, que aglutina um grupo significativo de pessoas em torno de comportamento ou características semelhantes, identificáveis numa série de tempo determinada".

Outro ponto a ser considerado nas questões de tendência se refere à evolução tecnológica, que em alguns casos são orientados para atender a essas necessidades. Hopkins (2013) levantou para a o relatório Forrest as dez tendências em tecnologia para 2014 a 2016, apresentando os aspectos relevantes que orientam o desenvolvimento de equipamentos e sistemas. Assim, os aspectos mais relevantes para os bancos de dados são:

- Convergência entre os serviços físicos e digitais: os usuários requerem uma uniformidade nos serviços, independente da plataforma, com



destaque para serviços em dispositivos móveis, com o uso de interface para o acesso programado às aplicações (API): possibilitam o acesso fácil e seguro, por qualquer plataforma, a serviços disponíveis na web;

- Os serviços disponíveis na web são a marca da instituição: a digitalização é uma realidade, então, os serviços devem representar a instituição, com isso valorizam-se as áreas de informação, que apoia a intermediação da informação com os usuários e com isso o negócio assume o controle dos sistemas que ofertam serviços na web;
- Big Data e Open Data: A necessidade de fazer uso de sistemas que consigam lidar com grandes quantidades de dados, com métodos mais eficazes de análise colaborativa de dados em tempo real;
- Aplicações em nuvens: infraestrutura e sistemas utilizando plataformas públicas, privadas ou híbridas em nuvens, possibilitando o armazenamento de grandes quantidades de dados, seu processamento e a oferta de serviços consolidados;
- Serviços colaborativos: Utilização de sistemas relacionados às redes sociais e de colaboração.

Revela-se que os bancos de dados precisam estar preparados para serem acessados por dispositivos móveis como os smart phones, tablets e fablets, com sistemas desenhados para se ajustarem a quaisquer dispositivos e integrados às redes sociais. Da mesma forma necessitam estar preparados para processamento e armazenamento em nuvens, para manipular grandes quantidades de dados.

### **3.3.1. BIG DATA**

O termo Big Data está associado a grande quantidade de dados, mas conforme Vieira et al (2012) e Breternitz e Silva (2013) não possui uma definição definitiva e totalmente aceita, mas refere-se ao processamento para análise de grandes quantidades de dados, efetuados de forma eficiente. Apresenta-se como uma nova abordagem com o propósito de extração de informação baseados no processamento de grandes volumes de dados.

Jewel et al (2014) descrevem o Big Data como um fenômeno resultante da rápida expansão da oferta de dados na internet, resultantes da conversão digital, seja pela criação de novas informações ou migração de informações antigas para a nova plataforma, apresentando um desafio e uma oportunidade à sociedade. Para tanto as soluções que atendem ao Big Data devem atender a(o):

- **volume:** as ferramentas devem ter capacidade de gerenciar e processar grandes quantidades de dados;
- **velocidade:** o processamento de grandes quantidades de dados deve atender à necessidade de entrega aos usuários, ou seja, deve ter velocidade de processamento compatível com as necessidades dos usuários;
- **variedade:** O Big Data se caracteriza pela heterogeneidade dos dados que compõem a base, logo, as ferramentas devem possibilitar o processamento de dados não estruturados em formatos diversos.
- **Veracidade:** os resultados dos processamentos devem ter acurácia, com validação automática, que oferte segurança para o uso.

Taurion (2013) descreve o Big Data como um conjunto de tecnologias e métodos que visam tratar grandes quantidades de dados, de forma a obter informações estratégicas ou para serem mais eficientes no gerenciamento do negócio. Alinha-se, dessa forma, o Big Data a uma finalidade, ou seja, explicar o porquê de ser tão importante para as organizações.

Ao contextualizar o Big Data ao ambiente acadêmico, tem-se um ambiente promissor. Há um montante de informação científica grande, ofertada por universidades e institutos de pesquisa em forma de artigos, apresentações, teses, dissertações, livros e tantos outros, em bancos de dados semiestruturados, portais institucionais desestruturados e outras iniciativas acadêmicas. Um cenário propício para a aplicação das ferramentas voltadas ao Big Data, com o propósito de analisar e gerenciar todo esse montante.

Atualmente a seleção das fontes se apresenta como um grande desafio visto a grande quantidade de documentos científicos digitais existentes. A aplicação do *Big*

*Data* na Ciência apoiaria esse processo, pois a informação existente é fonte para geração de novos conhecimentos. A possibilidade de seleção automática das fontes se apresenta como solução devido, em parte, à incapacidade de levantamento de toda a documentação relevante em uma pesquisa.

Outro ponto em que o *Big Data* apresenta-se como relevante está no processo de coleta de dados. Em alguns casos, pesquisas são baseadas em coleta de dados na internet, recolhendo dados disponíveis, como as pesquisas webométricas, por exemplo. O *Big Data* torna-se uma ferramenta útil nesse processo ao processar informações desestruturadas, recuperando informações baseadas em padrões.

A geração de indicadores torna-se administrativamente o ponto mais relevante do *Big Data*. Agências de fomento, governos, instituições de pesquisas, universidades entre outros podem fazer uso do *Big Data* para obter indicadores importante relativos à produção científica, dados estratégicos que se relacionam com o negócio da instituição.

### **3.3.2. MINERAÇÃO DE DADOS**

O termo inglês *data mining* não possui uma unanimidade na tradução para o Português, sendo geralmente traduzido para o Português do Brasil como mineração de dados, em Portugal como prospecção de dados ou extracção de dados. Independente da tradução, a mineração de dados apresenta-se como uma tendência para uso em bancos de dados, em particular em *Big Data*, principalmente nas empresas comerciais.

Para Monk (2013) mineração de dados é o processo informatizado que trata grandes quantidades de dados com o propósito de encontrar relações desconhecidas com o intuito de apoiar a tomada de decisão, com aspecto estratégico. Esta definição situa a finalidade da mineração de dados no âmbito da geração de informação estratégica, que pode se ampliado conforme a necessidade de processamento de grande quantidade de dados para obter novas informações.

Perrenoud (2013) destaca que a mineração de dados deve operar em bases de dados desestruturados estáticos, com forte viés na estatística, executando processos de descrição ou distinção de dados. Para tanto, opera em análises de aglomerações

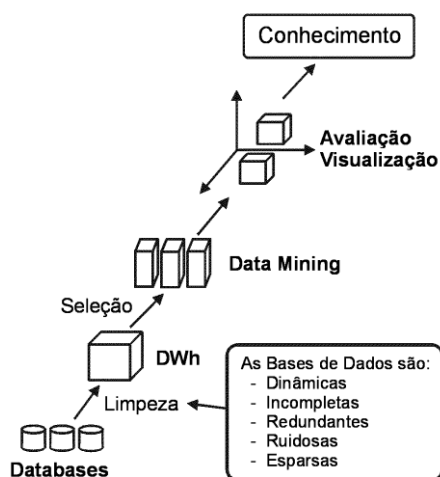
(clusters), associações e elementos isolados para encontrar as relações entre os dados, por isso não podem sofrer alterações online durante a mineração.

Turban et al (2008) estendem a utilização da mineração de dados para a descoberta de tendências e comportamentos, no que, pode ser utilizado na web, analisar listas de discussões, caixas de emails e tantas outras bases de dados mais pessoais. O processo de análise dos resultados obtidos pela mineração requer domínio da área, especialidade no assunto, pois em muitos casos, pode revelar resultados inesperados.

Na Ciência, a mineração de dados é altamente recomendável no processo de obtenção de resultados, como técnica que apoia o reconhecimento de padrões. Swift (2001) destaca que a mineração possui duas fases, a descoberta de conhecimento e o desenvolvimento do conhecimento, fases alinhadas com o processo científico, que requer a análise preliminar das fontes de entrada, conhecimento da área de abrangência para uma mineração mais eficaz.

O processo de mineração de dados para obtenção de conhecimento (figura 6), proposto por Navega (2002), pode ser utilizado em pesquisas acadêmicas ou em outra atuação. Revela ser um processo intermediário ao qual se devem ter previamente selecionado e preparado fontes, o que requer que os problemas a serem analisados tenham sido bem definidos. Da mesma forma que a avaliação dos resultados obtidos é independente do processo de mineração.

FIGURA 6 - DIAGRAMA DO PROCESSO DE MINERAÇÃO



FONTE: Navega (2002)

No fluxo proposto por Navega (200) as bases de dados são compostas por fontes de dados diversas, heterogêneas, que requerem tratamento (verificação de consistência, remoção de redundância e outros), de forma a gerar bases de dados organizados, denominados de “Data Marts” ou “Data Warehouses” (DWh). A mineração de dados é aplicada nesses DWh para extrair padrões que apoiem a criação do conhecimento.

A mineração de dados é um processo em que tem-se como fonte diversas bases de dados, distribuídas, desestruturadas e despadronizadas. Seleciona-se as bases desejadas, trata-as e aplica a mineração de dados, de forma a obter o conhecimento. Assim, é um processo que envolve seleção e tratamento para aplicação de ferramentas que extraem padrões, que requerem análise.

Destaca-se que a mineração de dados associa-se ao método para obter informações, uma ferramenta útil em qualquer atividade que requer processamento de grandes quantidades de dados para obtenção de novas informações e conhecimentos. Associa-se a outras tendências como o Big Data ou processamentos em nuvens, no propósito de tratar a grande quantidade de informação existente.

### **3.3.3 COMPUTAÇÃO EM NUVENS**

A evolução da tecnologia, em especial da infraestrutura da internet, possibilitou o suporte a várias tendências atuais. A computação em nuvens figura entre essas tecnologias, que apoia a resolução tecnológica para muitas ferramentas, ofertando a infraestrutura robusta necessária para plena operacionalização. Compõe-se de uma integração de vários recursos de software e hardware que os tornam transparentes aos usuários.

Baun et al (2011) afirmam que não há uma definição padrão para computação em nuvens, mas conceitos que indicam seus objetivos, que os definem como sendo o uso da virtualização, recursos de armazenamento e tecnologias web para oferta de um ambiente escalonável, confiável e centrado em rede. Esse ambiente fornece recursos independente da aplicação, ofertando-os sob demanda, ou seja, ajustando a necessidade dinamicamente.

Ampliando o entendimento sobre computação em nuvens, Furht (2010) destaca que trata-se de um novo estilo de computação, a qual pode se adaptar dinamicamente às necessidades das aplicações. Por integrar vários tipos de dispositivos e pela capacidade de virtualizar dinamicamente os recursos contidos na nuvem, configura-se numa mudança na forma de estruturar o ambiente da TIC.

Buyya, Broberg e Goscinski (2010), após análise de várias definições, levantaram as características comuns, a todas as definições contemplam, sendo: i) configuração sob demanda, serviço em que se paga o que se usa; ii) flexibilidade na oferta de recurso, em uma ilusão de que os recursos são infinitos; iii) configuração via interface, facilidade na reconfiguração das partições da nuvem; e iv) abstração e virtualização. Para as aplicações e usuários torna-se transparente onde a aplicação ou os dados residem.

Computação em nuvens forma, assim, abstrações, uma infraestrutura virtual que oferta recursos às aplicações, sejam quais forem. Jansa (2011) distingue as nuvens conforme a entidade mantenedora, podendo ser:

- **Nuvens privadas:** mantida por uma entidade para uso próprio, ou seja, somente aplicações de uma entidade reside na nuvem, possuem alto custo de implantação e manutenção, mas ofertam grande segurança às aplicações;
- **Nuvens públicas:** disponível para uso geral, compartilhado por várias entidades, geralmente mantido por grandes organizações que oferecem serviços de nuvens, uma solução mais barata, mas com segurança menor;
- **Nuvens comunitárias:** nuvens compartilhadas por entidades, de forma restrita, como em sociedades;
- **Nuvens híbridas:** integrações de nuvens de tipologia diferente, para formar nuvens maiores.

Revela-se que em muitos casos a utilização de computação em nuvens é totalmente transparente aos usuários. Entretanto, para os gestores de bancos de dados pode apresentar vantagens, no que a nuvem pode fornecer os recursos necessários ao

banco de forma dinâmica, pois a nuvem é virtual, podendo se ajustar às necessidades. Isso implica que espaço requerido para crescimento do banco, requerimentos de segurança, entre outros, são ofertados pela nuvem de forma transparente. Assim, as instituições podem implementar nuvens para manter os serviços ou utilizar serviços comerciais ou públicos.

### **3.3.4 MÍDIAS SOCIAIS**

As Mídias Sociais fazem parte do cotidiano de muitos usuários, principalmente como um espaço de interação pessoal, em que cada usuário participa de várias comunidades de interesse. Entretanto, as Mídias Sociais são mais que espaços de interação, pois possuem um poder de divulgação e abrangência grande, utilizando várias tendências tecnológicas, mesmo que os usuários não percebam.

O termo mídias sociais, segundo Newson, Houghton e Patten (2009), foi cunhado por Chris Shipley, co-fundadora e diretora executiva do grupo Guidewire. Este termo não define uma tecnologia, mas um conjunto de ferramentas e utilitários, que tem por objetivo a comunicação informal, participação e colaboração, na modalidade online, assistida por tecnologia da informação. Nesse ponto, é uma criação para web, explorando todas as possibilidades que esse ambiente proporciona.

Destaca-se que para cada tipo de mídia social existem ferramentas e utilitários, com regras, formas e formatos de informação. Para Tobin e Braziel (2008) existem treze tipos de mídias sociais, categorizadas conforme a finalidade (quadro 3). Entretanto essa classificação não é definitiva, visto que as mídias sociais são dinâmicas.

QUADRO 3 - TIPOS DE MÍDIAS SOCIAIS E SEUS OBJETIVOS

<b>Categoria</b>	<b>Objetivo</b>	<b>Exemplo</b>
Ferramentas de publicação	Publicação de informações, disseminando-as, com a possibilidade de receber comentários.	blogs, postcast, vlogs
Redes sociais	Criação de perfis possibilitando a participação de comunidades e a interação interpessoal	facebook, MySpace, LinkedIn
Ferramentas colaborativas	Criação da informação de forma colaborativa, possibilitando que vários usuários contribuam na escrita de um determinado tópico.	Wikis
Notificador de presença	Criação de publicações curtas, instantâneas, possuindo limite no tamanho da mensagem.	Microblogs, Twitter
Compartilhamento de fotos	Criação de álbuns de fotografias virtuais, com a possibilidade de criação de comunidades.	Flick
Compatilhamento de vídeo	Criação de coleção de vídeos para compartilhamento.	YouTube, Vimeo
Ferramentas de divulgação pessoal	Transmissão em tempo real de vídeos e/ou áudio, com a possibilidade do uso de conversa por meio escrito (chat).	UBroadcast, NowLive
Favoritos (bookmark)	Compartilhamento de favoritos	del.icio.us
Mundos virtuais	Simulação real em três dimensões ou aspectos da vida real	Second Life
Comercio eletrônico	Sistemas de compra online	Amazon, Submarino
Sites de compartilhamento	Criação de diretórios para compartilhamento de arquivos eletrônicos	Torrent, Napster
Widgets	PlugIns que ofertam serviços	Widgets, gadget
Monitoração de conteúdo	Facilidade que possibilita receber informações sobre determinadas fontes	RSS, BuzzMonitor

Fonte: Adaptação de Tobin e Braziel (2008)

No contexto em que as mídias sociais apresentam uma variedade de aplicativos, com uma ampla abrangência, Safko e Brake (2009) consideram as mídias sociais são uma evolução dos meios de comunicação. Revela-se, assim, a efetividade das mídias sociais, no estabelecimento de conexões virtuais entre usuários, podendo ser de ordem pessoal ou comercial.

Deve-se destacar, no entanto, a superficialidade das relações mantidas pelas mídias sociais. Tobin e Braziel (2008) destacam algumas características como a proximidade das mídias sociais com as reuniões informais, a disponibilidade, a assincronicidade, entre outras. Desenvolve-se uma etiqueta social, com regras conhecidas pelos integrantes da mídias, numa relação que apesar da superficialidade é seguida pelos usuários.



Nota-se que as mídias sociais também são amplamente utilizadas comercialmente, não apenas para estabelecer ou concretizar transações, mas para criar ou reforçar uma imagem positiva, como revela Evans (2012). O uso das mídias sociais para divulgar produtos, marcas e serviços é comum e pode fazer parte de estratégias que visam a estabelecer um canal entre os usuários das mídias sociais e os detentores dos produtos, marcas e serviços.

### **3.3.5. WEB-SCIENCE**

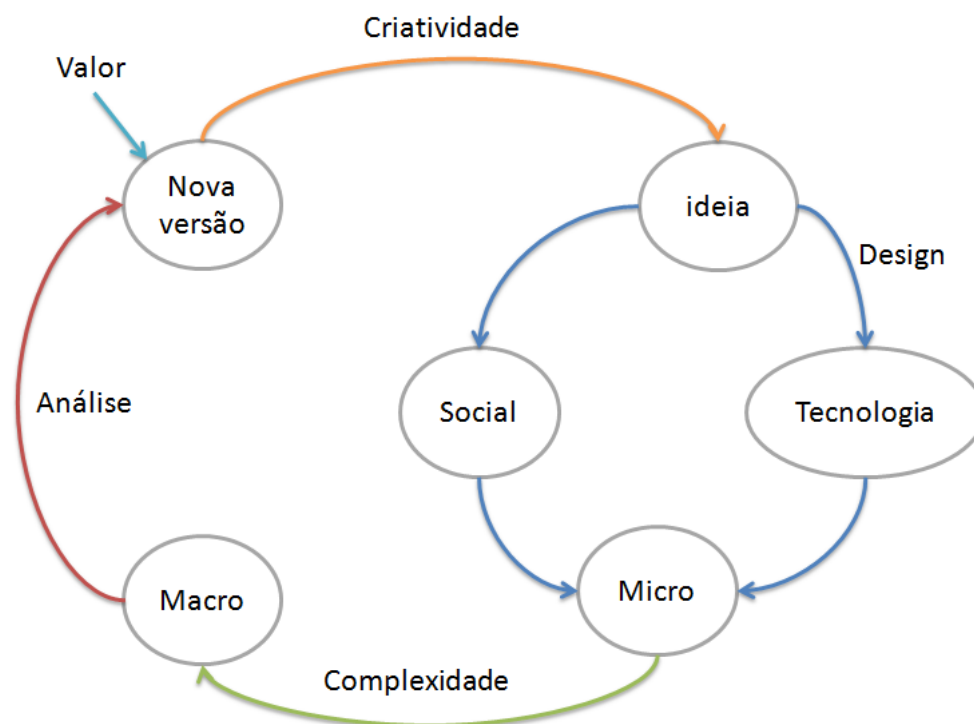
A World Wide Web (www) ou somente web, nascida na década de 90 do século passado, muitas vezes é confundida com a Internet, mesmo sendo coisas distintas. A internet é a infraestrutura no qual a Web opera. Em muitos casos, resalta-se as questões tecnológicas da Web, dos programas, aplicativos, portais e páginas. A web, também, é formada pelos usuários e suas interações, pelos padrões de comportamento, formando um ambiente complexo e fértil à pesquisa. Na convergência entre o tecnológico e o social apresentado pelo Web, tem-se a Web Science, que procura entender a Web em todos os seus aspectos.

Em uma visão mais ampla sobre o tema, Berners-Lee (2005) refere-se ao Web-Science como sendo uma mescla da análise física utilizada nas Ciências Rígidas com os algoritmos da Ciência da Computação. Com isso, alia as questões mais tecnológicas que estabelecem a web com as convenções humanas dos usuários e geradores de conteúdo. Torna-se, assim, intrinsecamente interdisciplinar, no que, pode-se entender o próprio crescimento da web e criar abordagens que permitam a criação de novos padrões, mais potentes, que tragam mais benefícios a todos.

Para o referido autor a Web-Science é mais que uma modelagem à web atual, relaciona-se com o desenvolvimento de novas tecnologias web e sobre o entendimento da sociedade que as utiliza, de forma a criar sistemas que ofertem maiores benefícios. Nesse sentido, agrega técnicas e conceitos provenientes de várias disciplinas para estudar desde as questões pontuais da web até o fenômeno global web e a relações entre esses dois conceitos. Nas palavra do autore “A Web Science refere-se a construção de ferramentas novas e poderosas para a humanidade, de forma a abrir o seus olhos” (BENDERS-LEE, 2005: p.771).

De forma a explicar a base conceitual da Web Science para o desenvolvimento de novos softwares, Hendler et al (2008) apresentam uma nova abordagem, no qual se destacam duas etapas, uma mais pontual denominada de Micro e outra mais ampla denominada de Macro (figura 7). Essa estrutura difere em alguns aspectos dos desenvolvimentos padrão focados na solução de um problema, ou seja, para informatizar processos, de forma a torná-los mais eficaz.

FIGURA 7 - BASE CONCEITUAL DA WEB-SCIENCE



FONTE: TRADUÇÃO DO AUTOR DE HENDLER ET ALL (2008).

Analisando a primeira parte do esquema, denominado de micro, tem-se que uma aplicação nasce de uma ideia e é desenhada conforme uma tecnologia, uma infraestrutura informática composta por ferramentas, sistema gerenciador de banco de dados, linguagem de programação, entre outros. Da mesma forma que uma aplicação é feita para uma comunidade, que dependendo da aceitação, pode determinar a evolução ou descarte do software. Assim, tudo ocorre num âmbito Micro, restrito comparado à escala mundial.

Para os *softwares* muito bem aceitos, ocorre um efeito viral que os levam ao âmbito Macro, ou seja, aumentando muito a comunidade usuária, que interage de maneira nem sempre previsível. Com isso, pode-se verificar problemas ou utilizações que não ocorrem em âmbito Micro, apresentando uma complexidade que não ocorre no âmbito Micro.

A sustentabilidade e continuidade, em muitos casos, é resultante do Macro, que pode, inclusive, desenvolver novas técnicas ou tecnologias. Novas versões são frutos da análise das interações ocorridas no âmbito Macro. O esquema apresentado na figura x propõe se entender a Web, de como desenvolver melhor os aplicativos, de forma a atender a expectativa dos criadores e do público alvo.

Halfords, Pope e Carr (2010), em estudo sobre a Web Science, apresentam quatro conceitos chaves, de forma a prover uma epistemologia para a disciplina, sendo:

- **Co-constituição:** Parte do princípio que a tecnologia atua no desenvolvimento da sociedade, que por sua vez, influencia no desenvolvimento da tecnologia, apresentando uma simbiose, em que tecnologia e sociedade são mutuamente dependentes, e colaboram para o aperfeiçoamento mútuo.
- **Redes heterogêneas:** Ressalta que as redes são formadas por diferentes tipos de usuários e que deve-se levar em conta a questão não-humana dos equipamentos que constituem a rede.
- **Performatividade:** Tanto a co-constituição, quanto as redes heterogêneas enfatizam as relações entre sociedade e tecnologia. A performatividade trata dos resultados dessas relações, sem conceitos pré-concebidos, estuda os impactos que a relação sociedade e web impõe.
- **Mutabilidade:** Destaca que a web está em constante mutação, que em rede, basta um elemento se modificar para que toda a rede mude, atores novos entrando na rede ou atores saindo são modificações na rede.

Para esses mesmos autores, a web science deve:

1. Ter aspectos interdisciplinares, envolvendo questões técnicas, tecnológicas e sociais;
2. Deve estudar como os humanos constroem a web e como a web atua na construção dos humanos;
3. Todos os atores são importantes, o indivíduo, o grupo, a sociedade, a tecnologia, e outros, de forma a entender como um afeta o outro na rede;

4. A web Science deve expandir as fronteiras nos estudos dos fenômenos Micro e Macro;
5. Os desenvolvedores e usuários terem uma postura mais crítica.

### **3.4. COMUNICAÇÃO CIENTÍFICA NA ERA DIGITAL**

A comunicação científica é o meio pelo qual a informação se torna científica, é um processo estabelecido e aceito pela comunidade a que faz parte, que confere a uma informação o status de científico, por meio da avaliação e disseminação a seus pares. A tradição da comunicação científica nasceu junto com a Ciência, mesmo antes da separação entre a Filosofia e Ciência. Ocorreram mudanças nos processos ao longo do tempo, mas é indiscutível que o surgimento do computador e da internet foram marcos para a Comunicação Científica.

Para Osburn (1986, p. 2) a Comunicação Científica é um "fenômeno social pelo qual a atividade criativa e intelectual é transmitida de um acadêmico a outro". Ziman (1984), por sua vez, considera a Comunicação Científica como a principal instituição social, no que ampara a geração da informação científica, apoiando a própria manutenção da Ciência. Destaca-se a questão social da Comunicação Científica, que opera na comunidade. Da mesma forma que salienta ser a atividade de disseminação do conhecimento científico, resultado de pesquisa.

Neste mesmo sentido, Kim (2008) enfatiza que a Comunicação Científica é composta por dois elementos, comunicadores e conteúdo. Os comunicadores são as partes envolvidas, não apenas os autores, mas os avaliadores, editores, associações científicas, universidades, leitores, entre tantos, revelando a abrangência da comunidade envolvida na Comunicação Científica. No que diz respeito ao conteúdo, envolve todos os dados de pesquisa e a documentação de apoio, além das publicações dos resultados da pesquisa, revelando que a Comunicação Científica permeia todo o processo de pesquisa.

Revela-se um cenário composto pela tríade formada pela Informação Científica, Pesquisa e Comunicação Científica, em que cada elemento é independente conceitualmente, mas mantém um forte vínculo um com outro. A Comunicação Científica não existe sem ter uma Informação Científica para comunicar. Assim como não existe Informação sem que haja uma Pesquisa, visto que a Informação é tanto fonte quanto resultados da Pesquisa. A interdependência dos elementos revela-se tão intensa como um grande sistema, em que a alteração em um elemento reflete em todo o sistema.

Nesse contexto, atualmente o uso da tecnologia é constante na Pesquisa e na Comunicação Científica, em que a Informação toma o formato digital. Borgman (2007) afirma que criar uma base tecnológica que ampare a Comunicação Científica é possivelmente fácil, pois a adoção da tecnologia baseia-se nas vantagens ofertadas, que devem superar as facilidades existentes, possibilitando uma contínua evolução nos processos. A autora também destaca que a relação da comunidade científica com a tecnologia não possui um aspecto único, envolve questões sociais, econômicas e políticas. As vantagens devem transcender a pura questão tecnológica, visto que tem que ser adotada por uma comunidade, exposta a vários fatores.

Johnson (2009) destaca que a tecnologia na comunicação científica transcende as questões relacionadas à publicação, envolve os novos modelos de disseminação da informação científica, destacando questões como: a mudança do modelo econômico das publicações, o movimento de acesso aberto, direitos autorais, preservação e a mudança no papel da biblioteca nesse cenário. Revela-se que a utilização de tecnologia é mais abrangente, com decorrências mais profundas que a simples mudança na forma de publicar os resultados de pesquisa, do impresso para o digital.

Nesse contexto, o estudo da tecnologia na comunicação científica requer desmembramento, no qual diversos aspectos podem ser analisados de forma a possibilitar um estudo mais completo. Com isso, pode-se levantar um cenário mais completo com a verificação mais profunda da influência da tecnologia nos processos que envolvem a informação, pesquisa e comunicação.

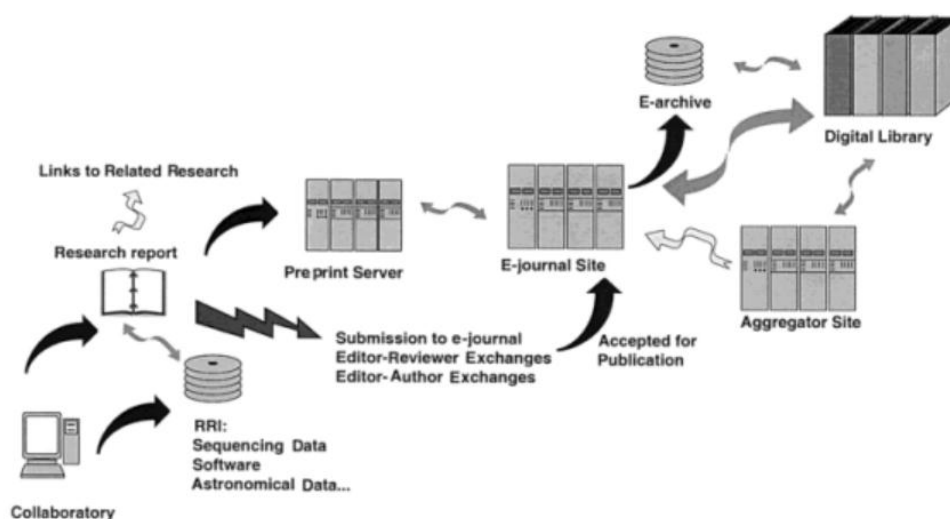
### **3.5. TENDÊNCIAS NA COMUNICAÇÃO CIENTÍFICA**

A Comunicação Científica, como todo processo humano e de cunho social, evolui conforme ocorrem mudanças na comunidade na qual se insere. Em muitos casos apresenta indícios das mudanças que estão por tornar-se aceitas pelos integrantes da comunidade como padrão. Outros fenômenos, como os tecnológicos, também influenciam no processo. Nesse sentido, levantar as tendências possibilita fundamentar modelos duradouros.

Hurd (2000) propondo um modelo para comunicação científica para o ano de 2020 (figura 8), revela tendências que tem se tornado cada vez mais presentes. Assim, apresenta algumas características importantes, tais como:

- Produção colaborativa, o fim da autoria única;
- Publicação em periódicos eletrônicos, fim das publicações impressas;
- Depósito em repositórios ou bibliotecas digitais, múltiplas cópias do artigo em diferentes locais;
- Conexão entre trabalhos correlatos;
- Repositórios de dados de pesquisa.

FIGURA 8 - MODELO DE COMUNICAÇÃO DE HURD PARA 2020



FONTE: Hurd (2000)

No modelo de Hurd para 2020, a tecnologia fortemente faz-se presente em todas as etapas, de ferramentas para construção colaborativa do conhecimento científico, até a

preservação digital dos documentos em agregadores, arquivos e bibliotecas digitais. Nota-se que esse modelo restringe-se a publicação de artigos, que neste caso, dá-se de forma digital, com disseminação do pré e pós-print, em que o pré-print é depositado em serviço destinado a esse tipo de documento e o pós-print em periódico eletrônico.

Nessa mesma linha das tendências na publicação de periódicos, Venkadesan (2009) destaca que os periódicos tem sido a cola que une todas as facetas da comunicação científica, mesmo que nas Ciências Sociais e Humanidades as obras monográficas façam esse papel. No entanto, a tecnologia tem afetado mais os periódicos que as obras monográficas, visto que, quantitativamente há um número maior de artigos digitais disponíveis na web que obras monográficas. Nesse contexto, apresenta um cenário para a Comunicação Científica que envolve:

- Ambiente rico em informação, no qual a internet tornou mais eficaz as formas de comunicação entre pesquisadores, incrementou a oferta de informação e facilitou a troca de informação entre pesquisadores de várias disciplinas;
- Ferramentas para criação colaborativa do conhecimento científico, em que pesquisadores distribuídos geograficamente com certa dispersão podem contribuir em uma mesma pesquisa;
- Agencias de fomento dando suporte ao compartilhamento de informações, com a implementação de políticas de disseminação dos resultados de pesquisa via web, como forma de prestação de contas das pesquisas fomentadas;
- Acesso aberto à informação científica expandindo o compartilhamento de documentação científica de forma gratuita, quebrando as barreiras econômicas, disciplinares, geográficas e outras;

Revela-se um cenário no qual a tecnologia faz-se presente e torna-se ferramenta que suporta todo o processo. Destaca-se a posição atual da adoção de políticas de disseminação dos resultados de pesquisa por parte das agencias de fomento, que utilizam a tecnologia ao longo de todo o processo, do pedido de financiamento até a publicação dos resultados. A questão dos financiamentos nem sempre é abordada nos modelos de comunicação científica.

O uso de tecnologia também é destaque na página dedicada às tendências na Comunicação Científica da Universidade Brandon<sup>1</sup> no Canadá, revelando algumas mudanças na comunicação científica, em que se destacam:

- Compartilhamento instantâneo do conhecimento gerado, por meio de postagem em blogs, notícias institucionais ou pessoais, anúncios em eventos, entre outros;
- Há uma interação maior entre pesquisadores e público em geral, por meio da internet e suas iniciativas, como redes e mídias sociais e das enciclopédias colaborativas;
- Uso de ferramentas para construção colaborativa do conhecimento via internet;
- Utilização de ferramentas educacionais para abranger um número maior de usuários;
- Uso de tecnologia para facilitar o acesso ao conhecimento científico;

Sumarizando as discussões apresentadas pelos autores citados, revela-se que a tecnologia é a tônica das tendências na Comunicação Científica, em que se faz presente em todas as etapas da pesquisa. A influência tecnológica na Comunicação Científica abre uma nova discussão do que é a nova Ciência. Nesse sentido, revela algumas tendências como o Movimento de Acesso Aberto e Ciência Aberta. Portanto, as tendências atuais na Comunicação Científica se apresentam como uma nova revolução aportada pela tecnologia.

### **3.5.1. ACESSO ABERTO**

O Movimento de Acesso Aberto centra-se na disseminação dos resultados de pesquisa, caracterizado pela possibilidade de acesso integral aos resultados de pesquisa sem ônus. Nascido no âmbito das disciplinas das Ciências Rígidas tem se expandido para outras áreas do conhecimento, principalmente para as Ciências Sociais, apresentando um grande impacto na Comunicação Científica.

---

<sup>1</sup> New Trends in Scholarly Communication, disponível em: <http://libguides.brandonu.ca/content.php?pid=223976&sid=1857707>, acessado em 10/10/2014.



O termo Acesso Aberto, do original inglês Open Access, também é conhecido por Acesso Livre, revelando uma variação linguística na Língua Portuguesa. Entretanto, há uma tendência de se utilizar o termo acesso aberto, visto que, em uma busca por termo exato no Repositório de Acesso Aberto de Portugal, que agrega a produção científica luso-brasileira de acesso aberto, revelou indícios dessa tendência. O termo "acesso aberto" retornou 125 documentos, sendo 18 de 2013 e 11 de 2014, já o termo "acesso livre" retornou 132 documentos, mas 17 de 2013 e nenhum de 2014.

O surgimento do Acesso Aberto repousa na chamada crise dos periódicos, em que os autores repassam os direitos patrimoniais aos editores das revistas científicas, que por sua vez cobram altos preços pela assinatura dos periódicos, sendo que na maioria das vezes não há um pagamento pelos direitos autorais. Este modelo de publicação, muitas vezes impedia a própria instituição do autor em ter acesso aos resultados de pesquisa efetuadas pelos seus pesquisadores.

Assim, o Acesso Aberto nasce como um novo modelo para o negócio das publicações de revistas científicas. Para Suber (2012) o Acesso Aberto é a literatura no formato digital, disponível online, de forma gratuita, livre de barreiras para acesso, remetendo a definição apresentada pela Declaração de Budapest<sup>2</sup>:

"Por "acesso aberto" a esta literatura, nos referimos à sua 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." (Declaração de Budapest, 1999).

Destaca-se duas questões, quebra das barreiras ao acesso à informação e a preservação dos direitos dos autores. A internet como instrumento para disseminação livre da informação científica.

---

<sup>2</sup> Disponível em: <http://www.budapestopenaccessinitiative.org/translations/portuguese-translation>

### 3.5.2. ARQUIVOS ABERTOS

Os arquivos abertos, originalmente em inglês *Open Archives*, muitas vezes é confundido com o *Open Access* (Acesso Aberto), pois tanto em português, quanto em inglês, possuem a mesma sigla OA (inglês) e AA (Português). Há também certa proximidade semântica, visto que, os dois termos são tratados no âmbito da Comunicação Científica, relacionando-se com as questões sobre fontes livres de documentos científicos e muitas vezes estão presentes num mesmo sistema.

A relação entre o arquivo aberto e o acesso aberto também pode ser contextualizada no que Costa (2006) denomina por "Filosofia Aberta", composta pelo uso de software livre, arquivos abertos e acesso aberto, com o propósito de criar bases de dados de documentos científicos sem custos de licenças, que ofereçam acesso livre de pagamento e interoperabilidade. Conforme a autora, acesso aberto e arquivo aberto, até certo ponto se complementam e devem ser implementados com software livre de forma a ter em uma plena formalização.

Se o acesso aberto possui um aspecto mais filosófico e político, o arquivo aberto é mais técnico e tecnológico, como apresentado na Convenção de Santa Fé<sup>3</sup>, Novo México, ocorrido em Outubro de 1999, a qual se definiu os parâmetros principais desse movimento. Os conceitos fundamentais do arquivo aberto, sumarizados no quadro 4, revelam a abordagem técnica e tecnológica para que sistemas implementem os conceitos dos arquivos abertos:

---

<sup>3</sup> Disponível em [http://www.openarchives.org/sfc/sfc\\_entry.htm](http://www.openarchives.org/sfc/sfc_entry.htm)

QUADRO 4 - QUADRO SUMARIZADO DOS CONCEITOS DO ARQUIVO ABERTO APRESENTADO NA CONVENÇÃO DE SANTA FÉ

Parâmetro	Tipo	Descrição
Componentes fundamentais	Mecanismo de submissão	O sistema que gerencia documentos digitais deve implementar sistemas de submissão que permita que o autor deposite sua produção intelectual sem a necessidade de intermediário.
	Mecanismo de preservação	O sistema que gerencia documentos deve implementar estratégia para preservar os documentos por longos períodos de tempo.
Tipos de sistemas	Provedor de dados	É o sistema que gerencia documentos digitais, ofertando as funcionalidades aos autores e usuários finais do sistema além de permitir que seus metadados sejam coletados pelo provedor de serviço.
	Provedor de serviço	Sistema que coleta metadados dos provedores de dados e oferta um mecanismo de busca consolidado
Provedor de dados	Identificador único	Todo recurso disponível em arquivo aberto deve possuir um identificador único, de forma a poder ser acessado diretamente,
	Conjunto de metadados	Os documentos disponíveis pelos sistemas gerenciadores de documentos devem descrever os documentos por um conjunto de metadados.
	Camada de protocolo	Os provedores de dados devem implementar uma camada de protocolo que responde à coleta de metadados proveniente do provedor de serviço.
Provedor de serviço	Manter o identificador único	O provedor de dados coleta os metadados e mantém o identificador único .
	Manter as restrições	Se o provedor de dados restringe o uso a um documento, o provedor de serviço mantém a restrição, ou seja respeita as políticas adotadas pelos provedores de dados.

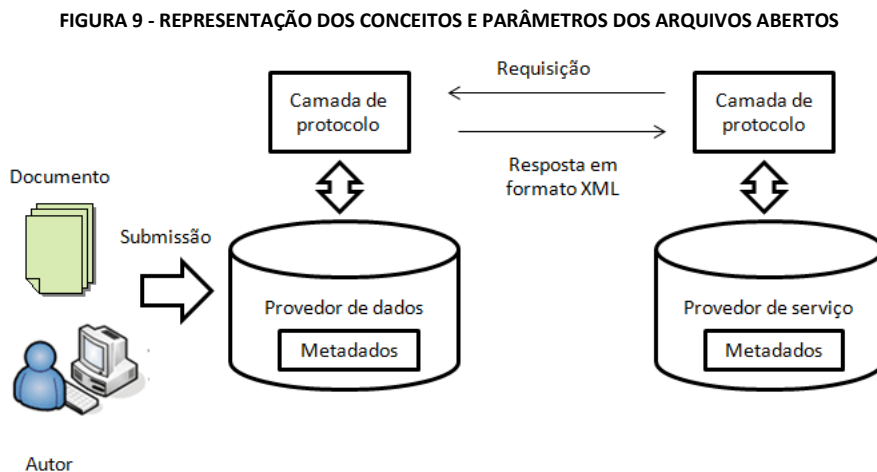
Fonte: o autor; 2014.

Analisando esses conceitos e parâmetros, Triska e Café (2001) apontam que a finalidade dos arquivos abertos é a interoperabilidade entre sistemas informatizados, sendo este processo a palavra-chave dos arquivos abertos, requerendo:

- **Uso de conjunto mínimo de metadados:** para que os sistemas interoperem, devem compartilhar de um esquema de metadados comum, sendo o Dublin Core o mais utilizado;
- **Uso do formato XML (eXtensible Markup Language):** Os metadados devem ser interoperados no formato XML, que facilita a troca semântica dos metadados;
- **Uso de protocolo de comunicação comum:** a comunicação entre o provedor de dados e o provedor de serviço é implementado por um protocolo de comunicação complementar.

De forma geral a figura 9 sintetiza as questões técnicas e tecnológicas dos arquivos abertos, em que dois sistemas (provedor de dados e provedor de serviços), pertencentes a instituições distintas interoperam por meio do uso de um protocolo

comum. Com isso, virtualmente institui-se um vínculo entre as duas instituições, havendo um intercâmbio de informação entre os dois sistemas.



Fonte: o autor

Nesse sentido, Jailet (2005) destaca que para o arquivo aberto funcione eficazmente é preciso que haja uma colaboração entre as instituições. A adesão aos arquivos abertos, no entanto, segundo o autor, ainda não se apresenta tão abrangente como o acesso aberto, no que, muitos estudiosos possuem maior apreço pela disponibilização dos trabalhos em acesso aberto, recuperáveis por motores de busca como o google, não tendo preocupação com a interoperabilidade dos sistemas.

Outro ponto relevante fundamenta-se na evolução tecnológica, em que os sistemas que implementam tanto os provedores de dados, quanto os provedores de serviços adotam novas funcionalidades, que ampliam os conceitos dos arquivos abertos e acesso aberto. Pode-se destacar, especificamente, nos arquivos abertos a questão das camadas de protocolos.

Nesse sentido, cabe salientar a distinção entre protocolos e suas implementações, na medida em que um protocolo é um conjunto de requisitos funcionais, que pode ser implementado de várias formas e em várias linguagens de programação. Por isso, cada ferramenta desenvolve a sua implementação do protocolo, que por ter requisitos funcionais bem definidos possibilita a comunicação de forma eficaz.

## OAI-PMH

Inicialmente, todas as redes que implementavam a interoperabilidade faziam uso do protocolo Open Archives Initiative – Protocol for Metadata Harvesting (OAI-PMH). Com isso, foram estabelecidas todos os parâmetros para a interoperabilidade entre sistemas que utilizam os preceitos dos Arquivos Abertos. Assim, todos os sistemas que declaram aderir ao movimento de arquivos abertos implementam o OAI-PMH.

Lagoze e Van de Sompel (2002), descrevendo o protocolo OAI-PMH, apresentaram os conceitos que orientam os arquivos abertos e o uso do protocolo, que podem ser sumarizados como no quadro 5. Esses requisitos orientam o desenvolvimento de ferramentas que implementam os preceitos dos arquivos abertos.

QUADRO 5 - SUMÁRIO DO CONCEITOS DO OAI-PMH

Nº	Conceito	Explicação
1	Harvester (coletador)	É uma aplicação que é implementada nos provedores de serviço para coletar os metadados dos provedores de dados
2	Repositório	É um servidor conectado à internet, que responde ao harvester, gerenciado por um provedor de dados, possibilitando várias configurações
3	Item	O Item pertence ao repositório e é a unidade coletada pelo Harvester. É um container que armazena ou gera os metadados sobre um recurso em múltiplos formatos em forma de registros.
4	Identificador único	Identifica um item no repositório sem que haja ambigüição, com um formato de Uniform Resource Identifier (URI), possibilitando a extração de metadados em qualquer formato de registro.
5	Registro	É a resposta a uma requisição do Harvester, fornecida no formato XML com as seguintes partes: <ul style="list-style-type: none"> <li>• Cabeçalho: Contém as informações que identificam o registro, como o identificador único, a data e hora da coleta e informações opcionais como o(s) set(s) e o status do registro.</li> <li>• Metadados: contém conjunto de metadado no formato em que o harvester selecionou, sendo que o Dublin Core não qualificado é o padrão.</li> <li>• Sobre: parte opcional do registro que contém informações sobre direitos ou procedência.</li> </ul>
6	Set	O acervo de um repositório pode estar organizado em partes chamadas Sets, com isso, há a possibilidade de fazer coletas seletivas.
7	Tipos de coleta	Uma coleta pode ser feita por diferentes níveis, conforme: <ul style="list-style-type: none"> <li>• Coleta total, coletando todos os itens de um repositório;</li> <li>• Coleta seletiva, coletando todos os itens de um ou mais Sets;</li> <li>• Coleta de um ou mais itens selecionando-os pelo identificador único</li> </ul>

Fonte: O autor com base em Lagoze e Van de Sompel (2002).

Outro ponto crucial no desenvolvimento do OAI-PMH que teve reflexo em toda a web estabelece-se em que todo recurso disponível em um sistema que implementa este protocolo possui uma URI. Com isso, pode-se acessar diretamente o recurso pelo endereço eletrônico. Esta implementação possibilitou que motores de busca, como o google, indexasse o documento e possibilitasse acessá-lo diretamente.

## **OAI-ORE**

A complexidade de alguns recursos presentes na web e que podem ser armazenados em bancos de dados cada vez se apresenta maior. Com isso, os sistemas que implementam os preceitos do acesso aberto enfrentam um desafio, no que necessitam interoperar recursos complexos, mantendo as suas características.

Esses recursos complexos são formados por unidades lógicas que agregam vários arquivos com uma determinada estrutura, a qual deve ser mantida para representar corretamente o recurso. Páginas hipertextuais são exemplos de recursos complexos, visto que são compostas por páginas enlaçadas, podendo conter imagens, folhas de estilos, *scripts* e outros.

Esses recursos agregados possuem as seguintes características:

- compõem-se de mais que um arquivo, podendo ou não estar no mesmo portal;
- os recursos possuem uma relação lógica entre sí;
- os recursos podem ter vários formatos e tipos;
- as relações entre os recursos podem se apresentar de variadas formas.

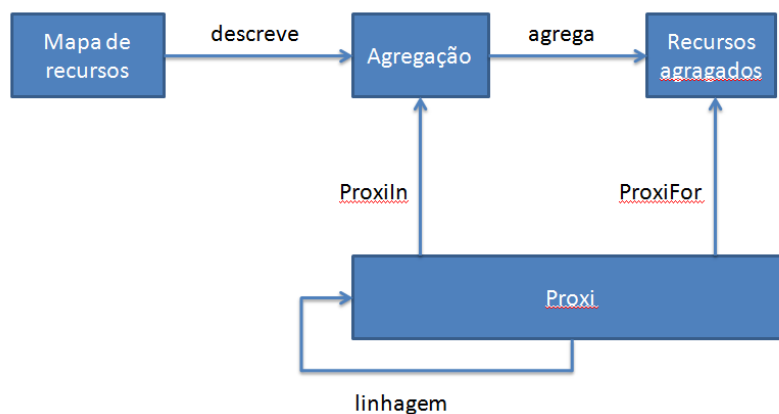
Afim de solucionar a interoperabilidade entre bancos de dados que armazenam recursos agregados, desenvolveu-se o Open Archives Initiative Object Reuse and Exchange (OAI-ORE). Este protocolo possibilita mapear e serializar as relações entre os recursos, de forma a manter as características dos recursos agregados.

Os conceitos envolvidos no OAI-ORE são os seguintes:

- agregação: é um conjunto de recurso identificado como uma unidade lógica;
- recurso agregado: é um dos recursos que compõem a agregação;
- mapa de recurso: é um recurso que contém as seguintes informações: a) descreve uma única agregação; b) enumera os recursos que compõem a agregação; e c) apresenta as propriedades da agregação e dos recursos agregados, tal como as relações entre os recursos.
- proxí: é um recurso que apresenta o contexto do recurso agregado para a agregação.

O mapa de recurso é expresso no formato de tripla RDF (Resource Description Framework), constituída por uma URI do tipo R, ou *URI Resource Map*, que descreve as relações entre os recursos. Com isso, possibilita apresentar semanticamente as relações entre os diversos recursos agregados que compõem uma agregação. A figura 10 apresenta um digrama da relação entre os conceitos do OAI-ORE.

FIGURA 10 - DIAGRAMA DAS RELAÇÕES ENTRE OS CONCEITOS DO OAI-ORE



FONTE: Autor, 2014.

O Proxi é uma instanciação de uma agregação e pode ter várias aninhadas, ofertando informações tanto a agregação, quanto aos recursos agregados, um contexto. Um mapa de recurso descreve as relações entre a agregação e seus recursos.

### 3.5.3. DADOS ABERTOS

O movimento de Acesso Aberto estabeleceu orientação para fomentar a difusão do conhecimento científico, principalmente para a disseminação dos resultados de pesquisa de forma gratuita. Entretanto, esse movimento atua, principalmente, no processo final da pesquisa, na disseminação dos resultados, em uma parte específica do ciclo informacional da comunicação científica, não atuando nas etapas intermediárias da produção do conhecimento.

Assim, os Dados Abertos expandem a abrangência dos Acesso Aberto, de forma a possibilitar o acesso dos dados produzidos nas etapas intermediárias de pesquisa facilmente, sem barreiras. Com isso fomentar o reuso de dados de pesquisa, tanto que Linde et al (2014) relatam que o acesso aos dados de pesquisa pode possibilitar um progresso mais rápido da Ciência, já que oferta aos pesquisadores uma quantidade grande de dados para re-análise, comparação, integração e validação.

Mesmo que tenha sido desenvolvido no âmbito dos dados governamentais abertos, as leis de Eaves (2009) possui grandes aplicabilidades aos dados científicos abertos, sendo:

- ❖ **Se o dado não pode ser encontrado e indexado na Web, ele não existe:** refere-se a questão da necessidade dos dados estarem disponíveis na web e de fácil recuperação, por exemplo, por ferramentas de busca como o Google. A necessidade de disseminar os dados, no caso, de pesquisa;
- ❖ **Se não estiver aberto e disponível em formato compreensível por máquina, ele não pode ser reaproveitado:** refere-se à questão do acesso ao dado, em que não deve estar criptografado ou ilegível por programas de computador, que aumenta substancialmente sua possibilidade de reuso;
- ❖ **Se algum dispositivo legal não permitir sua replicação, ele não é útil:** aqui se destaca a questão da licença do dado aberto, que deve explicitamente definir que tipo de ações é possível aplicar nos dados, devendo ser a mais ampla possível.



Gutierrez (2011) relaciona os Dados Abertos à produção e disseminação de dados em escala global, remetendo aos objetivos iniciais da Web. Nesse ponto, não se restringe a liberar o acesso ao dado, mas adota metodologias que facilitem a comunicação e a recuperação dos dados, revelando que o Movimento dos Dados Abertos é um movimento mais amplo, compondo-se de questões conceitos, produtos e serviços.

Refinando mais esse conceito, Piller (2014) considera que Dados Abertos se referem aos dados não confidenciais e sem características de privacidade produzidas por fundos públicos e que foram distribuídos sem restrição para uso. Para tanto o autor destaca que o dado aberto não pode ser confidencial, restrito ou privado e deve ser completo, primário, acessível online, processável por computador, não discriminatório, não proprietário e de licença aberta.

Essas definições alinham ao que o grupo de trabalho americano<sup>4</sup> discutiu sobre os dados abertos governamentais, definindo princípios orientadores para dados abertos. Assim, podem-se adaptar esses princípios de forma a apoiar a questão do que é e como devem ser disseminados esses dados. Assim, adaptando os preceitos ao novo contexto:

- ❖ **Completos:** Os dados de uma pesquisa devem ser disseminados na sua forma integral, eletronicamente, incluindo os dados primários e os resultados de processamento, se for o caso. Exceto àqueles que possuam caráter restritivo relacionado à confiabilidade entre pesquisado e pesquisador;
- ❖ **Acessíveis:** Os dados de pesquisa devem estar disponíveis na internet da forma mais ampla, sem a necessidade de identificação para acesso ou restrições licenciasais;
- ❖ **Processáveis por máquinas:** os dados devem ser estruturados, de forma a possibilitar processamentos automáticos;
- ❖ **Formatos restritos:** os dados estar preferencialmente disponíveis em formatos livres, sem codificações que restrinja o acesso ao seu conteúdo;

---

<sup>4</sup> Disponível em: <http://opengovernmentdata.org/>

- ❖ **Preservação dos direitos de autor:** estar livre de licenças restritivas não significa que não se deve manter os direitos de autor. Assim, deve-se manter questões de autoria e licença original.

Para o governo do Reino Unido, na publicação do *Livro Branco de Dados Abertos* (2012), os dados abertos, na concepção mais restrita do termo, tem o seu público alvo na comunidade científica. Mesmo compartilhando elementos com a transparência de dados, são conceitos distintos, em que um tem relação com dados governamentais e outros mais científicos.

#### **3.5.4. CIÊNCIA ABERTA**

A globalização do mundo, impulsionada pela internet, tem alterado consideravelmente a forma de se fazer pesquisa. O surgimento dos Dados Abertos, o Movimento do Acesso Aberto e dos Arquivos Abertos são exemplos dessas mudanças que tem elevando a pesquisa científica a outro patamar. Fortemente apoiada por ferramentas informatizadas eleva a pesquisa a uma nova dimensão, do compartilhamento, do reuso, colaboração, democratização do conhecimento científico, entre tantos outros aspectos que caracterizam a Ciência Aberta.

Com relação a todos os movimentos científicos voltados a abertura das informações produzidas no processo de pesquisa, Lima e Lima (2013) esquematizam esses movimentos como uma evolução, no qual a Ciência Aberta é o ponto culminante, porque agrega todas as ações dos movimentos anteriores, pertencendo a um mesmo contexto:

Fontes Abertas → Dados Abertos → Acesso Aberto → Ciência Aberta

Nesse sentido, Alexandre Hannud Abdo, cientista molecular e pesquisador da Faculdade de Medicina da Universidade de São Paulo (USP), em depoimento para Hourcade (2013) descreve o movimento da Ciência Aberta como uma evolução das práticas científicas com a adoção das novas tecnologias. Destacando dois pontos relacionados à Ciência Aberta, a questão das novas práticas e o uso da tecnologia, em que um ponto tem relação direta com o outro, no que as novas práticas são reflexos do uso da tecnologia.

Nesses sentido, como mudança de práticas, pode-se destacar a cooperação entre cientistas, em que a tecnologia elimina as barreiras geográficas, linguísticas e culturais. Cada vez é mais comum estudiosos trabalharem em colaboração separados por quilômetros de distância. Atualmente é raro ter pesquisas executadas por um único cientista, no que muitos estudiosos chamam de fim do autor único.

A questão do uso da tecnologia no processo de pesquisa está presente, em muitos casos, a todas as etapas, desde o início do estudo até a disseminação dos resultados. No caso da coleta e armazenamento dos dados, é muito comum o uso de ferramentas e bases de dados, ao ponto que aumentam a eficácia do processo e que podem disponibilizar os dados coletados. No caso da disseminação, tem-se repositórios que disponibilizam o pré-print e revistas eletrônicas e repositórios para o pós-print. Em ambos os casos, revelam uma mudança nas práticas científicas.

A liberação sem restrições de acesso aos dados, resultados e métodos de pesquisa possibilitam a verificação e validação do estudo, que aumenta a reputação dos autores juntos aos pares. Com isso, contribui com uma ciência mais ampla, mas robusta e rápida, visto que estamos em um momento tecnológico que requer transparência nos processos, que evita mal entendidos.

Numa abordagem econômica, Borgman (2007) destaca que a Ciência Aberta é baseada na premissa que a informação científica é um bem público, o qual se torna passível de compartilhamento sem que haja perda de valor e que não pode ser manter exclusivo. Essa afirmação impõe um caráter mais social à Ciência, visto fundamentar-se na questão da informação científica ser pública, pertencer a toda a sociedade sem ter uma exclusividade e nem por isso perder o seu valor.

A questão da informação científica como bem público revela questões mais complexas, pois tem relações com direitos autorais e patentes. Neste caso, Nelson (2006) comenta argumentando a questão da Ciência Aberta nas universidades e na indústria, que se uma universidade patenteia ou limita o acesso aos resultados de pesquisa não necessita ser financiada por dinheiro público. Assim, coloca a pesquisa científica fomentada por dinheiro público como geradora de bem público.

No entanto, Nilsen (2011) destaca que implementar a Ciência Aberta tornou-se um desafio, no que enfrenta questões relacionadas à cultura científica vigente, em que algumas áreas do conhecimento são mais tradicionais que outras. Disciplinas que fazem uso constante de tecnologia como as ciências rígidas aderem mais facilmente a movimentos amparados por tecnologia, no que as Humanidades, por exemplo, tradicionalmente aderem mais tardiamente. Com isso, as universidades geram grandes quantidades de informação científica, mas que nem sempre atingem o público alvo, sejam outros pesquisadores, indústrias, governo e comunidade.

Assim, a Ciência Aberta apresenta-se como uma ferramenta para a democratização da Ciência, na medida em que abrange questões tecnológicas, culturais, jurídicas e sociais. Na era da internet em que a colaboração e o compartilhamento se estabeleceram como *modus operandi*, a Ciência Aberta apresenta-se como uma questão irreversível, à qual a comunidade científica deve se adequar, independente da área do conhecimento.

### 3.5.5 CURADORIA DIGITAL

Atualmente, muitos estudos tem sua documentação e dados registrados no formato digital e processos informatizados tornam-se importante ferramentas de apoio à pesquisa. Como isso, tratar do gerenciamento de toda a informação digital gerada no processo de pesquisa apresenta-se como uma necessidade, na qual diferentes áreas do conhecimento precisam ser envolvidos, pois envolve toda uma adaptação nos processos de pesquisa e não apenas ofertar ferramentas informatizadas.

Nesse contexto, surge a Curadoria Digital, em que a Digital Curation Centre (DCC) define como um conjunto de procedimentos que envolve a manutenção, preservação e adição de valor aos dados digitais de pesquisa através de todo o ciclo de vida. Para o Centro, o ciclo de vida da Curadoria Digital compõe-se de:

- **Desenvolvimento:** etapa para conceber e criar objetos digitais, envolvendo métodos de captura e armazenamento;

- **Criação:** Etapa de produção dos objetos digitais e caracterizando-os com metadados descritivos, administrativos, estruturais entre outros.
- **Acesso e uso:** Ofertar aos usuários uma forma fácil de acesso aos objetos digitais de forma livre ou restrita;
- **Avaliação e seleção:** Avaliar os objetos digitais de forma a selecionar os que requerem curaço e preservação por longos períodos, alinhados com recomendações, questões legais, políticas e outros adotados pela instituição;
- **Descarte:** Utilizar formas de descarte de objetos digitais com base em recomendações, questões legais, políticas e outros adotados pela instituição;
- **Alimentação:** transferir objetos digitais para repositórios ou outro tipo de sistemas conforme as recomendações, questões legais, políticas e outros adotados pela instituição;
- **Preservação:** Empreender ações que assegurem a preservação dos objetos digitais, assim como as questões de autoria;
- **Reavaliação:** Por ser um ciclo, reavaliar os objetos digitais, a fim de validar as ações, visto que políticas, recomendações e outros aspectos são dinâmicos e os objetos digitais podem mudar o status com o tempo;
- **Armazenamento:** Adotar formas seguras de manter os objetos digitais conforme padrões relevantes;
- **Acesso e reuso:** Assegurar que os dados estarão disponíveis para os usuários, com o objetivo de uso e reuso, podendo estar abertos a todos ou restrito.
- **Transformação:** Criar outros formatos para o objeto digital

Curry, Freitas e O’Riáin (2010), discutindo sobre curadoria de dados, consideram a curadoria digital como um processo em que um conjunto de informações confiáveis sejam geridas em repositórios, de forma a proporcionar o uso corrente ou futuro, por diversos tipos de usuários, como os pesquisadores, historiadores, acadêmicos, entre outros. Para tanto, os autores elencam diversas atividades formando

um ciclo para a curadoria digital composto de: seleção, preservação, manutenção, coleção e arquivamento de ativos informacionais.

Para autores como Giaretta (2011), o termo mais amplo para curadoria digital seria preservação digital, ao ponto que a curadoria se estabeleceria na área de pesquisa e acadêmica, no ciclo de vida da informação científica, no que a preservação pode ser expandida para a área comercial ou jurídica, por exemplo. Neste ponto, revela-se certa distinção entre Curadoria e Preservação, no ponto de vista do autor, com a curadoria se restringindo ao contexto de acadêmico e de pesquisa.

McGinley (2007), discorrendo sobre questões legais para objetos digitais, afirma que a curadoria digital é ampla, semelhante a uma lei, em que as questões legais que envolvem a curadoria são dependentes de que estágio se opera, em que vários tipos de usuários acessam os objetos digitais curados, com níveis diferenciados. Autores diferem dos curadores, que por sua vez, diferem dos leitores (possíveis usuários da informação).

### **3.5.6. E-SCIENCE**

Muito provavelmente os livros de história do futuro, se realmente existirem livros de história no futuro como os conhecemos agora, apresentarão o momento atual como um período de adequação, em que a Tecnologia da Informação e Comunicação (TIC) é incorporada ao dia-a-dia das pessoas nas mais diversas atividades. Em muitos casos tornando os processos mais eficazes, no que se consegue fazer mais, melhor e em menos tempo. Com isso tem alterado muitos aspectos da sociedade.

Nesse contexto, surge o E-Science como um mundo novo à pesquisa, amparado pela TIC, revelando-se como uma revolução na forma de coletar, armazenar e processar grandes quantidades de dados, facilitando a geração de novos conhecimentos. A TIC possibilita que todo o processo de pesquisa possa ser compartilhado e que a contribuição possa ocorrer sem as barreiras físicas.

Segundo Hey e Hey (2006), O termo e-Science foi primeiramente utilizado por John Taylor para descrever um tipo distribuído e colaborativo de Ciência, em que muitas áreas da Ciência podem ser beneficiadas pelo uso de grandes infraestruturas de

TIC, como a poio a colaborações distribuídas e envolvendo múltiplas disciplinas. Assim, definiu o E-Science como: “E-Science refere-se a colaboração global em áreas-chave da Ciência e da possibilidade de uso da nova geração de infraestrutura”.

Para os mesmos autores, o E-Science não é uma nova disciplina científica, mas um conjunto de tecnologias que apoiam uma rede científica, uma infraestrutura global que potencializa a pesquisa, tornando os processos mais eficazes em diferentes formas. Apoia os cientistas nas tarefas de produzir novos conhecimentos com a oferta de funcionalidades cada vez melhores, com alta disponibilidade e poder de processamento e armazenamento.

Para Yang, Wang e Jie (2011) o E-Science é um processo inevitável na evolução tecnológica aplicada às Ciências. Baseados em níveis, o E-Science tem por base os avanços da TIC, no desenvolvimento do hardware e software que possibilitam avanços significativos no processamento de informação. Sob esse desenvolvimento, está à utilização de novos métodos de análise de dados, apoiado pela tecnologia. Tudo isso ampara uma Ciência mais social, cujos resultados podem focar problemas mais complexos e amplos, que envolvem mais que uma disciplina.

Sayão e Sales (2014), por sua vez, afirmam que o E-Science, com a possibilidade de tratar grandes quantidades de dados, altera os novos ambientes de pesquisa, afetando a gestão de pesquisa e promovendo mudanças significativas nos fundamentos da Ciência. Os autores alinham o pensamento que o E-Science é uma revolução na Ciência, já que alteram os processos, temas e a infraestrutura da pesquisa, criando possibilidades de pesquisa mais amplas.

De forma simplificada, após estudos conceituais sobre o E-Science, Medeiros e Caregnato (2012) compararam o E-Science com a pesquisa tradicional apresentados no Quadro 6. Revelam-se questões básicas como a eliminação das barreiras disciplinares e físicas para compor as equipes de pesquisa, o cuidado com a preservação e compartilhamento dos dados coletados e gerados na pesquisa, uso da tecnologia para processamento dos dados, uso intensivo da internet em todos as etapas da pesquisa.

QUADRO 6 - COMPARAÇÃO ENTRE O E-SCIENCE E A PESQUISA TRADICIONAL

<b>Característica</b>	<b>E-Science</b>	<b>Pesquisa tradicional</b>
Participantes	Diversamente qualificados, equipe de pesquisa distribuída	Pesquisador individual ou pequena equipe local de pesquisa
Dados	Gerados, armazenados e acessíveis de localizações distribuídas	Gerados, armazenados e acessíveis apenas localmente
Computação e instrumentação	Larga escala ou sob demanda ou acesso à informação compartilhada	Emprego da computação em lote, ou emprego do computador, ou instrumentos do próprio pesquisador
Rede	Confiança na internet e em mediadores	Não confiante na internet
Disseminação da pesquisa	Via websites e portais especializados	Via publicações impressas ou apresentações em conferências

Fonte Medeiros e Caregnato (2012).

A colaboração torna-se um aspecto fundamental ao E-Science, efetuada por meio da tecnologia. Assim, as ferramentas informatizadas tornam-se apoio à construção colaborativa do conhecimento científico, sem as usuais barreiras geográficas, disciplinares e linguísticas. Cada vez é mais comum que cientistas de origem de diversos países, áreas e idiomas trabalhem em um único projeto utilizando a TIC nas pesquisas.

No que diz respeito aos dados Grey (2009) destaca o envolvimento de três ações fundamentais: a coleta, a curadoria e a análise. A Coleta refere-se ao uso de tecnologia para obter os dados, seja por meio de simulação até uso de instrumentos de medidas acoplados às bases de dados, de forma a possibilitar o compartilhamento. A Curadoria trata, como visto na seção 3.5.5, da gestão dos dados em todo o seu fluxo, com ênfase na preservação. A Análise, por sua vez, envolve o uso de ferramentas informatizadas para extração de conhecimento dos dados existentes.

O uso intenso de computação em infraestrutura robusta e de alta-performance dá ao E-Science a eficácia desejada no processamento de grandes quantidades de dados. Incorpora ações voltadas ao Open Data e Big Data, da mesma forma que do Open Science, pois o uso da TIC, principalmente da Internet, são constantes. Com isso a colaboração e compartilhamento tornam possíveis o uso de ferramentas computacionais mais complexas.



A disseminação da informação via web democratiza o conhecimento, seja em canais mais tradicionais como os periódicos ou mais atuais como os repositórios e bancos de dados abertos. Com isso, a E-Science alinha-se a Ciência Aberta e outros canais menos tradicionais, como as redes e mídias sociais. Destaca-se que o E-Science é uma convergência de tendências científicas em que a tecnologia é o suporte aos processos científicos, criando novas práticas e mudando paradigmas.

### **3.6. REPOSITÓRIOS, REDES E FEDERAÇÕES**

A mudança de suporte do físico para o digital trouxe profundas mudanças, do acesso à forma de apresentação dos documentos. Muitos estudiosos comparam esse advento a outros acontecimentos importantes para a humanidade, como a invenção do papel ou o uso do tipo móvel. Assim, a tecnologia põe-se como um facilitador no processo de disseminação da informação.

Marcuschi (2003, PP. 9-12), numa análise sobre a relação de gêneros e suportes, fixa-se no formato textual para vincular um gênero literário a determinados tipos de documentos, e seus suportes. Dessa forma, confere ao suporte um significado que transcende o definido nos dicionários, ganhando um status de *locus* para registrar informações, tanto fisicamente como digitalmente. Assim, revela que se no suporte físico o gênero tem relação com o suporte, digitalmente, não há distinção.

Nesse sentido, revela-se, que como o papel é suporte para vários formatos de textos, o suporte digital, de forma ampla, também permite variados formatos. A mudança para o suporte digital independe do gênero. Um livro ou um jornal, classificado como gêneros distintos, podem ser transpostos para o suporte digital, sem grandes dificuldades, tornando-se documentos digitais sem que haja distinção.

Por outro lado, Bezerra (2010, p. 59) revela que a mudança para o suporte digital traz indício de mudanças nas práticas discursivas, pois o suporte interfere na forma de apresentação dos diversos gêneros, da forma que a informação é apresentada e recebida pelo leitor. Por esse motivo, permite a criação de novos gêneros, principalmente, com as possibilidades do hipertexto e das multimodalidades.

Dessa forma, se independente do gênero, todos são documentos digitais, o conteúdo da prática discursiva muda, mesmo para um mesmo gênero. Isso implica que um livro digital pode apresentar diferenças do original em papel. A hipertextualização e uso de diversas mídias permitem ao suporte digital prover facilidades impossíveis ao suporte físico. Assim, a tecnologia influencia a prática discursiva.

Dessa mesma forma, a mudança do suporte afetou não apenas a questão do formato ou forma de organizar os documentos, mas o seu acesso. A internet, com as sua oferta de informação digital afetou a forma de acesso aos documentos digitais.

Levacov (1997, p. 9), numa discussão sobre a evolução tecnológica no âmbito das bibliotecas, aponta que a mudança de suporte afetou as bibliotecas de várias formas. Dentre essas mudanças, ressalta a questão do paradigma da propriedade para o paradigma do acesso. Nesse sentido a biblioteca não se restringe a possuir o acervo e permitir o acesso local, mas pode oferecer acesso, independentemente do local, a um acervo físico e virtual.

Assim, no que se refere aos acervos virtuais, em conteúdo completo no formato digital, um dos sistemas possíveis de ofertar acesso, entre outros serviços, são os repositórios. Por isso, muitas bibliotecas universitárias ou de centros de pesquisas tem implementado esses sistemas como forma de facilitar o acesso à produção intelectual da instituição.

### **3.6.1. REPOSITÓRIO**

Os repositórios, como é visto atualmente, nasceram no âmbito do movimento de acesso aberto à informação científica. Nesse sentido, a Declaração de Budapest<sup>5</sup> e em 2002, apresenta os repositórios como uma estratégia para a implementação do acesso aberto, em que os autores são estimulados a autoarquivar sua produção científica em repositórios institucionais, como uma forma de facilitar a disseminação da informação científica.

---

<sup>5</sup> Disponível online em <http://www.opensocietyfoundations.org/openaccess/read>

Nesse sentido, Costa (2006, pp. 40-49) revela que a implantação dos acesso aberto impacta a comunidade científica, principalmente no processo de comunicação científica. Discute os impactos nos autores e as instituições vinculadas, os editores científicos e editoras comerciais e agências de fomento. Segundo a autora, a adoção plena do acesso aberto é iminente, mesmo com as diferenças disciplinares em relação à adoção de novas tecnologias.

Revela-se que os repositórios, ao contribuírem para facilitar o acesso à produção científica, aumentam a visibilidade da própria produção do autor, da instituição à qual o autor está vinculado, diretamente, e outras entidades, indiretamente. Assim, impactam o cerne da comunidade científica, ou seja, na produção do conhecimento. Visto que, ao facilitar a disseminação, aumentam a oferta, que serve de insumo à produção de novos conhecimentos.

Em estudo sobre o impacto do acesso aberto, Hanard et al (2004, p.5) associam o autoarquivamento nos repositórios ao sinal verde dos editores ao depósito, assim denominando os repositórios de via verde, diferenciando dos periódicos de acesso aberto, via dourada. Essa denominação faz menção à liberdade, sinal verde, ao depósito e acesso ao conteúdo completo.

Nesse ponto, um periódico de acesso aberto é denominado de via dourada ao conhecimento científico. No que, os repositórios, por conter uma cópia dos artigos publicados anteriormente, necessitam requerer o direito de dar acesso aos artigos, dado pelos editores. Assim, os repositórios podem conter artigos restritos, embargados e de acesso livre.

Nem todo documento depositado em repositório possui o acesso irrestrito. Muitos repositórios implementam políticas de acesso, que restringem ou impossibilitam o acesso a documentos sem o aval dos editores. Assim, reforçam a denominação de via verde, visto que embargam o acesso aos documentos em fase de negociação ou pelo tempo requerido pelo editor.

O depósito, então, não implica em dar acesso. Assim, um repositório institucional pode ser o grande agregador da produção intelectual da instituição,

principalmente da produção científica. Assim, revistas e repositórios promovem o acesso ao artigo, mas os repositórios podem fornecer a preservação do acesso ao documento.

Nesse ponto, Weitzel (2006 p.143), associa esse conceito de via dourada e via verde, a questão das fontes primárias e fontes secundárias da comunicação científica. Pondo, nesse sentido, os periódicos eletrônicos como fonte primária, como já é amplamente aceito, e os repositórios como fontes secundárias, uma republicação da documentação, anteriormente publicada nos periódicos.

Essa associação de fontes primárias e secundárias aos periódicos científicos e repositórios, moderniza o conceito, colocando-o no momento atual, em que o acesso digital oferta novas facilidades. Assim, os repositórios, como fonte secundárias, não se relacionam a um papel secundário, mas apenas revelam-se como uma outra possibilidade de acesso além da publicação original.

Nesse contexto, os repositórios, como apresentado por Bjork (2007, pp. 29-30), no modelo de comunicação científica, apresenta-se como um facilitador de acesso à informação científica, assegurando um acervo de documentos digitais para acesso amplo. Assim, compor um acervo com diversas tipologias documentais, que atendam as variações disciplinares em relação às fontes preferenciais de publicação e uso.

Da mesma forma, revela que os repositórios tornam-se fonte importante à produção de novos conhecimentos. Ao contemplar uma diversidade de tipologia documental, atende as peculiaridades das disciplinas, assim como, pode conter bases de dados de pesquisa, que apoiariam a reutilização, para comprovação de resultados ou novos estudos.

Segundo Crow (2002, p. 4), os repositórios podem definir a tipologia documental a compor o acervo, podendo ser exclusivamente científico ou estendido para outros tipos de documentação, como objetos de ensino, relatórios e outros. Independentemente da tipologia, o autor destaca que o repositório tem a finalidade de dar acesso à informação, podendo coletar, preservar e disseminar o conteúdo mantido no acervo.

Destaca-se, assim, a quantidade de finalidades possíveis aos repositórios, além da disseminação. Criando um fluxo informacional diferenciado, em que além da seleção, tratamento técnico e disseminação, não ocorre o desbaste e pode não ocorrer o descarte. A coleta e preservação tornam-se processos adicionais aos repositórios para apoiar a disseminação.

Para ofertar essas facilidades, Camargo e Vidotti (2009, p. 55) apontam que os repositórios são constituídos por sistemas de informação, assim inserindo-os em ambiente informacional e questões da arquitetura da informação. Dessa forma, apresentando os critérios de usabilidade, qualidade e acessibilidade nos repositórios, além dos próprios para repositórios.

Revela-se, assim, a flexibilidade dos repositórios em atender os critérios que os tornam mais eficazes. Acentua-se a questão da qualidade dos serviços ofertados pelos repositórios. Apresentando um sistema de informação, que ofereça um diferencial, no que concerne ao atendimento às necessidades informacionais dos usuários.

É importante mencionar que o público alvo do repositório tem grande influência nos critérios avaliadores dos repositórios. A orientação na forma de ofertar as facilidades, usabilidade e acessibilidade tem relação direta com esse público. Por exemplo, mesmo para um repositório científico, um repositório para informáticos tem diferenciação de um repositório para filósofos.

Nesse ponto, os repositórios, também podem ofertar serviços diferenciados dependendo do conteúdo a ser gerenciado. Um repositório tipicamente científico ofertará serviços diferentes de repositório com apenas dados de pesquisa. Da mesma forma, repositórios institucionais que preservam a memória técnica da instituição terá características diferenciadas .

No contexto da preservação da memória institucional, por meio da preservação da produção intelectual de uma instituição, Dodebei (2009, pp 104 e 105) aponta os repositórios institucionais como um modelo híbrido que contempla memória e transferência. Assim, ressaltando duas características de preservação e acesso, mas com

a visão da manutenção da memória organizacional. Amplia-se, nesse sentido, a abrangência dos repositórios.

É necessário, no entanto, observar que a preservação ofertada pelos repositórios refere-se a lógica, ou seja, relacionado a preservação do acesso. No caso de preservação, como explica Ferreira (2006, pp 31-38), a preservação digital possui vários níveis, sendo todos baseados na preservação física, que no caso dos repositórios, enquadra-se na preservação da integridade e acesso, com estratégia de normalização, que reduz os formatos para facilitar processos necessários à preservação física e lógica.

Esse ponto tem dois aspectos, a questão do usuário e a questão mais técnica. Preservação tecnicamente tem relações com os procedimentos que atenderão aos usuários. Manutenção da base de dados e outros procedimentos com características mais tecnológicas. Os aspectos mais tecnológicos dependem da implementação, tecnologias utilizadas, até mesmo políticas tecnológicas implementadas pela instituição.

Pelo viés dos usuários, os repositórios caracterizam-se pelos serviços ofertados e pelo uso. Características tecnológicas, no entanto, são observáveis apenas como reflexos dos serviços, tais como a interoperabilidade, que permite o intercâmbio de informação entre sistemas de informação.

### **3.6.2. INTEROPERABILIDADE**

A interoperabilidade tem raízes nos arquivos abertos, possuindo aspectos mais tecnológicos sobre o intercâmbio de informações entre sistemas de informação. Nesse sentido, Triska e Café (2001, p. 92) revelam que a interoperabilidade, juntamente com o autoarquivamento e a avaliação pela comunidade, formam os pilares dos arquivos abertos. Constitue-se basicamente do processo de troca de informações entre sistemas heterogêneos, utilizando um protocolo comum, sendo que as informações são codificadas no formato XML num padrão comum, ou seja, em um esquema de metadados adotado por ambos.

Nesse sentido, nota-se que o autoarquivamento é um procedimento a ser adotado. O depósito de material publicado anteriormente, em repositórios, pelo próprio autor, é um procedimento desejável, mas que não impede o processo de alimentar um

repositório de arquivos abertos, visto que em muitas instituições, a biblioteca fica com esse processo.

Da mesma forma, a avaliação pela comunidade é um procedimento desejável aos preceitos dos arquivos abertos. Esse procedimento parece ser mais importante para repositórios de pré-prints, pois nos repositórios científicos, todos os documentos depositados foram avaliados pelos pares.

Quanto à interoperabilidade, é o único processo tecnológico, que envolve mais de um sistema simultaneamente. Difere, dessa forma, dos outros pilares, principalmente, por ter tarefas internas, automatizáveis, e por ser requerido para a implantação de sistemas de arquivos abertos.

Por esse motivo, a interoperabilidade apresenta um modelo de implementação, pois, sendo um processo, pode ser implementado de várias formas. Requer-se, também requisitos que padronizem as tarefas ou etapas, as recomendações. Assim, sistemas heterogêneos podem implantar a interoperabilidade.

No que concerne as recomendações e mecanismos dos arquivos abertos, van de Sompel e Lagoze (2000) apresentam dois aspectos, sobre o modelo de interoperabilidade e sobre a coleta automática de metadados (harvesting). Esses dois aspectos revelam a estrutura na qual a interoperabilidade torna-se possível.

O modelo de interoperabilidade, apresenta-se estruturado em camadas, a camada de serviço, metadado e documento. Para as questões de harvesting, apresentam o uso de um conjunto de metadados, um protocolo de comunicação comum e o uso do formato XML. Esses aspectos permitem o mecanismo automático da interoperabilidade.

No que se trata de interconexão de sistemas de informação heterogêneos, Marcondes e Sayão (2002, p. 50) revelam que a interoperabilidade possui várias faces. Apresentam aspectos que tornam possível o compartilhamento de informações, para gerar novos serviços e agregar valor a informação consolidada.

Da mesma forma, Miller (2000), apresenta os principais aspectos da interoperabilidade. Revelando que para a interoperabilidade se constitua, requer mais

que a infraestrutura tecnológica (interoperabilidade tecnológica), como destacada nos arquivos abertos. Requer outras ações, como apresentadas pelo referido autor:

- **A interoperabilidade técnica**, que assegura, as questões de interconexão entre os sistemas, segurança da comunicação e meio de acesso, entre outros, refere-se as questões técnicas que aplicam sobre as questões tecnológicas;
- **A interoperabilidade política**, por sua vez, refere-se aos acordos formalizados para que a interoperabilidade se constitua. Da mesma forma que a gestão, a articulação, a coordenação e a cooperação que afetam a iniciativa que implementa a interoperabilidade e as instituições participantes;
- **A interoperabilidade semântica**, refere-se às questões de interpretação dos dados interoperados, uma tentativa de tornar mais eficiente o processo de troca de informação, proporcionar meios para que a informação seja interpretada no destino e fonte da mesma forma;
- **Interoperabilidade jurídica** refere-se aos vários aspectos relacionados às licenças e direitos autorais. A interoperabilidade, se verificada com potencial mundial, tem implicações legais internacionais, além da visão local.

Nesse âmbito, a interoperabilidade revela questões mais amplas, que envolvem vários aspectos, que não somente a questão tecnológica. Revelando uma complexidade maior, envolvendo outros elementos importantes na plena constituição da interoperabilidade. Não apenas para a troca de informação, mas para a prestação de serviços, também.

No que concerne às facilidades ofertadas pela interoperabilidade no âmbito dos estudos sobre a segurança e saúde do trabalho, Santos e Miraglia (2009, p. 93) revelam que essas facilidades são instrumentos preciosos ao desenvolvimento organizacional, alinhando às questões da Gestão da Informação. O intercâmbio de informação, temática ou não, aumenta a oferta de informação, que contribui para o aprendizado organizacional.



Ressalta-se, por fim, que a razão da interoperabilidade é consolidar a troca de informação entre uma rede ou federação. Apresenta-se como o canal que possibilita a interconexão da rede, dos aspectos políticos aos físicos. Assim, compõe-se como um dos elementos do cerne das redes ou federações.

### **3.6.3. REDES**

O termo rede tem origem no vocábulo *rete*, que em latim significa teia, entrelaçamento de fios, ou seja, rede, no sentido do objeto físico. A ideia de ligação entre nós, por meio de um elemento conectivo, transcendeu o conceito físico para denotar relacionamento entre indivíduos, instituições, entidades, sistemas e outros, em que, nem sempre há uma conexão física.

Nesse sentido, a palavra rede torna-se base para uma variedade de termos. Tomando aqui a composição da terminologia apresentada por Faulstich (2003, p.14 ), em que um termo geral, ou base, ganha especificidade com o acréscimo de qualificadores, mas que mantém a característica original, apenas restringindo o campo semântico.

Assim, há uma infinidade de termos com a base rede, em que há uma conexão física entre os elementos que compõem a rede, ou nós. Rede elétrica, telefônica, de gás, água, esgoto e outras, são exemplos de redes com conexão física. Em outros casos, as redes possuem elementos físicos não conectados, tais como redes de postos de gasolina, supermercados e outros, mas que mesmo assim, compõem-se em redes.

Rede de computadores, por exemplo, possui uma forte abordagem física. Por ser estruturante e ter por finalidade criar condições técnicas e físicas para permitir a comunicação entre computadores e, assim, compartilhar recursos. Assim, para que uma rede de computadores se realize será necessário uma estrutura física, computadores com canais interligados, e um estrutura lógica, programas e outros.

Nesse sentido, Gallo e Hancock (2003, p. 198) conceituam rede de computadores como "um conjunto de computadores e dispositivos que usam um mesmo protocolo de rede para compartilhar recursos através de um meio de rede". Acentua-se,

três pontos na definição, os elementos de rede (computadores ou dispositivos), o meio (protocolo e canal) e o objetivo (compartilhamento de recursos).

Revela-se que os pontos levantados são constantes em diferentes tipos de rede. Para construir uma rede requer-se que os elementos compartilhem de um meio comum, que permita a comunicação entre eles. Que a estrutura dê suporte a comunicação entre os elementos para atender a um objetivo.

Em um âmbito organizacional, O'Brien e Marakas (2013, p. 206), de forma ampla, consideram que nas redes referem-se a interconexão ou inter-relacionamento de sistemas ou entidades. Revela-se que rede, podem ocorrer dois tipos de atividades, a conexão ou o relacionamento entre os elementos da rede, ampliando o entendimento de redes.

Numa visão mais voltada a economia, Castells (2002) discorre sobre a relação do Ser e da Organização com as redes, destacando uma mudança. Nesse sentido o Ser se relaciona com a rede de forma a criar identidades coletivas baseadas nos conceitos compartilhados, ao passo que as organizações são alteradas pela rede, na medida em que altera a forma de estabelecimento e produção. O referido autor revela a importância das redes amparadas pela tecnologia na evolução da sociedade.

Nesse ponto, mesmo não mencionando explicitamente os objetivos, expande-se a visão dos elementos e do meio. No que se refere a entidade, destaca que podem não ser entidades físicas, mas podem ser sistemas, que incutem uma complexidade maior às redes, visto que sistemas apresentam uma gama maior de possibilidades. Na visão administrativa, sistema para Chiavenato (2005, p. 52) é "um conjunto integrador de partes, íntima e dinamicamente relacionados, que desenvolve uma atividade ou função e é destinado a atingir um objetivo específico".

Visto assim, um sistema apresenta a complexidade por contemplar mais que um elemento unitário. Imprime a rede a possibilidade de agregar sistemas, que possuem finalidades, mas que devem ter relação com os objetivos da rede. Assim, revela certo alinhamento de objetivos dos sistemas com os objetivos da rede.

No que concerne ao meio, apresenta as possibilidades de conexão ou relação. Revela-se graus diferenciados quanto a forma em que os elementos formam a rede. Um conexão apresenta-se com maior formalidade, enquanto a relação possui uma formalidade mais tênue. Essas possibilidades dão as redes uma flexibilidade maior em sua composição.

Para as redes sociais, Wasserman e Faust (1994, p. 20) definem redes sociais como um conjunto de atores e suas relações. Note que para a definição, atores contempla desde indivíduos até corporações, enquanto relações são enlaces entre os atores. Assim, uma rede social pode relacionar entidades e suas relações, tais como suas conexões ou atividades.

Nesse ponto, os elementos, denominados de atores, assim como na visão organizacional, podem assumir proporções diferenciadas, em uma linha contínua do individual ao coletivo, em suas mais diversos níveis. Quanto ao meio, revela uma nova possibilidade, a atividade como agregadora dos elementos. A atividade relaciona-se com os objetivos, presentes nesse caso pelo meio, sendo um meio termo entre a conexão e a relação.

Num aspecto mais pessoal, Falcão (2006, p. 328) apresenta uma questão próxima a rede social, mas com uma especificidade maior, a Rede de Apoio, que se sustenta em estrutura e funcionamento. Essa especificidade da rede social, tem por objetivo assegurar o fornecimento de apoio. Revela-se, assim, a questão das redes em assegurar a oferta de serviços.

Revela-se que em determinado tipo de rede a finalidade pode assumir, não o compartilhamento de recursos, mas a oferta de recurso. Coloca-se os objetivos de rede, também, com uma variedade de opções, numa abordagem mais complexa em que todas as características podem assumir diferentes níveis.

Retornando, a um ponto mais complexo, refere-se a redes de sistemas, principalmente, por integrar sistemas, que por si, engloba certa complexidade. Numa explicação tipicamente filosófica, Brito (2002, pp. 138-139) revela que os sistemas compreendem elementos e relações, de forma que torna-se uma rede. Assim, uma rede de sistemas é composta por várias redes interligadas, numa complexidade infinita.

Entretanto de forma mais prática, pode-se compreender que as redes de sistemas compõem-se de elementos complexos, que em cada nó há um conjunto de facilidades a serem compartilhadas. Da mesma forma que a visão filosófica, sistemas se compõem de elementos e relações para oferta de facilidades aos usuários. Dessa forma, sistemas em rede expandem seus domínios no que se refere a utilização das facilidade de uma rede.

Deve-se também ressaltar a questão da Internet, que compõe-se de uma rede de sistemas heterogêneos. Oferta facilidade para conectar sistemas diferentes, cada qual com a sua complexidade, estrutura, serviços, papel, sistemas operacionais e outros. Tim Benners-Lee, criador da Web, desenvolveu-o para ser autogerenciável e facilitar a troca de informação, colocando outras questões, como dos direitos autorais, em segundo plano.

Outro ponto a ser relevado em redes de sistemas heterogêneos, refere-se a processamento distribuído, que compõe um rede para compartilhamento de recursos, afim de permitir processamento paralelo. Nessas redes, recursos como a oferta de processamento e memória é compartilhada, cada elemento se torna parte de um todo, para processamento de grandes quantidades de dados ou processamentos muito complexos.

Nesse contexto geral, pode-se sintetizar as características da rede pelos seus componentes: os elementos, os enlaces ou relacionamento, o meio e o compartilhamento de recursos. Ressalta-se, porém, que em rede ocorre certa integração entre os componentes, em que os enlaces ou relacionamentos, pelo meio, regram o compartilhamento de recursos entre os elementos.

#### **3.6.4. FEDERAÇÕES**

A federação é um termo muito difundido nas Ciências Políticas, sendo a origem do termo. Etimologicamente tem origem no latim, na palavra *foederatio*, nominalização do verbo *foederare*, que significa ligar, unir, fazer alianças. Esse termo é, particularmente, utilizado nas Ciências Políticas relacionado à forma de governo federativa.

Nesse ponto, federação, no sentido das Ciências Políticas como governo, é considerada como um meio termo entre estado unitário e a confederação, como relata Rabat (2002, p.3), quanto à divisão dos poderes. Assim, mantém-se a soberania de um estado federado, mantendo a autonomia dos seus partícipes.

Nesse ponto, agrega entidades que possuem certa autonomia, mas que possuem forte relação com uma entidade superior. A autonomia dá-se por possuir sua própria gestão, ou governo próprio, independente do governo central, mas que devem obediência a uma constituição, que orienta o governo de cada unidade.

Discutindo as características da autonomia territorial, Suksi (2001, p. 107) relata que a federação é como um arranjo de duas ou mais entidades em um modelo constitucional. Nesse caso, cada entidade mantém seus poderes, mas um poder central possui poderes sobre as entidades. Nesse ponto, torna-se um modelo de camadas, como nos modelos de Federal e Estadual.

Reforça-se a característica de independência, mas com vínculo a uma entidade superior. A existência de um conjunto de regras que orienta a formação das federações revela uma formalização na sua constituição, apresentado tanto para as questões filosóficas quanto para as questões físicas.

Na Ciência da Computação, por sua vez, federação também é amplamente discutida, com vieses diferenciados e em vários contextos. Revela-se o extenso campo de estudo e aplicação da Ciência da Computação, que interage, como área meio, com várias outras disciplinas, pela tecnicidade apresentada, intrinsecamente, pela própria Ciência da Computação.

No que se refere a identificação de usuários e servidores em redes, Benantar (2006, p. 48) relata que a federação é responsável pela resolução de nomes e identificados em uma rede distribuída. Insere-se, a federação no bojo de serviço, ou de serviços federados, que apoiam o funcionamento das redes.

Outro ponto importante revela-se na colocação que a federação não se opõe a rede, mas se integra, como um serviço de rede. Apresenta-se uma relação de pertinência, em que a rede revela-se uma entidade maior, com aspectos mais teóricos,

enquanto a federação possui aspectos mais práticos, fortemente relacionados aos serviços.

No âmbito da segurança da informação, Paulus, Pohlmann e Reimer (2004, p. 90) afirmam que federação gera uma identidade federativa, uma integração para compartilhamento, mas mantendo a autonomia em sistema de parceria. Assim, estende a questão da pura identificação para uma contexto de gerenciamento de permissões de acesso aos recursos compartilhados.

Evidencia-se a questão de mais um serviço federado, da identificação e controle de permissão. Revela-se a questão da identidade federada e o controle aos recursos compartilhados em federação. Apresenta-se que em federação pode haver tipos diferentes de usuários, assim como níveis diferentes de permissão de acesso, aos recursos compartilhados pela federação.

Assim, para redes mais complexas, que são compostas por múltiplos domínios, os serviços de federação para identificação e gerenciamento de permissão de acesso aos recursos são primordiais. Minoli (2003, p. 115) considera as federações como comunidades que cooperam para identificar os domínios e gerenciar a conexão.

Da mesma forma, em *Reference Model of Open Distributed Processing* (RM-ODP) federação equivale a comunidades. Segundo Putmann (2001, p. 602), federação é como uma comunidade que contempla vários domínios, que se juntam para compartilhar recursos, mantendo sua autonomia sobre os recursos. Assim, têm como objetivo a cooperação para compartilhamento de recursos.

Pode-se resumir que para a Ciência da Computação, as federações equivalem a comunidades que operam para compartilhamento de recursos e oferta de serviços. Revela-se a cooperação entre os elementos envolvidos na federação, visto que cada elemento possui uma autonomia para o gerenciamento dos recursos. Por fim que federação relaciona a serviços ofertados em rede.

No que concerne à Ciência da Informação, Weitzel (2006 p.143) considera as federações como fontes terciárias. Em uma visão mais contextualizada com o movimento de acesso aberto, as revistas eletrônicas publicam como fonte primária, os

repositórios mantêm uma cópia, como fonte secundária, e as federações, coletando dos repositórios ou revistas, tornam-se fontes terciárias.

Nessa visão, os recursos a serem compartilhados, com ou sem controle de acesso, são os artigos científicos. Aproximando as federações dos sistemas de informação, que ofertam serviços referentes a um acervo coletado de fontes primárias ou secundárias. Portanto, apresenta-se, também, como prestação de serviço.

No mesmo sentido, federação está intimamente relacionado a sistemas federados, contemplando bases de dados heterogêneas, que, por meio de esquemas federados, integrando informações oriundas de outras bases de dados (NICOLLI, SIMON e YÉTONGNON, 2005, pp. 1651-1652). Nota-se que federação transcende de serviços federados para sistemas federados, escala que aumenta a complexidade e oferta de serviços.

Em visão mais global, envolvendo a Ciência da Informação e a Ciência da computação, na discussão de sistemas colaborativos, ElMaraghy (2006) declaram que, federação de repositórios é um *locus* que armazena informações, apoiando o desenvolvimento e distribuição de produtos informatizados. Essa assertiva, dá margem a dois entendimentos, um mais informático e outro mais para informação.

Para a visão mais informática da proposta acima, tem-se as federações de repositórios como os sistemas que armazenam e distribuem programas. Esses sistemas federados catalisam a distribuição, tanto para instalação, quanto para atualização, visto que atualmente pode-se comprar ou baixar um software ou atualização de um software de forma online, desses sistemas federados.

A visão da Ciência da Informação, por sua vez, revela sistemas que agregam informações oriundas de sistemas de informação distintos. Implica a questão da interoperabilidade do Arquivo Aberto, para o compartilhamento de informação, que em muitos casos, está no formato do conteúdo integral, seja texto ou outro formato de arquivo.

Em suma, as federações, assim como as redes, tem por finalidade o compartilhamento de recursos. No entanto, as federações de repositórios, na visão da

Ciência da Informação, com relação mais próxima a oferta de serviços, aproxima-se da oferta de compartilhamento e controle de acesso de informação, aproximando-o aos sistemas de informação.

### **3.6. CONCLUSÃO DO LEVANTAMENTO DE LITERATURA**

A revisão de literatura apresentada por este estudo fundamenta a pesquisa de modo a apoiar a discussão sobre o tema. Estrutura-se para atender os objetivos propostos, de forma a cobrir todos os aspectos requeridos a formalizar a identificação, análise e proposição de um modelo para federações de repositórios científicos, destacando as tendências tecnológicas no desenvolvimento de sistemas informatizados e tendências nas Ciências. Dessa forma, oferta subsídios a identificação, análise e criação de um modelo atual mais robusto, com possibilidade de maior durabilidade, mesmo com a dinamicidade imposta pela tecnologia.

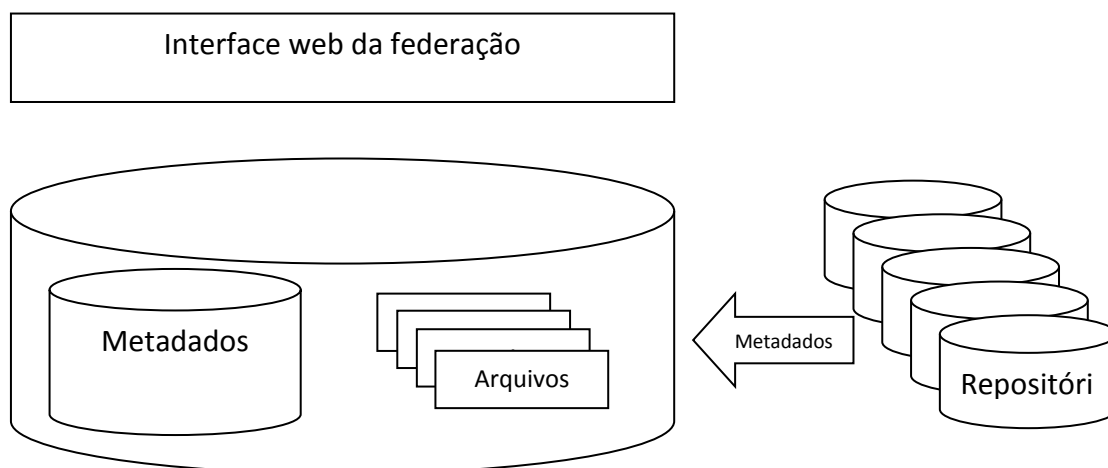
O levantamento de literatura fundamentou-se na tríade formada por Bancos de dados, Tendências Tecnológicas e Tendências científicas para aportar a identificação, análise e proposição de um modelo para federação. Com isso procurou-se abranger todos os aspectos necessários para alcançar os objetivos propostos.

Banco de dados entendido no âmbito da Ciência da Informação, como um serviço de armazenamento e recuperação de informação. Mesmo que o termo tenha acepções próximas em Ciência da Informação e Ciência da Computação nota-se certa complementaridade, enquanto a CI tem ênfase no serviço ofertado a CC restringe-se na estrutura e no funcionamento dos bancos de dados.

Nesse sentido, a federação de repositórios científicos pode ser consideradas como bancos de dados, que podem ser estruturados, se coletar apenas metadados, ou semi-estruturados, caso colem e armazenem arquivos digitais. Assim, possuem uma interface que oferta serviços informacionais e uma base de dados composta por metadados e arquivos, coletados de repositórios, com esquematiza a figura 11.



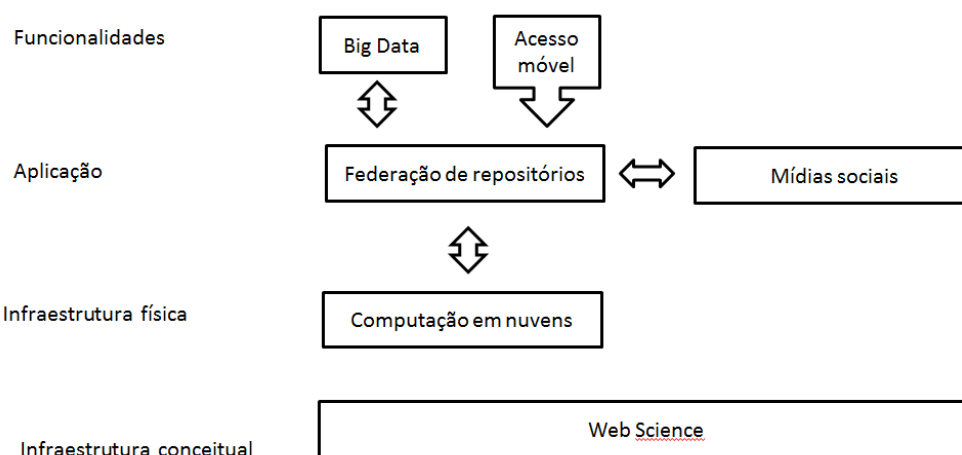
FIGURA 11 - ESQUEMA REPRESENTATIVO DA FEDERAÇÃO DE REPOSITÓRIOS CIENTÍFICOS



Fonte: Autor

Tendências tecnológicas, por sua vez, fornecem estrutura, relação e funcionalidade ao repositório. Pode-se esquematizar a relação entre as tendências tecnológicas e a federação conforme a figura 12, que apresenta a computação em nuvem ofertando armazenamento à federação; o Big Data sendo uma funcionalidade a mais; a federação sendo acessível por dispositivos móveis ou não, e tendo relação com as mídias sociais.

FIGURA 12 - RELAÇÃO ENTRE AS TENDÊNCIAS TECNOLÓGICAS E A FEDERAÇÃO DE REPOSITÓRIOS



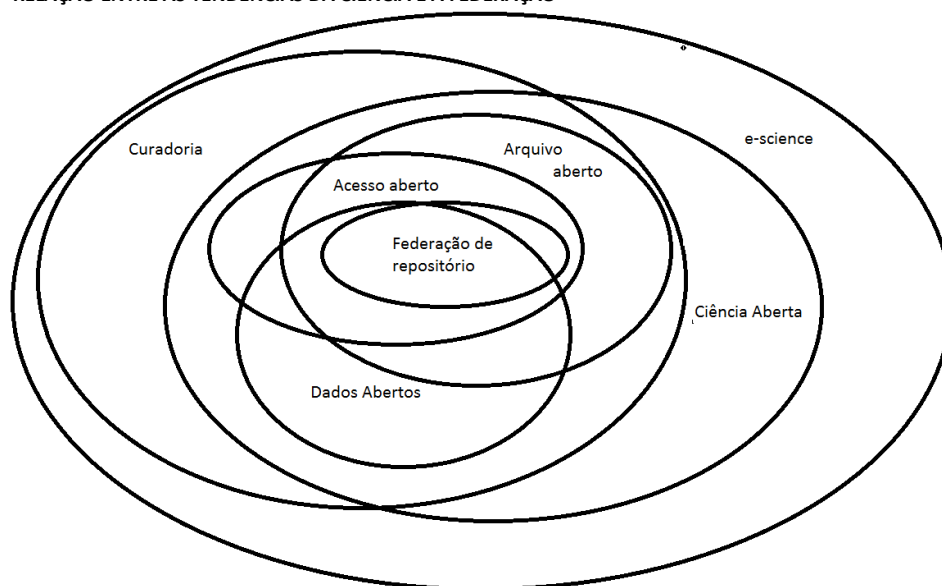
Fonte: Autor

A representação da figura 12 descreve a Web Science como base conceitual a outras tendências tecnológicas, atendendo estudos a toda a web. A computação em nuvens oferta infraestrutura aos repositórios, sendo implementada com ferramentas próprias e independentes da federação. As mídias sociais são sistemas de informação independentes no mesmo nível das federações, mas que podem intercambiar informações. O acesso móvel e o big data estão ao nível de aplicação, sendo que o acesso móvel pode ser aplicado ou integrado, como nos casos dos aplicativos, mas o big data é independente, uma adição de funcionalidades.

Nota-se que a web science ampara as outras tendências, por ter um aspecto mais conceitual, no que as outras tendências são mais práticas. A Computação em nuvens é infraestrutura, fornecendo um ambiente computacional seguro à federação e o Big Data um funcionalidade que pode ser integrado à federação, para ofertar serviços. Quanto às mídias sociais, são portais independentes que podem ser integrados às federações por web-services ou outros aplicativos. O acesso por dispositivos móveis é possível por tecnologia responsiva. Portanto, as federações podem fazer uso das tendências tecnológicas como forma de atender às necessidades atuais, tanto de infraestrutura, quanto de facilidades ofertadas.

As tendências da ciência, ao contrário das tendências tecnológicas, orientam o desenvolvimento ou agregação de funcionalidades à federação. Assim, em questão conceitual pode-se esquematizar a relação entre a federação e as tendências da ciência conforme um diagrama de Venn, no que possuem características de pertinência e intersecção, como apresentado na figura 13.

FIGURA 13 - RELAÇÃO ENTRE AS TENDÊNCIAS DA CIÊNCIA E A FEDERAÇÃO



Fonte: o autor

Nota-se que a federação é uma iniciativa fortemente vinculada ao acesso aberto, que por sua vez possui muito em comum com os arquivos abertos. Entretanto, nem todo repositório hospeda dados abertos. Por isso, a configuração inicial em que o repositório está contido totalmente no acesso aberto e nos arquivos abertos, mas apenas parcialmente nos dados abertos.

Dados Abertos, Acesso Aberto e Arquivos Abertos estão contidos na Ciência aberta e na Curadoria Digital. Entretanto, Ciência Aberta e Curadoria Digital, possuem características que as distinguem. De forma geral, pode-se apresentar que a e-science engloba todas as outras tendências, visto que todas possuem forte utilização de tecnologia, dados e informações no formato digital. Assim, o repositório é centralizado para o presente estudo.

Com isso, oferta-se um esquema conceitual, com todos os aspectos levantados na literatura, relacionando-os entre si e com os objetivos. Com isso, contribui-se para a discussão de um tema atual, criando um referencial teórico apropriado ao presente estudo.

## 4. METODOLOGIA

---

A federação de repositórios é um objeto de estudo típico da Ciência da Informação, que envolve muitos tópicos de interesse desta disciplina, principalmente relacionados à comunicação científica. Mesmo sendo associado à Ciência da Informação pode apresentar-se mais amplo no que diz respeito às possibilidades de estudo, tendo relações com a Ciência da Computação e outras áreas.

Freire (2006) discutindo sobre a Ciência da Informação, detectou três categorias que agregam os estudos relacionados à disciplina, nomeadamente: Recuperação da Informação, Comunicação da Informação e Estudos da Cognição. O autor, no entanto, revela que essas categorias não são isoladas e que muitos estudos transpassam mais que uma categoria. Neste quesito os estudos sobre a federação de repositórios científicos abordam todas as categorias.

No que diz respeito à Recuperação da Informação, a federação é uma ferramenta desenvolvida conforme os conceitos relacionados a este tema, no que, é mantida por centros de informação para atender aos usuários. Assim, relaciona-se com a representação da informação, processamento da informação, indexação para recuperação, entre tantos outros processos, que permeiam o tema da Recuperação da Informação.

A federação de repositórios científicos tem por objetivo disseminar a informação científica, posicionando-se na Comunicação Científica como ferramenta de apoio, principalmente para a utilização dessa informação. Atua na produção e uso da informação científica, apoiando os canais de publicação, no que diz respeito à potencialização de acesso, por meio da aplicação de Tecnologia da Informação e Comunicação.

Por fim, em relação aos Estudos da Cognição ampara a aprendizagem ao ofertar ferramentas de busca e fornecer acesso a textos integrais. Mesmo tendo como público alvo os docentes e discentes de cursos de graduação e pós-graduação em âmbito local, regional, nacional e internacional torna-se uma fonte de informação útil para a formação

profissional, para diversos segmentos da sociedade ou ainda para pessoas que individualmente busquem por informação científica.

Nesse contexto, revela-se a pertinência dos estudos relacionados à Federação de Repositórios Científicos no âmbito da Ciência da Informação, contribuindo para o avanço desta disciplina em várias áreas de interesse. Da mesma forma, pela interdisciplinaridade intrínseca da Ciência da Informação, em alguns aspectos, pode ser útil a outras disciplinas.

#### **4.1. REFERÊNCIAL TEÓRICO**

A revisão de literatura ofertou ao presente estudo a base necessária para a pesquisa. Enfatizada e orientada pelos objetivos, garantiu não somente os conceitos adotados à pesquisa, mas insumos ao próprio estudo, no que possibilitou relacionar tópicos de estudos que se complementam. Com isso, influenciou toda a pesquisa ao apontar os caminhos para as tarefas seguintes, ao revelar que um modelo para federação de repositório deve incluir principalmente as questões sobre as Tendências na Comunicação Científica e as Tendências Tecnológicas.

A base para o desenvolvimento de federações está nos repositórios científicos, que fornece os metadados para a formação da base de dados, pelo qual as federações podem ofertar serviços. Por isso, conceitualizar os repositórios e verificar suas características se torna essencial ao estudo, visto que, as características dos acervos dos repositórios tem influência direta nas federações.

A figura 14 apresenta uma estrutura de federação composta por vários repositórios, com características distintas, compondo uma rede heterogênea. Cada repositório é um sistema autônomo com seu acervo e interface, ofertando serviços restritos àquela base de dados. A federação, por sua vez, coleta metadados ou dados dos repositórios formando uma base de dados consolidada, a qual pode fornecer, por meio da sua interface, serviços consolidados.

FIGURA 14 – ESTRUTURA DE UMA FEDERAÇÃO DE REPOSITÓRIOS



Fonte: o autor

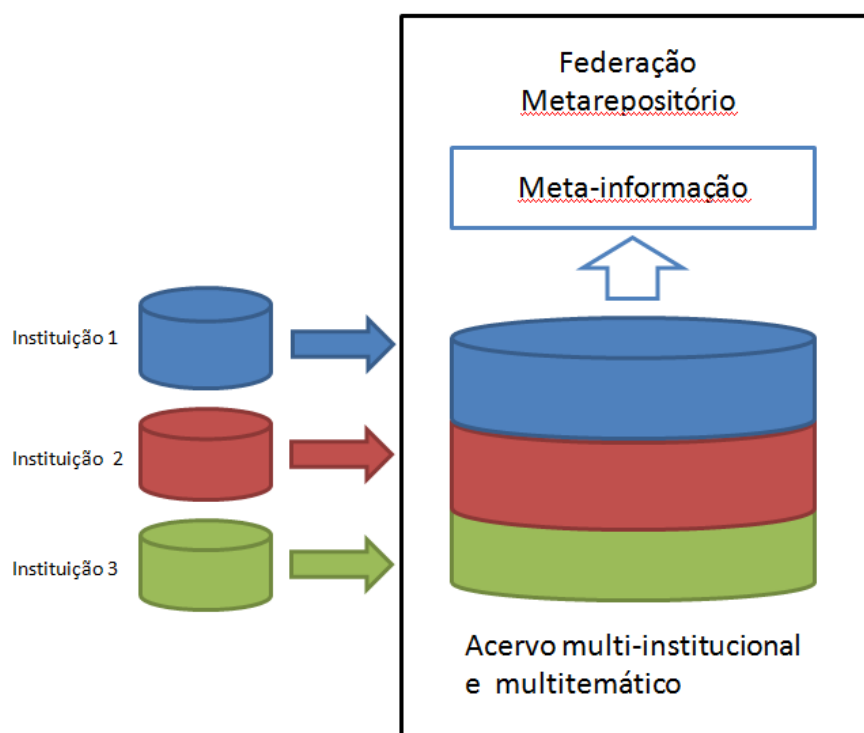
Cabe salientar que muitos dos serviços ofertados na federação dependem dos metadados provenientes dos repositórios, pois federações não possuem registros próprios. Entretanto, essa limitação não se apresenta como uma restrição, já que oferta quase os mesmos serviços dos repositórios que integra, com a vantagem de ser consolidado, ou seja, com informações de vários repositórios e, assim, disponibilizar serviços com maior amplitude.

É importante salientar que por integrar acervos de vários repositórios, uma federação de repositórios científicos torna-se um metarepositório. Ao coletar metadados de repositórios, forma um acervo de registros provenientes de várias instituições, ou seja, é multi-institucional. Da mesma forma, que seu acervo é variado, no que diz respeito aos temas, por isso é multitemático. Assim, a federação apresenta características únicas, no que se refere a fontes de informação.

Devido à formação do acervo, a federação pode ofertar não apenas acesso a uma grande variedade de documentos científicos provenientes de um grande número de instituições, mas indicadores consistentes. Compõe-se de um corpus significativo da produção científica consolidada dos repositórios integrados, em que a análise possibilita a extração de meta-informações.

A figura 15 sintetiza todos esses conceitos sobre federação, apresentando graficamente-a como um meta-repositório, em que o acervo é multi-institucional e multitemático, oriundo de repositórios distintos, e da qual se pode extrair meta-informação. Desta forma, consiste em um banco de dados colaborativo, com características únicas, nos sistemas de visam a apoiar a disseminação da informação científica.

FIGURA 15 – CONCEITUAÇÃO DE FEDERAÇÃO DE REPOSITÓRIO



Fonte: o autor

## 4.2. MODELOS

Os modelos são comuns nas ciências, pois podem representar uma parcela da realidade, apoiando a representação do objeto de estudo. Nesse sentido, com um modelo delimita-se o objeto a ser estudado, levantando todos os aspectos relevantes ao estudo. Além disso, possibilita a criação de uma infinidade de modelos, conforme a disciplina e a finalidade, por exemplo, de modelos matemáticos que posteriormente podem ser adaptados à física ou em outra disciplina.

Para Kuhn (2006, p. 43), os modelos são diretamente ligados aos paradigmas, relatando que "no seu uso estabelecido, um paradigma é um modelo ou padrão aceito". Nesse conceito, coloca-se os modelos no cerne das ciências. Assim a Ciência cresce com a geração e aceitação de novos modelos, colocando-a em processo constante de confrontação, renovação, revisão e evolução de modelos.

De Re (2000), sugere que os modelos são ferramentas mentais da ciência, em que o pensamento descritivo é representado pelos modelos físicos, enquanto o pensamento argumentativo é representado pelos modelos matemáticos. Evidencia-se, nesse ponto a visão dos modelos como representações mentais verificáveis e aceitos por uma comunidade.

Em discussão sobre Química Teórica, Tomasi (1999) afirma que os modelos estão presentes em todas as disciplinas científicas. Mesmo que modelos se apresentem incompletos na representação, são referências úteis nos processos das pesquisas. Assim, modelos podem ser reinterpretados para agregar novas informações de forma a melhorar a representação do objeto de estudo.

Quanto à tipologia de modelos, vários autores apresentam formas diferenciadas de organizá-los. Em suma, pode-se organizar a tipologia de modelos, de forma mais ampla, conforme:

- Modelo Icônico - representação fiel e dimensionável do objeto a ser estudado;



- Modelo Diagramático - representação simbólica, com intuito de verificar a funcionalidade e sem preocupação com a semelhança física;
- Modelo Matemático - idealização por construção lógica, representada por fórmulas;
- Modelo Estrutural - representa a estrutura e as relações do objeto a ser estudado;
- Modelo Conceitual - representa conceito, ação, procedimento, atitude e outros fenômenos do objeto estudado;
- Modelo de Representação Gráfica - representação visual do comportamento do objeto a ser estudado.

Nota-se que um objeto de estudo pode ser representado por vários modelos. Por exemplo, a água pode ser representada de forma icônica por uma figura com dois átomos de hidrogênio e um de oxigênio (modelo icônico); pela explicação de substância sem cheiro ou cor, solúvel universal (modelo conceitual); pela fórmula de  $H_2O$  (modelo matemático); pela estrutura H-O-H, com as ligações entre os átomos (modelo estrutural) etc.

Rego (2006, pp. 48-49) organiza os modelos tipologicamente baseado no modelo apresentado por Black (1962), da seguinte forma:

- Modelos por Escala - representam objetos, processos ou sistemas que apresentam proporções relativas semelhantes;
- Modelos por Analogia - representam objetos, processos e sistemas que, analogamente, reproduzem a estrutura ou rede de relacionamento de outro objeto, processo ou sistema;
- Modelos Teóricos - representam a extensão de um modelo aceito em um campo a outro campo.

No âmbito dos estudos da administração, para tomada de decisão, Moore e Weatherford (2005, p. 32) apontam que os modelos têm inúmeras finalidades, pois

propiciam uma estrutura lógica e consistente para análise. Independentemente do tipo do modelo, revelam os motivos que levam ao uso de modelos para apoiar a solução de problemas:

- Forçam a explicitação dos objetivos;
- Forçam a identificação e o registro dos tipos de decisões que influenciam os objetivos;
- Forçam a identificação e o registro das interações e concessões entre as decisões;
- Forçam a reflexão sobre variáveis a serem incluídas e sobre as definições, para que sejam qualificáveis;
- Forçam a consideração de que os dados são pertinentes para a qualificação das variáveis e determinação das interações;
- Forçam o reconhecimento de restrições;
- Fomentam a comunicação para disseminação de ideias.

Nesse ponto, os autores advogam que os modelos só cumprem seu papel, no âmbito da tomada de decisão, se a decisão tomada pela análise do modelo for melhor do que a tomada sem a utilização de modelo. Essa questão revela a capacidade dos modelos em ser utilizados como ferramentas de representação, com o intuito tanto de análise quanto de comunicação.

Independentemente do objeto de estudo, a definição de um modelo torna o estudo ou proposta mais clara. Para tanto, deve-se fazer levantamento dos aspectos relevantes, da mesma forma que a definição do tipo mais apropriado à representação do objeto a ser estudado é outro ponto importante para o êxito da pesquisa.

Nesse contexto, o objetivo do presente estudo é identificar, analisar e propor um modelo para federação de repositórios. Pela complexidade que o estudo apresenta - por envolver aspectos interdisciplinares, tecnológicos, técnicos, gerenciais e até jurídicos - entende-se que o desenvolvimento de um modelo tendo como foco o modelo conceitual,

mas agregando características de modelos análogos parece ser o mais adequado para este caso. Isso vai permitir, por ter similaridade com os repositórios ou outros sistemas informatizados, que se agreguem informações de sistemas heterogêneos.

#### **4.2.1 MODELO CONCEITUAL**

Os modelos conceituais são muito comuns na Ciência da Computação, pois são considerados como uma etapa no ciclo de desenvolvimento de projetos. Nesse ponto, os modelos conceituais têm por finalidade comunicar a ideia do desenvolvimento aos envolvidos no projeto, dos usuários finais aos gerentes, de forma que todos os envolvidos tenham a mesma compreensão do produto a ser desenvolvido.

Com isso, apoia-se a identificação das características de um projeto a ser desenvolvido, no que possibilita a formalização, com o registro de todos os elementos que compõem o projeto. Registra-se as pessoas envolvidas, os processos, os dados, tudo que influencia no desenvolvimento do projeto.

Da mesma forma, facilita-se a análise do projeto a oferta de um cenário bem estabelecido. Ao ponto que se torna adequado à identificação, análise e proposta de soluções a problemas existentes ou a objetos de estudos.

Analisando as proposições de Marjomaa, desenvolvidas em 2002, Daun (2003, p.44) revela uma lista de orientações para a criação de modelos conceituais. Com essa verificação, procura-se cobrir toda a complexidade do problema. Tal lista contempla vários aspectos, como um *check list*, apresentado a seguir:

- Conceituação - definir o domínio;
- Completude - conter todos os aspectos relevantes ao domínio;
- Formalização - formalizar para evitar ambiguidades;
- Inteligibilidade - facilitar as explicações;
- Correspondência – relacionar as entidades do domínio;

- Invariância – conceituar com base em entidades, que não variam, e;
- Decomposição - possibilitar a escalabilidade.

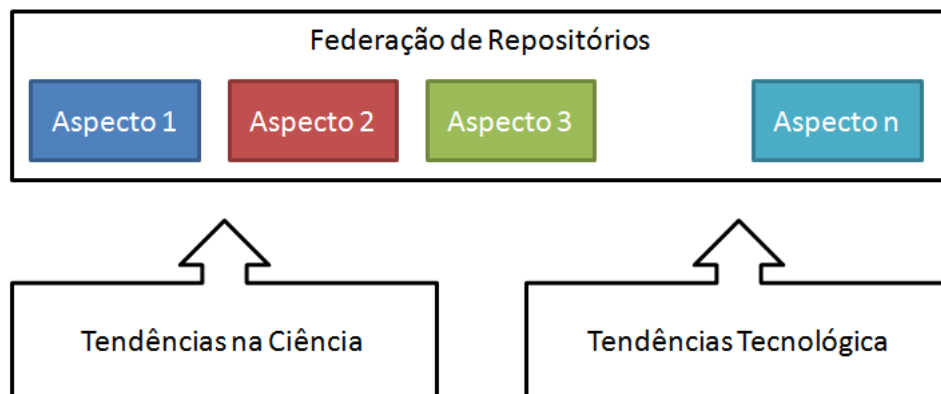
No contexto dos projetos de tecnologias interativas, Preece, Rogers e Sharp (2005) discorrem sobre os modelos conceituais. Revelam que o modelo conceitual é resultado das necessidades apresentadas pelos usuários e de outros requerimentos levantados. Para tanto, o modelo deve contemplar soluções concretas, envolvendo processos, interações e outros requerimentos.

Modelos conceituais de sistemas informatizados, por exemplo, podem apresentar atividades operacionalizadas por meio de interfaces, ou páginas. É importante mencionar que duas ou mais atividades podem estar vinculadas a propósitos diferentes. Mesmo que para isso tenham que ajustar os requisitos para atuar em mais que uma atividade.

Essa questão fica mais clara no contraste entre um modelo conceitual de instrução e um modelo de manipulação e navegação. O primeiro modelo tem por objetivo permitir que o usuário instrua o sistema para que execute um determinado processamento, essa instrução pode ser feita digitando comandos ou selecionando opções. Para a segunda opção, tem-se a manipulação e navegação por intuição. Arrastar um objeto ou selecionar um ícone são exemplos claros de manipulação e navegação. Note-se que em alguns casos, algumas atividades podem ser classificadas tanto como instrutivas quanto como navegacionais.

Um modelo contextual pode ser desenvolvido pela decomposição do objeto de estudo, conforme vieses diferentes, respaldados em alguns aspectos. Cada aspecto, nesse caso, refere-se a um conjunto de requisitos que embasam o modelo, de forma a proporcionar múltiplas dimensões ao problema. Assim, para a presente pesquisa pode ser esquematizado um modelo para federação de repositórios científicos que representa os vários aspectos que os compõem, bem como as tendências que o orientarão. Conforme a figura 16.

FIGURA 16 - ESQUEMA DE UM MODELO CONCEITUAL PARA FEDERAÇÃO DE REPOSITÓRIOS.



Fonte: o autor

Nesse ponto, pode-se representar um modelo de federação como um conjunto de aspectos influenciados pelas tendências tecnológicas e da ciência. Cada aspecto trata de um ponto específico, que se complementam para formar o todo, por exemplo, por ser um sistema informatizado disponível na internet, um dos aspectos refere-se a apresentação do portal, os requisitos que representam o aspecto.

#### 4.2.2 MÉTODO DE REPRESENTAÇÃO IDEF0

A concepção de um modelo, independente do tipo, requer uma formalização, uma representação gráfica que exprima o entendimento do objeto a ser modelado. Para a presente pesquisa o método de modelagem escolhida foi o IDEF0, que se presta a descrever processos ou sistemas, na medida em que possibilita representar processos e fluxos de informação.

Inicialmente IDEF era a abreviatura de ICAM (*Integrated Computer-Aided Manufacturing*) *Definition*, que em 1999 formalizou-se como *Integration DEFinition*. Esta mudança no significado da abreviatura deu-se pelo próprio desenvolvimento do método de representação, pois o ICAM tem origem nos projetos militares da Força

Aérea Americana dos anos setenta do século XX e o IDEF tornou-se mais abrangente, não somente para projeto militares.

O IDEF0 faz parte de uma família de linguagem para modelagem de sistemas informatizados denominada de IDEF, que contempla vários métodos de representação. Com isso, possibilita a formalização de uma quantidade variada de tópicos de estudo relacionados ao desenvolvimento de ferramentas informatizadas.

A família IDEF é composta, atualmente, pelos seguintes métodos de representação:

- IDEF0 : método para modelagem de funções para representação de decisões, atividades e informações próprios de uma organização ou sistema;
- IDEF1 : método para modelagem de comunicação e análise de requisitos;
- IDEF1X : método para modelagem de dados;
- IDEF2 : método para modelagem de interfaces;
- IDEF3 : método para modelagem de negócios;
- IDEF4 : método para modelagem de programação orientada a objeto;
- IDEF5 : método para modelagem de ontologias;
- IDEF6 : método para modelagem de relacionamentos;
- IDEF7 : método para modelagem de auditoria de sistemas de informação;
- IDEF8 : método para modelagem de interfaces de usuários;
- IDEF9 : método para modelagem de descoberta e análise de restrições em sistemas de negócios;
- IDEF10 : método para modelagem e implementação em arquitetura;
- IDEF11 : método para modelagem de artefatos de informação;
- IDEF12 : método para modelagem de organizações;
- IDEF13 : método para modelagem de mapeamento em esquema de árvore;
- IDEF14 : método para modelagem de redes.

No que tange ao IDEF0, a agência federal vinculada ao departamento americano de comércio, *National Institute of Standards and Technology* (NIST) apresentou em 1993 as regras e técnicas deste método, por meio de uma publicação oficial (FIPS

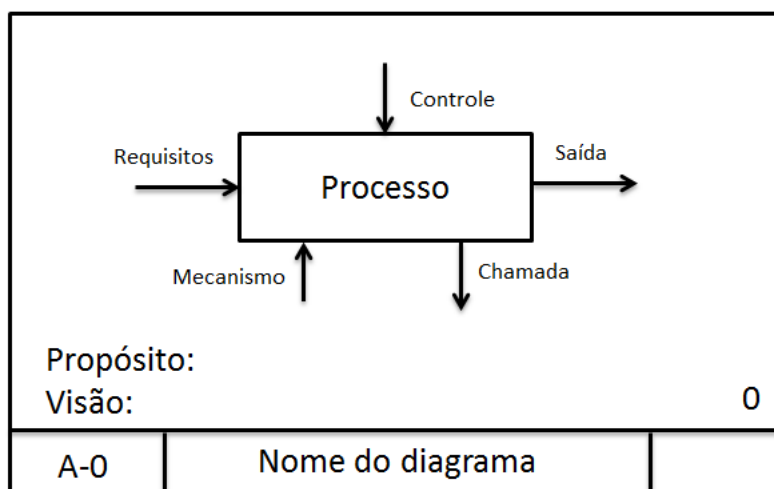
PUBS, 1993). Este documento apresenta as bases orientadoras dessa linguagem de modelagem, que foi a primeira a ser formulada desta série de métodos de representação de modelagem.

Com o entendimento de que um modelo é uma representação gráfica de um sistema, composto por elementos interligados, incluindo pessoas, processos, equipamentos, produtos entre outros elementos, O IDEF0 é um método de modelagem que combina representação gráfica e de texto para a representação de um sistema. Para tanto, este método apresenta seus diagramas no formato gráfico de refinamento hierarquizado e em forma textual.

O objetivo do IDEF0 é ser de fácil construção, compreensão, além de ser expressivo e flexível, de maneira a possibilitar a apresentação de sistemas complexos, de forma coerente e consistente. Com isso permite a comunicação fácil entre todos os atores envolvidos no desenvolvimento ou representação do sistema. Outro ponto importante é a simplicidade, no que deve ser de fácil representação, de forma que os diagramas podem ser desenvolvidos em uma grande quantidade de *softwares*, dispensando a criação de um específico para essa finalidade.

Todo modelo IDEF0 inicia em um diagrama geral denominado de A-0, que sintetiza todo o sistema, uma caixa com um termo ou frase (figura 17). A comunicação desse sistema é representada por setas, que possuem regras que indicam a sua função. As entradas devem se representadas por setas que apontam para a caixa, no lado esquerdo. As saídas são setas que saem da caixa, pelo lado direito, e indicam o resultado do processo. Setas na parte superior indicam controle, que apoiam a execução do processo indicado pela caixa. Setas na parte inferior indicam mecanismos e chamadas (dependendo da direção da seta), indicando quem executa o procedimento.

FIGURA 17 – REPRESENTAÇÃO DE UM DIAGRAMA IDEFO PADRÃO

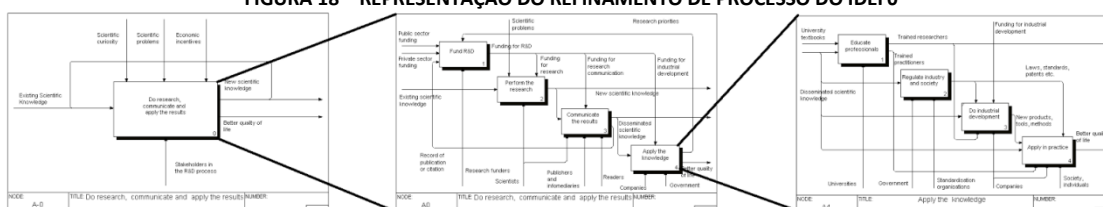


Fonte: o autor

O Diagrama A-0 deve conter a cláusula propósito e visão para apoiar o entendimento do sistema representado. Por ter que sintetizar todo o sistema a ser representado, não precisa apresentar todas as entradas, saídas, controle e mecanismos, apenas os principais. O detalhamento dos processos apresentará as minúcias de cada fase.

Uma das características mais marcantes do IDEF0 é o refinamento dos processos, em que cada caixa representante de uma fase pode ser detalhada em outro diagrama. Assim, o primeiro diagrama A-0 é refinado, detalhando os seus processos no diagrama A0. Os processos constantes em A0 são detalhados nos diagramas A1, A2, A3, A4, por exemplo. Os processos descritos em A1 podem ser detalhados nos diagramas A11, A12, A13. O Diagrama A11 pode ser detalhado nos diagrama A111 e A112 e assim por diante (figura 18). Essa hierarquização possibilita chegar a altos níveis de detalhes.

FIGURA 18 – REPRESENTAÇÃO DO REFINAMENTO DE PROCESSO DO IDEFO



Fonte: o autor



Neste contexto, as caixas são processos e precisam ser descritas com frases centradas em verbos. As setas são os dados requeridos ou produzidos pelos processos. Ao detalhar um processo, o nome desse processo dá o nome ao diagrama. Recomenda-se que apresente até quatro caixas em cada diagrama. Ao final deve-se criar um sumário com todos os diagramas que compõem o modelo de forma a apresentar, em forma de lista, todos os diagramas que compõem o modelo.

### **4.3 DESENHO DA PESQUISA**

É correto afirmar que a Ciência é pautada pelo método. É nesse fundamento que uma informação torna-se científica e, portanto, parte fundamental na pesquisa. Pilar da Ciência, o método fornece a legitimidade e a confiabilidade necessária aos estudos, orienta as formas de se obter os resultados e amplia a análise. Assim, é o norteador maior do estudo, que assegura o andamento da pesquisa.

Focado nos objetivos para a identificação, análise e proposição de um modelo, o presente estudo requer aprofundamento no conhecimento do problema, que de acordo com Gil (2006) se caracteriza como pesquisa exploratória. Com isso, procura explicitar os elementos que compõem o problema, de forma a possibilitar a formulação de um modelo, que atenda o contexto do problema. O estudo exploratório possibilita o aprimoramento de ideias como nos casos dos modelos, que são representações formais de um fenômeno ou problema.

É importante reforçar a ideia de que a identificação, análise e desenvolvimento de um modelo prevê um aprofundamento sobre o objeto de estudo revelando elementos muitas vezes desconhecidos ou pouco explorados. Com isso, contribui com a temática, atualizando conceitos ou características frente à evolução do tema e esclarecendo questões que muitas vezes são tratadas com superficialidade em outros estudos.

Espírito Santo (1992) revela que nas Ciências Aplicadas modelos são ferramentas para solucionar problemas, diferindo de teorias por serem mais práticos. Destaca-se o aspecto aplicado dos estudos voltados a apresentar modelos, na medida

que visam a desenvolver visões práticas de um sistema ou parte dele, com foco na solução de problemas.

Alinhando à questão mais aplicada do modelo, o presente estudo visa propor um modelo para federação de repositórios científicos. Assim, para o mesmo autor, os modelos podem ser classificados referenciando seu objetivo, podendo ser categorizados como descritivos, exploratórios, preditivos ou prescritivos. Entretanto, em grande parte dos estudos, os modelos podem ter finalidades diversas e serem classificados em mais de uma categoria.

Para fins de classificação, o presente estudo, que objetiva desenvolver um modelo para federação de repositórios, pode ser categorizado como um estudo para o desenvolvimento de um modelo prescritivo, porque apresenta uma situação real e desenvolve orientações para solução de adaptação às tendências atuais em tecnologia e nas ciências.

### **4.3.1 ABORDAGEM**

A abordagem é um norteador para o estudo, pois orienta a metodologia a ser utilizada na pesquisa. Para Richardson (2008) as pesquisas podem ser classificadas, quanto a abordagem, em quantitativas e qualitativas. Essas abordagens, no entanto, não são exclusivas nem antagônicas e, em alguns estudos, podem ser complementares, podendo ter uma orientação mista, em que apresentam características quantitativas e qualitativas.

A abordagem desta tese tem uma orientação mista, já que apresenta características quantitativas e qualitativas. Assim, aliando a precisão quantitativa com a profundidade qualitativa o estudo utiliza-se de técnicas de pesquisa complementares apoiadas por essas duas abordagens, com o propósito de torna-lo mais abrangente.

Parte dessa pesquisa tem um enfoque mais qualitativo, focado nas ocorrências de um fenômeno. Assim, mais alinhados aos estudos humanos e sociais, em que a quantidade de ocorrência de um fato nem sempre torna a variável de maior

representatividade, mas sim a importância de um fato para uma determinada comunidade.

Corroborando a questão de a abordagem qualitativa estar mais alinhada aos estudos sociais e humanos, Denzin e Lincoln (2006) enfatizam que a pesquisa qualitativa observa a natureza social contextualizada dos fenômenos. Visto que os objetos de estudo, como experiências sociais, podem assumir significados distintos dependendo da comunidade em que está inserida, ou seja, um fenômeno observado do ponto de vista social é multifacetado, o que requer um estudo mais profundo, próprio das pesquisas qualitativas.

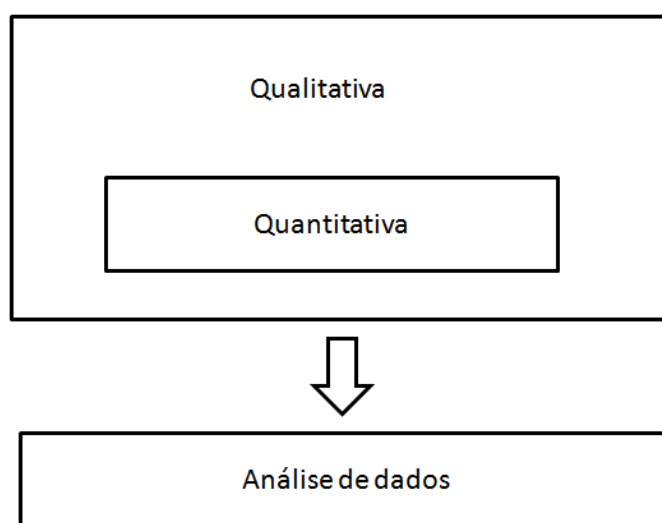
Outra parte da pesquisa, por sua vez, alinha-se mais com as pesquisas quantitativas, visto que procura confirmar as hipóteses formuladas por meio de coleta de dados quantificáveis, concordando com o que Richardson (2008) caracteriza como pesquisas quantitativas. Com isso, obtém-se maior precisão nos resultados, já que a análise de dados quantitativos é mais objetiva e possibilita obter características de grupos ou fenômenos, por meio de coleta e análise de dados em que a frequência se apresenta com uma importância maior, permitindo generalizações baseadas na quantidade da ocorrência de um fato.

A abordagem quantitativa é útil, da mesma forma, aos estudos sociais possibilitando uma complementaridade com a abordagem qualitativa, como descrevem Linhares e Alves (2014). Para as referidas autoras, as pesquisas sociais devem romper as barreiras e os antagonismos artificiais da dicotomia quantitativa *versus* qualitativa, de forma a desenvolver estudos mais abrangentes, em que cada abordagem oferece as ferramentas necessárias para a completude do estudo.

Creswell (2007) também aponta para a complementaridade entre as abordagens, pois entende que é possível combiná-las de várias formas em estratégias de pesquisa. Essas estratégias podem ocorrer nas diversas etapas da pesquisa, na coleta ou na análise dos dados, dependendo do objetivo do estudo, sendo que a integração na coleta pode-se dar sequencialmente ou concomitantemente, dependendo da necessidade.

Nesse ponto, esta tese se utiliza do que os referidos autores denominam de "Estratégia aninhada concomitante" (Figura 19), pois os dados quantitativos e qualitativos são coletados simultaneamente e não há uma necessidade de priorizá-los. Mesmo que se tenha uma predominância da abordagem qualitativa, parte das questões são resolvidas pela abordagem quantitativa, possibilitando uma perspectiva mais ampla.

FIGURA 19 - MODELO DE ESTRATÉGIA ANINHADA CONCOMITANTE (CRESWELL, 2007, P. 217)



FONTE: CRESWELL, 2007, PP. 217

Nesse contexto, a presente pesquisa se apoia em uma abordagem mista, ao ponto que a complexidade do tema estudado requer precisão em alguns pontos e a profundidade necessária para identificar, analisar e propor um modelo. Busca-se a flexibilidade que transcende as fronteiras, nem sempre tão bem marcadas das abordagens, utilizando técnicas de pesquisa características de cada abordagem.

### 4.3.2 TÉCNICAS DE PESQUISA

A presente pesquisa apresenta-se no âmbito dos estudos exploratórios, com abordagem mista, com o objetivo de identificar, analisar e propor um modelo para federação de repositórios acadêmicos. Para tanto, faz uso de duas técnicas para coleta de dados, a pesquisa documental e o levantamento, próprios das abordagens qualitativas e quantitativas, de forma a cumprir com os objetivos propostos, conforme quadro 7.

QUADRO 7 - SUMÁRIO REPRESENTATIVO DOS OBJETIVOS ESPECÍFICOS E TÉCNICAS DE PESQUISA

<b>Objetivo Específico</b>	<b>Fonte</b>	<b>Coleta de dados</b>
Levantar os conceitos, orientações, políticas, características, tendências e problemas que caracterizem às federações de repositórios	Bases de dados de literatura científica	Pesquisa bibliográfica
Mapear as principais características dos repositórios que influenciam na federação de repositórios	Diretório de repositórios	Análise dos repositórios e survey
Verificar a adaptação das novas tendências tecnológicas e da Ciência no desenvolvimento de federação	Administradores de repositórios	Pesquisa bibliográfica e survey
Propor um modelo para federação de repositórios que atenda as tendências tecnológicas e da ciência	Bases de dados de literatura científica	Pesquisa bibliográfica

Fonte: o autor; 2014.

A pesquisa documental tem por objetivo colher evidências que apoiem o estudo, coletar informações que assegurem os resultados, conforme os objetivos traçados para a pesquisa, utilizando como base os documentos existentes. Destaca-se a importância das informações contidas na documentação existente como fonte de pesquisa, no que são registros formais ou não de uma situação.

Neste ponto, discute-se a questão dos documentos, que podem ter variados tipos e origens, sendo que atualmente, discute-se a questão dos documentos digitais. Nesse sentido, páginas *web* podem ser consideradas como a unidade documental da *web*. Segundo Comber (1995, p. 194) "página web é definida como a informação recuperada por uma URL e, às vezes, referenciada como um documento". Assim, a informação contida em uma página não tem tipologia definida, mas se apresenta restrita à página, mesmo apresentando referências a informações de outras páginas.

A pesquisa documental, como técnica se apresenta muito apropriada a questões voltadas para a conceituação e verificação de relações, pertinente ao presente estudo. Richardson (2008) descreve a pesquisa documental como um conjunto de operações voltadas ao estudo de documentos, com o propósito de verificar os seus relacionamentos. Nesse ponto, a pesquisa documental tem nos mais diversos tipos de documentos as suas fontes, principalmente as que não apresentam tratamento analítico.

Nota-se que a pesquisa documental tem nos documentos a sua fonte principal, sendo que nos questionários o foco é nas pessoas. Assim, em muitos casos essas duas técnicas se complementam, como nos casos em que são requeridas evidências documentais e do posicionamento da população estudada. Este é o caso do presente estudo, que requer o levantamento do posicionamento dos gerentes de repositórios face às tendências tecnológicas e da ciência nos repositórios, que por sua vez tem reflexo nas federações e na pesquisa documental para aprofundamento dos conceitos.

Dessa forma outra técnica de pesquisa envolvida no presente estudo é o levantamento por meio de questionário (survey), visto que é apropriada a caracterização social, podendo abranger níveis amplos ou restritos de população ou amostras. Da mesma forma que pode-se ajustar o relatório para contemplar pesquisas quantitativas e qualitativas, revelando a flexibilidade da técnica, tanto em relação ao tamanho do universo pesquisado, quanto da orientação do levantamento.

O Survey possibilita verificar confirmações de fenômenos, por meio de questionários ou entrevistas realizados em canais como encontros presenciais, telefone ou questionários *on-line*, *chat* e videoconferência. Com isso, possibilita a coleta de dados mais estruturada, de forma mais restrita ou ampla, dependendo de como se faz o levantamento. Em questionários *on-line*, por exemplo, é possível utilizar tanto questões de múltipla escolha, quanto questões abertas, e com isso ter respostas padronizadas ou livres.

Nesse ponto, Freitas e colaboradores (2000) defendem que o survey é mais apropriado à pesquisa quando:

- a pergunta da pesquisa é do tipo: “o que?”, “como?”, “por que?” ou “quanto?”;
- quando o controle das variáveis é irrelevante ou não existentes;
- o estudo ocorre em ambiente natural;
- o fenômeno estudado é atual ou em passado recente.

Revela-se o aspecto mais focado na obtenção de dados sobre as características, razões, ações e pontos de vista de determinada comunidade ou parte dela. O survey

possibilita a realização de estudos mais complexos com coletas de grandes quantidades de dados, relacionando os resultados obtidos. Entretanto, as pesquisas que utilizam o survey restringem-se a estudos sincrônicos, sem uma grande abrangência temporal.

A associação das duas técnicas de pesquisas dá-se pela complementaridade, no que, os objetivos propostos requerem tantos os dados provenientes dos documentos, quanto de pessoas, de forma a obter um cenário mais completo. Considerando que o modelo deve ser amparado por teoria consistente e visão humana, a utilização da pesquisa documental e do survey parece muito adequada.

### **4.3.3 UNIVERSO DA PESQUISA E AMOSTRAS**

A pesquisa documental requer fontes condizentes, o universo para a aplicação dessa técnica de pesquisa no presente estudo se restringirá aos documentos disponíveis na *web*, em fontes de informação confiáveis como o Portal de Periódicos da Capes e a Biblioteca Digital de Teses e Dissertações. Portais de busca especializada como o google acadêmico também serão utilizados. A opção pela documentação digital dá-se pela atualidade do tema estudado e contempla as seguintes fontes:

- O **Portal de Periódicos da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES)** dá acesso a textos completos pertencentes a mais de 37 mil publicações periódicas, internacionais e nacionais. Mantido pelo Ministério da Educação (MEC), apresenta-se com um dos principais serviços de acesso à informação científica ofertados pelas bibliotecas universitárias, no âmbito do programa de incentivo assinatura de revistas científicas. Para o presente estudo, os periódicos consultados se restringirão apenas aos vinculados à Ciência da Informação e Informática.
- A **Biblioteca Digital Brasileira de Teses e Dissertações (BDTD)**, mantida pelo Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT), agrega mais de 250 mil registros de teses e dissertações oriundas de mais de 100 instituições. Possibilita o acesso ao texto integral por meio de ferramentas de busca simples e avançada, tornando-se uma fonte importante para a pesquisa.

- Os **portais das federações de repositórios científicos** disponíveis na internet, independente do país, são fontes primordiais à pesquisa, por conter auto-informações, que equivalem, na pesquisa humana a declarações autodescritivas, um documento autobiográfico. Ofertam informações que, até certo ponto, refletem os conceitos dos gestores das federações.
- Os **portais dos repositórios nacionais**, por sua vez, são fontes de informação cruciais a presente pesquisa por possibilitar a identificação de características do repositório que podem afetar a federação. A análise destes portais podem fornecer os indícios relacionados à interoperabilidade necessária à formação da federação. Pela orientação qualitativa, que requer coleta de dados mais aprofundada, o presente estudo restringirá a coleta e análise dos dez repositórios melhores colocados no *web ranking of repositories*.

No caso do Survey, os gestores de repositórios tem papel fundamental na definição das políticas dos repositórios. Essas orientações determinam as características do repositório e tem reflexo direto na federação. Por isso a importância do levantamento e da inclusão dessa comunidade como universo da pesquisa.

#### **4.3.4 OBJETO DA PESQUISA**

Fortemente amparado pelo objetivo de identificar, analisar e propor modelo para as federações de repositórios científicos, o objeto da pesquisa é a própria federação. Assim, prioritariamente este será o objeto de estudo, no que, toda a pesquisa volta-se ao estudo deste sistema de informação. Entretanto, outro objeto da pesquisa, mesmo que de forma secundária, merece destaque, são os repositórios científicos, que sustentam as federações com os metadados.

Um modelo de federação, como uma representação conceitual de um assunto objeto de estudo, requer uma caracterização desses sistemas de informação, de forma a possibilitar a sua identificação, análise e proposição. Com isso, é possível executar a coleta de dados, com base nas características, mesmo nos casos em que a federação se apresente com outra denominação.



Por ser um sistema de informação não muito comum na internet, serão estudados todos aqueles que cumprirem as características de agregarem registros que remetam a documentação científica em texto integral, coletados de repositórios científicos, sem a utilização de amostras. Pelas características qualitativas do estudo, levantar-se-á variáveis de cunho qualitativo nessas federações.

Para tanto, os portais das federações, como representação do sujeito da pesquisa, serão estudados como documentos web. A pesquisa documental, utilizada como ferramenta, levantará os indícios relacionados às variáveis, possibilitando a coleta de dados. O estudo complementar a coleta por meio dos documentos científicos relacionados às federações, tanto os conceituais, quanto os relatos de experiência, de forma a ter a completude necessária aos objetivos propostos.

Da mesma forma, os portais dos repositórios, como sujeitos da pesquisa complementar, terão o mesmo tratamento, diferindo por ser alvo de um survey, por intermédio de um questionário destinado aos gerentes dos repositórios. A motivação desta diferenciação para este sujeito da pesquisa deve-se que:

- ✓ as federações possuem apenas os metadados, os dados completos são mantidos pelos repositórios;
- ✓ as políticas dos repositórios influenciam diretamente o comportamento da federação;
- ✓ as formas de disseminação, relação com direitos autorais, entre outros comprometem as federações;
- ✓ a adesão a novas tecnologias podem afetar as federações;
- ✓ a utilização de padrões de comunicação facilita a integração por federações;
- ✓ o acompanhamento às tendências da ciência pode ter reflexo na federação.

Por conseguinte, os sujeitos da pesquisa se complementam, no que as federações não existem sem os repositórios. A coleta de informações sobre esses sujeitos possibilita a extração de conteúdos relativos às variáveis selecionadas, para a completude do estudo.

### 4.3.5. VARIÁVEIS

O presente estudo se orienta de forma mista, ou seja possui características quantitativas e qualitativas, fazendo uso associado de dois tipos de técnicas de pesquisa, cada qual apresentando parte do universo de pesquisa de forma complementar. Nesse sentido, apresenta categorias de variáveis distintas para cada técnica de pesquisa, visto se aplicarem a parte do universo de pesquisa diferente.

Para estudos quantitativos as variáveis tornam-se mais explícitas, pois podem ser representadas por frequência. Os estudos qualitativos, por sua vez, possuem aspectos mais de representatividade ou importância, não podendo ser suas variáveis expressas por valores numéricos. Entretanto, independente do tipo da variável, Richardsdon (2008) destaca que devem ser observáveis e apresentar variações ou diferenças em relação ao mesmo ou a outro fenômeno.

Neste ponto, a pesquisa documental, alinhada com as questões qualitativas pode categorizar as variáveis conforme os objetivos. Dessa forma, as variáveis qualitativas são apresentadas como:

- ❖ Para o objetivo específico: “levantar os conceitos, orientações, políticas, características, tendências e problemas que caracterizem as federações de repositórios” as variáveis qualitativas são:
  - ✓ **conceitos de federação de repositórios:** exprimem as diversas visões sobre o tema;
  - ✓ **políticas de federação de repositórios:** descrevem as orientações utilizadas nas federações;
  - ✓ **características de federação de repositórios:** apresentam os aspectos distintivos das federações.
- ❖ Para o objetivo específico: “mapear as principais características dos repositórios que influenciam na federação de repositórios” as variáveis qualitativas são:
  - ✓ **políticas de repositórios:** descrevem as regras que orientam o funcionamento dos repositórios;

- ✓ **esquemas de metadados:** descrevem a tipologia dos esquemas de metadados utilizados pelos repositórios no processo de interoperabilidade;
- ✓ **tipologia documental:** apresentam a variedade de tipos de documentos formadores do acevo mantido pelo repositório.
- ❖ Para o objetivo específico: “propor um modelo para federação de repositórios que atenda as tendências tecnológicas e da ciência” as variáveis qualitativas são:
  - ✓ **tendências tecnológicas:** apresentam as tendências tecnológicas aplicáveis nas federações de repositórios;
  - ✓ **tendências da ciências:** apresentam as tendências da ciência aplicáveis nas federações de repositórios;
  - ✓ **processos de federalização:** apresentam as atividades envolvidas nas federalização de repositórios;
  - ✓ **fluxo de dados:** apresentam o fluxo de dados envolvidos na federalização de repositórios.

No que se refere às variáveis quantitativas, coletadas por questionários, podem ser categorizadas como:

- ❖ Para o objetivo específico: “Verificar a adaptação das novas tendências tecnológicas e da Ciência no desenvolvimento de federação” as variáveis quantitativas são:
  - ✓ **uso das tendências da ciência nos repositórios:** apresenta a frequência no uso de funcionalidades relacionadas às tendências da ciência nos repositórios;
  - ✓ **uso das tendências tecnológicas nos repositórios:** apresenta a frequência no uso de funcionalidades relacionadas às tendências tecnológicas nos repositórios;

O conjunto total das variáveis combinam aspectos qualitativos e quantitativos se complementado com dados provenientes de documentos e gerentes de repositórios.

Com isso, possibilita uma visão ampla necessária a identificação, análise e proposição de um modelo de federação de repositórios científicos.

#### 4.3.6. MÉTODOS PARA ANÁLISE E INTERPRETAÇÃO DE DADOS

Pela aplicação de uma abordagem mista, a análise e interpretação de dados dão-se, prioritariamente, por análise de conteúdo, já que possui maior similaridade com as pesquisas qualitativas. Assim, visa comprovar as hipóteses propostas por meio da análise dos dados coletados. Por isso, como as variáveis para esse fim são em grande parte qualitativas, a análise deve ser basear na representatividade e não na frequência.

Nesse contexto, podem-se relacionar as variáveis com as hipóteses formuladas, conforme o Quadro 8, de forma a facilitar a análise e interpretação dos dados. Destaca-se o caráter qualitativo nesta etapa da pesquisa, que requer uma análise e interpretação mais aprofundada dos dados coletados de forma a responder as hipóteses.

QUADRO 8 - RELAÇÃO ENTRE AS HIPÓTESES E AS VARIÁVEIS

Hipótese	Variável
Considerando o quantitativo de repositórios científicos brasileiros é possível a construção de um modelo de federação que atenda a várias finalidades	<ul style="list-style-type: none"> <li>• conceitos de federação de repositórios</li> <li>• políticas de federação de repositórios</li> <li>• características de federação de repositórios</li> </ul>
Considerando o quantitativo de tipos de repositórios existente no país é fortemente viável que o modelo possa ser utilizado para criação de federações de repositórios não científicos	<ul style="list-style-type: none"> <li>• conceitos de federação de repositórios</li> <li>• políticas de federação de repositórios</li> <li>• características de federação de repositórios</li> </ul>
Com base nos resultados da pesquisa, o modelo apresentado pode amparar a o desenvolvimento de sistemas sustentáveis e estáveis, se utilizado para implementação de ferramentas informatizadas para a criação de federação de repositórios	<ul style="list-style-type: none"> <li>• conceitos de federação de repositórios</li> <li>• políticas de federação de repositórios</li> <li>• características de federação de repositórios</li> <li>• tipologia documental</li> <li>• tendências tecnológicas</li> <li>• tendências da ciências</li> </ul>
Considerando a federação de repositório como um sistema que organiza um cenário heterogêneo, é possível a aplicação de formas diferenciadas de representação, que contextualize a informação apresentada	<ul style="list-style-type: none"> <li>• conceitos de federação de repositórios</li> <li>• políticas de federação de repositórios</li> <li>• características de federação de repositórios</li> <li>• políticas de repositórios</li> <li>• tendências tecnológicas</li> <li>• tendências da ciências</li> </ul>
Considerando que o estudo inclui várias linhas de pesquisa da Ciência da Informação, assim como disciplinas correlatas, pode-se presumir que a pesquisa contribuirá para a Ciência da Informação, em suas questões epistemológicas no que diz respeito as relação entre tópicos de estudo e interdisciplinaridade	<ul style="list-style-type: none"> <li>• conceitos de federação de repositórios</li> <li>• políticas de federação de repositórios</li> <li>• uso das tendências da ciência nos repositórios</li> <li>• uso das tendências tecnológicas nos repositórios</li> </ul>

Fonte: o autor; 2014.

A análise do conteúdo coletado das variáveis extrapola a simples comprovação das hipóteses, visto que ampara os objetivos de identificar, analisar e propor modelo para federação de repositórios científicos. Fornece indícios que apoiam a pesquisa no que compõem um conjunto significativo de informações voltadas aos sujeitos da pesquisa.

## 5. RESULTADOS E DISCUSSÕES

---

A auto-identificação é utilizada nas Ciências Sociais, ao ponto que para minorias sociais ser juridicamente aceita o indivíduo se autodeclarar surdo, negro, branco e outros. Assim, para estudos sociológicos, por exemplo, a auto-identificação é extremamente importante, ao passo que caracteriza o elemento que liga o indivíduo ao grupo ou comunidade. Assim, é o elo entre o indivíduo e a parte da sociedade a que considera pertencer.

Da mesma forma que revela a auto-identidade, possibilitando aos estudiosos aprofundarem nas pesquisas sobre um indivíduo ou grupo. Nos sites da internet, como elementos transmissores de informação, essa auto-identificação ou auto-conceituação deve ser estudada e colocada como elemento de relevância, pois é a expressão do autor, mantendo o gerente sobre o site. Contextualiza o site em relação aos outros, identificando-o como pertencente a uma tipologia ou diferenciando-o dos demais.

Torna-se natural que em estudos que visam a identificar, analisar e propor modelos para sistemas informatizados disponíveis na internet, que o façam tendo como objeto de estudo os próprios sistemas. Ao passo que os sites e portais disponíveis na internet possuem públicos-alvo, finalidade e características que possibilitam ser categorizados de forma a apoiar os estudos. Assim, todos os serviços e estruturas voltam-se a atender as necessidades desse público.

Destaca-se que a identificação e análise desses sistemas, amparada por uma base conceitual, proporciona fundamentos para a proposição de modelos. Assim, as teorias complementam o estudo, fornecendo a base para a análise, de forma a obter resultados mais formalizados e robustos.

Alinhado aos objetivos, os resultados se organizam inicialmente na identificação e análise das federações, em 5.1. No apoio à proposição do modelo em 5.2 e na proposição do modelo em 5.3. Com isso, evolui na identificação dos elementos que compõem uma federação, a análise dos elementos dos repositórios que influenciam

a federação, com o objetivo de propor um modelo que inclua as tendências tecnológicas e da Ciência.

## **5.1 FEDERAÇÕES DE REPOSITÓRIOS: CONCEITOS, POLÍTICAS, CARACTERÍSTICAS E TENDÊNCIAS.**

Federações de repositórios científicos não se apresentam comuns na internet. O levantamento efetuado revelou dez federações de repositórios científicos, localizadas em vários países, sendo três no Brasil. Esta constatação reforça questões relacionadas à inovação do tema e os desafios técnicos e tecnológicos em agregar repositórios. Da mesma forma que indica que se tornam iniciativas quase únicas por país.

As federações identificadas no presente estudo estão apresentadas a seguir. O critério utilizado na seleção das federações para o estudo relacionou-se a tipologia dos documentos do acervo, em que apenas federações que agregasse repositórios científicos seriam selecionadas. Assim, federações como a Federação de Repositórios Educa Brasil (FEB) e o The China Digital Museum Project não fazem parte do estudo, visto que um federa repositórios de objetos educacionais e o segundo documentos museológicos.

Lista de federações estudadas no presente trabalho:

- ❖ **Networked Digital Library Theses and Dissertation (NDLTD):** federação americana disponível em: <http://www.ndltd.org/>, agrega repositórios e outras federações, restrita a trabalhos de conclusão de cursos;
- ❖ **Biblioteca Digital Brasileira de Teses e Dissertações (BDTD):** federação brasileira disponível em: <http://bdttd.ibict.br>, agrega repositórios de teses e dissertações;
- ❖ **Portal Brasileiro de Acesso Aberto à Informação Científica (oasisbr):** federação brasileira disponível em: <http://oasis.ibict.br/>, agrega a BDTD, e outras fontes de informações científicas brasileiras;
- ❖ **Repositório Científico de Acesso Aberto de Portugal (RCAAP):** federação pertencente a um projeto amplo de integração de repositórios

portugueses, disponível em: <http://www.rcaap.pt/>, agrega principalmente repositórios acadêmicos portugueses;

- ❖ **LaReferencia:** federação latinoamericana, formada por instituições nacionais de nove países, disponível em: <http://lareferencia.redclara.net/>, agrega a produção científica dos países membros;
- ❖ **Sistema Nacional de Repositorios Digitales (SNRD):** federação argentina disponível em <http://repositoriosdigitales.mincyt.gob.ar:8380/dnet-web-generic/>, agrega repositórios científicos da Argentina.
- ❖ **Biblioteca Digital Colombiana (BDCOL):** Federação Colombiana, disponível em: <http://www.bdcoll.org/>, agrega repositórios científicos da Colômbia;
- ❖ **Red Mexicana de Repositorios Institucionales (REMEDI):** Federação mexicana disponível em: <http://www.remedi.org.mx/>, agrega repositórios científicos mexicanos;
- ❖ **Acceso Libre a Información Científica para la Innovación (ALICIA):** federação peruana disponível em: <http://alicia.concytec.gob.pe/alicia/>, agrega repositórios científicos Peruanos;
- ❖ **Repositório da Produção Científica do Conselho dos Reitores das Universidades Estaduais Paulistas (CRUESP):** Federação restrita brasileira disponível em: <http://www.repositorio.cruesp.sp.gov.br/>, agrega os documentos dos repositórios Institucionais da Universidade de São Paulo, Universidade de Campinas e Universidade Estadual Paulista “Júlio de Mesquita Filho”.

Esse conjunto representa possivelmente grande parte das federações existentes, visto que a quantidade de terminologias apresentadas revela pouca padronização nas nomenclaturas dos sistemas de informação disponíveis na internet. A liberdade intrínseca da internet acarreta numa variação de serviços ofertados que, até certo ponto, dificulta a categorização dos sistemas. Com isso, apresenta certa dificuldade no presente levantamento.



### 5.1.1. NETWORKED DIGITAL LIBRARY OF THESES AND DISSERTATION

A Networked Digital Library of Theses and Dissertation (NDLTD) nasceu da necessidade de adequação dos acadêmicos à era digital, tendo as suas raízes nas discussões sobre teses e dissertações eletrônicas no encontro da University Microfilms International (UMI) em 1987, em Ann Arbor, Michigan. Entretanto, o projeto teve que esperar até 1991 para ter início, na Universidade Virgínia Tech, com a coordenação do reitor Gary Hooper e pelos colaboradores Ed Fox e John Eaton. Entretanto, a NDLTD teve início propriamente dito com o desenvolvimento de um software para a submissão de teses e dissertações eletrônicas, um tipo de repositório exclusivo para esse tipo de documento, sendo finalmente disponibilizado à comunidade em 1996.

Como federação, a NDLTD começou em um encontro na VirginiaTech em 1994, no qual se iniciou os estudos sobre padrões, a quais seriam a base para interoperar repositórios, então chamados bibliotecas digitais. Em 1996 algumas universidades adotam o software ETD db, desenvolvido pela Virginia Tech para criação de um repositório de tese e disponibilizado pela NDLTD, formando-se assim a federação.

Neste sentido, Fox e colaboradores (1997) consideram a NDLTD como um esforço internacional com o propósito de apoiar os acadêmicos com o depósito dos trabalhos finais de graduação, mestrado e doutorado em bibliotecas digitais. Agrega uma vasta tipologia de documentos, visto que dependendo do país denominam-se diferentemente os trabalhos de conclusão de curso. Com isso, reuni esforços individuais para a disseminação de estudos, aumentando a visibilidade dos trabalhos, autores e instituições de defesa. Para tanto, os autores revelam que a NDLTD possibilita:

- apoio para que os estudantes aprendam mais sobre publicação em bibliotecas digitais;
- assegura o depósito eletrônico dos documentos;
- potencializa o acesso aos documentos por meio das bibliotecas digitais.

Hussain e colaboradores (2001) ampliam a conceituação, afirmando que a NDLTD é um esforço colaborativo de universidades por todo o mundo, com o objetivo de criar, arquivar, distribuir e dar acesso livre a teses e dissertações eletrônicas. Com isso, adiciona os processos que descrevem os objetivos da NDLTD.

Ressalta-se a questão das federações serem colaborativas e voluntárias, em que cada elemento participante oferta metadados à federação sem uma obrigatoriedade. Da mesma forma que, requer-se apenas que os repositórios participantes atendam aos critérios estabelecidos pela federação, sendo a única restrição. Torna-se um ambiente em que a tônica é a promoção direta dos documentos pertencentes ao acervo e seus autores e indiretamente às instituições dos autores.

Quanto à tipologia documental, mesmo que apresente certa restrição voltadas aos trabalhos de conclusão de curso, revela a diversidade de terminologia. Enquanto nos Estados Unidos um trabalho de conclusão de doutoramento é denominado de dissertação, aqui no Brasil é tese. Assim, como política para formação de acervo a NDLTD aceita trabalhos monográficos de conclusão de cursos de doutorado, mestrado, bacharelado, até mesmo, relatórios que visam cumprir parcialmente graus, cursos ou pesquisas.

Outro ponto importante é relacionado à possibilidade de acesso ao texto integral dos documentos federados pela NDLTD. É fortemente recomendado que os registros oriundos dos repositórios remetam ao texto integral, com o acesso sem custos. A política de acesso aberto da NDLTD, até certo ponto, se alinha a tipologia documental, visto que as teses, dissertações e Trabalhos de Conclusão de Curso (TCC) não são oficialmente publicada, facilitando as questões relacionadas aos direitos autorais.

No que diz respeito aos arquivos abertos, mais precisamente à interoperabilidade, a NDLTD foi uma das iniciativas pioneiras, ofertando modelos e ferramentas para o desenvolvimento de uma infraestrutura para federalização. Destacam-se, nesse sentido os seguintes pontos, abarcando as principais características dos arquivos abertos:

- Desenvolvimento de uma ferramenta que possibilitasse o autoarquivamento para a criação de provedores de dados;
- Desenvolvimento de um provedor de serviço, com portal de busca consolidada;
- Utilização do protocolo OAI-PMH para intercâmbio de dados;
- Desenvolvimento do padrão de metadados Electronic Theses and Dissertation - Metadata Schema (ETD-MS).

Para o Portal de Busca Consolidada foi utilizando o software de gerenciamento de biblioteca Virtua, desenvolvido pela Visionary Technology in Libray Solutions (VLTS), especialmente customizado para a NDLTD, possivelmente pela proximidade desta empresa com a Virginia Tech. Apresenta a preocupação em aproximar esse serviço de busca aos sistemas de informação, ofertando uma interface de busca próxima aos sistemas de gerenciamento de bibliotecas.

Com relação aos resultados de busca, um problema existente tem base linguística, por ser uma base internacional com documentos oriundos de instituições de ensino e pesquisa de várias partes do mundo. Uma busca simples com termos em idiomas diferentes revela a indexação física ou sintática, em que não há uma relação entre termos em línguas diferentes, conforme quadro 9, com o quantitativo do resultado de busca para o termo “Brasil”, em português, inglês, francês, chinês e japonês.

QUADRO 9 - RELAÇÃO ENTRE RESULTADOS DE BUSCA PELE MESMO TERMO EM IDIOMAS DIFERENTES

Idioma	Termo	Resultado
português	Brasil	54.847
inglês	Brazil	46.775
francês	Bresil	993
chinês	Baxi (巴西)	130
japonês	Burajiru (ブラジル)	11

Fonte: o autor; 2015.

Por fim, revela que a NDLTD inspira a criação de bibliotecas digitais de teses e dissertações pelo mundo, visto que em dezembro de 2014 agregava mais de 3,8 milhões de documentos, nas mais diversas línguas, provenientes de vários países. Da mesma forma que apresenta desafios para integrar as diferenças linguísticas, mesmo com a

restrição na tipologia documental. Apresenta um modelo, baseado na interoperabilidade utilizando o protocolo OAI-PMH e restrito a apenas um esquema de metadados.

### **5.1.2 BIBLIOTECA DIGITAL BRASILEIRA DE TESES E DISSERTAÇÕES**

A Biblioteca Digital Brasileira de Teses e Dissertações (BDTD) teve os trabalhos iniciados em 2002, fortemente inspirados na NDLTD, pelo Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT). Com isso, procurou ofertar uma infraestrutura informacional e tecnológica com o propósito de dar visibilidade à produção científica no âmbito da pós-graduação brasileira.

Diferentemente da NDLTD, a BDTD nasce como uma federação, contando com a colaboração da Universidade de São Paulo (USP), Universidade Federal de Santa Catarina (UFSC) e a Pontifícia Universidade Católica do Rio de Janeiro (PUC-Rio), em um projeto piloto de implantação da interoperabilidade, utilizando o protocolo OAI-PMH e ferramentas disponibilizadas pela NDLTD. A BDTD é uma federação, não possuindo serviços para submissão local, apenas coletando e indexando metadados provenientes das bibliotecas digitais parceiras.

Nesse sentido, Kuramoto (2011) considera a BDTD como um trabalho de unificação de iniciativas isoladas, com o propósito de potencializar a visibilidade da produção acadêmica brasileira, no nível de pós-graduação. Destacam-se as questões relacionadas à integração e visibilidade por meio da federalização de repositórios locais de teses e dissertações.

Southwick (2006) destaca que a BDTD fomenta a colaboração entre as universidades e institutos de pesquisa em prol da disseminação de produção científica nacional. As instituições parceiras podem torna-se proativas e desenvolverem as bibliotecas digitais locais e, assim, assumirem um papel mais ativo na disseminação das teses e dissertações. A BDTD é uma catalisadora na disseminação aberta da produção científica de teses e dissertações por parte das universidades e institutos de pesquisa.

Nesse ponto a BDTD apresenta certa restrição, em relação à NDLTD, ao passo que agrega apenas trabalhos de conclusão de cursos de mestrado e doutorado. Com isso restringe a sua abrangência, ao passo que nem todos os departamentos das universidades

possuem cursos de pós-graduação *strictu sensu*, deixando de agregar a grande produção nacional de trabalhos de conclusão de cursos de graduação e extensão.

No que concerne à interoperabilidade, no entanto, a BDTD se apresenta mais flexível que a NDLTD. Mesmo tendo o MTD2-BR como principal esquema de metadados, a BDTD permite a coleta em outros esquemas, possibilitando a participação de uma vasta gama de repositórios (bibliotecas locais de teses e dissertações), sendo os principais o DSpace e o Teses e Dissertações Eletrônicas (TEDE), como destaca Macedo e colaboradores (2014a)

Nesse mesmo ponto, ressaltam-se os problemas enfrentados pela BDTD com a qualidade dos dados coletados, que afetam as federações. Macedo e colaboradores (2014b) ressaltam que a BDTD apresenta certa falta de normalização dos dados, por possibilitar a agregação de diferentes ferramentas. Da mesma forma, padrões diversos são utilizados na descrição de teses e dissertações por instituições distintas. Este ponto se apresenta sensível, principalmente para indexadores mais físicos e menos semânticos, requerendo ações para aumentar a acurácia na recuperação.

Além de coletar, a BDTD disponibiliza os metadados coletado em diversos esquemas para outras federações, sendo que entrega os metadados em Dublin Core para o OasisBr e em ETD-MS para a NDLTD. Para tanto faz uso de tecnologia *Extensible Stylesheet Language Transformations* (XSLT) para conversão de um esquema de metadados para outros, normalizando alguns campos, para fornecer maior qualidade dos dados.

No que se refere à interface de busca consolidada, pode-se destacar três fases: a implantação com uso de ferramenta com desenvolvimento próprio do Ibict, baseados no gerenciador de banco de dados Oracle; a consolidação com o uso da ferramenta proprietária *OmniFind* da *International Business Machine* (IBM) e; por fim, a modernização com o uso do software livre de descoberta e entrega *VuFind*. Com isso, revela o alinhamento da BDTD com o uso de ferramentas mais apropriadas a recuperação da informação, constituídos por banco de dados com grandes acervos.

Em 2014, a BDTD agrega mais de 260 mil registros de teses e dissertações provenientes de mais de 100 universidades e institutos de pesquisa, sendo o grande fomentador para a criação de repositórios locais de teses e dissertações no Brasil e pioneiro na implementação dos preceitos dos arquivos abertos. Uma grande federação de repositórios de teses e dissertações com o propósito de dar visibilidade à produção acadêmica nacional.

### **5.1.3 PORTAL BRASILEIRO DE ACESSO ABERTO À INFORMAÇÃO CIENTÍFICA**

O Portal Brasileiro de Acesso Aberto à Informação Científica (oasisbr) é uma federação que tem por finalidade integrar toda a produção científica brasileira em acesso aberto, por meio dos repositórios institucionais de universidades, instituições de pesquisa, bibliotecas locais de teses e dissertações e periódicos científicos. Integra, inclusive, a BDTD, de forma a conter no acervo prioritariamente composto de artigos, teses e dissertações, ou seja, de documentos tradicionalmente científicos, avaliados pelos pares.

Segundo o portal, o oasisbr se conceitualiza como uma ferramenta de busca multidisciplinar de documentos científicos de acesso livre, cujos autores são vinculados a institutos de ensino e pesquisa. De certa forma, destaca a característica da federação em ofertar busca consolidada, com a possibilidade de acesso ao texto integral dos registros indexados, revelando um dos principais serviços ofertados pela federação.

Silva, Alencar e Souza (2014), apresentam o oasisbr por meio de vários aspectos, de forma a revelar o cenário abrangente sobre essa federação. No que diz respeito à implantação, segundo os autores, destaca-se o apoio financeiro da Financiadora de Estudos e Projetos (Finep) para o desenvolvimento do portal, que se contextualiza na Expansão e Consolidação do Sistema Nacional de Ciência, Tecnologia e Inovação, tendo como linha direta de ação o Fomento à Tecnologia da Informação e Comunicação do Ibiict;

No que diz respeito à política de indexação apresenta-se bem restrita, coletando apenas documentos avaliados pelos pares. Utiliza o protocolo OAI-PMH e o esquema

de metadados Dublin Core para compor um acervo de documentos provenientes da BDTD, repositórios, revistas e anais de eventos, conforme critérios de indexação rígidos. Com isso, o oasisbr mantém um acervo consolidado da produção científica brasileira.

O oasisbr também interoperava com outras federações, sendo tanto provedor de dados como provedor de serviços. Assim oferta os metadados para o LaReferência, sendo o provedor oficial brasileiro para essa federação. Quanto ao RCAAP, torna-se um espelho, ou seja, possuindo o mesmo acervo, na medida em que coleta metadados e é coletado pelo RCAAP, sincronizando os acervos.

Nesse contexto, por coletar metadados provenientes de várias fontes, um dos problemas é a duplicidade de registros. É bastante comum que um artigo publicado em uma revista seja depositado no repositório institucional da instituição do autor. Com isso se a revista e o repositório forem coletados, o registro apresentará duplicidade. Esse problema se agrava com a tendência atual de colaboração, que autores de instituições diferentes desenvolvem estudos, publicando artigos em revistas e depois depositando nos respectivos repositórios, gerando múltiplas cópias e múltiplos registros no oasis. A figura 20, apresenta um detalhe da página de resultados de busca, apresentado a duplicidade de registros.

FIGURA 20 - DETALHE DA PÁGINA DO OÁSISBR APRESENTANDO RESULTADO DE BUSCA

The screenshot shows the Oasisbr search results page. At the top, there is a navigation bar with links for 'Página Inicial', 'Sobre o oasisbr', 'Política', 'Busca', 'Diretório', 'Estatísticas', 'Contato', and 'Ajuda'. Below the navigation bar, there is a search filter section with 'Filtro personalizado: Documentos do Brasil' and 'Classificar por: Relevância'. The search results show two entries for the same document, indicating duplication. The first entry has a title 'Webometria dos repositórios institucionais acadêmicos, Webometrics of the academic institutional repositories' and an author 'Shintaku, Milton Robredo, Jaime Baptista, Dulce Maria'. The second entry has a title 'Webometria dos repositórios institucionais acadêmicos' and an author 'Milton Shintaku, IBICT Jaime Robredo; UnB Dulce Maria Baptista; UnB'. Both entries include metadata such as dates, identifiers, and URLs.

FONTE: Portal oasisbr; 2014.

O oasisbr apresenta-se como a principal federação brasileira, com mais de 600 mil registros, provenientes de fontes diversas. Torna-se o ponto focal na agregação de registros científicos no Brasil. Ao integrar-se ao RCAAP e a LaReferencia oferta visibilidade a produção científica nacional, ao passo que contribui significativamente com essas outras federações.

#### **5.1.4 REPOSITÓRIO CIENTÍFICO DE ACESSO ABERTO DE PORTUGAL**

O Repositório Científico de Acesso Aberto de Portugal (RCAAP) é uma federação desenvolvida pela Universidade do Minho e apoiada pela Fundação para a Ciência e Tecnologia (FCT), com o propósito de coletar, agregar e indexar conteúdos digitais de acesso livre oriundos dos repositórios institucionais portugueses. O RCAAP faz parte do projeto português para disseminação da informação científica de acesso aberto, como forma de dar visibilidade a produção científica em língua portuguesa, por isso a ligação com o oasisbr.

O Portal do RCAAP foi lançado em julho de 2008, numa iniciativa da Agência da Sociedade do Conhecimento – UMIC, com o apoio da Fundação para a Computação Científica Nacional (FCCN) e execução e coordenação técnica da Universidade do Minho. Com isso, não apenas agregar a produção dos repositórios portugueses existentes, mas também fomentar a criação e hospedagem de novos repositórios.

Da mesma forma que o oasisbr, o RCAAP define-se como um ponto único para a pesquisa, localização e acesso a milhares de documentos de caráter científico e acadêmico. Reforça-se a questão da busca consolidada nas federações, como um dos principais serviços ofertados aos usuários, caracterizando os portais das federações, pois grande parte dos processos da federação volta-se a formar o banco de dados consolidado para ofertar a busca.

Ribeiro e Pinto (2009) consideram o RCAAP como catalizador para a criação de repositórios e adesão ao movimento de Acesso Aberto em Portugal. Nesse mesmo sentido, Carvalho e colaboradores (2010) ressaltam que o RCAAP tem por objetivo:

- **Aumentar a visibilidade, acessibilidade e difusão dos resultados das pesquisas portuguesas:** Integrar em uma única base



informações que possibilitem acessar os textos completos de literatura científica dispersa em várias iniciativas e dar visibilidade a literatura cinzenta;

- **Facilitar o acesso à informação sobre a produção científica:** ao interligar outros sistemas de informações voltadas a Ciência e Tecnologia, como o De Gois, por exemplo, possibilita obter informações sobre a produção científica portuguesa;
- **Integrar Portugal num conjunto de iniciativas internacionais:** Interligar a federação em outras iniciativas congêneres existentes no mundo, como forma de aumentar a visibilidade da produção científica em língua Portuguesa.

Além de coletar metadados para a federação, O RCAAP hospeda repositórios, como forma de fomento ao acesso aberto, pelo Serviço de Alojamento de Repositórios Institucionais (SARI). Com isso, aumenta a oferta de serviços, pois além de dar suporte técnico e tecnológico aos repositórios garante infraestrutura para as instituições criarem seus repositórios e, assim, disseminar a informação científica.

A interligação com sistemas de informação científico distintos apresenta uma tendência, ao passo que, possibilita compartilhamento de informações, que em muitos casos se complementam. O RCAAP, por exemplo, se integra com o DeGóis, cadastro de currículos, semelhante e Plataforma Lattes brasileiro, diminuindo o retrabalho e complementando as informações prestadas.

A adoção das Diretrizes do Digital Repository Infrastructure Vision for European Research (DRIVER) por parte da federação RCAAP visa padronizar alguns campos de metadados para facilitar a interoperabilidade, em um ambiente multilíngue europeu. Com isso, os conteúdos de campos como tipo de documento ou idioma são padronizados por terem possibilidades restritas de preenchimento.

Cabe ressaltar, no entanto, que um problema existente na busca do Portal RCAAP tem origem nas diferenças linguísticas entre o Português do Brasil e de Portugal. Esse problema ocorre também no oasisbr, porém é mais visível no RCAAP pela representatividade dos registros brasileiros, que contribui com 64% dos registros

vindo do oasisbr, fora os documentos de brasileiros depositados nos repositórios portugueses. Assim, diferenças de grafias como "projeto" e "projecto" retornam quantidade de documentos diferentes, pois documentos mais antigos portugueses não adotaram a padronização de grafia do Português.

Acima de tudo, o RCAAP apresenta-se como uma federação, que oferta uma quantidade de serviços importante, não apenas aos usuários, mas aos repositórios que federaliza. Assim, se dispõe mais que um agregador de metadados e uma interface de busca, uma federação que apoia as entidades parceiras. Com isso, torna-se um expoente como federação.

### **5.1.5 LAREFERENCIA**

O LaReferencia é um projeto financiado pelo Banco Interamericano de Desenvolvimento (BID), coordenado pela *Red Cooperação Latino Americana de Redes Avanzadas* (CLARA), com o propósito de agregar informações científicas da América Latina, de forma a aumentar a visibilidade da produção científica desta região. Desenvolver uma federação em que cada país é representado por uma instituição que mantenha ou crie uma federação nacional.

Nascido em 2010 com a participação de oito países latino-americanos, tem como representante do Brasil o Ibict, com o oasisbr. Em 2014 foi lançado o portal de busca consolidado, com metadados coletados das instituições dos países membros, em que o Brasil tem papel de destaque com 75% dos registros disponíveis. Atualmente o LaReferencia é composto por nove países, nomeadamente: Argentina, Brasil, Chile, Colômbia, Equador, El Salvador, México, Peru e Venezuela.

O LaReferencia se intitula rede federada de repositórios científicos, uma iniciativa regional com o propósito de compartilhar e dar visibilidade à produção científica gerada nas instituições de ensino e pesquisa dos países membros. Para tanto, desenvolve parcerias e acordos para se estabelecer com outros países e manter-se sustentável, operando por meio da cooperação entre os seus membros.

Para formar a rede federada o LaReferencia executa os seguintes procedimentos:

- **Acordar:** os países membros acordam em criar a federação local (do país) para ofertar metadados ao LaReferencia, definindo políticas e padrões alinhadas com a federação, tornando o ponto focal do país;
- **Coletar:** o LaReferencia coleta os metadados dos países membros por meio da federação local, caso o país membro ainda não possua uma federação, um repositório expressivo será eleito como ponto focal;
- **Validar:** com o propósito de aumentar a qualidade dos dados o LaReferencia valida a estrutura dos metadados e alguns de seus conteúdos, assegurando certa padronização e normalização.
- **Federar:** realiza-se o processo de integração e indexação dos metadados coletados de forma a ofertar um serviço de busca consolidado;
- **Ofertar:** um portal de busca consolidada e estatística é ofertado para os usuários.

Para a constituição do acervo, o LaReferencia coleta-se principalmente artigos, teses, dissertações e relatórios técnicos, descartando os registros coletados que não atinja esses critérios. Com essa política gerencia um acervo tipicamente científico e textual. Revela certo tradicionalismo, visto que atualmente pode ter outras tipologias documentais de interesse científico, como dados de pesquisa, objetos educacionais e outros.

Destaca-se nessa federação o Portal de busca implementado com um software livre de descoberta e entrega. Com isso, revela uma tendência na integração e indexação de sistemas de informação científica, possibilitando buscar ou navegar pelo acervo, em que o sistema facilita a recuperação, por listas ou facetamento.

Problemas de indexação também ocorrem nessa federação, principalmente por questões linguísticas. Como o acervo é praticamente composto por documentos em Português e Espanhol, a busca pelo mesmo termo em idiomas diferente recuperarão quantidades distintas de documentos. Por exemplo, o termo "ciência e tecnologia" recuperam 3.912 registros, ao passo que "ciencia y tecnologia" recupera 2.506 registros.

Claro que isso ocorre pelo quantitativo de registros brasileiros. Entretanto, destaca-se que oitos dos países membros são hispano falantes.

Por fim, o LaReferencia torna-se uma federação *sui generis*, por não ter uma instituição responsável, mas uma colaboração entre instituições representativa de países. Com mais de 938 mil registros, forma uma federação promissora, pois fornece visibilidade a produção científica regional. Da mesma forma que pode crescer, não apenas com os países latino-americanos, mas com os países ibéricos, o qual compartilham os idiomas.

### **5.1.6 SISTEMA NACIONAL DE REPOSITARIOS DIGITALES**

O *Sistema Nacional de Repositorios Digitales* (SNRD) é uma iniciativa do Ministério da Ciência, Tecnologia e Inovação Produtiva da Argentina, juntamente com o *Cosejo Interistitucional de Ciencia y Tencologia* (CICYT), para criar uma rede interoperável de repositórios digitais. Para tanto, o SNRD estabelece padrões e políticas com o propósito de orientação à adesão a rede.

Criado em 2011 por decreto presidencial foi resultado das orientações do LaReferencia, visto que é o ponto focal argentino para esta federação, inclusive, utiliza a mesma tecnologia. Assim, o SNRD coleta metadados dos repositórios de instituições de ensino e pesquisa argentino e oferta esses metadados para o LaReferencia, utilizando as mesmas ferramentas para coleta, indexação e oferta.

Bongiovani e Nakano (2011) relatam que a finalidade do SNRD, Bongiovani e:

- Promover a adesão ao Acesso Aberto das instituições e, assim, tornar acessível gratuitamente a produção científica e tecnológica argentina;
- Promover o intercâmbio entre as instituições e aumentar o acesso por meio de uma rede interoperável;
- Gerar políticas comuns por meio de colaboração, de forma a favorecer a sustentabilidade dos repositórios;
- Desenvolver estratégias com o propósito de garantir os direitos autorais dos documentos incluídos nos repositórios;

- Definir padrões gerais que orientem o funcionamento da rede;
- Adotar métodos de proteção internacional para a produção científica e tecnológica argentina;
- Fomentar a formação de recursos humanos;
- Contribuir para desenvolver condições adequadas para a gestão e preservação dos documentos dispostos em repositórios;

O respaldo oficial dessa federação possibilita ações mais amplas e apresenta uma preocupação governamental com formas de seleção, coleta, indexação, preservação e disseminação de documentos científicos e tecnológicos produzidos pelas instituições de ensino e pesquisa. Com isso, garante legitimidade na definição de procedimentos, criação de padrões e políticas que asseguram o funcionamento da federação.

O Portal não apresenta a opção de busca avançada e a navegação apresenta poucas opções. Nota-se que a recuperação dos documentos pertencentes ao acervo é um dos principais serviços das federações. Entretanto, o portal do SNRD não oferta muitas opções para a recuperação, o facetamento, por exemplo, existente em grande parte dos portais de federação não foi implementado nesse portal.

Por fim, mesmo que ainda seja uma federação recente apresenta um grau de formalização e oficialização importante a sua manutenção e sustentabilidade. Atualmente integra mais de 40 mil documentos provenientes de oito repositórios, representando parte da produção científica argentina. Com isso, oferta certa garantia aos usuários na disponibilização de conteúdo por longo tempo.

### **5.1.7 BIBLIOTECA DIGITAL COLOMBIANA**

A *Biblioteca Digital Colombiana* (BDCOL) é uma iniciativa da *Red Nacional Académica de Tecnología Avanzada* (RENATA), de forma a constituir uma rede de repositórios com o propósito de dar visibilidade à produção acadêmica colombiana. Com isso ofertar uma infraestrutura para indexação da produção acadêmica, científica e cultural de instituição de ensino superior e pesquisa, entre outros.

A BDCOL nasceu em 2007 e seu desenvolvimento foi dividido em duas fases, sendo a primeira fase responsável pela implementação da infraestrutura tecnológica para a criação do portal de busca consolidada com metadados coletados de repositórios. A segunda fase é a consolidação da primeira fase com a criação de padrões, políticas e orientações que embasasse a federação. Assim, se define como uma rede de repositórios e bibliotecas digitais com a finalidade de aumentar a visibilidade da produção intelectual colombiana.

Otero e Gutiérrez (2011) descrevem a BDCOL como um sistema coletador, em forma de rede, que possibilita o acesso livre os documentos em texto completo por meio da internet. Assim, salienta algumas características das federações, estabelecida por meio de uma rede, possibilitando o acesso ao texto integral e dar aceso via internet. No caso da BDCOL, agregar a produção científica, acadêmica, cultural e patrimonial das instituições voltadas ao desenvolvimento da Colômbia.

A BDCOL é o ponto focal colombiano do LaReferencia, ofertando a essa federação os registros colombianos. Entretanto, a BDCOL agrega uma tipologia documental mais diversificada que o LaReferencia, visto que agrega documentos históricos e objetos de aprendizagem, entre outros que não possui cunho científico. Da mesma forma integra objetos multimídia, enquanto o LaReferencia possui característica mais textual.

Nesse ponto, vê-se que a BDCOL adota conjuntos diferenciados de metadados para cada tipologia documental específica, ajustando conforme as características colombianas. Mesmo que todos os padrões para descrição de coleções sejam baseados no Dublin Core, cada coleção possui o seu próprio padrão, adequando as necessidades colombianas na descrição de objetos digitais.

Dentre as coleções pertencentes à BDCOL destaca-se a presença dos objetos educacionais, que se apresenta como tendência no compartilhamento e reuso de objetos digitais. Com isso alinha a BDCOL ao chamado movimento de objetos educacionais livres, em que universidades têm disponibilizado cursos, aulas e objetos de ensino gratuitamente à comunidade, fomentando a disseminação gratuita e ao reuso dos objetos educacionais.

A amplitude da BDCOL se apresenta relevante, transcendendo a questão da documentação científica. Ao incorporar documentos patrimoniais e históricos de relevância, coloca a discussão sobre os próprios repositórios institucionais e suas políticas de acervo, ao passo que muitos desses documentos possuem importância científica para disciplinas como a história, por exemplo. Assim, revela ser uma federação mais flexível, apropriada a ofertar serviços diferenciados aos usuários.

### **5.1.8 RED MEXICANA DE REPOSITARIOS INSTITUCIONALES**

A *Red Mexicana de Repositorios Institucionales* (REMERI) é uma federação mexicana desenvolvida por instituições membros da *Corporación Universitaria para el Desarrollo de Internet* (CUDI), com apoio do *Consejo Nacional de Ciencia y Tecnología* (CONACYT). Esta rede federada tem por objetivo a integração da produção científica e documental mexicana, fomentando a colaboração e apoiando a disseminação e acesso de conteúdos em acesso aberto.

O REMERI teve início em 2012, baseando-se em um levantamento efetuados com as universidades, como forma de verificar o cenário para a construção da federação. Assim, foi desenvolvida coletando metadados de dez instituições que possuíam repositórios, ou seja, 38% das instituições mexicanas, em 2013 já coletava de 27 instituições, um crescimento significativo.

Os documentos oficiais da REMERI a definem como uma rede constituída por instituições de ensino públicas e privadas coordenadas por um conselho, com o objetivo de agregar, difundir, preservar e dar visibilidade a produção científica, acadêmica e documental. Da mesma forma que, por ser interoperável, integra-se às redes internacionais, contribuindo a colaboração e divulgação de conteúdos de acesso aberto.

Tapia (2013) descreve a REMERI como sendo um projeto com restrição à produção mexicana, de forma a atender a necessidade de infraestrutura tecnológica para integrar e dar visibilidade a produção educacional e de pesquisa, contidas nos repositórios de acesso aberto. Reforça o conceito de federação com restrição a um país e a questão da infraestrutura ofertada por esses sistemas, que transcendem a pura questão tecnológica.

O REMERI é o ponto focal mexicano que responde ao LaReferencia. Nota-se, assim, a forte influencia desta federação no REMERI, nas orientações e políticas definidas, assim como nos padrões adotados, como no uso do metadados alinhados com as diretrizes DRIVER. No que diz respeito à tipologia de documentos coletáveis, por exemplo, a REMERI se apresenta flexível compondo-se de cinco grandes categorias:

- **produção científica:** documentos tradicionais científicos acrescido dos dados de pesquisa e softwares;
- **produção acadêmica:** objetos educacionais e outros produzidos no âmbito das universidades, como nota de aula, trabalho de conclusão de cursos e posters;
- **produção cultural e documental:** documentos de valor histórico e cultural como mapas, folhetos e manuscritos, que apresentem significancia;
- **produção administrativa:** documentação de cunho administrativo relacionado a projetos e programas.

Com essa diversidade documental apresenta-se como solução e problema. Como solução, a amplitude de tipos de documentos, assuntos, idiomas e outros que atendem a variados tipos de usuários. Como problema revela-se nos resultados de busca e na falta de navegabilidade. Não há uma categorização nos resultados de busca, apenas uma grande lista, que dificulta a seleção dos documentos desejados. O facetamento, por exemplo, facilitaria o refino de busca e a categorização, mas o REMERE ainda não o implementa. O REMERI também não oferta formas de navegação pelo acervo.

Um ponto se destaca no REMERI, a questão da sustentabilidade definida um forma de política. Por não ter uma instituição mantenedora, o REMERI adota várias formas de arrecadação para manter-se. Com isso, a conta é dividida por quotas dos membros, por órgãos governamentais e pela iniciativa privada, por meio patrocínio, em um modelo menos centralizador.

Mesmo com poucos anos de existência, o REMERI apresenta um desenvolvimento significativo. Da mesma foram que se torna uma fonte relevante de informações, principalmente pela importância do México na America Latina. Contribui-



se para a representatividade da pesquisa nesta região, que ainda apresenta pouca representatividade no mundo.

### **5.1.9 ACCESO LIBRE A INFORMACIÓN CIENTÍFICA PARA LA INNOVACIÓN**

O *Acceso Libre a Información Científica para la Innovación* (ALICIA) é uma federação peruana administrado pelo *Consejo Nacional de Ciencia, Tecnología e Innovación Tecnológica* (CONCYTEC), de forma a constituir uma rede peruana de repositórios interoperáveis. Para tanto, estabelece orientações, políticas, padrões, protocolos e outros de forma a conformar a rede.

Formalizado por lei datada de 2013, o ALICIA é recente, mas bem institucionalizado, contextualizado no chamado governo aberto para gestão e armazenamento do conhecimento. Revela o alinhamento do governo peruano ao acesso aberto, fomentando a disseminação da produção científica em repositórios, de forma a agregar essa produção na federação ALICIA.

Segundo o próprio Portal, o ALICIA se denomina como um repositório nacional, um site centralizado, que mantém a informação digital resultado da produção científica, tecnológica e de inovação, com o propósito de preservar e disseminar em acesso aberto. Mesmo coletando metadados de outras instituições, como uma federação, faz referencia a repositório central, mesmo que não mantenha os objetos digitais. Como grande parte das federações, o ALICIA coleta e indexa os metadados, remetendo ao texto integral que fica nas fontes coletadas.

Cabe salientar que o ALICIA é mais que uma federação, por fazer parte de um esforço peruano para disseminação da produção científica em acesso aberto formalizado por lei, que determina a publicação em acesso aberto a toda instituição pública ou financiada por fundos públicos. Com isso, abriga várias iniciativas voltadas ao fomento de criação de repositórios e depósito da produção.

Da mesma forma que as outras federações latino-americanas posteriores ao LaReferencia, o ALICIA possui fortes influencias deste sistema de informação, sendo o ponto focal do Peru. As políticas adotadas são próximas ao do LaReferencia, assim como os metadados e padrões. Nesse sentido, quanto aos documentos disseminados

pelo ALICIA, revelam-se com tipologia diversa, dos tradicionais científicos avaliados pelos pares, dados de pesquisa, trabalhos técnicos e outros.

Destaca-se que o ALICIA utiliza os mesmas ferramentas para o portal do LaReferencia, fazendo uso das mesmas tecnologias. O uso de softwares livres de descoberta e entrega ajuda na recuperação das informações, apoiando os usuários, sendo uma opção apropriada a esses sistemas. Da mesma forma que o uso de softwares livres diminui o custo dos projetos.

O ALICIA ainda apresenta um acervo pequeno, com pouco mais de 16 mil registros, apresentando um desafio manter o crescimento com a adesão de novos parceiros e manter-se sustentável, mesmo que mantido por instituição pública. Este parece ser um problema das federações manter-se técnica e tecnologicamente atualizadas, de forma a ser sustentáveis e atrativas aos parceiros.

Por fim, cabe salientar a importância do ALICIA no âmbito da comunicação científica peruana, visto que esse país teve a primeira instituição de ensino superior das Américas, a Universidade de San Marcos de 1530. O alinhamento do governo ao acesso aberto revela a preocupação na visibilidade da produção científica peruana, dando ao ALICIA uma importância impar.

#### **5.1.10 REPOSITÓRIO DA PRODUÇÃO CIENTÍFICA DO CONSELHO DE REITORES DAS UNIVERSIDADES ESTADUAIS PAULISTAS**

O estado de São Paulo mantém três das mais importantes universidades brasileiras, a Universidade de São Paulo (USP), Universidade Estadual de Campinas (UNICAMP) e a Universidade Estadual Paulista Júlio de Mesquita Filho (UNESP), abrigando a Conselho de Reitores das Universidades Estaduais Paulistas (CRUESP). Assim, reunindo a produção em uma federação restrita às universidades no âmbito do conselho, para preservar e proporcionar acesso aberto, público e integrado à produção científica de docentes, pesquisadores, alunos e servidores dessas universidades.

A Federação da CRUESP tem início em 2012, por meio de uma reunião de representantes das três universidades com a finalidade de discutir o acesso aberto à produção intelectual, como forma de promover as instituições, para melhoria nos

critérios utilizados nos rankings internacionais de avaliação de universidades. Em 2013 o Portal do Repositório da Produção Científica do CRUESP foi lançado com serviço de busca consolidada implementada com um sistema de descoberta e entrega, como relata Ferreira (2014).

Apoiada pelo conselho de representantes das três universidades teve como parceria a Fundação de Amparo à Pesquisa do Estado de São Paulo (Fapesp). A Federação CRUESP se denomina repositório, enaltecendo a oferta de uma ferramenta de busca consolidada executada por ferramenta atual de descoberta e entrega, propiciando um ambiente apropriado a disseminação da informação.

O acervo da federação é coletado dos repositórios institucionais das três universidades e possui um cunho mais científico, com registros de documentação tradicional avaliada pelos pares. Entretanto, como relatam Ferreira e colaboradores (2013), futuramente outras tipologias serão agregadas. Utiliza padrões internacionais, de forma a assegurar a interoperabilidade com outros serviços de disponibilização de conteúdo. Assim, adotam esquemas de metadados como o Dublin Core e o LOM, normalizado campos conforme as diretrizes DRIVER.

Mesmo utilizando uma ferramenta potente e inovadora para recuperação, o portal da federação da CRUESP não oferta a opção de navegação pelo acervo. Possui busca e facetamento que possibilita refinamento e certa navegabilidade, somente com os resultados de busca. Não a nenhuma opção na página inicial que possibilite navegar pelos registros do acervo, utilizando algum critério, que poderia ser os mesmos do facetamento, ou da organização dos repositórios.

A restrição na participação de repositórios torna essa federação única, pois grande parte das federações busca por novas adesões. A federação da CRUESP, por sua vez, é fechada a somente as instituições estaduais paulistas. Com essa iniciativa, a CRUESP oferta à comunidade uma federação que agrega parte de destaque da produção científica brasileira, na medida em que, em grande parte das avaliações as três universidades estaduais paulista ocupam lugar de destaque.

### **5.1.11 SOBRE AS FEDERAÇÕES**

A coleta de informações sobre as federações possibilitou levantar os dados das variáveis qualitativas relacionadas aos objetivos específicos. Assim, possibilitou discutir questões relacionadas às federações, de um ponto de vista mais analítico, de forma a levantar as questões seguintes:

#### **5.1.11.1 SOBRE OS CONCEITOS:**

O estudo revelou uma variada presença de nomenclatura, como biblioteca digital, rede, sistema, repositório e portal, que reflete a própria condição da internet, em que a variação terminológica pode apresentar um problema, mas também apresenta uma riqueza a ser estudada. Com isso, apresenta conceitos diversos, mas similares e complementares.

Em relação a conceitos, tem-se que as federação pode ser considerada como: *esforço internacional (NDLTD), regional (LaReferencia) ou unificação de iniciativas isoladas (BDTD) formando colaborativamente uma rede de repositórios ou bibliotecas digitais (LaReferencia, SNRD e REMERI), utilizando um sistema para coletar os metadados (BDCOL), para ofertar um ponto único de acesso (RCAAP) a documentação, com a utilização de uma ferramenta de busca (oasisbr, CRUESP).* Revela-se, dessa forma, que na sua auto-conceituação cada federação destaca pontos que se complementam.

Independente da nomenclatura ou ponto de destaque, todas as federações têm por objetivo potencializar a visibilidade dos documentos que compõe o seu acervo, dos autores e das instituições parceiras. Para tanto, coletam, indexam, preservam e dão acesso às informações coletadas de repositórios por meio de um portal de serviços e ofertando esses metadados para outras federações.

#### **5.1.11.2 SOBRE AS CARACTERÍSTICAS**

Quanto às características as federações estudadas apresentaram uma variedade de pontos. Nota-se que em grande parte dos casos as características são as mesmas por serem federações com objetivos muito próximos, praticamente similares. Entretanto,

cada federação apresenta uma contextualização que a difere. Assim, podem-se apresentar algumas características distintas, como:

- ❖ ser criada por regulamentação governamental (ALICIA e SNRD);
- ❖ ser mantida por instituição pública (ALICIA, SNRD, BDTD, oasisbr, RECAAP, CRUESP);
- ❖ ter uma organização própria para manutenção (NDLTD, REMERI, LaReferencia, BDCOL)
- ❖ utilização de padrões para interoperabilidade como: Dublin Core (LaReferencia, oasisbr, SNRD, ALICIA, REMERI, RECAAP, CRUESP), ETD-MS (NDLTD), baseados em Dublin Core (BDTD e BDCOL);
- ❖ fomentar a criação de repositórios (ALICIA, BDTD, BDCOL, NDLTD, RECAAP, REMERI, SNRD, LaReferencia, oasisbr)
- ❖ instituições parceiras são colaborativas e voluntárias, requerendo apenas atender aos critérios determinados (ALICIA, BDTD, BDCOL, NDLTD, RECAAP, REMERI, SNRD, LaReferencia, oasisbr);
- ❖ possuem portal de busca consolidada (ALICIA, BDTD, BDCOL, NDLTD, RECAAP, REMERI, SNRD, LaReferencia, oasisbr, CRUESP);
- ❖ ofertam dados para outras federações (ALICIA, BDTD, BDCOL, RECAAP, REMERI, SNRD, oasisbr);
- ❖ hospedar repositórios (ALICIA e RECAAP).

A colaboração aparece como a tônica das federações. Esse é um dos pontos que mais caracterizam as federações, visto que a adesão a uma federação é totalmente voluntária e que cada membro federado mantém a sua independência. Com isso, tem-se um modelo em que a federação é constituída de membros independentes e autônomos, formando uma rede colaborativa, ofertando serviços que aumentam a visibilidade mútua.

Assim, o estabelecimento da federação se dá pro meio da interoperabilidade e um portal de busca consolidada, que também a caracterizam no que tange a questões mais técnicas e tecnológicas. Nesse sentido, o uso de esquemas mais simplificados como o Dublin Core ainda se faz presente na interoperabilidade com o uso do protocolo

OAI-PMH, revelando o mesmo modelo de operacionalização do início das federações com a NDLTD.

Um ponto de destaque se dá ao passo que as federações se tornam promotoras dos repositórios ou bibliotecas digitais, fomentando a criação e colaboração. Com isso, esses sistemas apoiam a disseminação da informação, não apenas por agregar a informação coletada, mas no apoio ao desenvolvimento de repositórios institucionais. Assim, tornam-se centros de divulgação do acesso aberto, apoiando os repositórios membros.

Nesse contexto, as características das federações estudadas revelam pontos comuns a algumas federações. Em poucos casos há certa unanimidade nas características, mas é nas diferenças que apresenta a riqueza desses sistemas, visto que são as contextualizações de cada um. Em cada federação, mesmo que possuam muito em comum, sempre há uma diferença que a torna única.

### **5.1.11.3 SOBRE AS POLÍTICAS**

As políticas são orientações de cada uma das federações, que remete ao seu funcionamento. Nesse sentido, a política mais presente nas federações diz respeito à tipologia documental a ser coletada e indexada. Nesse ponto, mesmo sendo federações de cunho científico, há certa divergência, visto que atualmente não há consenso do que é documentação científica. Assim a tipologia documental segundo as políticas apresentadas pelas federações foi:

- ❖ documentos científicos avaliados pelos pares (ALCIA, BDTD, BDCOL, NDLTD, RECAAP, REMERI, SNRD, LaReferencia, oasisbr, CRUESP);
- ❖ teses e dissertações (ALCIA, BDTD, BDCOL, NDLTD, RECAAP, REMERI, SNRD, LaReferencia, oasisbr, CRUESP);
- ❖ dados de pesquisa (ALICIA, REMERI);
- ❖ documentação técnica (ALICIA, LaReferencia, REMERI, BDCOL, SNRD, CRUESP);
- ❖ documentos culturais e históricos (BDCOL, REMERI);
- ❖ objetos educacionais (REMERI, BDCOL).

A documentação científica tradicional avaliada pelos pares impera nas federações estudadas. Entretanto outras tipologias são coletadas e disseminadas por parte das federações apresentando um cenário mais flexível, o que reflete diretamente na forma de coleta e aos tipos de repositórios membros. Com isso levantando questões sobre o que é documentação científica atualmente.

Algumas federações já coletam registros sobre dados de pesquisa, alinhados ao movimento de Dados Abertos, da mesma forma que coletam documentação técnica e objetos educacionais que possuem menor relevância científica no sentido tradicional. Entretanto, essas outras tipologias documentais revelam-se importantes como produção acadêmica e intelectual. Essa questão coloca em cheque algumas políticas mais restritivas das federações.

Nota-se, no entanto, que a política da federação relacionada à tipologia documental pode ser revista e alterada a qualquer momento incluindo novas tipologias, adequando-se a tendências ou necessidades e, assim, afetar os repositórios. Da mesma forma, alterações nas políticas dos repositórios podem, em alguns casos, afetar as federações, ao passo que passam a gerenciar e ofertar registros de outros tipos de documentos. Esse sistema de inter-relacionamento revela-se útil a ambos.

#### **5.1.11.4 SOBRE OS PROBLEMAS**

O levantamento dos problemas apresentados nas federações possibilita planejar modelos ou ajustes que proponha soluções adequadas. Assim, contribuir com as federações existentes, apresentando soluções a problemas, que em alguns casos são comuns. Entretanto, na maioria dos casos, muitos dos problemas apresentados são oportunidades de estudos ou desafios a serem suplantados. Os problemas levantados foram:

- ❖ Qualidade dos dados coletados (BDTD, NDLTD);
- ❖ Questões de indexação linguísticas (NDLTD, LaReferencia, RECAAP);
- ❖ Navegabilidade (REMEDI, CRUESP);
- ❖ Problema de busca (SNRD);
- ❖ Duplicidade de registros (oasisbr).

Como as federações possuem um fluxo de informação simplificado composto pelos processos de coleta, indexação e recuperação, podem-se categorizar os problemas conforme essas etapas. Com isso, facilita a análise, ao ponto que os problemas podem ter implicações se estiverem relacionados a um mesmo processo.

Nota-se que não foi apresentado nenhum problema relacionado à coleta de metadados, visto que é um processo simples. Por meio do uso de um protocolo de comunicação comum, a federação coleta os metadados dos repositórios membros, num processo de transferência. Assim, os problemas que afetam essa etapa geralmente impedem a operacionalização da federação.

A indexação, por outro lado, revela problemas, pois tem reflexo na recuperação. Três dos principais problemas apresentados tem relação com esse processo, sendo:

- ✓ **Quanto à qualidade dos dados:** a padronização de campos e a normalização de conteúdos se apresentam como fatores de problemas na indexação. Registros que não apresentam campos obrigatórios, por exemplo, impedem a plena indexação, gerando descarte do registro em muitos casos. Variações de conteúdos em muitos campos influenciam negativamente na recuperação ou categorização. Campos que necessitam de formatos específicos (data, números etc) ou restrição de valores (sim/não, tese/dissertação etc) muitas vezes apresentam grandes variações, influenciando na indexação.
- ✓ **Quanto às questões linguísticas:** a diversidade de registros com idiomas distintos revela-se como problemas às federações, que intrinsecamente possuem registros de várias fontes. Da mesma forma que as homônimas são problemas complicados, que mesmo com ferramentas de web semântica ainda não tem solução condizente. Assim, indexadores mais semânticos resolveriam em parte o problema, mas atualmente grande parte dos indexadores é sintática.
- ✓ **Quanto à duplicidade de registros:** Esse problema ocorre geralmente quando dois registros iguais veem de fontes distintas, por isso não sendo problema do processo de coleta. Nesse sentido o indexador pode resolver com ferramentas mais sofisticadas de indexação que detectam a



duplicidade e apresenta como um único registro, com mais que uma fonte.

A recuperação na forma de busca e apresentação é o processo de maior interação com os usuários e apresentou os seguintes problemas:

- ✓ **Quanto à busca:** geralmente os portais de busca ofertam a busca simples e a avançada, cada qual com suas especificidades.
- ✓ **Quanto à navegabilidade:** os portais devem ofertar a possibilidade dos usuário navegar pelo acervo, ofertando listas categorizadas e facetas que apoiem o usuário a navegação pelo acervo.

Os problemas de recuperação, independente do tipo, são resolvidos com ferramentas de descoberta e entrega que, na maioria dos casos, ofertam todas as funcionalidades para o processo de recuperação. Por isso, grande parte das federações tem adotado essas ferramentas em seus portais.

Por fim, um dos problemas não mencionado explicitamente, mas que se apresenta latente em todas as federações refere-se a manter-se tecnologicamente atualizado, visto a dinamicidade da Ciência da Computação. Requer, na maioria dos casos, equipe de informática dedicada à federação, com processo de prospecção contínua. Assim, esse problema tem relação íntima com a sustentabilidade da federação.

#### **5.1.11.5 SOBRE AS TENDÊNCIAS**

As tendências são altamente dinâmicas devido às mudanças técnicas e tecnológicas. Da mesma forma, políticas adotadas podem apoiar certas tendências, visto que muitas vezes possuem alinhamento teórico. Nesse ponto, o levantamento e análise das tendências possibilitam verificar o cenário mais provável e qual o caminho que algumas federações estão tomando. Assim, as tendências levantadas são:

- ❖ Uso de sistema de descoberta e entrega em software livre para o desenvolvimento do portal de busca (ALICIA, BDCOL, LaReferencia, CRUESP);

- ❖ Manter objetos de aprendizagem alinhando ao movimento de objetos educacionais livres (BDCOL, REMERI);
- ❖ Uso de mais de um esquema de metadados, apresentando flexibilidade na interoperabilidade (BDTD);
- ❖ Operação em ambiente heterogêneo, com uso de padrões internacionais (ALCIA, BDTD, BDCOL, NDLTD, RECAAP, REMERI, SNRD, LaReferencia, oasisbr e CRUESP);
- ❖ Modelo de sustentabilidades amplo (REMERI, LaReferencia).

O uso de software livre tem se tornado comum e, por isso, uma grande quantidade de tipos de ferramentas tem sido disponibilizada, incluindo os softwares de descoberta e entrega que as federações utilizam. Com isso, os portais de busca consolidada das federações ofertam formas de recuperação mais adequadas a sistemas de informação que tratam com grandes quantidades de registros.

Outro ponto importante é a integração em cenário heterogêneo, em que cada repositório membro da federação pode ser desenvolvido por um tipo de ferramenta. A federalização é possível utilizando ou não padrões mais rígidos, requerendo apenas que sejam conhecidos. Essa questão permite a flexibilização necessária à operacionalização mais ampla, com a possibilidade de adesão de uma variedade de tipos de ferramentas.

Dentre todas as tendências, provavelmente uma das mais controversas repousa na questão da sustentabilidade. Mesmo que as federações sejam mantidas por instituições públicas, há vários pontos relacionados à sustentabilidade que merecem discussão. No que se refere à questão financeira, a REMERI e LaReferencia revelam modelos de sustentabilidade em que membros e entidades públicas e privadas contribuem com cotas. A NDLTD e BDTD, por exemplo, que possuem mais tempo revelam sustentabilidade, no sentido de longevidade e ser interessante aos membros, por ofertar apoio, modelo que as outras federações têm seguido.

Independente dos contextos de cada federação vê-se que cada uma se adequa às necessidades, de forma a criar modelos de funcionamento que podem ser seguidos. Com isso, revela-se certa cooperação entre federações, como ocorreu entre a NDLTD e BDTD e mais recentemente com a LaReferencia, REMERI, ALICIA e SNRD. Essa

parece ser a maior das tendências entre esses sistemas de informação, mesmo que não esteja explicitado nos documentos.

## 5.2 CARACTERÍSTICAS DOS REPOSITÓRIOS QUE INFLUENCIAM NA FEDERAÇÃO DE REPOSITÓRIOS

As federações e os repositórios possuem uma relação de dependência, em que não há federação sem que haja repositórios, ofertando os conteúdos para a federação. As federações, por sua vez, amplia a visibilidade dos repositórios. Nesse caso, as características dos repositórios influenciam nas federações, mesmo que preencham os requisitos para a participação. Por isso, o estudo das características torna-se uma necessidade na identificação, análise e proposição de um modelo para federação de repositórios.

Cabe ressaltar que tanto as federações, quanto os repositórios são sujeitos da pesquisa. Assim, com o propósito de analisar os repositórios, levantando as características que influenciam na federalização, esta etapa da pesquisa foi dividida em quatro partes, sendo:

- ❖ **A análise dos portais dos repositórios:** levantar as políticas, tipologia documental e esquemas de metadados, variáveis qualitativas que influenciam na participação de um repositório em uma federação;
- ❖ **A análise da adesão dos repositórios às tendências tecnológicas:** levantar o quanto os repositórios tem aderido às tendências tecnológicas, que podem indiretamente, apoiar os processos de integração com as federações.
- ❖ **A análise da adesão dos repositórios às tendências da ciência:** levantar o quanto os repositórios tem aderido às tendências da ciência, que podem influenciar no comportamento das federações, visto que as federações precisam refletir os repositórios federados.
- ❖ **Análise da oferta de meta-informação pelos repositórios:** Levantar o quanto os repositórios acadêmicos tem gerado e ofertado informações estratégica para tomada de decisão.

As quatro etapas se complementam, ao formar o cenário com características que influenciam a formação de federações. Em cada etapa observa-se algumas das

características dos repositórios, que até certo ponto, vão influenciar nas participações dos repositórios nas federações.

### 5.2.1 ANÁLISE DOS PORTAIS DOS REPOSITÓRIOS

A análise dos portais dos repositórios possibilita levantar características dos repositórios que influenciam na formação de uma federação, levantando os detalhes dos repositórios verificáveis pelo portal e por literatura. Nesse sentido, a presente seção analisa de forma qualitativa os seguintes pontos:

- ❖ **Políticas dos repositórios:** como as políticas são orientações para as atividades do repositório, direcionam questões básicas como tipo de documentos e acesso, processo de depósito entre outras;
- ❖ **Tipologia documental:** como as federações coletam registros sobre os documentos que compõem o acervo dos repositórios, a tipologia documental de um repositório tem um caráter importante;
- ❖ **Esquemas de metadados:** a interoperabilidade dá-se por meio de um esquema de metadados, um padrão utilizado pela federação e pelo repositório.

Para a análise dos Portais de repositórios selecionou-se os dez repositórios brasileiros melhores posicionados na Web Ranking of Repositories no segundo semestre de 2014, sendo:

- Biblioteca Digital de Teses e Dissertação da Universidade de São Paulo;
- Repositório Digital da Universidade Federal do Rio Grande do Sul;
- Repositório Acesso Livre à Informação Científica da Embrapa;
- Biblioteca de Teses e Dissertações da Universidade Federal do Paraná;
- Repositório Institucional da Universidade Federal da Bahia;
- Repositório Institucional Universidade Federal de Santa Catarina;
- Sistema Maxwell da Pontifícia Universidade Católica do Rio de Janeiro;

- Repositório da Universidade de Brasília;
- Acervo Digital da Universidade Estadual Paulista Júlio de Mesquita Filho;
- Repositório Institucional Fundação Oswaldo Cruz.

A amostra é representativa, visto que atende a critérios de ranqueamento, levando em conta o tamanho do acervo do repositório e sua visibilidade, entre outros pontos. Mesmo que esses critérios não representem unanimidade entre os estudiosos, classifica os repositórios levando em conta pontos variados, conforme pode ser observado pela própria nomenclatura e distribuição geográfica.

### **Biblioteca Digital de Teses e Dissertações da Universidade de São Paulo**

A Biblioteca Digital de Teses e Dissertações da Universidade de São Paulo (BDTD da USP) é um repositório exclusivo de teses e dissertações mantidas pela Universidade de São Paulo (USP) para disseminar a produção dos programas de pós-graduação desta instituição. Nascida em 2001 e fazendo parte do projeto piloto da federação BDTD, atualmente possui o maior acervo brasileiro de teses e dissertações de uma instituição, em texto completo disponível na internet em acesso aberto. Uma importante fonte de informação científica, tanto pelo quantitativo, quanto pela relevância da instituição.

A BDTD da USP foi desenvolvida totalmente pela USP, utilizando o modelo fornecido pela ND LTD, como relatam Masieiro e colaboradores (2001), destacando que levou um ano para se completar. Há uma integração entre o sistema de gerenciamento de biblioteca "Dedalus" com a BDTD da USP, de forma que o processo de catalogação é único, envolvendo o autor e a equipe da biblioteca. Por causa dessa integração a recuperação da tese pode ser feita em vários sistemas da USP.

Mesmo com a restrição da BDTD da USP por teses e dissertações, Masieiro e colaboradores (opus cit) revelam que não hospedam apenas textos, mas teses ou dissertações mais complexas envolvendo tipos de arquivos multimídia. Há uma orientação ao acesso aberto por todo o acervo da BDTD da USP, exceto é claro, os

protegidos por patentes ou outras situações similares. A BDTD da USP participa da federação da BDTD entregando metadados.

Quanto à interoperabilidade a BDTD da USP responde em oai-dc (padrão para a interoperabilidade via OAI-PMH), mtd2-br e etd-ms. Com esses esquemas de metadados, a BDTD da USP ratifica ser um repositório restrito a teses e dissertações visto que tanto o MTD-BR, quanto o ETD-MS são exclusivos para interoperar registros de teses e dissertações. O MTD-BR é utilizado pela BDTD do Ibict e ETD-MS pela NDLTD, federações restritas a documentação de final de curso.

Assim, a BDTD da USP apresenta visibilidade pela representatividade na quantidade de teses e dissertações, mas não se conforma com um repositório institucional, por ser restrito a apenas uma tipologia de documentos. É importante mencionar que a USP possui outro repositório, que dissemina a produção acadêmica desta instituição. Entretanto, possivelmente pelo acervo e tempo de vida, a BDTD da USP possui maior visibilidade, visto que o repositório da USP data de 2012, mais de dez anos depois da BDTD da USP.

### **Repositório Digital da Universidade Federal do Rio Grande do Sul**

O Repositório Digital da Universidade Federal do Rio Grande do Sul, denominada de LUME, é o repositório institucional da Universidade Federal do Rio Grande do Sul (UFRGS). Tem como propósito reunir, preservar, divulgar e garantir o acesso confiável e permanente aos documentos acadêmicos, científicos, artísticos e administrativos gerados nesta instituição, dando visibilidade a essa produção.

Nascida em 2007 como forma de modernizar e ampliar a biblioteca de teses e dissertações da UFRGS, o LUME foi desenvolvido utilizando o software livre DSpace, como relata Pavão e colaboradores (2008). Com isso, o LUME assume o papel de reunir a documentação distribuída pela universidade, catalogando-a de forma padrão e disseminando de forma organizada. A utilização do Dspace facilitou a implementação do LUME e atender aos seus objetivos.

A obrigatoriedade de depósito das teses e dissertações é regulamentada por resolução da instituição, por meio do decanato de pós-graduação. Ampara-se o

repositório na disseminação dessa literatura, que há pouco tempo era denominada de cinza, principalmente pela disseminação. Regulamenta o seu depósito, formalizando a obrigatoriedade de depósito no LUME, pois por não ser publicado é o tipo de literatura própria dos repositórios.

Quanto à documentação existente no acervo do LUME se apresenta mais flexível, ao passo que contém não apenas documentação científica tradicional avaliada pelos pares, mas documentos históricos, de patrimônio cultural e administrativo, representando a produção acadêmica da instituição. Apresenta uma organização voltada ao acervo, com coleções compostas por tipos de documentos variados no formato e tipo, imagens, vídeos, áudio e texto, em um amplo acervo.

No que concerne a interoperabilidade, o LUME responde com vários esquemas de metadados, como o oai-dc, mtd2-br, qdc, rdf, ore e mets. Com isso, reforça-se a questão de ter um acervo diversificado e poder interoperar com várias federações, utilizando o esquema de metadados mais condizente, do padrão de interoperabilidade oai-dc ao ore e rdf que possibilitam interoperar com objetos digitais complexos compostos de vários arquivos e que requerem uma hierarquia.

Dessa forma, destaca-se que a importância do LUME para a caracterização dos repositórios institucionais pela sua formalização, por meio de políticas regulamentadas pela instituição. Da mesma forma que preserva e dá acesso à importante documentação cultural e histórica da instituição, sendo referência para outras instituições, visto a inovação técnica e tecnológica apresentada.

### **Repositório Acesso Livre à Informação Científica da Embrapa**

O Repositório Acesso Livre à Informação Científica da Embrapa, denominado de ALICE, é o repositório institucional científico da Empresa Brasileira de Pesquisa Agropecuária (EMBRAPA), visto que possui outro repositório restrito a produção técnica da instituição o Informação Tecnológica em Agricultura (INFOTECA). Assim, o ALICE tem por objetivo, organizar, armazenar, preservar e disseminar, na íntegra, informações científicas produzidas por pesquisadores da Embrapa.



Com lançamento feito em 2011, o ALICE e o INFOTECA foram desenvolvidos com o software livre DSpace, de forma a acelerar a suas implementações, mesmo que por equipes diferentes da EMBRAPA. Castro, Pelufê e Arruda (2013) relatam que a utilização de tecnologia aceita internacionalmente facilita a interoperabilidade com uma rede global de informação científica, aumentando a visibilidade da instituição e seus autores. O INFOTECA integra a federação oasisbr e, por conseguinte o RCAAP.

O ALICE possui um acervo tradicionalmente científico avaliado pelos pares, composto por artigos de periódicos e eventos, livros, capítulos de livros, teses, dissertações entre outros de autoria dos pesquisadores da EMBRAPA. Toda a produção é catalogada, mas somente os documentos que não apresentem restrição são disponibilizados na íntegra, aderindo totalmente ao acesso aberto. Todo o processo de catalogação é feito pela equipe da biblioteca sem a intervenção dos autores, apresentando um histórico da produção científica da EMBRAPA.

Quanto à interoperabilidade o ALICE interopera nos seguintes esquemas de metadados: oai-dc, rdf e mets. Mesmo contendo teses e dissertações não interopera em mtd2-br ou etd-ms, pois as teses e dissertações que compõe o acervo do ALICE foram defendida em outras instituições, a EMBRAPA não possui programa de pós-graduação. Assim, os esquemas de metadados são para disseminação geral da produção científica.

A EMBRAPA figura entre as instituições brasileiras de pesquisa de maior visibilidade, por isso, da importância do seu repositório institucional. Dissemina toda a produção científica, que possui os direitos patrimoniais abertos da instituição, em texto integral. Com isso, contribui com a democratização da informação científica na área em que o Brasil é expoente, elevando a posição da ciência brasileira no cenário do acesso aberto.

### **Biblioteca de Teses e Dissertações da Universidade Federal do Paraná**

A Biblioteca de Teses e Dissertações da Universidade Federal do Paraná tornou-se Repositório Digital Institucional - UFPR como forma a ampliar a sua participação em disseminar a produção acadêmica da Instituição. Anteriormente era restrita a disseminar a produção de teses e dissertações defendidas nesta instituição, como um

parte do serviço ofertado pelo Sistema de Bibliotecas da Universidade federal do Paraná (UFPR).

O Repositório Digital Institucional da UFPR tem seu início em 2004 ainda como Biblioteca de Teses e Dissertações da Universidade Federal do Paraná. Desde o início foi desenvolvida com a ferramenta livre DSpace, que é amplamente utilizada no Brasil para implementar repositórios, sendo uma das primeiras instalações dessa ferramenta por universidades brasileiras.

A tipologia documental é bem ampla, porém bem diferente. Destaca-se que este repositório não disponibiliza artigos, ou seja, possui uma característica menos científico tradicional. Disponibilizam teses e dissertações, trabalhos de conclusão de cursos, objetos de aprendizagem, livros, relatórios e outros, não apenas texto, mas arquivos multimídias como áudio, imagem e som. Com isso se adéqua a várias finalidades, mesmo que não possua o alinhamento requerido por muitas federações. O Repositório Digital Institucional da UFPR participa da BDTD do Ibict.

Destaca-se a quantidade de esquemas de metadados que o Repositório Digital Institucional – UFPR, interoperando em doze esquemas de metadados (uketd\_dc, dim, oai-dc, marc, etd-ms, xoi, qdc, rdf, ore, mods, mets e didl). Com isso, possibilita intercambiar com várias federações, das mais restritas como a NDLTD, por meio do etd-ms , até as mais abrangentes com o oai-dc, muito apropriado a um repositório com a gama documental ampla.

O Repositório Digital Institucional - UFPR apresenta-se diferente com menor viés científico, mas com uma tipologia documental ampla e diversificada. Torna-se importante por ter sido uma das primeiras instalações do DSpace no Brasil, tornando-se referência nacional na utilização dessa ferramenta. Uma iniciativa inovadora, condizente com uma das primeiras universidades brasileiras.

### **Repositório Institucional da Universidade Federal da Bahia**

O Repositório Institucional da Universidade Federal da Bahia (RIUFBA) tem por missão "Armazenar, preservar, divulgar e dar acesso a produção científica e acadêmica da Universidade, seja na área das artes, das ciências, das humanidades, da

tecnologia, da inovação ou da criação". Reuni em um único bando de dados a produção acadêmica da instituição de forma a aumentar a visibilidade, o impacto das pesquisas e preservar a memória intelectual da instituição.

O RIUFBA foi lançado em 2007, tendo sido desenvolvido com a ferramenta livre DSpace e foi resultado de um estudo de doutoramento, como relata Rosa, Meirelles e Palácios (2011). Assim, entre tantas outras divisões da instituição, teve a adoção pela Editora da Universidade Federal da Bahia (EdUFBA), que possui uma coleção e disponibiliza livros em acesso aberto, disseminando livremente a produção da editora, das publicações fora de linha ou antigas, mas mesmo assim, importante fontes de informação .

O depósito de documentos é regulamentado por uma política oficial da instituição implantada por forma de portaria, que orienta os autores dessa instituição a publicar suas pesquisas e depois depositar no repositório. Esta orientação visa a institucionalizar o funcionamento do repositório, aconselhando os autores a publicarem em veículos de acesso aberto, de forma a facilitar o depósito no repositório.

A tipologia documental é variada e contempla documentos de cunho científico, cultural, artística, técnica e tecnológica. Com isso abriga parte importante da memória da instituição, ao passo que a UFBA possui o título de ter o primeiro curso superior brasileiro, ainda como escola de Medicina, fundada durante o período do Império. O RIUFBA tem disponibilizado livremente documentos daquele período, tornando o repositório com documentos mais antigos do Brasil.

A interoperabilidade se dá por meio dos esquemas de metadados oai-dc, rdf e mets. Nesse sentido, apresentam-se mais apropriado a interoperar materiais científicos, mesmo que se apresente mais amplo em termos de tipo de documentação. O RIUFBA interopera com o oasisbr e a BDTD.

Em 2011 durante a II Conferência Luso-brasileira de Acesso Aberto (CONFOA), o RIUFBA ganhou o premio de repositório brasileiro mais representativo pela comissão organizadora do evento. Este prêmio corroborou com a importância desse

repositório no cenário do acesso aberto brasileiro, que possui a maior coleção de livros dentre os repositórios, além de documentos de cunho histórico e científico.

### **Repositório Institucional Universidade Federal de Santa Catarina**

O Repositório Institucional da Universidade Federal de Santa Catarina (RIUFSC) tem por propósito reunir a produção da universidade em um acervo digital único de forma de melhorar a visibilidade da instituição na Internet. Tonini (2011) revela que o RIUFSC reuniu os cinco repositórios existentes na UFSC, todos desenvolvidos com DSpace, numa tentativa de integração de forma a manter um acervo central, com o máximo de documentos digitalizados da instituição.

Nesse ponto, a tipologia de documentos que compõe o acervo se apresenta variada sendo classificada pelo repositório em duas principais categoria, os documentos científicos e acadêmicos. Os científicos contam com os documentos tradicionalmente avaliados pelos pares adicionados de dados de pesquisa, software e arquivos multimídia de cunho científico. Os documentos acadêmicos possuem uma gama maior de tipologia, de administrativos como os editais e atas de reunião até gravações de programas televisivos.

No RIUFSC destaca-se a quantidade de esquemas de metadados que o RIUFSC interopera, com doze esquemas de metadados (uketd\_dc, dim, oai-dc, marc, etd-ms, xoai, qdc, rdf, ore, mods, mets e didl). Com isso, possibilita interoperar com várias federações, ofertando o variado acervo, nos mais diferentes esquemas de metadados, conforme as características da federação. O RIUFSC participada federação oasisbr .

O RIUSC apresenta pouca informação relevante em seu portal sobre o seu desenvolvimento. Entretanto, por causa de seu acervo variado possui uma boa visibilidade, mesmo que com grande parte de cunho não científico tradicional avaliado pelos pares. Essa questão levanta discussões sobre o próprio repositório institucional, que na sua essência deveria ser somente científico.

## **Sistema Maxwell**

O repositório da Pontifícia Universidade Católica do Rio de Janeiro (PUC-Rio), denominada de Sistema Maxwell, foi criado para fornecer a infraestrutura necessária aos docentes e discentes dessa instituição. Lopes (2002) descreve o Sistema Maxwell como uma ferramenta integradora do ambiente de ensino e da biblioteca digital, operando na internet e apoiando os professores e alunos.

Como um dos repositórios do projeto piloto da BDTD do Ibict, o Sistema Maxwell nasceu em 2001. Foi totalmente desenvolvido pela equipe do Laboratório de Automação de Museus, Bibliotecas Digitais e Arquivos (LAMBDA) da PUC-Rio utilizando os modelos disponibilizados pela NDLTD e BDTD. Com isso, tem nas teses e dissertações parte importante do acervo, ao passo que participa das federações BDTD e NDLTD, entregando os metadados.

Nesse ponto, a documentação mantida pelo Sistema Maxwell apresenta-se diversificada com mais que teses e dissertações, com os trabalhos de conclusão de curso em todos os níveis, objetos de aprendizagem e algumas documentações de cunho mais técnico-administrativo. Destaca-se ainda como um repositório de documentação de final de curso (graduação, extensão, mestrado e doutorado), visto ser a parte mais significativa do acervo.

a interoperabilidade se dá apenas pelos esquemas de metadados oai\_dc e mtd2-br, que ratifica ser um repositório de teses e dissertações. Nota-se, no entanto, que mesmo mantendo e dando acesso aos trabalhos de conclusão de cursos de graduação e extensão não se interopera com a NDLTD, entregando à BDTD apenas os registros de teses e dissertações.

Por fim, destaca-se a visibilidade do Sistema Maxwell pela iniciativa do Laboratório LAMBDA, pertencente ao departamento de Engenharia Elétrica. Dessa forma esse sistema transformou-se no repositório da instituição, principalmente, no que concerne à disseminação dos trabalhos de conclusão de cursos, interoperando os registros de teses e dissertações, dando visibilidade à instituição.

## **Repositório da Universidade de Brasília**

O Repositório da Universidade de Brasília (RIUnB) foi criado com o propósito de armazenar, preservar, disseminar à produção científica e acadêmica da Universidade de Brasília (UnB), de forma a dar visibilidade à produção científica da instituição. Com isso, apoiar as atividades de pesquisa e ensino, por meio da disseminação da documentação científica, alinhando à missão da instituição.

Criada em 2008 pela Biblioteca Central da UnB com apoio do Ibict, o RIUnB é considerado como um serviço ofertado pela BC para a gestão e disseminação da produção científica e acadêmica da UnB. O RIUnB foi desenvolvido com a ferramenta DSpace tornando-se modelo para outras instituições, principalmente por outras universidades federais, que compartilham características semelhantes.

O depósito de documentação científica no RIUnB é regulamentada por uma resolução da reitoria, no qual determina que o depósito dos artigos publicados em veículos de acesso aberto, teses e dissertações, mas desobriga o depósito de documentação com restrição dos editores ou patente. Determina, também a gestão do repositório pela BC e o alinhamento ao acesso aberto, ao passo que determina o livre acesso ao texto completo do acervo do RIUnB.

A tipologia constante no RIUnB é restrita aos documentos científicos tradicionais avaliados pelos pares, compondo-se de apenas oito categorias (artigos, livro, capítulo de livro, relatório de pesquisa, trabalho de pós-doutorado, trabalho em evento, tese e dissertação). Outras documentações acadêmicas são depositadas em outros repositórios, como o repositório de objetos educacionais e o de monografias. Assim é um repositório institucional puramente científico.

Quanto à interoperabilidade o RIUnB responde em oai-dc, mtd2-br, rdf e mets. Esses esquemas são condizentes com o acervo do repositório, revelando que a interoperabilidade está conformada a interagir com federações de cunho científico. O RIUnB interopera com o oasisbr e a BDTD.

O caráter científico do RIUnB, assim como a presença de outros repositórios na instituição, revela o tradicionalismo dos gestores, ao passo que muitos repositório

mantém toda a produção científica em um mesmo repositório. Entretanto, o RIUnB alinha-se as definições mais tradicionais de repositórios institucionais e tornou-se modelo para outras instituições de ensino, pela trajetória e formalismos.

### **Acervo Digital da Universidade Estadual Paulista Júlio de Mesquita Filho**

O Acervo Digital da Universidade Estadual Paulista Júlio de Mesquita Filho (Unesp) e tem por objetivo abrigar todos os conteúdos gerados ou adquiridos pela instituição. Com isso adequar-se aos avanços tecnológicos no que concerne ao tratamento, gerenciamento e manutenção resultante da digitalização. Assim, dar visibilidade a produção da Unesp por meio da disponibilização livre pela internet.

Nascido em 2008 e desenvolvido com o software livre DSpace, o Acervo Digital da Unesp foi coordenado pelo Grupo de Tecnologia da Informação (GTI). A utilização de uma ferramenta existente e largamente adotada no Brasil, no entanto, não é tão perceptível pela adaptação de layout efetuado pelo GTI, tornando-a mais intuitiva para encontrar os documentos mantidos no acervo.

A documentação constante no acervo digital se apresenta variada e categorizada por histórico-cultural, científico, objetos científicos e objetos de aprendizagem. Em cada categoria apresenta vários tipos e formatos de documentos. Esta organização ajuda os usuários e possibilita o Acervo Digital da Unesp a participar de várias federações, ofertando os registros conforme a característica da federação, científica, educacional e outros.

Interoperando em doze esquemas de metadados (uketd\_dc, dim, oai-dc, marc, etd-ms, xoi, qdc, rdf, ore, mods, mets e didl), o Acervo Digital da Unesp se apresenta condizente com o seu acervo. Com isso possibilita participar de várias federações no padrão de metadados da federação, por exemplo, Acervo Digital da Unesp interoperava com a NDLTD, ofertando registros das teses e dissertações em etd-ms, e com o oasisbr, com registros de documentação científica em oai-dc.

Condizente com a própria estrutura da Unesp, que é distribuída por todo o estado de São Paulo, o Acervo Digital se apresenta muito abrangente em assunto e tipologia documental. Mesmo que a Unesp possua outro repositório, de cunho mais científico, o Acervo Digital possui uma visibilidade proporcional a instituição, uma das mais importantes universidades do país.

### **Repositório Institucional Fundação Oswaldo Cruz**

O Repositório Institucional Fundação Oswaldo Cruz, denominado de Arca, é considerado pelos gestores como o principal instrumento para a disseminação em acesso aberto da produção da Fundação Oswaldo Cruz (Fiocruz). Assim, o Arca tem a missão de reunir, hospedar, preservar, disponibilizar e dar visibilidade à produção intelectual da Instituição.

O Arca foi criado em 2007 utilizando a ferramenta livre DSpace, mas só foi lançado oficialmente em 2011 com os 25 anos do Instituto de Comunicação e Informação Científica e Tecnológica em Saúde (Icict), passando por três fases, como relata Carvalho, Silva e Guimarães (2012), a incubação, a integração e o estágio atual. Essa divisão não é rígida, mas apenas para determinar etapas no desenvolvimento e alinhamento com o movimento de acesso aberto por parte da instituição.

Mesmo tendo sete anos, política do Arca só foi oficializada em 2014 por meio de portaria, como relata Maranhão (2014). Com isso, institucionalizou e formalizou o depósito de documentos no repositório, tornando mandatório o depósito das teses e dissertações defendidas no programa de pós-graduação da Fiocruz. Da mesma forma que orienta os seus pesquisadores a publicar em veículos de acesso aberto.

A tipologia documental é bem ampla contendo não apenas os tradicionais documentos avaliados pelos pares, mas manuais, fotografias, vídeos e outros. Mesmo com toda a diversidade e por se tratar de uma instituição que possui cursos técnicos, especialização e de pós-graduação, não há a presença de objetos de aprendizagem.

Assim, como vários repositórios mais abrangentes, o Arca responde à interoperabilidade com doze esquemas de metadados (uketd\_dc, dim, oai-dc, marc, etd-ms, xoai, qdc, rdf, ore, mods, mets e didl). Com isso possibilita a responder a várias



federações, diretamente com o esquema de metadados mais apropriado à federação. O Arca responde às federações BDTD utilizando o dim e ao oasisbr com o oai-dc.

A Fiocruz é uma das instituições de pesquisa brasileira de maior prestígio mundial, com pesquisas de ponta e programas de pós-graduação bem avaliados. Assim, o Arca, com a produção científica dessa instituição, torna-se extremamente importante, ao ponto que, dissemina um acervo composto por documentação representativa da instituição. Nesse sentido, apresenta-se como um bom representante do movimento de acesso aberto brasileiro.

### **Quanto às políticas**

As políticas, antes de tudo, são orientações para o funcionamento do repositório. Entretanto, mais que isso legitima as ações formalizadas por meio de uma norma promulgada pelo dirigente da instituição. Com isso, respalda-se questões como a tipologia documental, tipo de acesso e outros existentes no repositório.

A pesquisa revelou que ainda nem todos os repositórios melhores avaliados possuem políticas publicadas, mesmo que as tenham de forma não institucional. Para as instituições que as tem, destaca-se o alinhamento ao movimento de acesso aberto, ao passo que incentivam a publicação em canais de acesso aberto e o depósito no repositório. Defendem o respeito aos direitos patrimoniais dos editores, mas indicam a necessidade de que toda a produção da instituição esteja depositada no repositório, requerendo contato com os editores de revistas e livros de acesso restrito.

Ratificam-se, dessa forma, os repositórios como canais próprios do acesso aberto. A via verde na qual os autores recebem permissão dos editores para depositar nos repositórios e dar acesso irrestrito e sem custos. Revela-se que as instituições que possuem repositório alinham-se a essa orientação, visto a visibilidade ofertada por esse instrumento.

### **Quanto à tipologia documental**

Repositórios institucionais nasceram como forma de disseminar documentação científica tradicional avaliada pelos pares. Entretanto, verifica-se que muitas instituições

tem ampliado a tipologia que compõe o acervo do repositório. Assim, pode-se classificar os repositórios em três categorias principais em relação à documentação: os restritos, os científicos e os amplos.

Para os repositórios estudados apresentou-se uma relação em que quanto mais antigo é o repositório mais restrito é, como o caso da Biblioteca Digital de Teses e Dissertações da USP (2001) e Sistema Maxwell (2001). Em alguns casos, os repositórios nasceram restritos e foram ampliando a acervo com a modernização. Outros nascem com documentação variada conforme a necessidade da instituição em ter em um único ambiente toda a produção acadêmica da instituição.

Com isso, revela-se a tendência dos repositórios mais modernos integrarem uma documentação mais variada, mais condizente com as necessidades da instituição em reunir e dar acesso a toda a sua produção. Da mesma forma que atende aos usuários, que possuem diferenças no tipo de produção dependendo da sua disciplina. Assim o repositório torna-se, verdadeiramente, o canal representativo da produção acadêmica de uma instituição.

### **Quanto aos esquemas de metadados**

A interoperabilidade é um dos aspectos importantes nos repositórios, pois permitem intercambiar informações, principalmente com as federações. Por isso os esquemas de metadados são importante, pois algumas federações requerem um esquema típico para interoperar, como a NDLTD. Dependendo da finalidade da federação alguns esquemas se tornam mais apropriados, mesmo que interoperem com vários esquemas.

Nos repositórios estudados apresentou-se uma relação, quanto mais abrangente era o acervo maior era a quantidade de esquemas de metadados atendidos pela interoperabilidade. Da mesma forma que repositórios mais novos ou mais atualizados respondem a mais esquemas de metadados. Nota-se que a flexibilidade para responder à coleta é uma tendência, ao passo que a tecnologia permite essa conversão de esquemas com certa facilidade.

O esquema mais utilizado, por ser padrão para o protocolo OAI-PMH é o oai-dc. Entretanto é o mais problemático para as federações devido a sua simplicidade, na

medida em que possui poucos elementos e se não for qualificado, pode gerar ambiguidades, por exemplo: dc.contributor é qualquer contribuidor na criação do documento, podendo possuir papéis muitos distintos (editor, ilustrador, revisor, etc). Conforme os repositórios estudados, o rdf e o mets são os segundos esquemas de metadados mais utilizados e mais apropriados a interoperar uma vasta tipologia de documentos, sendo que o rdf possibilita a interoperabilidade de objetos digitais complexos.

## **5.2.2 ANÁLISE DA ADOÇÃO DAS TENDÊNCIAS TECNOLÓGICAS PELOS REPOSITÓRIOS**

A análise da adoção das tendências tecnológicas pelos repositórios possibilita verificar a infraestrutura tecnológica adotada pelos repositórios, como e se os repositórios estão enfrentando o desafio das mudanças tecnológicas atuais. Para tanto, foram levantadas a adesão às tendências denominadas de "nexus da forças", anunciadas pelo Gartner Group como as tendências que orientarão o desenvolvimento de sistemas informatizados nos próximos anos, sendo: mídias sociais, mobilidade, big data e computação em nuvens.

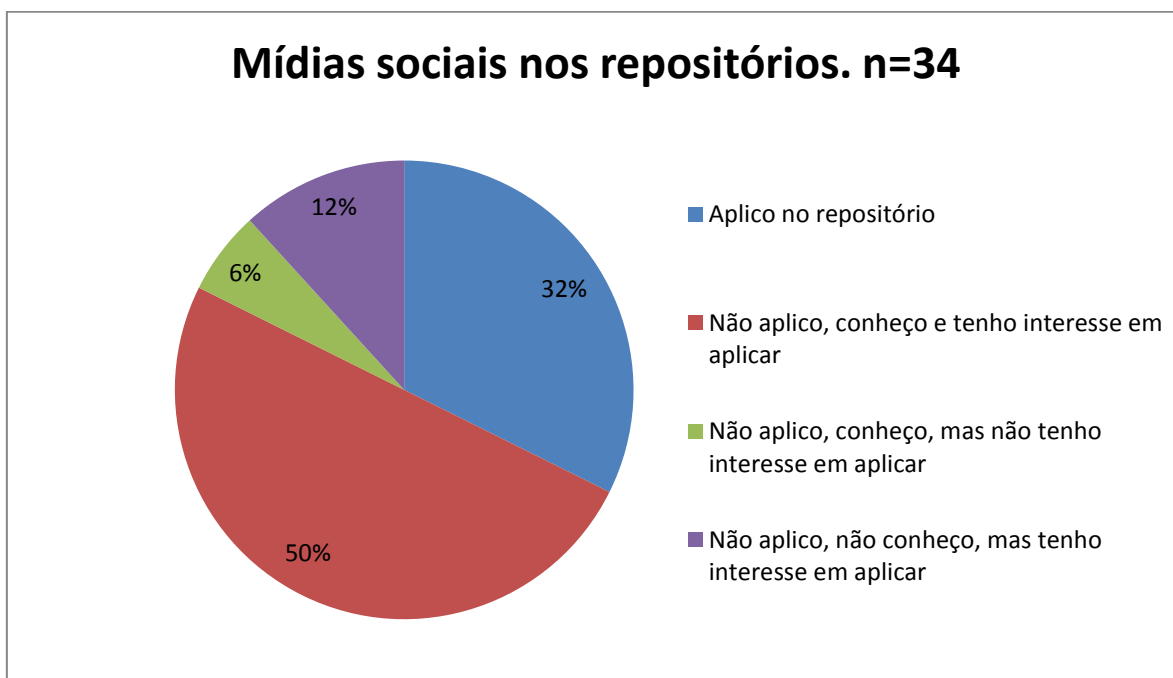
O estudo foi feito por meio de um questionário eletrônico padronizado com múltiplas respostas, disponibilizado pelo Google Docs, enviado para 50 gestores de repositórios. Teve-se 34 respostas, que significam 68%. Com isso, procurou-se levantar a adesão em cada uma das denominadas nexus das forças, de forma a levantar um cenário sobre a utilização das tendências tecnológicas por parte dos repositórios.

### **5.2.2.1 MÍDIAS SOCIAIS E OS REPOSITÓRIOS**

As mídias sociais são amplamente utilizadas pelas pessoas na vida social e profissional, sendo um canal útil para divulgação, com poder viral. Facebook, twitter, youtube são ferramentas úteis para diversos fins e compõem uma enorme gama de opções a serem utilizadas pelos repositórios em dar visibilidade ao acervo. Nesse sentido a presente pesquisa um cenário promissor a utilização das mídias sociais nos repositórios.

Foi levantando que as mídias sociais são amplamente conhecidas, com 82% dos gestores conhecendo a tendência (figura 23), entretanto somente 32% já a utilizam para divulgar o repositório. Essa tendência também apresentou baixa rejeição, com apenas 6% de gestores que não possuem a intenção em fazer uso das mídias sociais nos repositórios e 62% tem interesse em aplicar.

FIGURA 21 - MÍDIAS SOCIAIS NOS REPOSITÓRIOS



Fonte: O autor; 2014.

Não foi possível verificar alguma relação entre tempo de existência do repositório e a utilização das mídias sociais pelos repositórios, visto que existem repositórios antigos e novos que já aderiram a tendência, da mesma forma que repositórios antigos e novos não a utilizam.

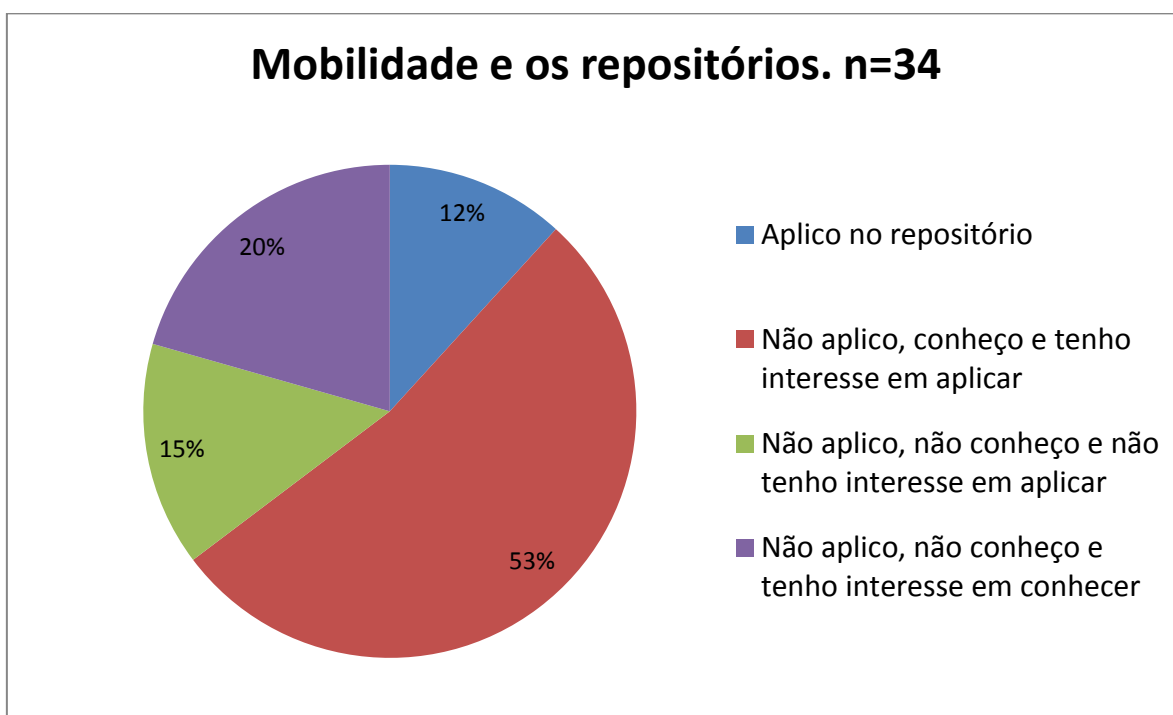
Com isso, revela-se um cenário promissor ao desenvolvimento técnico e tecnológico voltado a aplicação das mídias sociais nos repositórios. Pela quantidade de tipos de mídias sociais, pode-se as associar de várias formas, tornando a divulgação do repositório, seu acervo, serviços e outros mais eficientes, ofertando inúmeras possibilidades aos gestores. Com isso, as mídias sociais transcendem o papel dos repositórios, transpassando as fronteiras da ciência no processo de divulgação científica mais atual.

### 5.2.2.2 MOBILIDADE E OS REPOSITÓRIOS

A mobilidade atualmente pode ser utilizada para acessar diversos serviços disponíveis na internet, tornando parte do cotidiano de várias pessoas. Cada vez mais, dispositivos móveis estão fazendo parte dos processos, no trabalho, educação e lazer. Com isso, muitos sistemas informatizados tem se adaptado para ser acessível vias dispositivos móveis, utilizando tecnologias responsivas.

O levantamento revelou que somente 65% dos gestores conhecem a tendência, sendo que apenas 12% a aplicam no repositório (figura 22). Nota-se que 15% dos gestores não apresentaram interesse em fazer uso dessa tecnologia, mas também não a conhece. Essa rejeição pode se apresentar menor ao passo que a conhecerem mais. Como 20% dos gestores não conhecem a mobilidade, mas tem interesse em conhecer e 53% conhecem e possuem o interesse em aplicar, cria-se um ambiente favorável às iniciativas voltadas a esse segmento.

FIGURA 22 - MOBILIDADE NOS REPOSITÓRIOS



Fonte: o autor; 2014.

Mesmo que a mobilidade esteja tão presente a muitos usuários, revela-se pouca adesão a essa tendência, mesmo que apresente um cenário promissor. Pode revelar

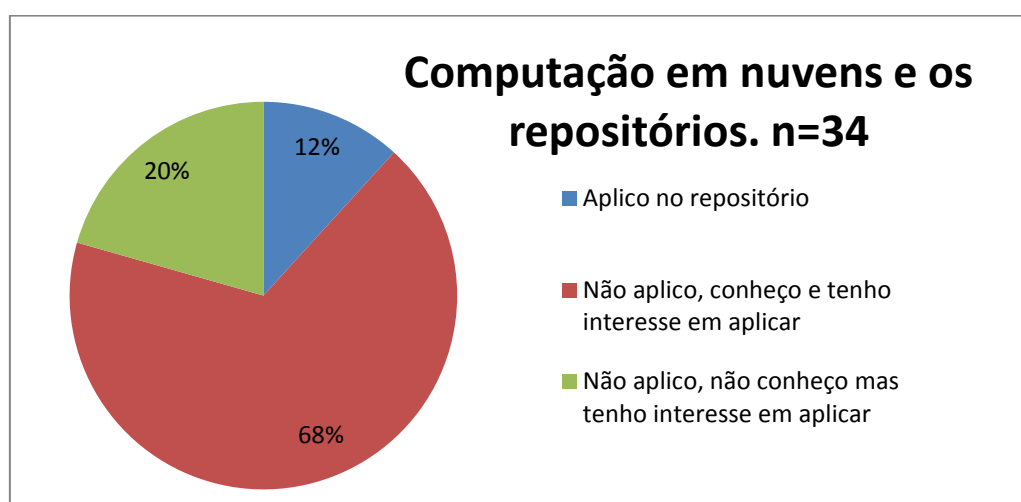
pouca preocupação dos gestores com as tecnologias para acesso com dispositivos móveis, possivelmente, por ser repositórios científicos ter um público alvo mais acadêmico. Entretanto, com a evolução dos dispositivos móveis e a intenção na aplicação torna-se um campo fértil a trabalhos.

### 5.2.2.3 COMPUTAÇÃO EM NUVENS E OS REPOSITÓRIOS

A computação em nuvens revelou ser uma das grandes tendências atuais na construção de infraestrutura de informática, ao passo que oferta um ambiente computacional seguro aos usuários. O que tem levado a muitos gestores de TI a migrarem seus parques computacionais a essa nova tecnologia. Dessa forma tende ser uma opção institucional a utilização dessa tecnologia como parte da atualização tecnológica.

O interesse por essa tendência parece ser uma unanimidade, seja por quem conhece ou não, mesmo que a aplicação ainda seja baixa, com apenas 12% (figura 23). Esse é um das mais promissoras tendências levantadas por este estudo, visto à propensão a adesão, não apenas pelos gestores dos repositórios, mas principalmente pela equipe de TI. Cabe, no entanto, orientar as equipes do repositório e de TI sobre como os repositórios podem ser hospedados na nuvem.

FIGURA 23 - COMPUTAÇÃO EM NUVENS E OS REPOSITÓRIOS



Fonte: o autor; 2014.

Destaca-se que os repositórios requerem grande interação entre equipe de TI e de gestão dos repositórios, para a instalação, configuração e manutenção. Assim,

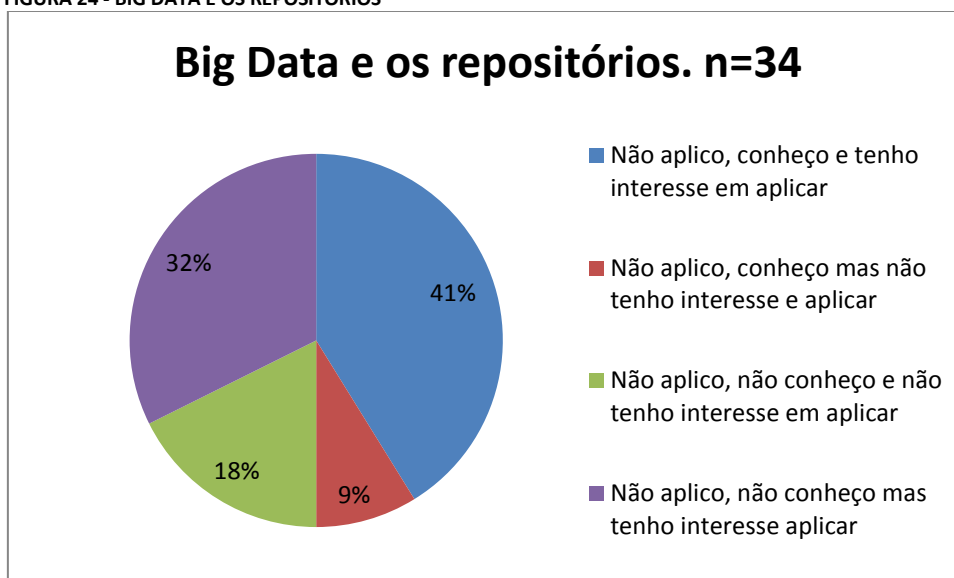
mesmo com toda a transparência ofertada pela computação em nuvens, em relação aos ambiente tradicionais tecnológicos, repositórios operando em nuvens apresentam certas peculiaridades, apresentando desafios a ambas equipes de TI e do repositório.

#### 5.2.2.4 BIG DATA E OS REPOSITÓRIOS

Os repositórios são fontes importantes de informação científica, porém por agregar a produção da instituição pode ser utilizado para obter meta-informação, que apoie a tomada de decisão. Em uma universidade, por exemplo, abriga informações sobre as pesquisas de toda a instituição, como um grande banco de dados semi-estruturado, requerendo ferramentas que consigam extrair as informações. Assim, o Big Data se apresenta como uma boa alternativa a essa tarefa.

O presente estudo verificou que o Big Data ainda é a tendência menos conhecida de todas, pois 50% dos gestores não a conheciam e nenhum repositório ainda a utiliza (figura 24). Com isso, apresenta um desafio, não apenas pelo desconhecimento, mas pela rejeição apresentada por 18% dos gestores. Em muitos casos por não entender que o repositório pode ser fonte importante de indicadores, por meio da análise de seu acervo. Entretanto, 59% dos gestores tem interesse em utilizar o Big Data, menor índice de todas as tendências, mas mesmo assim, um grande percentual.

FIGURA 24 - BIG DATA E OS REPOSITÓRIOS



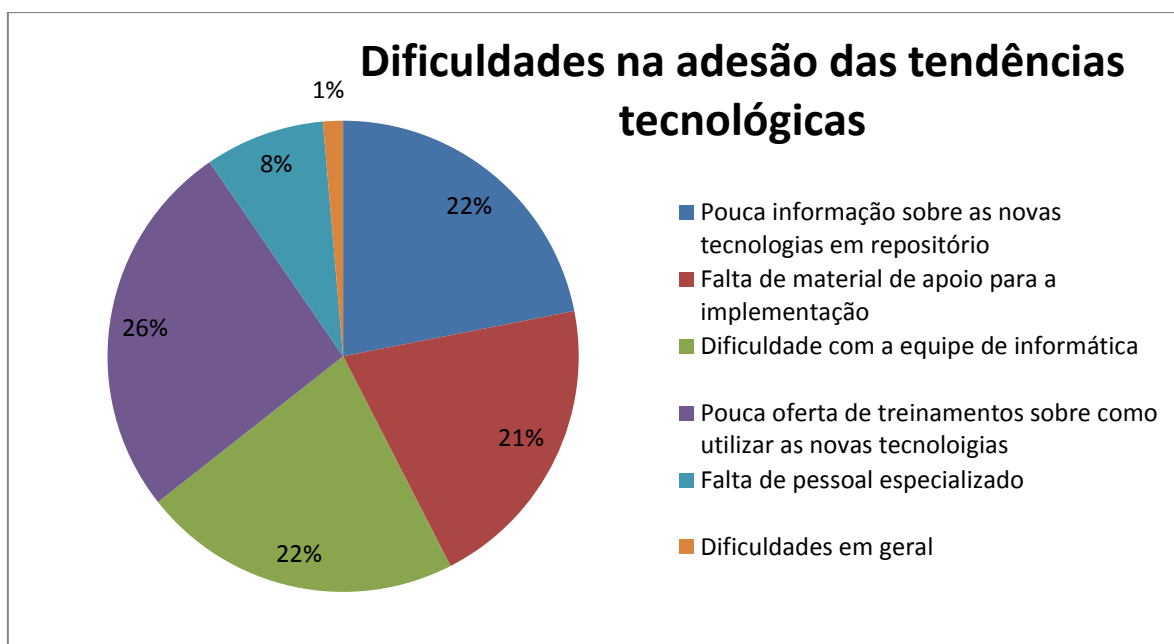
Fonte: o autor; 2014.

O Big Data tem aplicações mais alinhadas com os negócios, comércio eletrônico por exemplo, por isso, seu desconhecimento por parte dos acadêmicos é esperado, mesmo que seja uma ferramenta de pesquisa apropriada a operar com grandes quantidades de dados como nos repositórios. Revela ser um grande desafio a adoção do Big Data pelos repositórios, visto a necessidade de estudos para prospectar e selecionar ferramentas mais apropriadas para cada caso.

### 5.2.2.5 DIFICULDADES NA ADESÃO DAS TENDÊNCIAS TECNOLÓGICAS

No que concerne às dificuldades da adesão às tendências levantou-se várias razões de forma categorizada (figura 25) em que cada gestor podia apresentar quantos problemas fossem adequados, sendo apresentados 72 itens. Destaca-se que os gestores de repositórios indicam a pouca oferta de treinamento como um fator decisivo para a não adesão. Com isso, revela-se um desafio a oferta de cursos sobre o tema, direcionado à equipe gestora de repositório, indicando a necessidade de treinamento, ou seja, repasse de tecnologia de forma interativa, revelando a importância dos treinamentos.

FIGURA 25 - DIFICULDADES NA ADESÃO DAS TENDÊNCIAS TECNOLÓGICAS



Fonte: o autor; 2014.



A pouca informação sobre as novas tendências em repositórios (22%) e a falta de material de apoio para a implementação (21%) completam a questão da falta de treinamento, totalizando 69%, ou seja, a grande razão da não adesão repousa no fato da pouca disseminação das tendências nos repositórios. Essa questão se dá, possivelmente, por estar ainda restrita a Ciência da Computação e não ter se propagado para as outras disciplinas.

A questão de recursos humanos aparece por duas vertentes a falta de pessoal especializado e a relação com a equipe de informática, apareceram em um total de 30% das respostas. Esse é um problema conhecido, principalmente no que tange ao relacionamento entre informáticos e usuários. Xexeo (2007) destaca o problema de comunicação entre essas duas equipes afeta drasticamente o desenvolvimento do trabalho.

Destaca-se que não foi revelado problemas técnicos ou políticos para a pouca adesão às novas tendências tecnológicas, mas majoritariamente na falta de informação apropriada. Essa questão remete ao que Lara e Conti (2003) discutem, sobre os problemas na transferência de informação e nos desafios de criar um ambiente propício ao fluxo de informação, desde aspectos mais práticos aos de adaptação de linguagem.

Por fim, destaca-se a baixa adesão às tendências tecnológicas por parte dos repositórios, devido, principalmente, a questões de divulgação tecnológica. Revela-se um cenário desafiador, visto que grande parte dos gestores tem interesse em aderir, apresentando um ambiente propício, visto não ter problemas técnicos ou políticos, que facilitam a implantação das tendências nos repositórios.

### **5.2.3 ANÁLISE DA ADOÇÃO DAS TENDÊNCIAS DA CIÊNCIA NOS REPOSITÓRIOS**

A análise da adoção das tendências da ciência pelos repositórios possibilita verificar como os repositórios estão se ajustando as novas tendências da ciência, visto já serem consideradas ferramentas de apoio a disseminação da documentação científica. Para tanto, foram levantadas a adesão a seis tendências, por meio de questionário desenvolvido pelo google docs, enviado a 56 gerentes de repositórios brasileiros,

vinculados às universidades e institutos de pesquisa obtendo 35 respostas, ou seja, pouco mais de 60% do universo delimitado.

Foi levantada a adesão das seguintes tendências da ciência pelos repositórios brasileiros:

- Acesso Aberto
- Arquivo Aberto
- Dados Abertos
- Ciência Aberta
- Curadoria Digital
- E-science

Com isso levantou-se o cenário atual brasileiro dos repositórios, no que concerne a adoção das tendências brasileiras. Um cenário dinâmico, ao passo que

### **5.2.3.1 ACESSO ABERTO E OS REPOSITÓRIOS**

Os repositórios institucionais são instrumentos nativos do movimento de acesso aberto, denominado por Harnad e colaboradores (2004) como a via verde. Entretanto a adesão ou não de todos os preceitos do Acesso Aberto por um repositórios é questão da implementação, visto que o repositório é um instrumento que pode ser utilizado de forma mais restrita. Repositórios restritos são comuns na disseminação seletiva de objetos digitais, via pagamento ou outra forma de acesso.

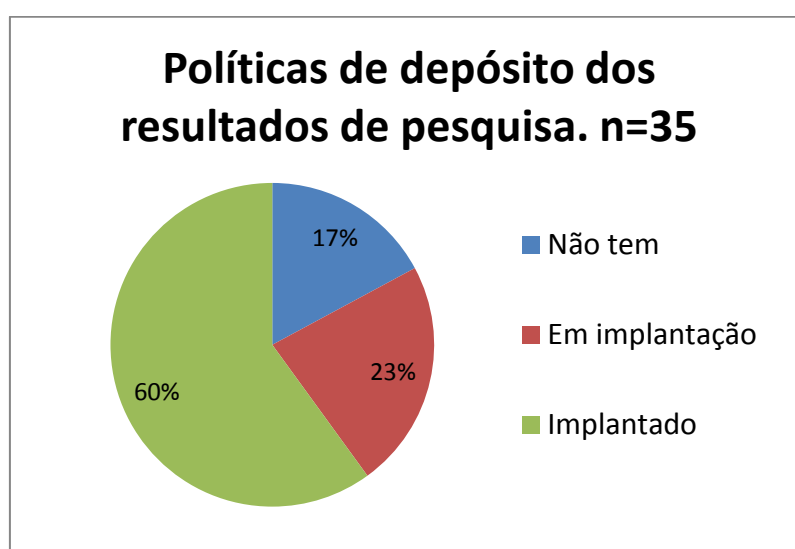
Nesse sentido, 32 dos gestores dos repositórios assinalaram conhecer e aplicar os preceitos do Acesso Aberto nos seus repositórios institucionais, ou seja, pouco mais de 90%. Esse resultado revela a forte adesão dos repositórios ao movimento, sendo que os quase 10% dos gestores restante declararam não ter conhecimento sobre o acesso aberto, mas com intenção em conhecer e aderir, que até certo ponto representa quase a unanimidade.

Estendendo um pouco a questão do acesso aberto, o estudo levantou o quanto o acesso aberto está formalizado na instituição por meio da implantação de políticas para depósito dos resultados de pesquisa (Figura 26). Revelou-se que 60% dos repositórios

já possuem políticas para depósito dos resultados de pesquisa, que institucionaliza o repositório e formaliza os seus processos. Com os 23% em implantação, futuramente 83% dos repositórios terão políticas oficializadas.

O tempo de existência do repositório não implica na existência de uma política oficial, ao passo que repositórios com até um ano possuem políticas oficiais e repositórios com oito anos ainda não a possuem. Assim o tempo de existência não se apresentou como um fator decisivo na implantação das políticas, não apresentando um padrão.

FIGURA 26 - POLÍTICAS DE DEPÓSITO DOS RESULTADOS DE PESQUISA



Fonte: o autor; 2014.

Os resultados do presente estudo ratificam os repositórios brasileiros como instrumentos de acesso abertos, com formalização por meio de oficialização de políticas. Apresentam um alinhamento com o movimento, ofertando acesso à documentação em texto integral, sem restrição, assegurado institucionalmente por política formalizada, um panorama bem estabilizado.

### 5.2.3.2 ARQUIVOS ABERTOS E OS REPOSITÓRIOS

Os repositórios geralmente implementam facilidades oriundas dos conceitos dos arquivos abertos, como o autoaquívamento e a interoperabilidade. Entretanto, nem sempre os gestores dos repositórios fazem uso dessas facilidades, pois como ressalta

Jaillet (2005), os arquivos abertos não são tão conhecidos, ao passo que muitos acadêmicos visam os aumentos do impacto dos resultados de pesquisa, por meio da disponibilização via internet e não de questões técnicas como os dos arquivos abertos.

Nesse sentido a presente pesquisa levantou que 60% dos gestores conhecem e utilizam os preceitos dos arquivos abertos nos repositórios. O desconhecimento sobre o tema não é alta, visto que só foi mencionado por 17% dos gestores e 23% indicaram conhecer, mas não aplicar. Mesmo com a adesão acima da metade dos repositórios, tem espaço para crescimento devido a pouca rejeição ao uso.

A questão da adesão remete aos dois dos principais elementos do arquivo aberto, a interoperabilidade e o autoarquivamento, conforme destacam Triska e Café (2001). Em relação à interoperabilidade, vê-se que grande parte dos repositórios brasileiros a implementam, principalmente por fazerem parte da federação oasisbr. Entretanto, como apresenta Assis (2013), o autoarquivamento nos repositórios brasileiros ainda é baixo. Essa questão se confirma, principalmente, porque grande parte dos repositórios é mantida pelas bibliotecas, como relatam Shintaku, Duque e Suaiden (2014). Assim, as bibliotecas fazem a catalogação da documentação dos repositórios.

Assim, no que concerne a adesão dos preceitos dos arquivos abertos pelos repositórios brasileiros, se restringirmos apenas para a interoperabilidade, tem-se grande parte de adesão, ao passo que participam da federação oasisbr, que integra 42 repositórios, ou seja, três-quartos dos repositórios pesquisados. A interoperabilidade é o que permite a formação da federação, por isso, essa tendência está bem resolvida.

### **5.2.3.3 DADOS ABERTOS E OS REPOSITÓRIOS**

Os dados abertos ampliam a disseminação da informação, estendendo esse processo para os dados da pesquisa. Por isso, podem-se ter repositórios só para esses tipos de documentação ou repositórios mistos, em que os dados e o resultados são depositados juntos, formando um conjunto. Independente se juntos ou separados, a oferta dos dados de pesquisa se torna possível pelo depósito em algum repositório, que ofereça, entre outras facilidades, formas de recuperação apropriadas.

Nesse ponto, o estudo focou nos repositórios institucionais, com isso restringiu a adoção dessa tendência por esses tipos de repositórios, visto que existem repositórios exclusivos de dados abertos. Assim, revelou que nove dos 35 repositórios estudados implementam os preceitos dos dados abertos, sendo que três já possuem políticas para o depósito dos dados de pesquisa e dois estão em fase de implantação. Dos 26 repositórios restantes, 18 gestores indicaram conhecer a tendência, mas não a utilizam no repositório e oito não conhecem a tendência, mas tem interesse em conhecer.

Mesmo que a adesão seja percentualmente baixa não apresentou rejeição, visto que não houve respostas que indicasse desinteresse. Esse ponto revela um cenário promissor, pois pode ser que a adesão aumente. Nesse sentido, di Foggi e Furnival (2013), em análise aos profissionais de informação brasileiros, revelam o alinhamento desses profissionais com o depósito de dados de pesquisa em repositórios e como os repositórios são mantidos pelas bibliotecas, confirma-se o cenário promissor.

#### **5.2.3.4 CIÊNCIA ABERTA E OS REPOSITÓRIOS**

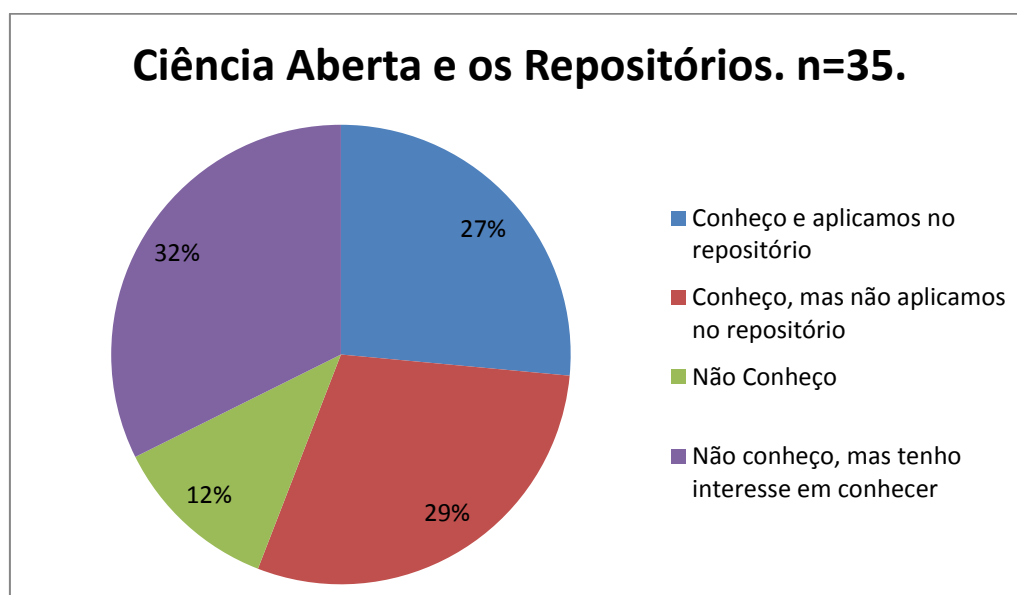
É correto afirmar que a Ciência Aberta é a evolução das tendências vistas anteriormente e que agrega todos os seus preceitos, como em um processo de evolução voltada a democratização da ciência. Com isso, os repositórios tornam-se parte importante nessa tendência, ao passo que, podem ser instrumentos de disseminação aberta de informações e dados, se apresentando como veículos de apoio para a Ciência Aberta.

Nesse sentido, conforme apresentado na figura 27, a adoção da Ciência Aberta nos repositórios brasileiros ainda é tímida, com apenas nove repositórios, uma percentagem ainda relativamente baixa. Este resultado, até certo ponto se confirma, ao ponto que muitos repositórios ainda se mantêm fiel às políticas de depósito de documentação científica tradicional avaliada pelos pares. Toma-se como exemplo a UnB, que possui cinco repositórios: a) O RIUnB, denominado de repositório institucional, agrega documentação científica tradicional avaliadas pelos pares; b) A Biblioteca Digital de Monografias (BDM), repositório de Trabalho de Conclusão de Curso (TCC); c) A Biblioteca Digital Sonora (BDS), repositório de objetos digitais acessíveis; d) O Livros Eletrônicos da UnB (LE-UnB), repositório de livros publicados

digitalmente pela UnB; e) Repositório de Objetos Digitais de Aprendizado (RODA), repositório de objetos educacionais digitais. Assim, o repositório digital fica restrito ao acesso aberto e arquivos abertos.

Nesse mesmo sentido que parte dos gestores de repositórios não implementam a Ciência Aberta, por destoar de certos preceitos, que iriam contra as definições de repositórios institucionais, visto que autores como Crow (2002) consideram que o acervo de um repositório deve ser composto apenas por documentação avaliada pelos pares. Isso pode apresentar um obstáculo à adoção desta tendência por repositórios.

FIGURA 27 - CIÊNCIA ABERTA E OS REPOSITÓRIOS



Fonte: o autor; 2014.

Nota-se que a adoção dos preceitos da Ciência Aberta por parte dos repositórios depende do resultado de uma mudança conceitual sobre as definições de repositórios institucionais. Uma ampliação no que concerne a documentação que compõe um acervo, relacionado a produção científica institucional, para que inclua outras tipologias, mais adequadas à Ciência Aberta.

### 5.2.3.5 CURADORIA DIGITAL E OS REPOSITÓRIOS

Com a adoção ampliação da tipologia documental por parte dos repositórios e a agregação de dados e outros tipos de documentação intermediária da pesquisa, requer-se ações para a gestão que garanta a integridade dessa documentação. Como ressalta

Sayão e Sales (2013) os dados de pesquisa são únicos e requerem cuidados de curadoria, visto a aos problemas existentes no ambiente virtual. Com isso a curadoria torna-se essencial ao aumento do uso das tecnologias da informação na pesquisa.

A adesão dos repositórios às práticas de curadoria ainda são baixas, possivelmente por ainda ser pouco divulgada. Apenas sete dos 35 gestores relataram aderir a essa tendência (quadro 10), enquanto 14 relatam não a conhecer. Esse desconhecimento se torna mais evidente, pois alguns desses repositórios que relatam não conhecer essa tendência, implementam políticas relacionadas à curadoria.

QUADRO 10 - RESUMO DA UTILIZAÇÃO DA CURADORIA PELOS REPOSITÓRIOS BRASILEIROS

Resposta dos gestores	Qde.	Políticas					
		de preservação física do acervo			seleção, depósito e descarte de documentos		
		Sim	Em implantação	Não	Sim	Em implantação	Não
Conheço e aplicamos no repositório	7	1	2	4	1	2	4
Conheço, mas não aplicamos no repositório	14	5	4	5	6	0	8
Não Conheço	6	0	0	6	0	1	5
Não conheço, mas tenho interesse em conhecer	8	1	2	5	1	2	5
<b>Total</b>	<b>35</b>	<b>7</b>	<b>8</b>	<b>20</b>	<b>8</b>	<b>5</b>	<b>22</b>

Fonte: o autor, 2014.

O cenário levantado pelo estudo apresenta-se promissor, pois nota-se que muitas instituições tem se preocupado em implementar políticas voltadas à curadoria. Mesmo que alguns gestores não relatam aderir a essa tendência formalizam ações relacionadas à curadoria por meio de políticas, o que facilita a implementação. Revela uma preocupação com o acervo, ponto principal das atividades de curadoria.

### 5.2.3.6 E-SCIENCE E OS REPOSITÓRIOS

É correto afirmar que a era digital, com suas novas práticas, tem alterando a forma de trabalho em muitas atividades. Não tem se apresentado diferente nas formas de fazer ciência, em que muitas das atividades têm sido apoiadas por ferramentas informatizadas, gerando grande quantidade de dados em formato digital, que pode ser armazenada de forma organizada em repositórios, como forma de preservação da memória da pesquisa e para compartilhamento e colaboração. Assim, os repositórios

podem ser ferramentas de apoio ao e-science, por manter dados antes e depois de serem processados.

Entretanto, como esperado, a adoção dos preceitos do e-science ainda é modesto, dos 35 gestores, apenas quatro indicaram aplicação em seus repositórios. Essa constatação ratifica com o que van Westrienen e Lynch (2005) constataram, que repositórios institucionais tem se restringido a manter cópias de documentação publicada e literatura cinzenta. Revela, que repositórios têm se comportado ainda em momento de transição, na representação digital do impresso e não na forma totalmente virtual.

O desconhecimento da tendência também se fez presente de forma marcante. Quase 50% dos gestores desconhecem o e-science, que revela pouca divulgação. Muitos dos gestores que tem conhecimento não a aplicam, 14 dos 35 gestores relatam conhecer a tendência mas não aplicar, que revela um grande percentual. Revela-se, dessa forma, um cenário desafiador para a adoção do e-science por parte dos repositórios.

Discorre-se da necessidade de alterações nas praticas dos repositórios para o atendimento do e-science, como afirmam van Westrienen e Lunch (2005), ampliando a tipologia documental de forma a atender as novas necessidades da ciência. Incluir os repositórios nas praticas intermediárias de coleta e processamento de dados e não apenas como apoio, em uma posição mais ativa nas pesquisas.

#### **5.2.4 ANÁLISE DA OFERTA DE META-INFORMAÇÃO PELOS REPOSITÓRIOS**

As universidades brasileiras, na sua grande parte, são organizadas de forma hierárquica em que o gestor maior é a reitoria. Assim, os decanatos, pró-reitorias ou estrutura semelhantes estão na posição intermediária, enquanto os departamentos são bem específicos. Por isso, cada qual teria certa responsabilidade hierarquicamente escalonada, do mais geral ao mais específico.

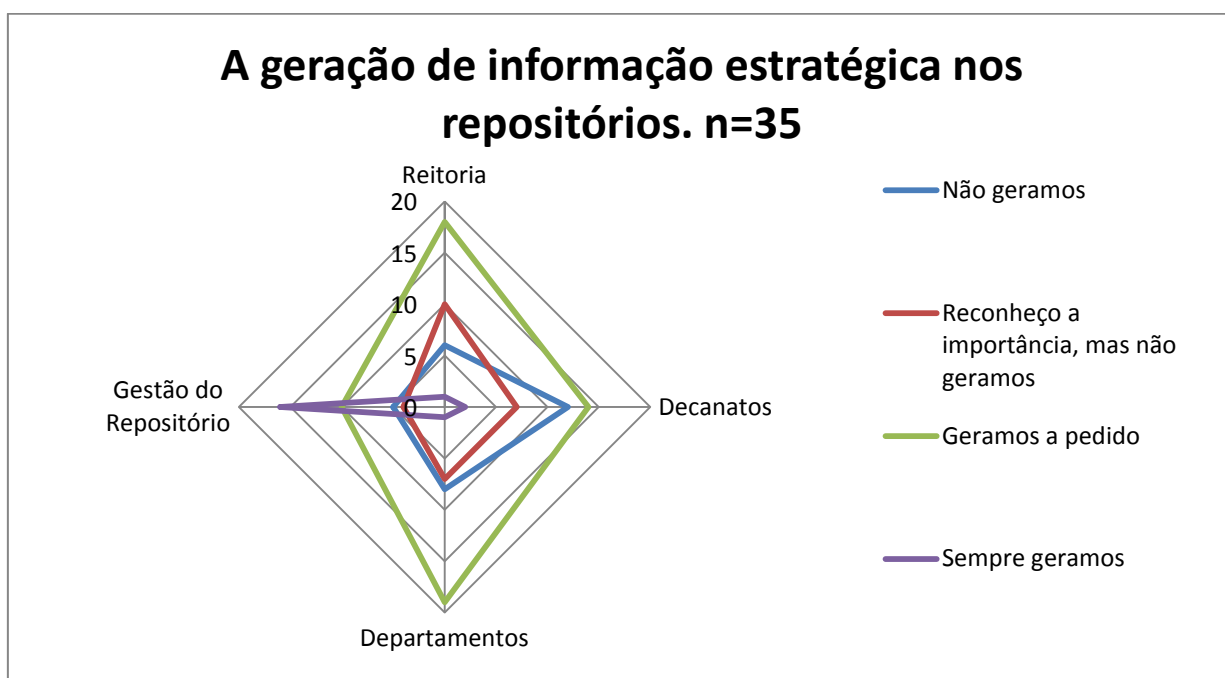
Nesse contexto, os repositórios acadêmicos por conter um acervo multi-departamental pode ofertar meta-informações extraídas desse acervo, desde o mais geral ao mais específico. Em um repositório com acervo mais diversificado pode ofertar meta-informações em vários níveis, dos mais gerais representativos da universidade



como um todo, até mais específico, como a quantidade de defesas de um programa de pós-graduação.

O estudo revelou que os repositórios reconhecem a potencialidade em geração de informação estratégica para a instituição, ao passo que apresentou a maior ocorrência na geração de informação estratégica por demanda (figura 28). Entretanto, esse procedimento ainda não é rotina, que pode significar a falta de percepção dos gestores acadêmicos na potencialidade dos repositórios como geradores de informações estratégicas.

FIGURA 28 - A GERAÇÃO DE INFORMAÇÃO ESTRATÉGICA NOS REPOSITÓRIOS



Fonte: o autor; 2014.

Os decanatos, pró-reitorias ou unidades equivalentes são estruturas intermediárias ligadas à reitoria, responsáveis pela graduação, pesquisa e extensão. São unidades administrativas que coordenam essas atividades. Utilizou-se o termo decanato apenas por questões gráficas, por ser menor, sem juízo de valor.

Destaca-se, no entanto, que para o autogerenciamento a geração de informação estratégica é constante, que reafirma a potencialidade dos repositórios. Esta constatação revela parte do comportamento dos gestores dos repositórios, no que concerne a obtenção de informações para tomada de decisão e como os repositórios se inserem

nesse processo. Com isso estendem-se as possibilidades de aplicação dos repositórios, para além da disseminação de informação, para a geração de informação.

### 5.3. TENDÊNCIAS TECNOLÓGICAS E DA CIÊNCIA NO DESENVOLVIMENTO DE FEDERAÇÕES

A representação de uma federação, ou modelo de federação, para apoiar seu desenvolvimento, assim como outros sistemas informatizados complexos, pode-se dar por meio de um modelo conceitual. Com isso levantam-se todos os requisitos, processos, pessoas envolvidas e fluxo de dados, apresentando todas as características do sistema. Este modelo é muito utilizado na Informática para evitar problemas no desenvolvimento de sistemas, além de fornecer uma documentação técnica.

Como forma de organização, esta parte da pesquisa foi dividida em três partes:

- ❖ Desenvolvimento de Federações: formalização de um modelo conceitual de federações com base na identificação e análise da BDTD (federação de teses e dissertações), oasisbr (federação nacional) e LaReferencia (federação internacional). Com isso analisa diferentes federações com características diferentes, abrangendo os requisitos necessários para a representação de uma federação de forma mais ampla;
- ❖ Tendências Tecnológicas e Federações: adaptação do modelo formalizado no item anterior para atender às tendências tecnológicas. Análise das tendências tecnológicas proposta pelo Gatner Group denominada “nexus das forças” em relação ao repositório, com o intuito de verificar as formas de adaptação da federação para atender essas tendências.
- ❖ Tendências da Ciência e as Federações: adaptação do modelo proposto às tendências da ciência. Análises das tendências atuais da ciência e verificar a possibilidade das federações em adotá-las, gerando uma proposição de alteração no modelo proposto.

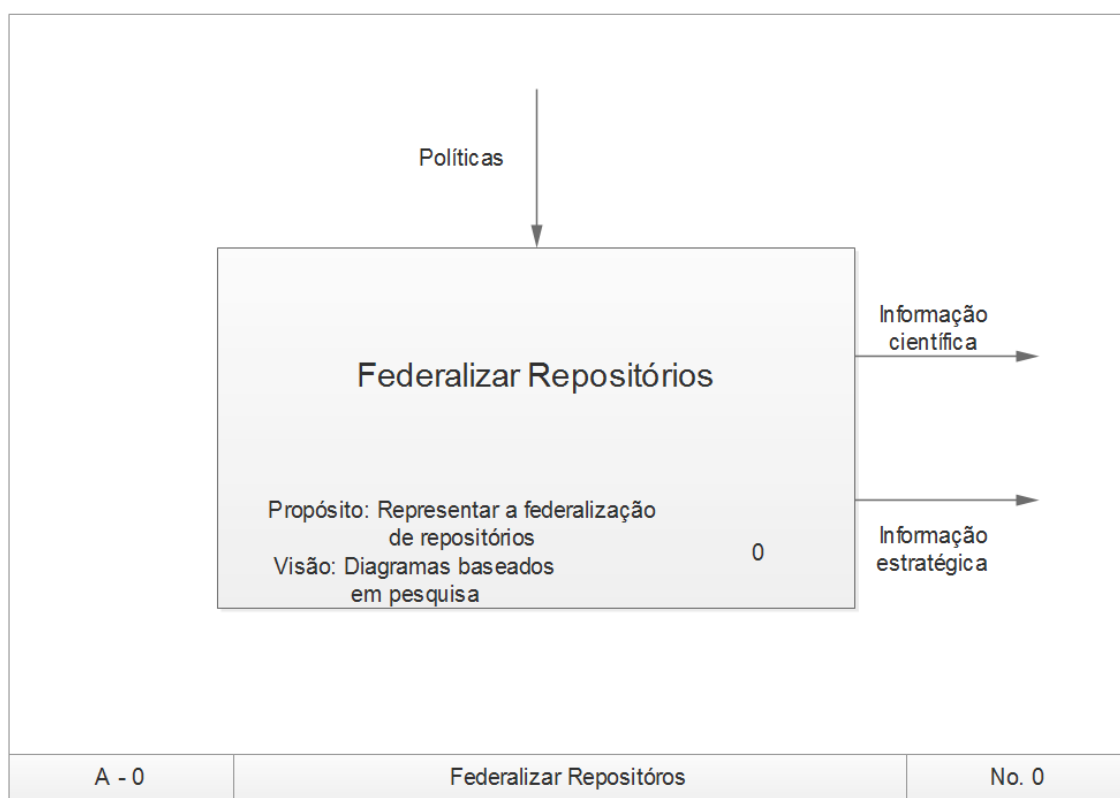
As três partes dessa seção se complementam, na medida em que, desenvolve a formalização de um modelo e o adapta às tendências tecnológicas e da ciência. Nesse sentido, completa-se a os objetivos em identificar, analisar e propor modelo de federações que atendam as tendências tecnológicas e da ciência.

### 5.3.1 DESENVOLVIMENTO DE FEDERAÇÕES

Estruturalmente, as federações destacam-se por serem bancos de dados no sentido de sistemas de informação, com alimentação automática proveniente dos repositórios (BDTD, ALICIA, RIMERI, SNRD, BDCOL) e outras federações (oasisbr, NDLTD, LaReferencia, RCAAP), por meio do uso de um protocolo comum (OAI-PMH e OAI-ORE), com regras bem formalizadas. São bancos estruturados compostos por metadados oriundos dos repositórios e outras federações, que ofertam serviços por meio de um portal, como buscas consolidadas e indicadores.

Assim, entendendo que as federações se configuram como meta-repositórios, pode-se esquematizar a federação em um diagrama IDEF0, em que o processo principal é federalizar repositórios (figura 31). As federações apresentam-se escalonáveis, ou seja, é um provedor de serviços que coleta de metadados, mas também oferta os metadados coletados para serem coletados por outros repositórios, no que é denominado de agregador (BDTD, oasisbr, RCAAP).

FIGURA 29 - DIAGRAMA 0 DE UMA FEDERAÇÃO DE REPOSITÓRIOS.



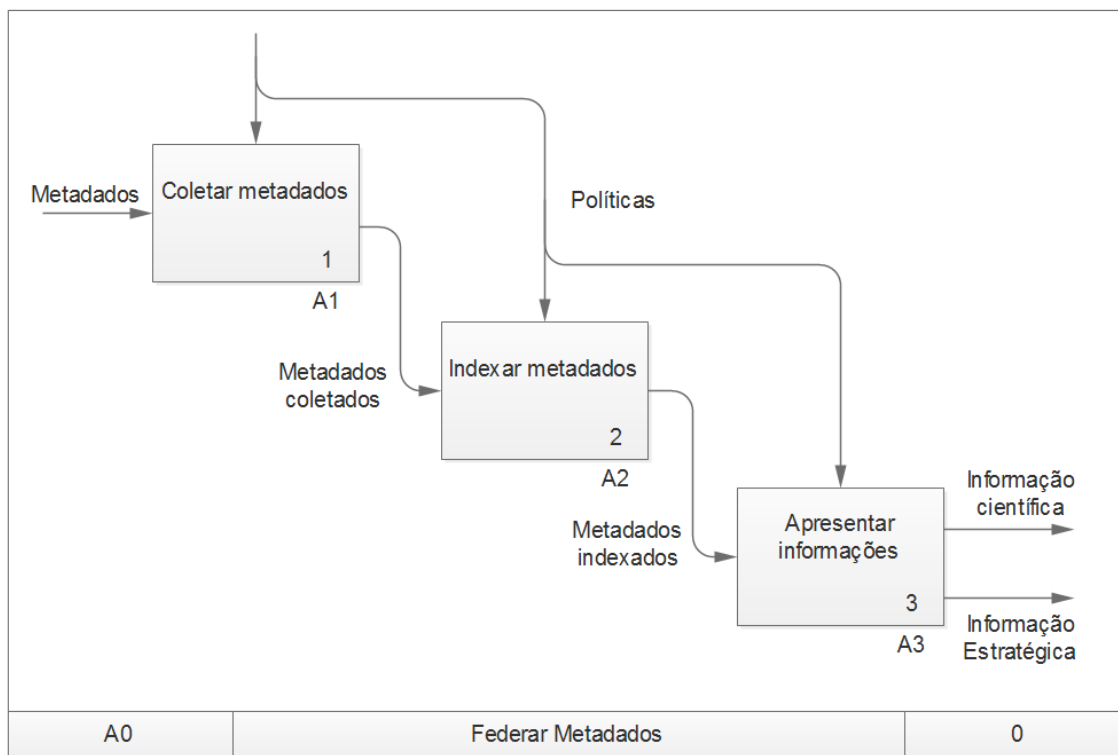
Fonte: o autor; 2014.

Federalizar, nesse contexto, agrupa todos os processos necessários para integrar os repositórios e os seus metadados, uma série de etapas que envolvem tarefas e pessoas. Com isso, possibilita o processamento dos metadados coletados, de forma que as saídas principais desses processos são as informações científicas consolidadas e informações estratégicas, como os indicadores. O principal controle é a política que orienta todo o processo.

Nesse sentido, para federalizar é preciso que se defina, inicialmente as políticas da federação, que orientarão todo o funcionamento do sistema. Um conjunto de regras que indicam como os processos devem operar. Em muitos casos, as políticas apoia a tomada de decisão nos processos da federação. Com isso procura-se assegurar um bom funcionamento da federação, abarcando todos os detalhes necessários para que os processos estejam amparados por determinações claras, evitando problemas.

O processo de federalização pode ser refinado em três grandes processos, coletar metadados, indexar e apresentar informações (figura 32). Cada processo, por sua vez, pode ser refinado, apresentando detalhes do seu funcionamento. As políticas devem ser aplicadas em todos os três processos maiores da federalização, visto que orientam todo o processamento, em todas as etapas. As linhas sobre as tendências foram omitidas de forma a facilitar a representação, pois no detalhamento oportuno em seções separadas.

FIGURA 30 - DIAGRAMA IDEFO DE FEDERAR REPOSITÓRIOS



Fonte: o autor; 2014.

### 5.3.1.1 POLÍTICAS PARA FEDERAÇÕES

As federações são meta-repositórios, visto que seu acervo é formado por metadados oriundos de vários repositórios. Desse modo, suas políticas devem estar em consonância com os repositórios agregados, para que não ocorra conflitos. Nesse sentido, no âmbito dos repositórios, Gilman (2013) enumera nove conjuntos de critérios a quais as políticas devem abranger, sendo:

- ❖ **responsabilidade editorial:** geralmente os administradores do repositório não assumem nenhuma responsabilidade editorial sobre o acervo, ou seja, assume o papel de distribuidor. Entretanto, em alguns casos o repositório disponibiliza documentos não publicados e únicos, como teses e dissertações, assumindo o papel de editor local. Assim, as políticas devem deixar isso claro;
- ❖ **propriedade intelectual:** as políticas devem indicar os detentores dos direitos de uso intelectual e patrimonial de todos os documentos

mantidos no repositório. Assim, assegurar os direitos de distribuir e preservar os documentos;

- ❖ **pesquisas relativas ao ser humano:** as pesquisas em que o ser humano é o sujeito da pesquisa requer aceite da comissão de ética e essa cláusula deve ser respeitada e indicada nas políticas, ou seja, trabalhos dessa natureza, não publicados, necessitam de aval da comissão de ética para serem depositados;
- ❖ **bases de dados:** as políticas deve ressaltar sobre os conteúdos das bases de dados depositadas no repositório, pois os conteúdos podem envolver outros critério, como a confiabilidade, confidencialidade, pesquisa com humanos e outros.
- ❖ **privacidade:** a política deve assegurar os direitos de privacidade dos usuários que acessam o repositório, depositam documentos e outros, de forma a não divulgar dados que possam ser sensíveis, com conteúdos privados sobre os usuários;
- ❖ **difamação:** a política deve ressaltar aos autores que conteúdos difamatórios não devem ser depositados, assegurando a instituição de problemas jurídicos por disseminar informações difamatórias;
- ❖ **trabalhos colaborativos:** trabalhos colaborativos, muitas vezes, envolvem autores de instituições diferentes. Nesse caso as políticas devem assegurar a questão permissão de depósito de autores vinculados a outras instituições;
- ❖ **retiradas e correções:** os conteúdos dos repositórios podem ser categorizados em formalmente publicados e os não publicados. Com isso, a política deve assegurar que um trabalho, publicado ou não, possa ser retirado ou que um trabalho corrigido possa substituir o vigente;
- ❖ **restrições de acesso:** a política deve assegurar níveis diferenciados de acesso para determinados documentos contidos no acervo;

As recomendações acima asseguram que detalhes importantes dos processos mantidos pelos repositórios estejam pautados por regras. São recomendações úteis, que orientam teoricamente a formulação de políticas de repositórios. A Análise das políticas

adotadas pelos repositórios apresentadas na seção anterior, por sua vez, possui um aspecto mais pragmático, visto que são aplicações reais, sendo que destacam:

- ❖ **tipologia documental:** a política deve determinar quais os tipos de documentos devem ser depositados no repositório, podendo ser mais restritos, somente para teses e dissertações, por exemplo, ou mais amplos com maior diversidade de tipos documentais;
- ❖ **formato dos arquivos:** repositórios podem restringir formatos para facilitar as políticas de preservação, ao passo que formatos muito específico, geralmente de software proprietário, pode não ser acessível após algum tempo, ou por falta de licença ou por estar em versão muito defasada. Por isso, em muitos casos são colocados nas políticas;
- ❖ **campos descritivos (metadados):** como forma de garantir qualidade nos metadados, muitas políticas de repositórios definem os campos obrigatórios, vocabulários controlados, esquema de metadado padrão, como forma de assegurar preenchimentos mais condizentes com padrões estabelecidos;
- ❖ **mandato:** Como forma de assegurar alimentação constante da base de dados políticas podem definir mandatos de depósitos, criando obrigatoriedade de depósito de determinadas tipologias documentais, como teses e dissertações para as universidades a não ser que tenha segredos de patente;
- ❖ **objetivos do repositório:** com fins de assegurar a orientação geral do repositórios, muitas políticas iniciam as políticas descrevendo os objetivos do repositório;
- ❖ **organização do repositório:** os repositórios mantêm um acervo organizado artificialmente de forma virtual, que oferta certo contexto aos documentos depositados. Portanto, muitas políticas descrevem a organização ofertada pelo repositório.
- ❖ **fluxo de depósito:** muitas políticas descrevem o processo de depósito, relatando quais as etapas existem e quem as executam, se o autoarquivamento é ofertado ou se a equipe do repositório o faz os depósitos.



Muitas dessas orientações podem ser úteis às federações, pois requerem que sejam propagadas para níveis superiores, assegurando que as recomendações sobre documentos se mantenham, dos repositórios para as federações. Outras, no entanto, são tipicamente de repositórios, restritas a determinado processo do repositório, por exemplo, nas federações não existe o processo de depósito de objetos digitais.

As federações possuem políticas mais abrangentes, visto que requerem maior flexibilidade, pois agregam vários tipos de repositórios e necessitam compatibilizar-se com as suas políticas. Com isso, adotam uma quantidade menor de orientações, em muitos casos, voltados diretamente para os seus processos, orientando o funcionamento de cada um deles.

Os maiores processos da federação, como visto da figura x, configuram uma tríade que processam os metadados oriundo dos repositórios, em um fluxo serial. Cada processo pode ser detalhado, revelando os subprocessos e as peculiaridades de cada um. Assim, as políticas para cada processo podem ser categorizadas da seguinte forma:

- ❖ **Coletar Metadados:** os metadados dos repositórios são coletados automaticamente pelo uso comum de protocolos de comunicação. Para tanto, os repositórios são cadastrados antecipadamente e a coleta (harvesting) dá-se em períodos determinados.
  - **tipo de repositório:** as políticas de federação devem indicar quais os tipos de repositórios são coletáveis pela federação, quais os critérios a serem avaliados pela federação para que um repositório seja aceito;
  - **esquema de metadados:** as políticas devem indicar em quais esquemas de metadados pode-se efetuar a coleta, visto que, internamente pode utilizar o esquema de metadados mais apropriado para descrever os documentos do acervo, mas o processo de interoperabilidade requer certa padronização, mesmo que seja flexível, adotando vários esquemas de metadados;
  - **formas de adesão:** para que a coleta seja feita, o repositório deve ser cadastrado, por isso a política deve indicar a forma em que o repositório passa a fazer parte da federação, se por convite,

submissão, prospecção. Caso requeira assinatura de convênio, deve ser disponibilizado um modelo de documento, além da indicação na política;

- **Atualização:** As coletas são manuais ou automáticas, podendo ser aplicada de forma geral ou específica para cada repositório. Assim, a política pode indicar a periodicidade das coletas, que tem influência na atualização do acervo;

❖ **Indexar Metadados:** os metadados coletados são indexados por algoritmos configurados, utilizando ferramentas informatizadas. Assim, indexação aqui não é o processo relacionado à Biblioteconomia, mas a Ciência da Computação, por ser um sistema informatizado. Por isso, deve-se indicar quais os critérios para configuração do indexador.

- **Campos de metadados indexados:** as políticas deve apresentar quais dos campos de metadados serão indexados, de forma a proverem serviços como facetamento e listagens;
- **descarte:** as políticas devem indicar os critérios para descarte de registros coletados, indicando o processo necessário para processar a retirada de registros, visto que não é um procedimento normal. As federações tendem a agregar registros e não removê-los;
- **vocabulários controlados:** o processo de indexação pode fazer uso de vocabulários controlados para categorizar, padronizar ou outro processo de indexação, mas deve deixar claro nas políticas.
- **idiomas:** muitos indexadores implementam facilidades para tratar certos idiomas, como questão de gênero e número. Entretanto, para repositórios multilinguais, nem sempre isso é possível, mas é bom colocar nas políticas, que idiomas são tratados.

❖ **Apresentar Metadados:** a política de apresentação de metadados deve indicar as formas de apresentação, explicando os critérios utilizados.

- **Formato de apresentação dos resultados:** As políticas podem apresentar orientações para a apresentação dos resultados, se em forma de lista ou gráfico. Qual a forma de apresentação da lista,

ordem alfabética, ranqueamento, qual o campo padrão de ordenamento;

- **campos de metadados:** geralmente se coleta mais informações que se apresenta, pois nem sempre interessa aos usuários ver todas como resultado de busca e apresentação do documento.
- **refinamento e sugestões:** atualmente os sistemas de descoberta e entrega utilizados na implantação dos portais apoia o usuário a descobrir documentos por meio de refinamentos de busca ou sugestão, por isso, as políticas podem registrar as formas que essa facilidade é ofertada.
- **Opções de busca:** as ferramentas de busca possui certas orientações, como o uso de operadores booleanos, caractere curinga, busca precisa, entre tantos outros. Da mesma forma que pode indicar formas de ranqueamento e outras formas de apresentar resultados.
- **formas de navegação pelo acervo:** as federações por serem meta-repositório podem se organizar-se virtualmente de várias formas, possibilitando navegação pelo acervo de modos distintos. As políticas podem descrever essa facilidade ofertada.

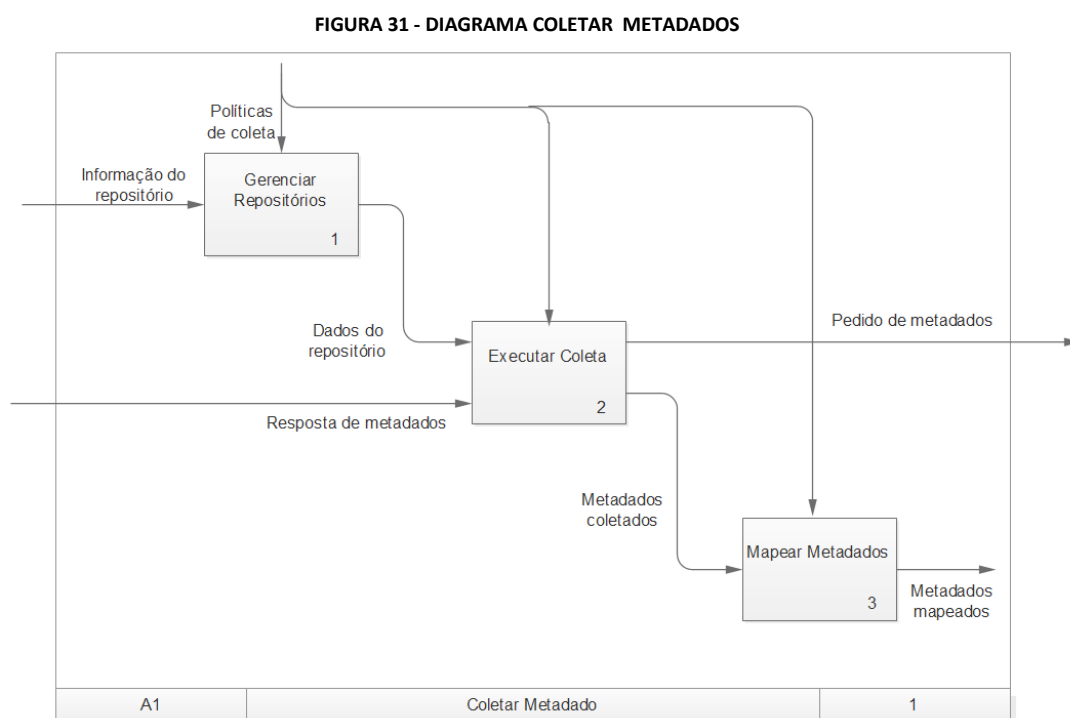
As observações acima são resultado da análise das políticas das federações apresentadas em 5.1.1. Compilação das principais características apresentadas, com adição de observações das análises dos portais das federações, visto a pouca informação apresentada nas políticas sobre os processos iniciais e intermediários. Inferência resultante das funcionalidades apresentadas implicitamente nos portais das federações.

### 5.3.1.2. COLETAR METADADOS

A federação possui um acervo composto por metadados coletados automaticamente dos repositórios, por um processo denominado de Coleta de Metadados ou harvesting, que possui um aspecto mais tecnológico. Para sua operacionalização utiliza um protocolo de comunicação, executando nas camadas mais altas da internet e assim não interferir nos repositórios, apenas responder à requisição de

coleta com os metadados apropriados. Por isso, em alguns casos esse processo fica delegado às equipes de informática.

O processo Coletar Metadados é executado em três etapas (figura 31), Gerenciar Repositórios, Executar Coleta e Mapear Metadados. É nesse processo que ocorre a interoperabilidade com os repositórios, orientada pela política de coleta. O processo pelo qual a federação se conecta automaticamente com os repositórios, é feito via internet, coletando metadados para o processo de indexação.



Fonte: o autor; 2014.

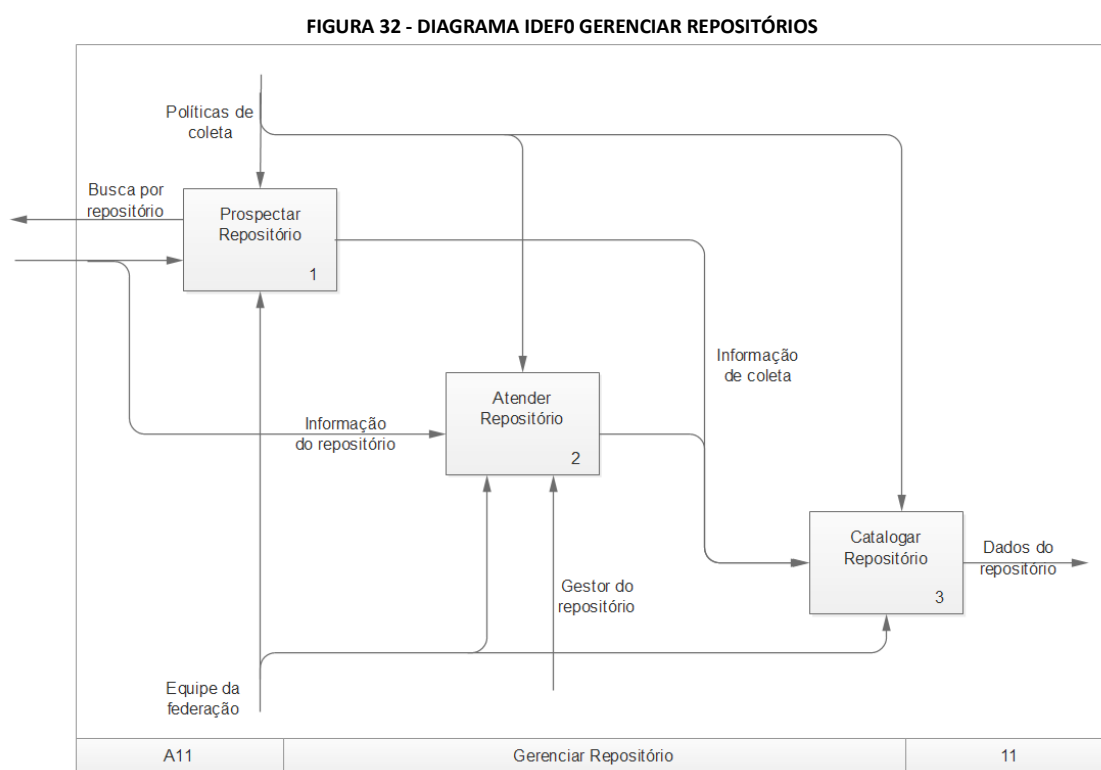
As políticas orientam os processos da coleta de metadados de forma diferenciada, conforme a finalidade de cada tarefa. Na maioria dos casos essas políticas apoiam a configuração dos processos ou a tomada de decisão na sua execução. Por isso está presente em todos os processos, mesmo que com orientações próprias para cada um.

### **Gerenciar Repositórios**

O processo de Gerenciar Repositórios cria e mantém registros de um banco de dados com os cadastros dos repositórios vinculados à federação. Para tanto recebe as informações dos repositórios, processando-as conforme as políticas, por exemplo, se

atendem aos critérios descritos na política de tipo de repositório aceito pela federação. Geralmente esta tarefa fica a cargo de pessoal com perfil mais tecnológico, visto que requer conhecimento de informática para executá-lo.

Detalhando o processo de Gerenciar Repositório (figura 32), vê-se que é composto por três processos: Prospectar Repositórios, Atender Repositório e Catalogar Repositório. Os dois primeiros processos possuem características próximas, gerando as informações a serem processadas pelo último processo, na medida que há duas formas de se vincular repositórios à federação, pela prospecção ou por demanda.



Fonte: o autor; 2014.

No processo de Prospectar Repositórios a equipe da federação busca por repositórios que atendam os critérios da política de tipo de repositório, geralmente por meio de levantamento pela internet. Da mesma forma que esse processo é orientado pela política de tipos de adesão. Por isso. Dependendo das políticas da federação adotada, esse procedimento requer que o gestor da federação negocie com o gestor do repositório a adesão, visto que, mesmo o repositório disponibilizando os metadados por meio dos preceitos dos arquivos abertos é recomendável o contato.

A demanda por sua vez, os gerentes de repositórios submetem seus repositórios para avaliação, por meio do processo Atender Repositório. Assim, a negociação é iniciada pelo gestor do repositório. Esse processo também é orientado pelas políticas de tipos de repositórios e formas de adesão. Em alguns casos, o repositório precisará ajustar-se às condições da federação. Nos dois processos, mesmo com diferenças de procedimentos, o resultado é o mesmo, obtêm-se as informações necessárias para o cadastramento do repositório.

O processo de catalogar repositórios é um processo manual e simples, basicamente inserir em um sistema informatizado os dados do repositório. Na maioria dos casos, faz-se testes de acesso via protocolo OAI-PMH para ver a disponibilidade dos dados. Nesse processo corrige-se os erros de conexão, requerendo revisão constante para evitar erros de coleta, por problemas de informação desatualizada dos repositórios.

### **Executar Coleta**

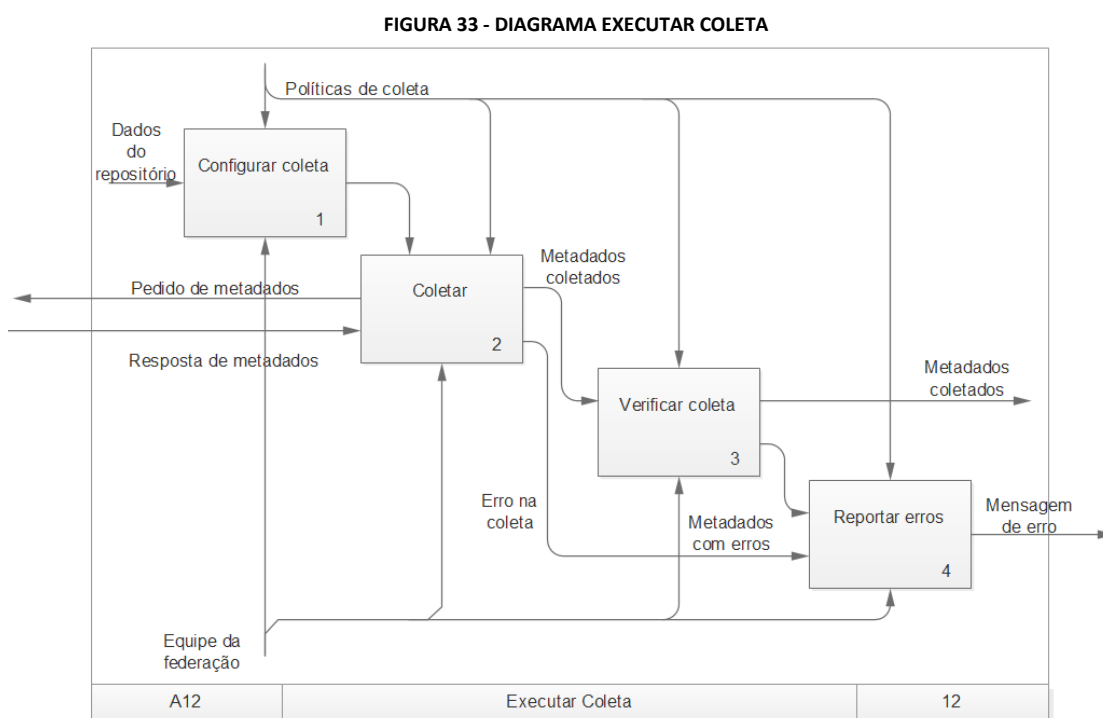
Na execução da coleta tem-se um aspecto mais automatizado, no que concerne a interoperabilidade com os repositórios. É um processo configurável amparado pelas orientações das políticas de coleta, como por exemplo, a periodicidade em que a coleta é executada, que tem implicações na atualização da federação e os tipos de esquemas de metadados aceitos. O aspecto mais tecnológico deste processo requer conhecimentos de Ciência da Computação para execução, assim requerendo um perfil diferenciado.

Mesmo que grande parte do processo de Coletar Metadados seja automatizada, este processo se detalha em quatro etapas (figura 33). Na configuração da coleta são ajustados os procedimentos de coleta de metadados, conforme as políticas. Parte das configurações é geral, como o que fazer com os registros com erro na coleta, ou seja, para todas as coletas, enquanto outras configurações são específicas para cada repositório, como o tipo de esquema de metadados utilizado na coleta, possibilitando uma flexibilidade nas configurações.

Geralmente o processo de coleta é executado automaticamente conforme as configurações ofertadas pela política de atualização. Entretanto esse processo pode ser ativado manualmente, quanto for o caso. É comum a reexecução manual de uma coleta,

por erro ou quebra de conexão. Entretanto essa reexecução requer análise de erros e contato com os gestores dos repositórios se for o caso.

Nesse sentido, o processo de verificação da coleta é essencial, principalmente para a coleta automática, visto que mesmo com o tratamento automático de erros, algumas políticas determinam o descarte dos registros com erro. Assim a federação não reflete totalmente os registros dos repositórios. A análise de erros pode levar a correção dos problemas, aumentando a fidelidade entre os repositórios e a federação.



Fonte: o autor; 2014.

## Mapear Metadados

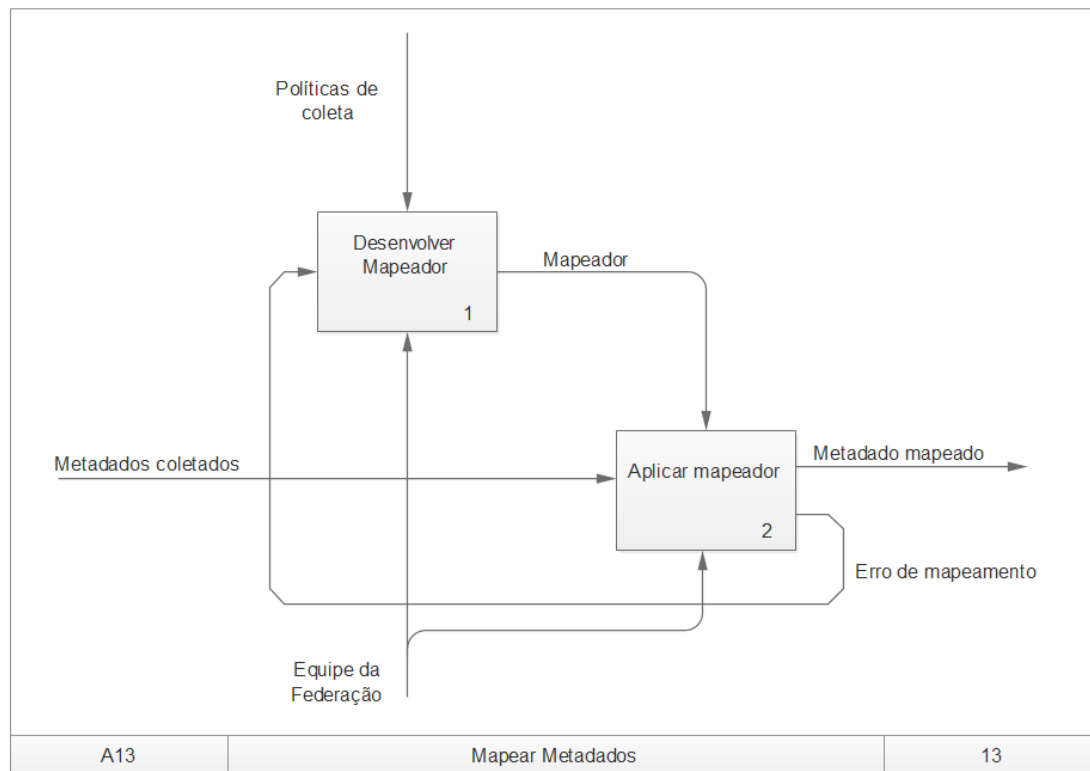
Normalmente, as federações agregam repositórios, que interoperam em diversos esquemas de metadados. Entretanto, essa facilidade implica na necessidade de conversão dos diversos esquemas para um padrão definido nas políticas. Essa troca de esquema dá-se por meio de um processo denominado de Mapear Metadados. Nesse sentido a coleta, seguindo os preceitos dos arquivos abertos utiliza o formato XML para interoperar. Essa forma facilita a conversão de esquemas ao utilizar um programa desenvolvido em XSLT, ou seja, transformar um arquivo em um formato XML para

outro. Esses mapeadores são desenvolvidos conforme as orientações da políticas, que indicam quais os esquemas de metadados são coletados.

O processo de Mapear Metadados, também, oferta certa padronização e normalização aos metadados coletados (figura x). No processo de mapeamento, o conteúdo de alguns campos pode ser alterado, de forma a ofertar maior qualidade nos dados, evitando variações para um mesmo conteúdo. Idioma, por exemplo, é um campo que pode ser normalizado facilmente, com base das políticas adotadas pela federação, utilizando o nome da língua por extenso, sigla como no padrão ISO-639 ou outra forma.

Para tanto, o processo de mapear metadados requer dois processos: desenvolver mapeador e aplicar mapeador (figura 34). Esses dois processos são totalmente tecnológicos e requer conhecimentos de informática, o que reafirma a necessidade de equipes interdisciplinares na manutenção das federações. Mapear metadados tem influência direta na qualidade e acurácia nas recuperações.

FIGURA 34 - DIAGRAMA DE MAPEAR METADADOS



Fonte: o autor; 2014.



O processo de mapear metadados tem relação com a política de esquema de metadados, pois a federação pode executar a coleta em vários esquemas de metadados, mas necessita transformar num esquema único para o processo de indexação. Assim, desenvolve-se um mapeador para cada esquema de metadados aceito pela federação. Com isso, ao aplicar o mapeador, tem-se a conversão do antigo esquema de metadados para o esquema de metadados interno da federação, possibilitando o processo de indexação.

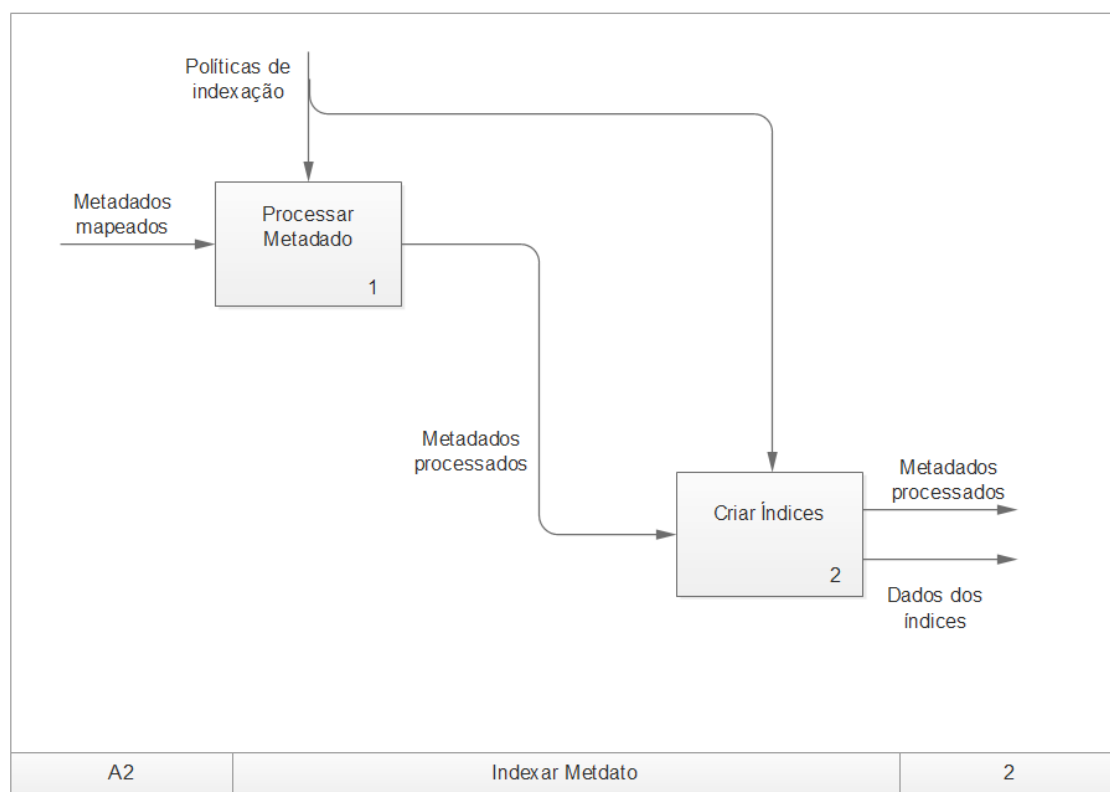
### **5.3.1.3 INDEXAR METADADOS**

As federações são implementadas com bancos de dados, estruturas informatizadas que requerem índices para facilitar a recuperação os registros armazenados. Nesse caso, a indexação não é o processo relacionado à Bibliotecnomia, mas sim à Ciência da Computação, em que índices são criados para possibilitar a recuperação de informação em bancos de dados. Esse processo é orientado pelas políticas de indexação, que indica como o processo deve ocorrer.

Assim, as políticas orientam a criação do índices, indicando quais dos campos de metadados armazenados no banco de dados devem ser indexados. Indica-se quais os campos devem ser indexados para fornecer serviços como listagem para a navegação. Geralmente campos de metadados como título e autor são indexados, campos com menor variação, como tipo de documento e idiomas são indexados para categorizar os documentos.

A indexação dá-se por meio de dois processos: processar metadados e criar índices (figura 35). Nesse sentido, o metadados é processado como preparação para a indexação, ao ponto que esse é um processo totalmente informatizado e automático. Com isso, os metadados podem ser adequados com políticas de indexação como descarte, para retirada de registros indexados, uso de vocabulários controlados.

FIGURA 35 - DIAGRAMA INDEXAR METADADOS



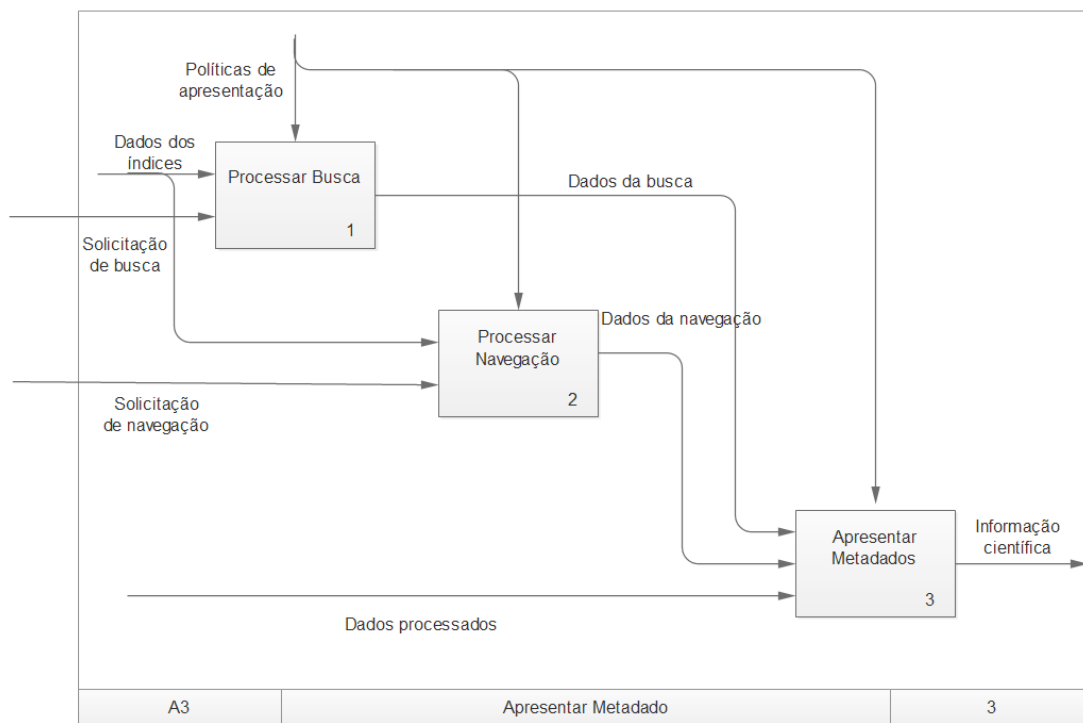
Fonte: o autor; 2014.

Outro ponto relacionado a indexação se refere a funcionalidades ofertadas por alguns software que apoia a recuperação. Uma das mais comuns não indexa as chamadas stop words (artigos, preposições, conjunções, termos esvaziados de semântica), além de não fazer distinção de gênero e número exclusivamente para alguns idiomas. Geralmente, bancos multi-idiomáticos não fazem uso dessa funcionalidades por gerar problemas de recuperação, pois regras de plural nos idiomas diferem, o mesmo para o masculino e feminino.

#### 5.1.3.4 APRESENTAR METADADOS

O Portal de busca se apresenta como o serviço mais importante da federação, visto que proporciona uma recuperação de informação consolidada e a interação com os usuários finais. Todos os processos anteriores são executados para ofertar esse serviço. Por isso, o processo de apresentar metadados se apresenta como a finalidade da federação. Assim, compõe-se de três processos, sendo dois relacionados a formas de recuperação e um para a apresentação da informação (figura 36).

FIGURA 36 - DIAGRAMA APRESENTAR METADADO



Fonte: o autor; 2014.

Geralmente nos portais de federações pode-se recuperar a informação de duas formas, por meio de busca ou navegação. Independente da forma a apresentação da informação é a mesma, visto que é o mesmo processo final para ambos os processos, diferindo a forma de seleção do registro a ser apresentado. Essa questão destaca a principal característica dos portais das federações de facilitar o acesso ao acervo.

A busca é ofertada por meio das opções simples e avançada, com uso de operadores booleanos e outras facilidades orientadas pela política de opções de busca. O resultado da busca, independente se na ferramentas simples ou avançada será a mesma, uma lista organizada conforma as orientações da política de busca.

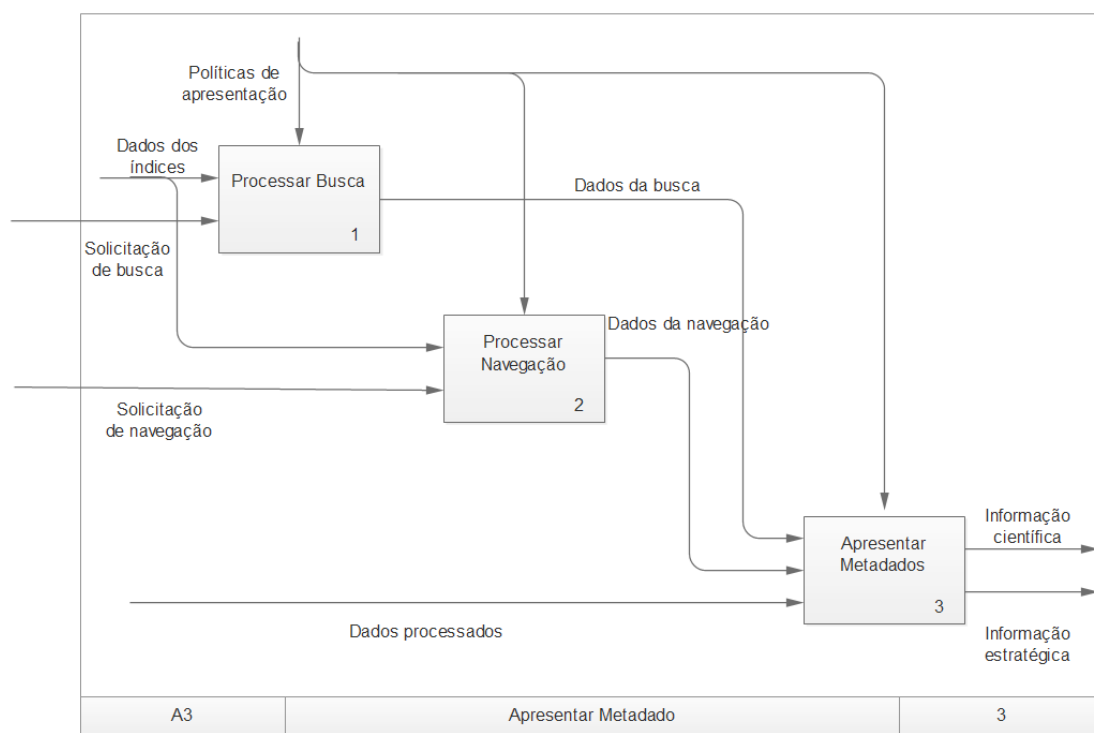
A navegação, por sua vez, pode ser ofertada de várias formas, visto que é possível gerar formas diversificadas de organização com base nos metadados. Assim, pode-se criar listas para navegação por unidades acadêmicas, autores, datas, tipos de documentos entre outros definidos pelas políticas de navegação. A navegação é

importante para àqueles usuários que preferem navegar pelos sites levantando os itens do acervo.

Ao selecionar um item, independentemente se por meio de busca ou navegação, o portal apresenta os metadados do item. A forma de apresentação e os metadados apresentados são definidos pela política de campos de metadados. Atualmente pode-se ofertar forma mais inovadoras, com a presença de miniatura, acesso via texto, vídeo ou áudio streaming, por exemplo. Entretanto, nas federações essa apresentação fica restrita ao que está presente nos metadados coletados, indexados e processados.

Outra opção para a apresentação dos metadados inclui a geração de indicadores (figura 37). De fácil implementação, a geração de indicadores básicos sobre o acervo e os acessos, como forma de estatística, podem embasar tomadas de decisão e assim ser considerado informações estratégicas. Mesmo não sendo uma opção comum nas federações, é mais comum nos repositórios, essas estatísticas apresentam-se muito representativas nas federações por serem extraídas de acervo consolidado.

FIGURA 37 - DIAGRAMA APRESENTAR METADADOS COM OS INDICADORES



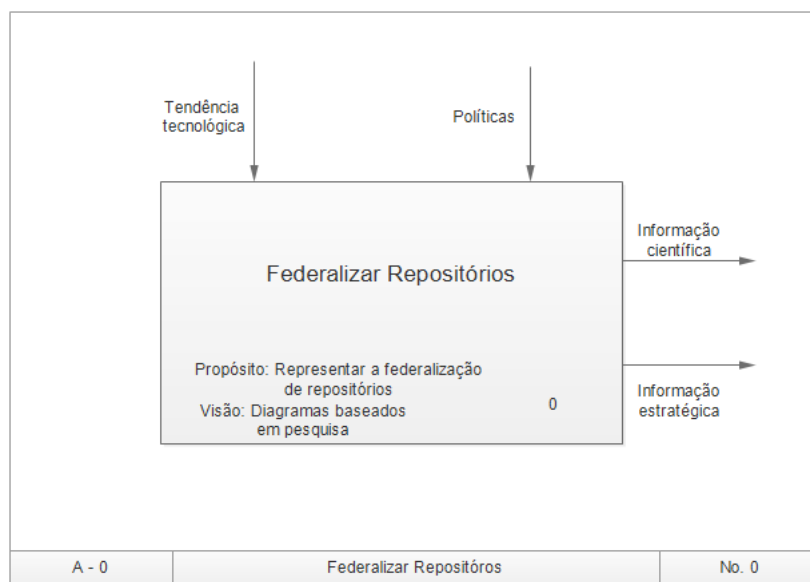
Fonte: o autor

### 5.3.2 TENDÊNCIAS TECNOLÓGICAS E AS FEDERAÇÕES

As federações, assim como todos os sistemas informatizados, estão sujeitas a evolução tecnológica, que propõem novas funcionalidades. Ofertam atualizações alinhadas com as tendências que visam melhoria nos sistemas, de forma a atender melhor os usuários. Muitas dessas tendências refletem a própria evolução da tecnologia, na medida em que nem sempre as necessidades dos usuários podem ser atendidas por questões tecnológicas atuais, mas que podem ser atendidas futuramente.

Nesse sentido, pode-se esquematizar uma federação, utilizando o diagrama IDEF0 (figura 38). Com isso, assegurar operação e atualização com as tendências tecnológicas e as políticas da federação. Essa proposta tem por base a análise das tendências tecnológicas e as possíveis adaptações para as federações. Com isso, apresentar um modelo de federação mais atual, agregando as tendências tecnológicas ao modelo identificado e analisado em 5.1, deste estudo.

FIGURA 38 - DIAGRAMA FEDERALIZAR REPOSITÓRIOS COM TENDÊNCIAS TECNOLÓGICAS



Fonte: o autor; 2014.

Nesse sentido, esta seção analisa e propõe mudanças no modelo apresentado para as tendências sugeridas pelo Nexus das Forças (nexus of forces), proposto pelo Gartner Group, enumeradamente: mídias sociais, mobilidade, big data e computação em

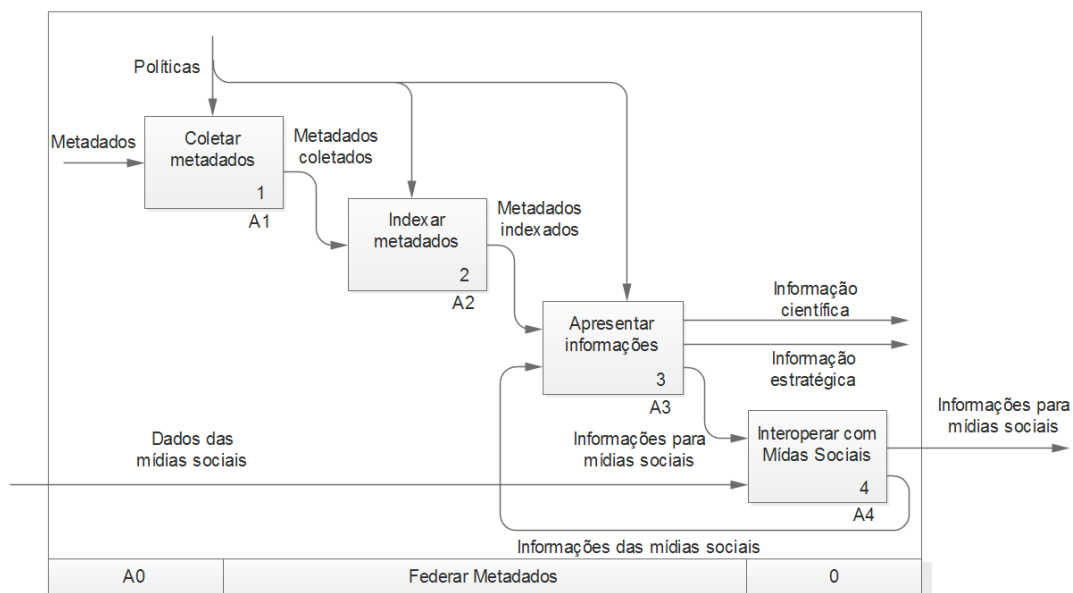
nuvens. Com o proposito de não se tornar extenso, apresenta-se apenas os diagramas afetados pela tendência.

### 5.3.2.1 MÍDIAS SOCIAIS

Atualmente a internet tem promovido interações entre usuários, entre outras formas, por meio das mídias sociais, um fenômeno que tem se apresentado inovador em alguns aspectos, devido à quantidade de opções ofertadas.

A federação e as mídias sociais são sistemas distintos, mas ambas interagem com os usuários por meio de interfaces disponíveis na internet. Assim, a integração entre esses sistemas dá-se no processo de apresentar informações, pois é o processo pelo qual ocorre a interação entre os usuários. No entanto, essa integração deve-se dar sem que requeira alterações nas mídias sociais, pois são mantidos por outras instituições, da mesma forma que não deve alterar as funcionalidades da federação, como uma nova facilidade ofertada.

FIGURA 39 - FEDERAR REPOSITÓRIOS COM MÍDIAS SOCIAIS



Fonte: o autor; 2014.

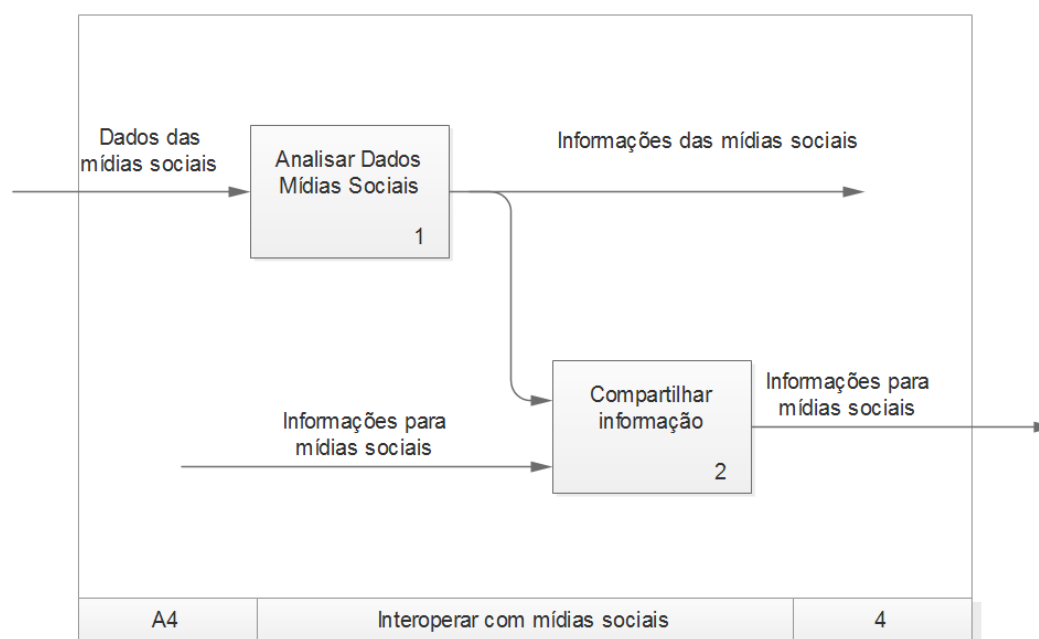
A figura 39 apresenta a proposta de integração das mídias sociais à federação, por meio da agregação de um novo processo, que adiciona funcionalidades à federação. Com isso não requer alterações significativas nos processos da federação, utilizando

algumas facilidades ofertadas pelas mídias sociais. Entretanto, as mídias sociais englobam uma grande variedade de sistemas, o que requer formas distintas de integração, revelando certos desafios aos desenvolvedores da federação.

Outra opção é a utilização de ferramentas desenvolvidas para esse fim, visto a tendência na integração das mídias sociais em sistemas informatizados. Algumas empresas comerciais de informática têm desenvolvido pacotes que oferta a integração de sistemas às mídias sociais mais comuns, como o Facebook, LinkedIn, Twitter, G+ entre tantos outros. Com isso evita-se desenvolvimento ou alterações na federação, visto que essas ferramentas integram os sistemas às mídias sociais de forma fácil, sem requerer alterações significativas.

Independente da forma de interoperar, nota-se que pode ser composta por dois processos como apresentado na figura 40. Um dos processos visa apoiar o compartilhamento de informações do repositório, com as mídias sociais, enquanto o outro processo analisa as informações fornecidas pelas mídias sociais para ser apresentada na interface da federação. Com isso atende a duas tendências, ofertar aos usuários informações sobre a federação tramitando nas mídias sociais e possibilidade de compartilhar informações pertencentes aos repositórios nas mídias sociais.

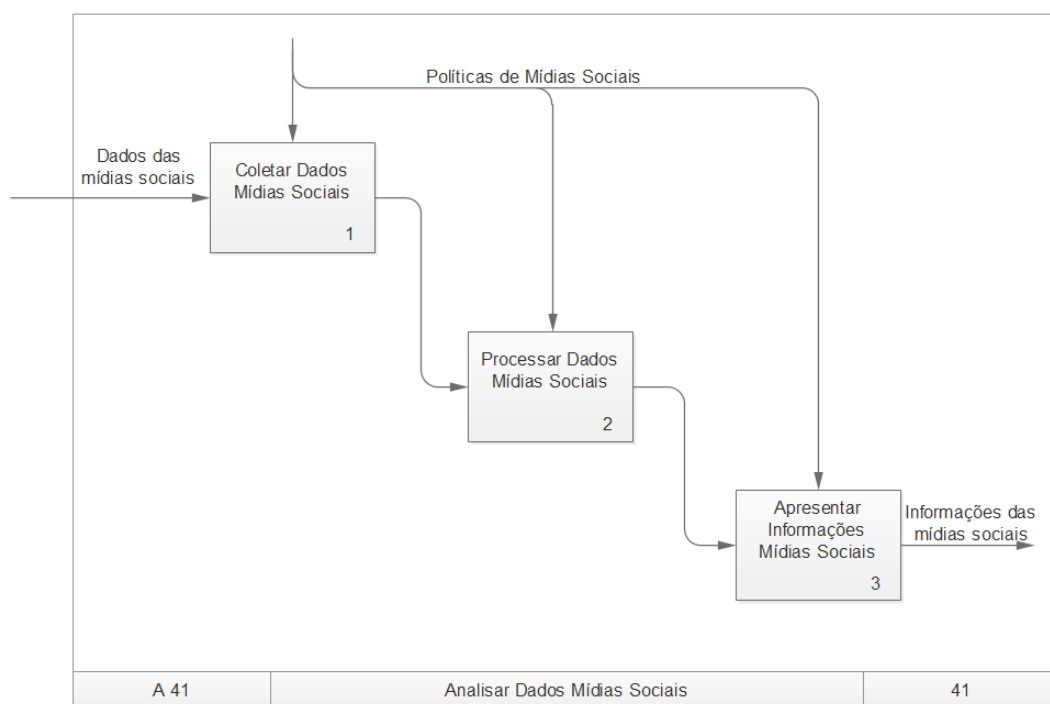
FIGURA 40 - INTEROPERAR COM MÍDIAS SOCIAIS



Fonte: o autor; 2014.

O processo de Analisar Dados Mídias Sociais é composto pela coleta e processamento de dados provenientes das diversas mídias sociais e a apresentação das informações processadas, orientados pelas políticas de mídias sociais, que indicam como os processos devem ser executados (figura 41). Com isso possibilita que os dados coletados das mídias sociais selecionadas possam ser processados conforme critérios especificados e apresentadas da forma desejada aos diversos tipos de usuários.

FIGURA 41 - DIAGRAMA ANALISAR DADOS MÍDIAS SOCIAIS



Fonte: o autor; 2014.

O processo de coletar Dados Mídias Sociais recebe dados das mídias selecionadas pelas políticas, na medida em que nem todas as mídias podem ofertar dados que interessem à federação. Assim, é um processo automatizado e configurável, pois deve ser flexível para se ajustar as necessidades de coleta de dados.

O Processamento de Dados das Mídias Sociais executa procedimentos que possibilita a contabilização e análise dos dados provenientes das mídias sociais. É o processo pelo qual os dados desestruturados são convertidos em formas estruturadas que possibilitam a análise automática e contabilização, de forma a gerar indicadores de mídias sociais.



A apresentação deve ofertar vários níveis de informação, conforme o usuário. Pode- ofertas os dados de forma bruta, iguais aos das mídias sociais a quais foram publicadas, para que a análise seja executada por quem acesse ou apresentada de forma mais tratada como gráficos ou tabelas. A informação adequada a cada usuário pode ofertar:

- **Gestores da federação:** podem obter informações estratégicas das mídias sociais em formas sumarizadas, pré-analisadas, que o apoie a tomar decisões, principalmente na reformulação das políticas, tais como dados demográficos dos usuários que comentam, tipo de interação, horário, itens mais discutidos, taxa de crescimento de interações, entre tantos outros;
- **Gestores dos repositórios vinculados à federação:** podem verificar informações sobre a participação do repositório na federação, filtrando informações exclusiva da seu repositório ou comparando as informações do seu repositório com outros;
- **Estudiosos:** levantar e analisar informações sobre a federação e os repositórios que a compõe, além dos usuários que comentam sobre a federação, como os dados demográficos (gênero, faixa etária, entre outros).
- **Público em geral:** ter informações consolidadas provenientes de várias mídias sociais.

Nesse ponto, revela a importância das mídias sociais na gestão da federação, ao passo que se torna o locus de interação entre usuários, possibilitando discutir abertamente sobre vários pontos da federação. Com isso, gestores e estudiosos podem utilizar essa funcionalidade para extrair dados que apoiem as suas tarefas. Tanto que, Montalvo (2011) revela que as organizações necessitam de um profissional para gerir as mídias sociais, visto a influência desses sistemas nas organizações.

### 5.3.2.2 MOBILIDADE

Cada vez mais se requer que os portais disponíveis na internet ofertem novas funcionalidades, tornando-os sistemas complexos. Serviços online são ofertados para

executar uma grande variedade de operações, impondo desafio ao desenvolvedores. As informações apresentadas nos portais são dinâmicas e, muitas vezes, possibilitam interação entre usuários (clientes) e os gestores dos portais, permitindo a criação de relacionamentos.

Da mesma forma, cada vez mais os portais são acessados por dispositivos móveis, visto a evolução desses equipamentos. O Centro de Estudos sobre as Tecnologias da Informação e da Comunicação (Cetic.br), no relatório de 2013 sobre o acesso à informação das organizações, confirmou essa tendência ao revelar que o acesso móvel nas grandes empresas passou de 33%, em 2012, para 42%, em 2013. Esse aumento ratifica o aumento do uso de dispositivos móveis para acesso à internet para diversos fins, inclusive trabalho e negócios.

Nesse contexto as federações devem estar acessíveis tanto por equipamentos fixos quanto móveis e para sê-lo requer adaptações, que podem ser implementadas de duas formas, adicionando um aplicativo para acesso móvel ou pela adoção de tecnologias responsáveis nas interfaces. Ambas as soluções possuem vantagens, entretanto a segunda opção apresenta-se mais adequada as federações, pois fornecem serviços de busca, recuperação, descoberta e entrega de informação científica, não requerendo o desenvolvimento de aplicativo móvel.

O desenvolvimento de portais com os preceitos da tecnologia responsiva possibilita que possa ser acessado por diversos dispositivos, sem perda de qualidade do layout, ao ponto que as páginas se ajustam ao tamanho da tela. A tecnologia responsiva, também conhecida por web design responsivo ou layout responsivo, é uma técnica de desenvolvimento de interfaces baseada no protocolo HyperText Markup Language (HTML) versão 5 e folhas de estilo Cascading Style Sheets (CSS) versão 3, que possibilitam esse ajuste automático no tamanho da tela.

Marcotte (2010) cunhou o termo web design responsivo, com base na chamada arquitetura responsiva, em que os espaços físicos se adaptam à presença humana. Para o referido autor, não se deve encarar a grande diversidade de dispositivos móveis de forma separada, pelo contrario, deve-se vê-los como se fossem facetas de um mesmo

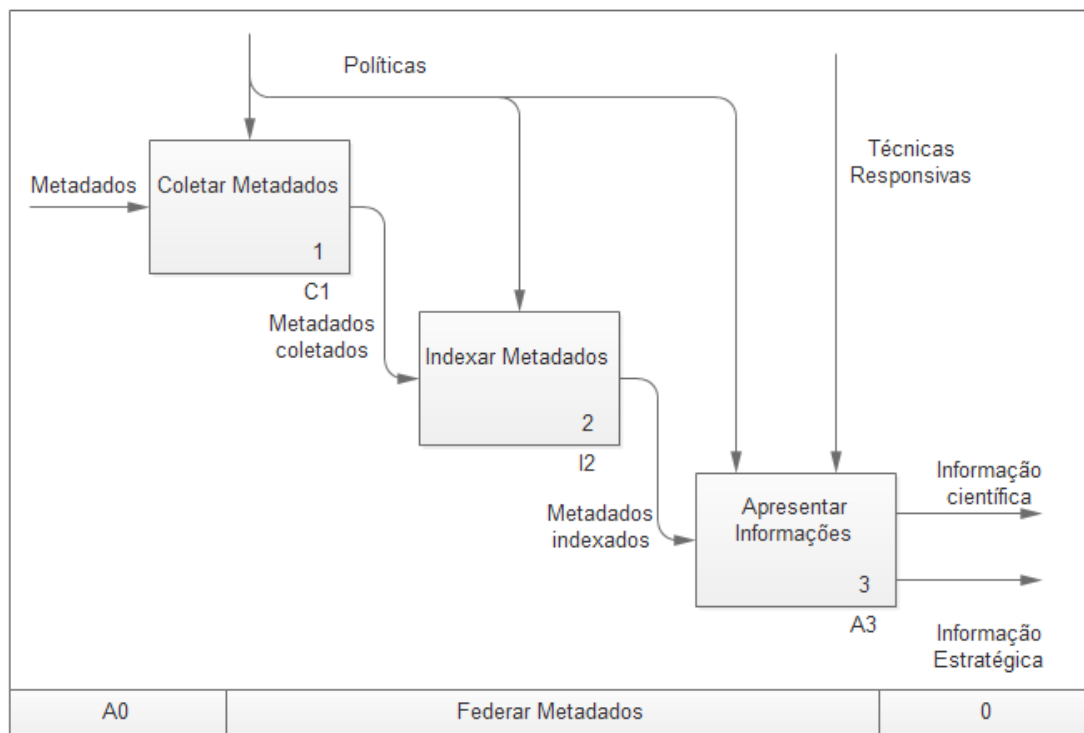
problema. Com isso, utilizar a tecnologia para tornar as interfaces mais flexíveis e adaptáveis, na web design responsivo.

Robbins (2012), por sua vez, destaca que a web design responsivo é uma estratégia para ofertar layouts ajustáveis a qualquer dispositivo, pela utilização de páginas desenvolvidas com o HTML, mas com CSS específico para cada dispositivo. Dessa forma, pode-se garantir que a mesma página web se adapte a tamanhos diferentes de tela, apresentando um grande diferencial para economia em manutenção de portais e atendendo a diversos equipamentos.

Assim, com a utilização dessa tecnologia possibilita a federação de ser acessada por meio de diversos tipos de equipamentos, tornando mais eficaz na sua finalidade de disseminar informação científica, visto que o acesso via dispositivos móveis tem se tornado comum. Não precisando alterar ou adicionar processos, mas apenas utilizar as orientações da tecnologia responsiva no desenvolvimento ou ajuste das interfaces do repositório.

Nesse quesito, mesmo que a tecnologia responsiva seja aplicada a todas as páginas da federação, o maior diferencial será no processo de Apresentação da informação, ao passo que nesse processo ocorre a interação maior entre a federação e seus usuários finais. Pode-se representar essa questão como na figura 42, em que as técnicas da tecnologia responsiva são orientadoras das páginas desse processo.

FIGURA 42 - FEDERAR REPOSITÓRIOS COM TÉCNICAS RESPONSIVAS



Fonte: o autor; 2014.

Evidentemente as políticas orientam na determinação do uso das tecnologias responsivas, mas não indicam como fazer, por isso a separação em duas linhas na representação. As políticas de apresentação como vistas anteriormente tem por objetivo orientar o funcionamento e, em muitos casos, servir para tomada de decisão. Nesse caso específico as Técnicas responsivas indicam como as páginas do processo de apresentação da informação devem ser desenvolvidas.

Essas técnicas podem ser aplicadas a todas as páginas de todos os processos e possibilitar que os integrantes da equipe de gestão da federação possam executar suas tarefas por meio de dispositivos móveis. Com isso, aumentar a oferta de acesso, ao passo que, cada vez mais os dispositivos móveis são utilizados no trabalho, visto a disponibilidade de acesso. Assim, tendo todas as páginas do portal da federação responsivas.

Com isso, a web desing responsivo ou tecnologias responsivas proveem à federação de interfaces acessíveis para ser acessado por qualquer dispositivo, estando em consonância coma a acessibilidade móvel. Entretanto, questões de usabilidade e

acessibilidade à deficientes requerem estudos mais profundos de forma a atender às necessidade dos usuários. Nesse ponto o estudo apenas propõe a adoção dos preceitos das tecnologias responsivas para atender à acessibilidade móvel, como forma de atender às tendências tecnológicas.

### **5.3.2.3 BIG DATA**

Das tendências estudadas na pesquisa no âmbito dos repositórios, o Big Data se apresentou como a tendência tecnológica menos conhecida. Entretanto, possivelmente essa tendência representará o maior impacto para a gestão dos repositórios quando totalmente implementada por esses sistemas. Da mesma forma, o Big Data utilizado nas federações proporcionará uma grande diferencial, principalmente, na geração de indicadores e informações estratégicas.

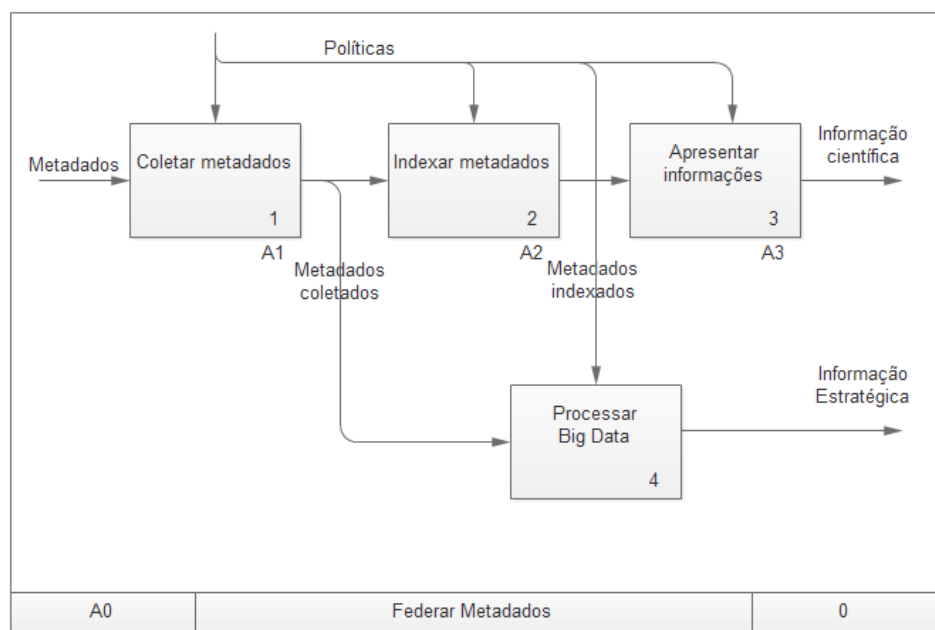
Nesse sentido o Big Data tem relação à geração de informações estratégicas em tempo real, a partir de grandes quantidades de dados, geralmente não estruturados, por meio de um conjunto de métodos e ferramentas. Muitas das pesquisas como as dos genomas, estudos astrofísicos e monitoramento ambiental geram grandes quantidades de dados utilizam essas técnicas e ferramentas para processar as informações, assim como as empresas comerciais para apoiar seus negócios.

Nesse ponto que as federações se apresentam como possíveis usuários do Big Data, por manter uma grande quantidade de informações, mesmo que estruturados. A NDLTD, por exemplo, possui quase quatro milhões de registros em fevereiro de 2015, oriundos de vários países e diversos assuntos, que representam as pesquisas das universidades efetuadas na graduação e pós-graduação. Revelam-se, assim, oportunidades e desafios às pesquisas, ao ponto que a base se apresenta disponível para estudos.

Nesse contexto, o Big Data pode ser adicionada à federação, como visto na figura 43, em que recebe os metadados coletados para gerar informações estratégicas. Nesse sentido, informações estratégicas podem assumir uma grande variedade de tipos de informações oriundas do processamento dos metadados, de indicadores simples sobre o acervo, até estatísticas, mas elaboradas. Mesmo que os metadados estejam de

forma estruturadas a grande quantidade e a variedade de temas apresenta-se como desafio ao processamento tradicional.

FIGURA 43 - FEDERAR REPOSITÓRIOS COM BIG DATA



Fonte: o autor; 2014.

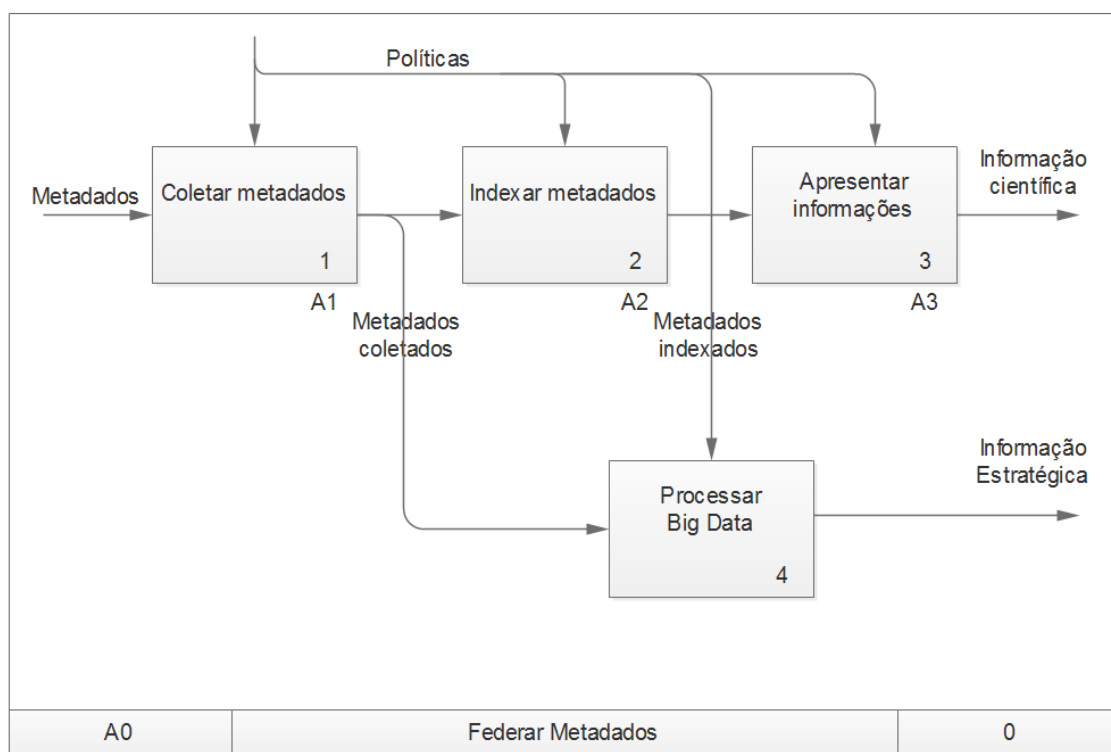
As oportunidades de uso das orientações e ferramentas do Big Data para extração de informação estratégica, de acordo com os metadados, dependem evidentemente dos repositórios e os metadados utilizados. Mesmo que apresente certa restrição, que geralmente não ocorre em bases desestruturadas, a riqueza dos temas e a relevância das informações tornam as federações importantes fontes de informação. Assim, algumas opções de processamento dos metadados:

- **Resumo:** dos campos de metadados, o resumo possui uma características textual, que impõe desafios aos processamentos tradicionais, mas que o processamento do Big Data pode extrair informações relevantes, pela própria característica dos resumos, que nos documentos científicos sintetizam de forma concisa o estudo apresentado;
- **Assunto (palavras-chave):** as palavras-chaves são campos muito estruturados, porém apresentam grande variação terminológica, que impõem dificuldades nas análises automáticas. Problemas para desambiguações de termos pertencentes a áreas do conhecimento, por exemplo, ainda se apresenta com problemas em muitas federações;

- **Título:** nos documentos científicos, o título muitas vezes sintetiza o estudo apresentado, por isso a grande presença de subtítulos. Não sendo incomuns as teses e dissertações possuírem títulos longos e ainda com subtítulos. Por isso torna-se um campo promissor ao processamento com as técnicas e ferramentas do Big Data;
- **Autoria:** esse é um campo simples que não tem como processá-lo sozinho, mas ao realizar cruzamentos com outros campos, possibilitam uma grande quantidade de análises:
  - **Autoria:** possibilita verificar as relações entre autores, como as parcerias e contribuições;
  - **Assunto:** possibilita levantar as comunidades que estudam determinados assuntos, além das relações interdisciplinares dos autores.

Outra utilização do processamento do Big Data se apresenta como apoio à indexação (figura 44). Os resultados das extrações de meta-informações e relações por meio do processamento do Big Data podem ser incorporados pelos indexadores, de forma a melhorar os processos de recuperação. Esse processamento pode-se apresentar como uma opção às federações, pela variedade de temas, idiomas e termos. Nota-se que pela análise das federações, a recuperação ainda apresenta problemas (seção 5.3.1).

FIGURA 44 - FEDERAR REPOSITÓRIOS COM BIG DATA PARA INDEXAÇÃO



Fonte: o autor; 2014.

As informações geradas pela incorporação do Big Data à federação poderá ser útil a vários tipos de usuários, além de potencializar alguns processos internos. Entretanto, ainda apresenta-se como um desafio ao gestão acadêmica a utilização de informação estratégica. Para órgãos de fomento, por exemplo, as informações estratégicas das federações seriam de enorme valia, por conter a informação consolidada oriundas de várias instituições. Da mesma forma que pode ofertar informações que revelem o cenário da pesquisa por diversos critérios.

Assim, o Big Data, mesmo não sendo muito conhecido e aplicado em determinadas áreas, sendo mais utilizado no âmbito das organizações e negócios. Nas Ciências é utilizada como ferramenta de apoio à pesquisa, de forma restrita ao processamento de grandes quantidades de dados. Entretanto, revela-se promissor para geração de informação estratégica para a gestão científica e, principalmente, para a geração de indicadores para a federação.

#### 5.3.2.4 COMPUTAÇÃO EM NUVENS



A computação em Nuvens tem se apresentado como uma opção à infraestrutura para diversas instituições e serviços. Muito provavelmente, grande parte das pessoas faz uso de serviços mantidos em nuvens sem ter conhecimento dessa tecnologia. Serviços como os do Google, por exemplo, residem em nuvens. Os usuários dos equipamentos Apple utilizam o iCloud residente em nuvens. A transparência, robustez e continuidade de serviços tem sido a tônica dessa tecnologia.

No caso das federações a computação em nuvem pode ofertar a infraestrutura necessária ao seu funcionamento, no que é denominado de Infraestrutura como Serviço (Infrastructure as a Service - IaaS ), em nuvem da própria instituição, visto questões de segurança. Com isso, distribui a responsabilidade de manutenção e sustentabilidade da federação entre diversas equipes da instituição.

O IaaS possibilita ofertar à federação os recursos necessários, de forma virtual, ao seu funcionamento, no que Mell e Grance (2011) descreve como a infraestrutura ofertada pelo IaaS, que inclui poder de processamento, acesso à rede, capacidade de armazenamento, além de softwares básicos para o funcionamento adequado ao sistema. Nurmi e colaboradores (2009) destacam que a nuvem é uma estrutura virtual que está distribuída por uma variedade de recursos físicos, mas que se apresenta aos sistemas como uma estrutura única.

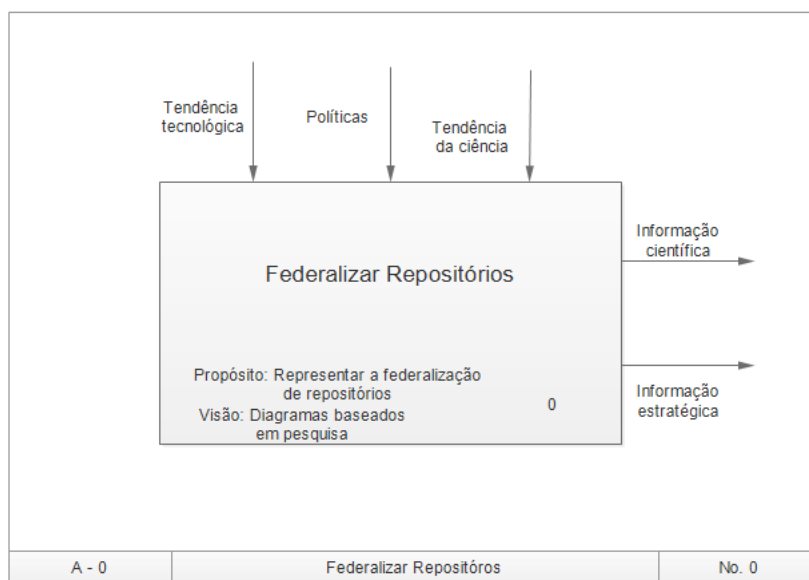
Das tendências tecnológicas estudadas, essa é, possivelmente, a que os gestores da federação terão a menor interferência, visto que, na maioria dos casos é de responsabilidade dos gestores de TIC por ter relação maior com a infraestrutura computacional. Nesse ponto, as federações apresentam vantagens tecnológicas ao residirem nas nuvens, mas nem sempre é decisão dos gestores da federação, mas uma questão política da instituição, amparada pela equipe de TIC.

### **5.3.3 TENDÊNCIAS DA CIÊNCIA E AS FEDERAÇÕES**

Assim como os repositórios, as federações já fazem parte da comunicação científica, no que diz respeito a facilitação ao acesso à informação. Com isso, devem estar em consonância com as tendências da ciência, de forma a atender melhor seus usuários, que em grande parte pertencem à comunidade científica. Nesse sentido pode-

se propor uma federação de repositórios científicos, que atenda as tendências da ciência, como no esquema apresentado na figura 45.

FIGURA 45 - FEDERAR REPOSITÓRIOS COM TENDÊNCIAS DA CIÊNCIA



Fonte: o autor; 2014.

Assim, os processos da federação passam a ser orientados pelas tendências da ciência, da mesma forma que as políticas e as tendências tecnológicas, numa tríade que torna a federação mais atualizada. Essas questões apoiam a proposta do modelo baseados nos estudos das federações, apresentados na seção 5.1, e nos repositórios, apresentados na seção 5.2.

### 5.3.3.1 ACESSO ABERTO

O Movimento de acesso aberto tem sua origem na chamada crise dos periódicos, que torna a aquisição das revistas científicas inviáveis pelo valor cobrado pela assinatura, criando um novo modelo para a publicação de artigos. Com isso fomenta a criação de revistas de acesso aberto, ou seja, que publicam os conteúdos sem barreiras para o acesso integral, sem, no entanto, romper com questões tradicionais como a avaliação pelos pares. Assim, com ênfase nos periódicos que publicam os resultados de pesquisa, ou no que Hanard e colaboradores (2004) denominam de via dourada.

Os repositórios, por sua vez, complementam a ação dos periódicos, expandindo a ação do acesso aberto para outras tipologias de documentos científicos, como os livros, teses e dissertações. Para tanto, requer a permissão dos editores e editoras para depositar e fornecer acesso, no que Hanard e colaboradores (2004) denominaram de via verde. Assim, a complementaridade da via dourada e verde se destacam na medida em que os periódicos publicam os novos conhecimentos e os repositórios preservam e facilitam o acesso, agregando outras tipologias documentais científicas.

Assim, os repositórios tornam-se o locus principal, facilitador do fornecimento de acesso ao conteúdo integral de documentação científica. Por isso, a questão das federações coletarem diretamente dos repositórios. Entretanto, nem todos os registros mantidos pelos repositórios remetem ao texto integral, assegurados pelas suas políticas e mandatos, que resguardam documentação restrita por motivos econômicos e de segurança. Assim, os documentos mantidos pelos repositórios podem estar com o acesso livre, restrito a determinados usuários ou para todos ou embargados (restritos por tempo determinado).

Essa variação de permissões de acesso não gera problemas ao repositório e são assegurados pelas políticas e mandatos, pois a finalidade do repositório é, em parte, agregar a produção acadêmica da instituição e não, necessariamente, fornecer o acesso a toda essa produção. Entretanto, para as federações essa questão podem ser discutidas e formalizadas nas políticas, para evitar frustrações dos usuários por não ter acesso aos documentos indexados pela federação.

Ao se alinhar aos preceitos do acesso aberto, as federações podem orientar a configuração do processo de indexação para evitar indexar documentos restritos ou embargados. Nas orientações do OpenAire, por exemplo, há uma normalização por meio de um vocabulário controlado para o preenchimento do campo `dc:rights`, com os seguintes conteúdos:

- **info:eu-repo/semantics/closedAccess:** para documentos restritos a todos;

- **info:eu-repo/semantics/embargoedAccess:** para documentos embargados, requerendo a inserção da data de liberação em outro campo de metadado;
- **info:eu-repo/semantics/restrictedAccess:** para documentos restritos, mas com acesso fornecido a uma determinada comunidade;
- **info:eu-repo/semantics/openAccess:** para documentos totalmente abertos ao acesso.

Assim, pode-se o processo de indexação pode descartar os registros com base no conteúdo do campo dc:rights, que não sejam de acesso aberto. Com isso, pode-se assegurar que todos os registros indexados pela federação remetem ao conteúdo integral, mantido nos repositórios. Assim, o alinhamento ao preceito principal do movimento de acesso aberto fica garantido e aumentando o grau de satisfação dos usuários.

Essa questão, no entanto, é política, ao passo que, as federações podem escolher ser uma representação fiel dos repositórios agregados ou alinhar-se a determinados pontos. Por isso as políticas de federação são importantes e devem ser explicitadas para orientar o seu funcionamento e servir, em grande parte, para tomada de decisão na operacionalização dos processos.

### 5.3.3.2 ARQUIVOS ABERTOS

Da mesma forma que Movimento e Acesso Aberto, os Arquivos Abertos possuem proximidades com as federações. Enquanto o Acesso Aberto é mais técnico com questões mais voltadas às orientações, os Arquivos Abertos possuem aspectos mais pragmáticos relacionados a definir como a federação deve ser implementada, com foco mais tecnológico. Nesse ponto, a federação só é possível pela interoperabilidade definida por esse movimento, que possibilita a coleta automática de metadados.

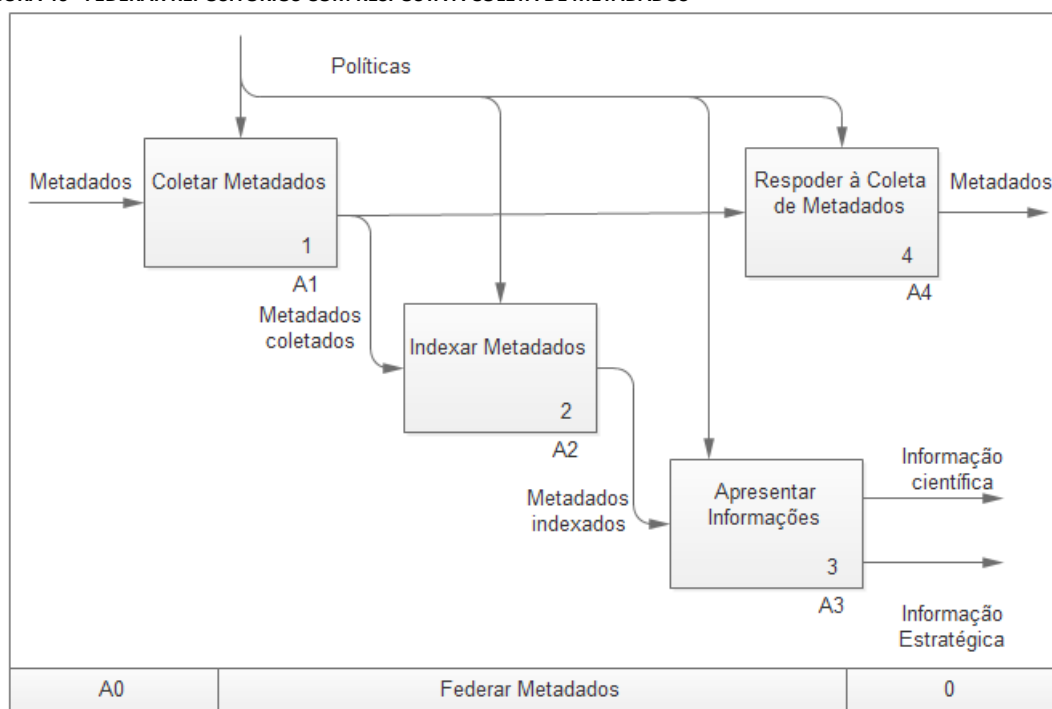
Nesse caso, a interoperabilidade se refere a aplicação de três princípios que permite que dados sejam coletados de um repositório (provedor de dados) pela federação (provedor de serviços):

- **uso de um padrão de metadados (esquema de metadados):** para que a coleta seja executada, requer-se que tanto o repositório, quanto a federação conheça a estrutura do dado, visto que as federações, em grande parte, são banco de dados estruturados. Assim, devem compartilhar o padrão de descrição, ou seja, conhecer o esquema de metadados a qual estão interoperando. No entanto, deve-se destacar que federação se repositórios pode adotar vários esquemas de metadados, não precisando ser único;
- **uso de um protocolo de comunicação comum:** para que a comunicação se efetive entre os repositórios e a federação é necessário que utilizem protocolos de comunicação semelhantes, em que a federação colete dos metadados e os repositórios repondam à coleta;
- **uso do formato XML na comunicação:** a informação interoperada entre o repositório e a federação deve estar no formato XML, tanto na requisição quanto na resposta. Com isso, a informação intercambiada integra conteúdo e estrutura, possibilitando validação por meio de programas que verificam a conformidade do dado, baseada no esquema de metadados utilizada.

Essas questões se restringem ao processo de coleta de metadados (tratada neste documento na seção 5.3.1), no qual os metadados de um repositório são coletados e copiados para a base de dados da federação, passando por validação e mapeamento. Assim, em parte os arquivos abertos estão presentes na federação na formação da base de dados, ou seja, na própria finalidade do repositório.

Entretanto, outro ponto a ser discutido em relação aos arquivos abertos e a federação, ampara-se na questão do processo em que a federação responde à coleta de outra federação, na medida em que grande parte desses sistemas o faz (BDTD, oasisbr, RCAAP, REMERI, ALICIA, BDCOL e outros). Essa é uma questão opcional, determinada pelas políticas da federação, não interferindo nos processos vigentes, mas na adição de mais um processo à federação (figura 46).

FIGURA 46 - FEDERAR REPOSITÓRIOS COM RESPOSTA À COLETA DE METADADOS



Fonte: o autor; 2014.

Não apenas o processo de responder às coletadas de metadados deve ser adicionado, mas as suas políticas também, visto que é um processo que precisa de orientação para a sua execução.

- ❖ **tipos de federações a ser atendidas:** essa questão política deve ser bem explicitada, pois muitas empresas comercializam bases de dados, que podem ser de acesso aberto, coletadas por meio da interoperabilidade. Assim, pode-se restringir o acesso aos metadados por meio da coleta apenas para federações de acesso aberto ou deixar que todos colem os metadados sem restrição;
- ❖ **esquemas de metadados para resposta:** as políticas devem definir quais os esquemas de metadados que a federação pode responder à coleta, visto que muitas federações são mais restritas, requerendo determinado esquema. Pos isso, assim como os repositórios, as federações devem ofertar metadados em vários esquemas de metadados, orientados pelas políticas.
- ❖ **sets ofertados:** a coleta de metadados pode ser total ou parcial, sendo que as parciais podem ser relacionada a tempo ou parte do acervo.

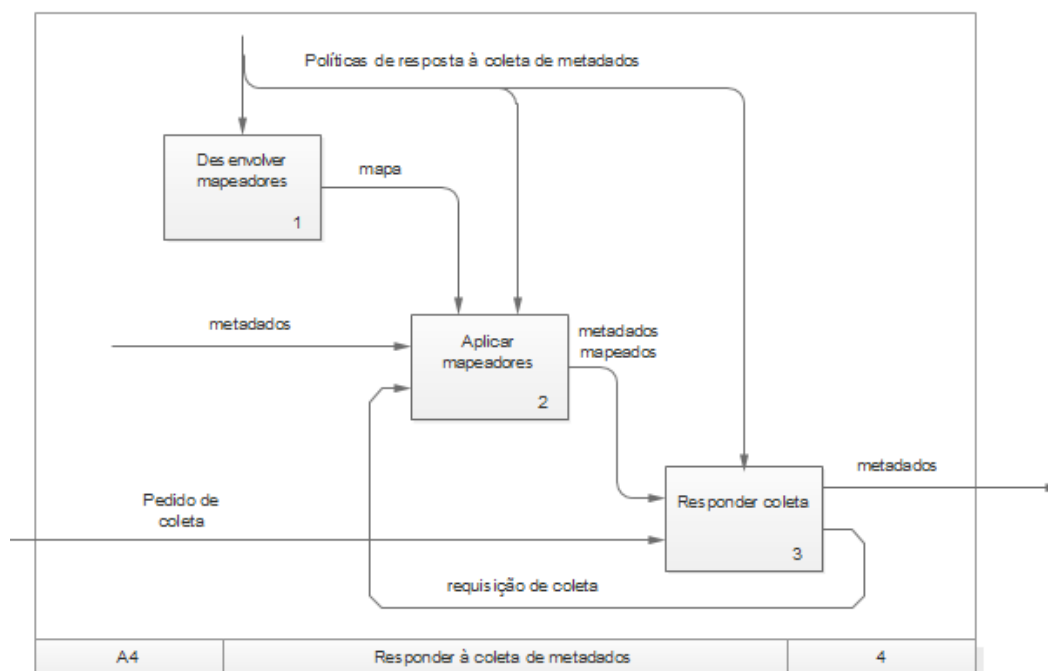
Assim, os sets tem relação com a coleta parcial, em que a federação pode criar partições do acervo conforme determinadas características presentes nos metadados. Com isso os sets podem ser organizados, por exemplo, das seguintes formas por exemplo:

- **origem dos registros:** organização do acervo para coleta pela origem dos registros em que cada partição reflete os repositórios coletados;
- **tipo de documentos:** organização do acervo para coleta por tipologia documental, criando partições de artigos, livros, teses e dissertações, entre outros. Assim, não importa a origem dos registros, mas o tipo de documento que descreve. Por exemplo a NDLTD só coleta registro de trabalhos de conclusão de cursos e a BDTD só coleta metadados de teses e dissertações;
- ❖ **normalização de conteúdo:** muitas federações requerem normalização de conteúdo de metadados. Assim, a federação pode mapear conforme essas requisições, criando partições adequando aos critérios de determinadas federações. Por exemplo, federações que aderem às diretrizes DRIVER requerem esse tipo de normalização;
  - tipo de acesso: partições podem ser criadas para evitar a propagação de registros de documentos embargados ou restitos. Com isso pode-se ter partições de registros de documentos de acesso aberto, restrito e embargados;
  - idiomas: nas federações de repositórios é muito comum ter registros de documentos escritos em diversos idiomas. Assim, podem-se criar partições por idiomas. Com isso, possibilitar as federações coletar apenas no idioma que interessa.
- ❖ protocolos de comunicação: atualmente a interoperabilidade pode ser implementada com os protocolo OAI-PMH e OAI-ORE. Assim, as políticas podem definir quais os protocolos a serem utilizados.

As políticas, dessa forma, orientam o funcionamento do processo de responder às coletas provenientes de outras federações, apoiando as decisões em cada etapa do processo. Como apresentado na figura 47, esse processo é composto por: Desenvolver

mapeadores, Aplicar mapeador e responder à coleta. Com isso, é possível responder às requisições de coleta, em vários esquemas de metadados, apresentando a flexibilidade necessária para atender federações que interoperam de forma diferenciada.

FIGURA 47 - RESPONDER À COLETA DE METADADOS



Fonte: o autor; 2015.

Para que uma federação possa interoperar em mais que um esquema de metadados é preciso que mapeadores (crosswalks) sejam desenvolvidos. É necessário um mapeador para cada esquema em que se deseja interoperar. Os mapeadores são programas que convertem os metadados de um esquema para outro. Da mesma forma, se os esquemas mapeados possuem vocabulários controlados restritivos, é possível alterar alguns conteúdos para se adequar ao esquema, aplicando pequenas normalizações de conteúdo.

Os sets também podem ser criados pelos mapeadores. Assim, a federação pode atender as peculiaridades de outras federações, por exemplo, a NDLTD só coleta trabalhos de conclusão de curso, restringido a coleta a tipo de documentos específicos. No caso do RCAAP a coleta é mais ampla, mas requer que os metadados estejam em consonância com as diretrizes DRIVER, requerendo um set diferenciado, relacionado à normalização do conteúdo.



Assim, as requisições e respostas de coleta de metadados chegam via protocolo de comunicação OAI-PHM ou OAI-ORE ao processo responder coleta. a requisição é enviada ao processo de aplicar mapeadores e retorna os metadados no formato e set requerido para a resposta. Tudo isso é feito de forma automática, entretanto requer configurações e ajustes sejam executadas de forma orientada pelas políticas.

Com a opção de responder à coleta, a federação adota plenamente os preceitos dos arquivos abertos, ao passo que, os arquivos abertos para as federações se restringem à interoperabilidade, pois alguns dos seus conceitos são exclusivos para os repositórios. Assim, a federação torna-se um agregador, um provedor de dados que opera, também, como provedor de dados, dois papéis importantes na disseminação da informação científica.

### **5.3.3.3 DADOS ABERTOS**

Dados abertos, de forma ampla, se referem a todos os dados que possam ser acessados livremente sem barreiras, tanto que o Open Knowledge Foundation destaca que os dados são considerados abertos se estão livres para acesso, reuso, distribuição, garantindo apenas os direitos de autores e as licenças. Assim, dados governamentais, civis, organizacionais e outros podem assumir esse status ao colocarem-se disponíveis ao acesso amplo.

No caso específico da ciência, a ideia de compartilhar dados de pesquisa para reuso não é novidade, mas somente com o advento da internet pode se difundir de forma mais ampla, pela amplitude que esse meio possibilita. Com isso os repositórios tornam-se uma possibilidade de depósito dos dados de pesquisa, ou em repositórios exclusivos de dados ou em repositórios institucionais/temáticos. Assim, atendendo a necessidade de estar disponível na internet em ambiente organizado e de fácil recuperação.

Sayão e Sales (2014) destacam que a pesquisa podem ter os dados coletados armazenados em repositórios de dados e vinculados aos documentos finais de pesquisa, sejam artigos ou trabalhos monográficos como teses e dissertações. Com isso, tornando fontes compartilhadas para o reuso, de forma a permitir a revalidação e a novas

descobertas. Nota-se que os repositórios podem conter vários dados sobre o mesmo assunto ou temas semelhantes, de forma a proporcionar estudos mais abrangentes.

No mesma linha, mas com pequenas diferenças, os repositórios institucionais podem manter dados de pesquisa junto aos documentos finais ou separados. Assim os dados de pesquisa podem estar juntos aos documentos finais, em itens complexos compostos por vários arquivos, com descrição única. Estando separados, em coleções distintas, requerem descrições complementares em que o documento final remeta aos dados da pesquisa. Repositórios mais atuais, como visto na seção 5.2.2, já contemplam os dados de pesquisa em suas políticas de acervo.

Independente da forma de armazenamento, o aspecto mais importante é a relação entre os dados de pesquisa e os documentos finais, seja um artigo, livro, tese ou dissertação, por exemplo. Com isso, os documentos se tornam mais eficientes, porém mais complexos, ao conter internamente enlaces (links) que remetam aos dados. Um gráfico, por exemplo, pode remeter à tabela que o deu origem, assim como, a análise dos dados pode ter links para os dados, possibilitando que o leitor confirme ou extrapole a análise do autor.

Os dados mantidos nos repositórios de forma aberta impõem à federação desafios para manter as relações geradas nos repositórios. Deve-se notar que o processo da interoperabilidade coleta os metadados, de forma que cada registro é independente, formando um conjunto único de dados. Caso exista um relacionamento entre os arquivos no repositório, muito provavelmente essa relação não será preservada no processo da interoperabilidade utilizando o padrão de metadados Dublin Core, por exemplo. Em alguns casos, as relações entre itens dão-se pela estrutura do repositório, que não é preservada no processo de interoperabilidade.

Nesse sentido Marcondes (2012) e Santos Neto e colaboradores (2013) discutem sobre a questão dos dados abertos e a interoperabilidade, com a opção de ser efetuada pelo uso do esquema de metadados Resource Description Framework (RDF), em uma estrutura denominada de dados abertos interligados (linked open data). Com isso, alinhado ao que Bizer e colaboradores (2009) denominade web de dados, propondo uma forma de interligar dados abetos e estruturados.

Essa questão envolve o que Berners-Lee (2006) apresenta como princípios para dados abertos na web:

- ❖ Identificadores persistentes no formato de Uniform Resource Identifier (URI) para identificar cada recurso disponível na internet;
- ❖ Uso do protocolo Hypertext Transfer Protocol (HTTP) como forma de facilitar a localização dos recursos;
- ❖ Uso do formato RDF para representar a informação recuperada por um URI;
- ❖ Os recursos devem remeter a outros recursos apresentando a relação entre eles, por meio de links.

Nesse contexto, revela-se que para acomodar os dados abertos a federação deve interoperar com RDF para manter as relações entre os dados (entre si) e os documentos resultados de pesquisa, o que não apresenta problemas, visto que grande parte dos repositórios já está preparado para interoperar com esse esquema de metadados, pelos menos os mais novos ou os atualizados. Essa questão impõe certa divulgação, no sentido em que o Dublin Core é o esquema de metadados padrão da interoperabilidade com o uso do protocolo OAI-PMH.

As URIs dos arquivos de dados presentes nos metadados RDF podem remeter ao artigo ou tese, por exemplo. Apresentando as relações. De forma mais específica, internamente em um documento mantido por um repositório, pode haver várias URIs que indicam recursos disponíveis na federação. O tratamento das URIs presentes nos metadados devem ganhar específico para manter os relacionamentos.

Impõem-se, assim, orientações sobre os relacionamentos entre os recursos interoperados no processo de indexação, de forma a preservar àqueles definidos nos metadados. Requerem-se estudos mais profundos sobre o assunto, pois ainda é um tema a ser explorado, apresentando oportunidades interdisciplinares, na medida em que envolvem profissionais da informação e informática.

Assim, diferente do Dublin Core, o RDF permite a estruturar hierarquias, apropriado às relações entre dados, como os dados primários e os dados processados.

Da mesma forma esses arquivos de dados possuem relação com os documentos resultados da pesquisa. Os arquivos abertos nas federações requerem ações que realcem todas as possibilidades, enriquecendo a federação com a oferta de informações relacionada com dados.

Da mesma forma, pode ofertar à federação facilidades relacionadas a web semântica, apropriada ao RDF. A inclusão da questão semântica nas federações apresenta-se como inovação, com base nas federações estudadas que possuem características mais sintáticas no processo de recuperação. Com isso atende as questões levantadas sobre os dados abertos, adicionando questões de semântica.

#### **5.3.3.4. CIÊNCIA ABERTA**

Os movimentos de abertura da ciência têm impactado de diversas formas, por exemplo, os dados abertos tem ofertado economia na coleta de dados, possibilitando reuso para novas descobertas, atuando nas etapas de coleta de dados. Já o acesso aberto tem alterado, em grande parte a etapa de disseminação dos resultados de pesquisa, criando novos modelos de negócio para a publicação. Os arquivos abertos tem uma atuação diferente, mas não menos importante, em que possibilita a interoperabilidade entre iniciativas, por exemplo, de acesso aberto.

Nesse contexto, a Ciência Aberta se posiciona como um agregador dessas iniciativas, uma forma de evolução dos movimentos abertos na ciência. Destaca-se por uma mudança nas práticas de pesquisa, fortemente apoiada pela tecnologia. Numa visão de que a ciência deve, como destaca Björk (2005), gerar novos conhecimentos científicos e melhorar a qualidade de vida. Uma visão mais social da ciência.

Assim, as federações se apresentam como a ferramenta adaptada a esse novo movimento, na medida em que, como visto antes, possibilita integrar dados abertos, acesso aberto e arquivos abertos. Mesmo não ofertando todas as funcionalidades que atendam as propostas da ciência aberta, torna-se contextualizada com a ciência aberta, principalmente no que tange a integração e disseminação de dados e informação científica.

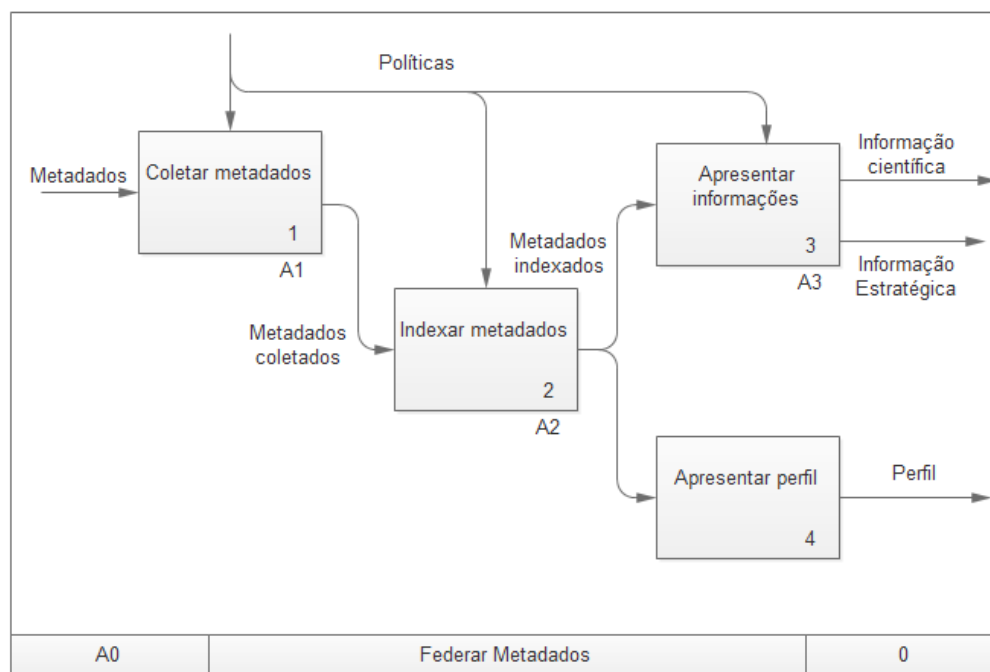
Em muitos casos nem sempre é eficaz agregar todas as funcionalidades em uma única ferramenta, sendo melhor criar ferramentas que tenham finalidades específicas, como nos casos de periódicos e repositórios, que são ferramentas distintas para finalidades diferentes, mesmo que os dois estejam em consonância com o acesso aberto.

Nesse sentido, discute-se algumas das características da ciência aberta em relação à federação, com o intuito de propor um modelo mais adequado as tendências atuais da ciência. Adição de funcionalidades que alinham ao movimento da Ciência Aberta, tornando a federação mais eficaz no papel de disseminar informação.

O reconhecimento pelos pares sempre foi uma das características da ciência, que é baseada na meritocracia. No Brasil, por exemplo, o Currículo Lattes, mantido pelo Conselho Nacional de Desenvolvimento em Ciência e Tecnologia (CNPq) apresenta perfis de pesquisadores e sua produção. Entretanto, o Currículo Lattes não remete, nem valida as informações apresentadas. Nota-se que na Ciência Aberta, o reconhecimento é uma característica, tendo relação com a produção e citações. Assim, as federações são ferramentas apropriadas para esse fim, ao ponto que as informações de produção estão presentes no acervo.

A criação de páginas dinâmicas de perfis de estudiosos torna-se relativamente fácil (figura 48), na medida em que as informações da produção já estão presentes, podendo ser pontos de validação de currículos como o Lattes. Da mesma forma, pode ofertar outros serviços como as páginas de perfis do google scholar, serviços de citação, assuntos estudados e relação entre autores. Torna-se uma ferramenta apropriada a essa característica, com adição da funcionalidade de perfis.

FIGURA 48 - FEDERAR REPOSITÓRIO COM OFERTA DE PERFIL



Fonte: o autor; 2014.

Outro ponto muito característico da Ciência Aberta é a colaboração, em que alguns autores consideram o fim dos pesquisadores solitários, ou dos autores únicos. Assim, espaços colaborativos estão alinhados a essa tendência que utiliza a tecnologia para quebrar barreiras geográficas, podendo ser de aspecto mais restritivo como o google docs, ou mais amplo como as wikis. Colocando, assim, a tecnologia no apoio a construção do conhecimento científico.

As federações não se apresentam apropriada a atender a essa característica, sendo mais adequado nos repositórios, que podem receber contribuições para pré-prints ou comentários para os pós-prints, em campos disponíveis na apresentação do documento. Da mesma forma não se apresentam apropriados a ofertar espaços para construção de conteúdos. Funcionalidades para interações entre leitores e autores se mostram ineficientes nas federações, pois os documentos estão depositados nos repositórios, sendo apresentados na federação apenas os metadados.

Possivelmente, uma das maiores discussões no âmbito da Ciência Aberta está no sistema tradicional de publicação, visto as facilidades de divulgação e disseminação ofertada pela internet. Nesse ponto, os repositórios podem apresentar modelos diferenciados, como nos casos das teses e dissertações, que os repositórios atuam como

editoras dessas documentações. Nesse caso as federações apoiam essa nova modalidade, mas não atua diretamente.

As federações tornam-se ferramentas adequadas à Ciência Aberta no propósito de ofertar acesso a informação e dados científicos, mas não a todas as suas necessidades. Coloca-se, assim, entre as ferramentas que estão em consonância com seus fundamentos, reforçando a necessidade do desenvolvimento de outras que atendam outros aspectos desse movimento.

### **5.3.3.5 CURADORIA DIGITAL**

O ambiente digital tem proporcionado muitas facilidades, mas também desafios, entre eles o da preservação, visto que as mudanças tecnológicas, ataques cibernéticos, vírus digitais e outros tornam esse ambiente altamente instável. Com isso, requer ações das instituições para manter seus serviços digitais da internet operacionais, assim como os dados digitais preservados, ofertando continuidade desses serviços.

Mais especificamente à comunidade acadêmica, a necessidade de manter os serviços ativos tem gerado interesse para a chamada curadoria digital, visto que manter a produção intelectual da instituição disponível via internet, fomenta a visibilidade da instituição. Garantir por meio de um conjunto de procedimentos a continuidade de serviços, além de ofertar facilidades para todo o ciclo de vida dos documentos digitais.

No que concerne à federação, preliminarmente a curadoria não parece ser adequada, visto que o acervo da federação é coletado dos repositórios, ou seja, o acervo é composto de cópias dos registros oriundos dos repositórios. Assim, a curadoria seria uma atividade para os repositórios e outros sistemas mantidos pelas universidades e institutos de pesquisa, como as revistas científicas e outros.

Note que os registros mantidos pela federação podem ser recomposto a qualquer momento, por meio de uma coleta total. Em muitos casos, a federação pode coletar todos os registros de um repositório por motivos diversos, ou seja, a curadoria não se apresenta como uma atividade primordial à federação, por não ter os arquivos dos documentos, apenas os metadados.

No entanto, entre os objetivos da Curadoria Digital a preservação pode ser discutida no âmbito da federação, pois como destaca Sayão e Sales (2014), a Curadoria Digital contempla metodologias para vários processos, entre eles a questão do arquivamento e preservação digital apoiado por tecnologia. Nesse ponto o presente estudo propõe uma opção que apoia a curadoria, restritamente alinhada ao objetivo da federação em ofertar acesso à documentação.

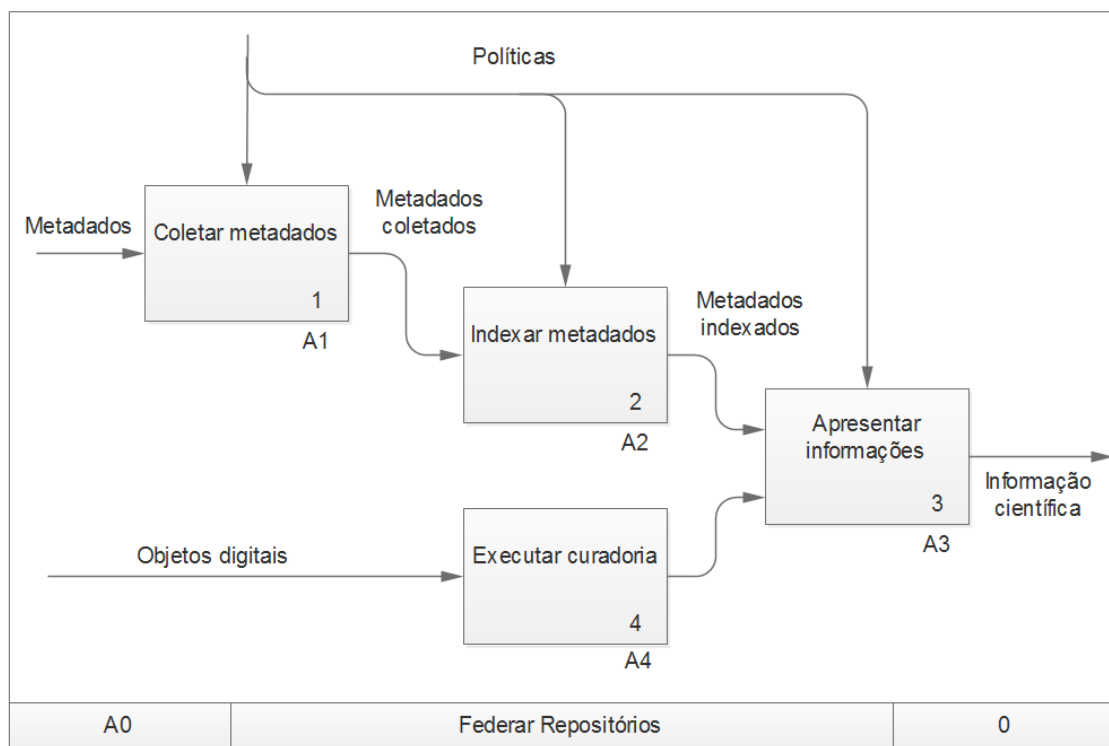
Normalmente as federações coletam apenas os metadados dos repositórios, sendo que os arquivos digitais ficam no repositório. Por isso a questão da curadoria ser mais apropriada a esses sistemas. Entretanto, A federação pode fazer uma cópia dos arquivos, como forma de preservação, como se fosse cópias de segurança, automaticamente por meio do protocolo OAI-PMH. Assim, em caso de sinistro os arquivos ficam preservados na federação.

Essa opção funciona como backup em nuvem para os repositórios, em que as cópias dos arquivos são armazenadas na federação, para que em caso de indisponibilidade do repositório, o documento possa ser acessado diretamente da federação. Esse ponto se aproxima à técnica de que múltiplas cópias distribuídas aumentam a possibilidade de preservação digital.

Essa opção, no entanto, requer alterações na federação para que adote medidas para preservação dos arquivos coletados, ou seja, adote medidas de curadoria, já que passa a manter arquivos (figura 49). Nota-se, no entanto que as ações de curadoria necessitam ser adaptado à federação ao passo que opera com documentos finais, podendo ser dados de pesquisa ou resultados finais de pesquisa, mas documentos finalizados.



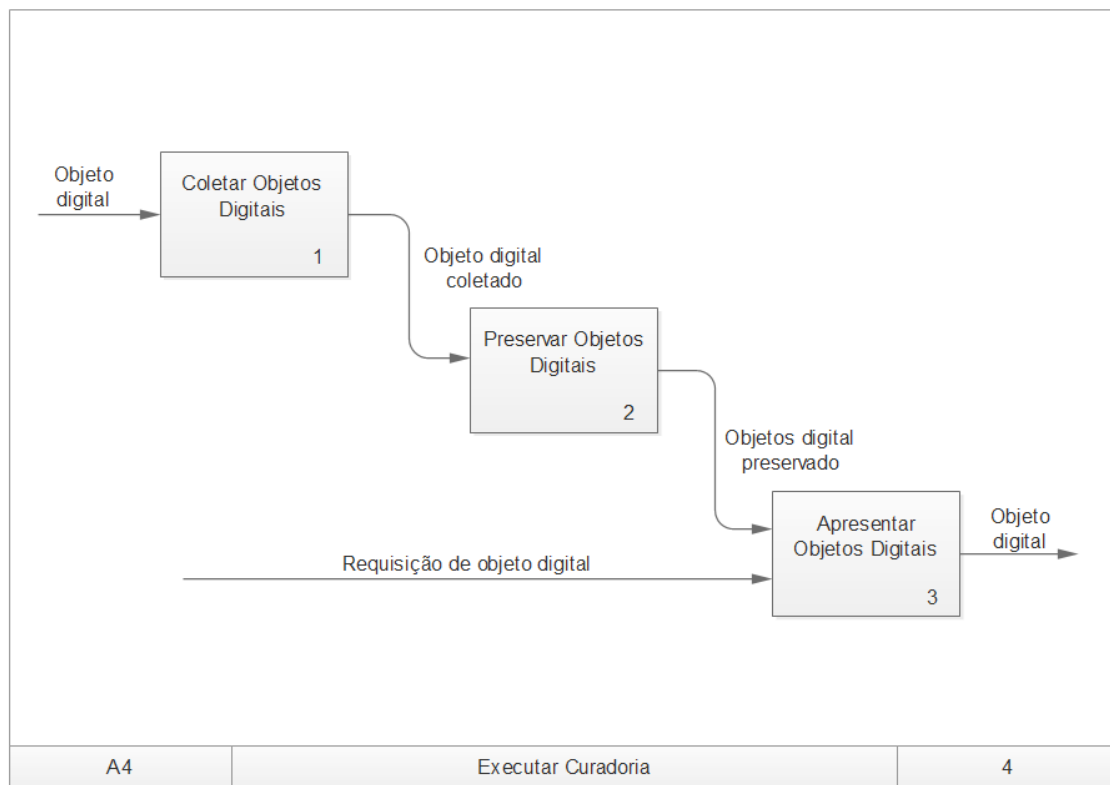
FIGURA 49 - FEDERAR REPOSITÓRIO COM CURADORIA DIGITAL



Fonte: o autor, 2015.

O processo de curadoria é adicionado ao processo de Federar Repositórios, ofertando serviços aos objetos digitais mantidos pela federação. Diferentemente do ciclo de vida natural geridos pela curadoria, para a federação é composta pela coleta dos objetos digitais, preservação e apresentação, como visto na figura 50. Assim os objetos digitais são coletados via protocolo OAI-ORE, juntamente com os metadados, depois são executados os procedimentos para preservação e por último são apresentados quando for o caso.

FIGURA 50 - EXECUTAR CURADORIA



Fonte: o autor; 2015.

Portanto a adoção de ações de curadoria altera os processos da federação de forma significativa, ao passo que a federação passa a ter um acervo composto de objetos digitais e não apenas os metadados. Com isso, requer maior cuidado com o acervo, pois assume a questão da curadoria desses objetos, compartilhando responsabilidades com os repositórios, de forma a suprir o acesso em caso de falha dos repositórios.

### 5.3.3.5 E-SCIENCE

O e-science, ou cyberscience como também é conhecido, tem sido visto por alguns estudiosos como um novo paradigma da ciência, fortemente apoiado na colaboração e uso de tecnologia para processar e analisar grandes quantidades de dados. Essa prática científica substitui a empiricidade pela simulação, embasadas por modelos em espaços virtuais, podendo ser distribuído e com participantes alocados em várias partes do mundo.

Tudo isso é possível pelos avanços tecnológicos e teóricos, em que modelos são criados para simular uma realidade. Em contrapartida com as ciências naturais em que se descrevem os fenômenos naturais por meio de experimentos práticos, a computação possibilita a ciência artificial, em que modelos são utilizados para simular um fenômeno, com a utilização de infraestrutura de grande capacidade de processamento e grande quantidade de dados.

Nesse contexto, as federações se apresentam como ferramentas que ofertam informações e dados para processamento, ao passo que os resultados das pesquisas são publicados em periódicos ou disponibilizados em repositórios. Assim, as federações não se apresentam com um papel ativo nesse novo paradigma, apenas como facilitador na recuperação das informações e dados.

Atualmente a discussão em relação ao e-science e as ferramentas informatizadas estão restritas à:

- ❖ Onde armazenar os dados e informações;
- ❖ Quais as ferramentas utilizar para processar os dados;
- ❖ Como visualizar os dados de forma mais eficaz;
- ❖ Quais as ferramentas mais eficazes para manter a colaboração entre pesquisadores;

Nesse contexto, as federações de repositórios se apresentam como fontes de dados e informações, agregando repositórios de dados e repositórios institucionais. Cabe salientar que muitas pesquisas podem ter características de pesquisas documentais, o qual os repositórios podem ajudar na localização dos documentos utilizados no processamento.

Essas questões se alinham ao papel das federações na Ciência Aberto e Big Data, destacando a contemporaneidade das federações. Da mesma forma que apresenta o desafio de ajuste ao novo cenário para as Ciências Sociais e Humanas, na utilização desse novo paradigma como apoio aos seus estudos. Possivelmente, no futuro com a adesão dessas ciências, a federação pode ter uma participação maior nesse cenário.

## 6 CONSIDERAÇÕES FINAIS

---

A criação de federações de repositórios científicos revelou-se uma tendência, visto indícios levantados no presente estudo, no que concerne a data de início de funcionamento, revelando que vários países têm criado federações de abrangência nacional mais recentemente (BDCOL, ALICIA, REMERI e SNRD). Mesmo que a NDLTD e BDTD estejam mais estabelecidas e com mais de dez anos, nota-se que a integração de repositórios ainda é, até certo ponto, uma questão atual, requerendo estudos que apoie a sua evolução.

Nesse sentido, tem-se federações de abrangência mundial (NDLTD), regional (LaReferencia, oasisbr e RCAAP), nacional (ALICIA, BDTD, BDCOL, SNRD e REMERI) ou mesmo mais restrita como a das universidades paulistas (CRUESP), cada qual com as suas características. Independente da abrangência, todas tem por finalidade integrar repositórios científicos por meio do uso dos preceitos da interoperabilidade.

Um ponto que destaca nas federações é a questão da sua gestão, que tem relação com a sustentabilidade do sistema. Grande parte das federações é mantida por instituições públicas (ALICIA, BDCOL, SNRD, BDTD, oasisbr, RCAAP), mas há federações mantidas por universidades (CRUESP) ou tem instituições próprias (NDLTD, LaReferencia e REMERI). Essa questão revela a preocupação com a sustentabilidade de sistemas acadêmicos de acesso livre.

Nesse sentido, o apoio governamental dado para o SNRD (pelo governo argentino) e o ALICIA (pelo governo peruano) apresenta um passo maior, visto que formaliza e legitima ações voltadas à disseminação de documentação científica em acesso aberto. Modelos mais independentes como o NDLTD, LaReferencia e REMERI, apresentam certa inovação, com a distribuição das responsabilidades de manutenção da federação. Assim, apresentam modelos de gestões diferenciados, revelando flexibilidade e diversidade.

Nas federações, mesmo com o viés da disseminação da informação científica, ressalta a grande variação documental que constituem o acervo. Há federações mais seletivas e outras que abrangem uma tipologia maior de documentação. Esse ponto põe

em discussão a própria questão atual do que são realmente documentos científicos. Uma questão mais contextualizada com as tecnologias atuais de disseminação da informação científica.

## **6.1 MODELOS DE FEDERAÇÕES**

As federações se apresentaram altamente colaborativas, visto que a BDTD teve influências e apoio da NDLTD e o LaReferencia fomentou a criação da BDCOL, SNRD e ALICIA, ofertando orientações. O oasisbr e o RCAAP possuem relação de compartilhamento de acervo, numa relação bem mais interativa. Revela um ambiente, até certo ponto, amistoso, pois possuem a mesma finalidade de facilitar o acesso à documentação científica.

A Identificação e análise das federações e repositórios, que embasou a proposição do modelo, possibilitou levantar informações que apoiaram a verificação das hipóteses propostas e atender aos objetivos. Com isso, contribuir com novos conhecimentos e validar outros existentes, mas não formalizados.

### **6.1.1 CONSIDERAÇÕES SOBRE AS HIPÓTESES**

O estudo propôs algumas hipóteses amparadas pelos objetivos, de forma a avaliar e validar proposições sobre as federações, muitas delas baseadas nos repositórios existentes. Assim, o presente estudo constatou:

- I. Considerando o quantitativo de repositórios científicos brasileiros é possível a construção de um modelo de federação que atenda a várias finalidades:
  - a. O estudo levantou numa amostra pequena, mas significativa, repositórios com características diferentes, algumas com acervo mais restrito e outros mais amplos. Com isso, podem-se ter federações com acervo variado, composto de metadados referentes a uma grande variedade de documentos científicos, possibilitando ofertar vários serviços, para várias finalidades.

Assim, o modelo propõe ser útil a uma variedade de usuários e finalidades;

- b. A aplicação das tendências tecnológicas e das ciências, também, revelou a possibilidade de um modelo que atenda a finalidade principal de disseminação da informação científica, alinhando a outros serviços como informação estratégica.
- II. Considerando o quantitativo de tipos de repositórios existente no país é fortemente viável que o modelo possa ser utilizado para criação de federações de repositórios não científicos
- a. O estudo não levantou os repositórios não científicos, entretanto, desde que possam aderir à interoperabilidade por meio do uso do protocolo OAI-PMH ou OAI-ORE, o modelo proposto pode ser adaptado para federações que integram repositórios não acadêmicos;
  - b. O modelo é adequado para gerar informações estratégicas para tomada de decisão, com base nos registros dos documentos mantidos na federação, característica muito apropriada a federações não acadêmicas.
- III. Com base nos resultados da pesquisa, o modelo apresentado pode amparar o desenvolvimento de sistemas sustentáveis e estáveis, se utilizado para implementação de ferramentas informatizadas para a criação de federação de repositórios:
- a. Estar em consonância com as tendências, principalmente às relacionadas à tecnologia, garante até certo ponto estabilidade tecnológica no sentido em que oferta serviços mais condizentes com outros portais, tornando a federação mais atualizada tecnologicamente;
  - b. No que diz respeito às tendências da ciência, o modelo propôs a adoção de tendências que atenderiam a principal comunidade usuária das federações. Assim, tornando-a mais estável e sustentável, visto que a sustentabilidade depende diretamente da federação atender aos seus propósitos;

- c. Da mesma forma, o estudo levantou as formas de gestão das federações, revelando como as federações se mantêm. Da gestão governamental, com ou sem apoio legal, até a formação de instituições para a manutenção da federação, com a divisão das questões financeiras entre os parceiros.
  - d. Destaca-se, no entanto, que a oferta de serviços e a evolução da federação são pontos cruciais para a sustentabilidade e estabilidade. Todo sistema da internet necessita de usuários para a sua sobrevivência e manter-se atualizado, tanto no conteúdo, quanto nos serviços, revela-se um atrativo para os usuários.
- IV. Considerando a federação de repositórios como um sistema que organiza um cenário heterogêneo, é possível a aplicação de formas diferenciadas de representação, que contextualize a informação apresentada:
- a. O modelo proposto apresentou que o ambiente virtual é possível criar organizações artificiais, que facilitem a recuperação. Facetamentos, por exemplo, facilitam a contextualização e refinamento;
  - b. Da mesma forma, formas mais atuais de apresentar a informação contextualiza melhor. Sistemas de descoberta e entregam atualmente utilizadas em federações, torna essa hipótese viável.
- V. Considerando que o estudo inclui várias linhas de pesquisa da Ciência da Informação, assim como disciplinas correlatas, pode-se presumir que a pesquisa contribuirá para a Ciência da Informação, em suas questões epistemológicas no que diz respeito às relações entre tópicos de estudo e interdisciplinaridade:
- a. O estudo revelou a relação intensa entre as tecnologias e a ciência, principalmente nas tendências da ciência, destacando a questão o uso da tecnologia nos processos científicos, estritamente relacionado aos interesses da Ciência da Informação, principalmente na Comunicação Científica;
  - b. Da mesma forma, ao relacionas as tendências tecnológicas com as federações revelou-se a proximidade com as questões mais

organizacionais e gerenciais alinhadas a questão da informação estratégica, relacionado a administração e disciplinas correlatas.

### **6.1.2 CONTRIBUIÇÕES DA TESE**

Federações são sistemas atuais e mesmo que algumas possuam mais de dez anos, revelam-se ainda pouco estudadas e formalizadas. Muitos dos estudos levantados descreviam a iniciativa, com pouca discussão teórica sobre o assunto. A falta de análise e contrastividade dos estudos restringem a discussões, na medida em que focam em uma única iniciativa.

Assim, as contribuições do presente estudo residem na identificação e análise envolvendo as federações existentes, que embasaram a proposição de um modelo mais contextualizado com as tendências tecnológicas e da ciência. Com aspecto mais qualitativo no levantamento de questões que envolvem a federação, identifica e analisa conceitos, estrutura e outros elementos, formalizando aspectos relevantes na forma de propostas.

#### **6.1.2.1 CONCEITUAÇÃO SOBRE FEDERAÇÕES**

O estudo revelou que mesmo tendo as federações ter mais de dez anos de existência, muitos conceitos ainda não estão totalmente estabelecidos, não se apresentando como um caso isolando na internet. A evolução tecnológica e conceitual tem imposto certa necessidade de adaptação, requerendo estudos contínuos para atender as mudanças.

Identificação e análise apresentadas na seção 5.1, revelaram as questões conceituais, por meio dos estudos qualitativos das características, problemas, tendências e outros das federações selecionadas como amostra. Nesse sentido podem-se conceituar federações de repositórios científicos como:

*"Esforço internacional, regional ou unificação de iniciativas isoladas, formando uma rede de repositórios ou bibliotecas digitais colaborativa, utilizando um sistema para coletar os metadados, para ofertar um ponto único de acesso à documentação, com a utilização de uma ferramenta de busca."*



Da mesma forma que revelou certa mudança, no sentido que federações mais novas têm se tornando mais flexíveis no que diz respeito à documentação coletada, com políticas mais abrangentes. Objetos educacionais livres tem se tornado comuns em alguns repositórios e esse fenômeno tem influenciado as federações, ao passo que essa documentação não é tradicionalmente científica.

Nesse sentido o estudo revelou que as federações estão se tornando mais flexíveis não apenas no tipo de documentação coletada, mas também na interoperabilidade e modelos de sustentabilidade. A tendência em torna-se mais flexível reflete certa tendência na própria internet, que não impõe formatos muito rígidos para integração.

#### **6.1.2.2 CENÁRIO DOS REPOSITÓRIOS COM O FOCO NAS FEDERAÇÕES**

As federações são totalmente dependentes de fontes abertas. No caso de presente estudo são restritas aos repositórios científicos. Por isso, levantar o cenário dos repositórios se torna essencial para determinar um modelo de federação, apresentando as características, tendências e restrições que influenciam diretamente nas federações.

Nesse sentido, o presente estudo revelou pontos interessantes sobre os repositórios, que podem ter reflexos em federações. A maior flexibilidade em relação à tipologia de documentos impacta na coleta, pois oferta metadados de vários tipos de documentos, inclusive os não avaliados pelos pares. Para federações mais restritas em relação à tipologia de documentos será preciso descartar registros. Outras federações, no entanto, apresentará um acervo mais diversificado, com maior opção de registros.

Outro ponto a destacar como contribuição refere-se ao levantamento sobre a adoção de tendências tecnológicas pelos repositórios. Revelou certa adoção às tecnologias por parte dos repositórios. Entretanto, revelou também certo desconhecimento dos gestores por essas tendências e suas possibilidades. Até certo ponto, indica que os diversos tipos de usuários ainda não vislumbraram tudo que um repositório pode ofertar, além de disseminar informações.

Nesse sentido, revela que alguns gestores já utilizam o repositório como fonte para gerar informações estratégicas para tomada de decisão, mesmo que restritamente

na própria gestão do repositório. Possivelmente as instituições ainda não observaram todo o potencial do repositório como representação da produção intelectual da instituição e, por isso, fonte de meta-informações sobre essa produção.

Tendências da ciência, por sua vez, mostraram-se menos conhecidas pelos gestores, revelando certa oportunidade de estudos e divulgação. Apresenta certo distanciamento entre os repositórios e as tendências mais novas como o e-science e a curadoria digital, mesmo que os repositórios possam ser ferramentas úteis.

### **6.1.2.3 PROPOSTA DE POLÍTICAS**

As políticas são orientações que apoiam a criação, desenvolvimento, manutenção e gerenciamento das federações. Mesmo que não seja explicitamente formalizado ou divulgado, todos os sistemas acadêmicos possuem políticas que apoiam o seu funcionamento. Nas federações estudadas, essa questão revelou-se uma questão mais crítica, ao passo que, nem todas explicitam todas as suas políticas.

Entretanto, na identificação e análise dos processos da federação possibilitou levantar as principais políticas, relacionadas aos seus principais processos, baseados nas orientações necessárias para a sua operacionalização. Assim, de forma resumida (a forma completa é apresentada na seção 5.3.1.1) as políticas das federações são apresentada no quadro 11, sendo umas das contribuições do presente estudo.

QUADRO 11 - SUMÁRIO DAS POLÍTICAS DE FEDERAÇÃO

Macro processo	Política	Resumo da descrição
Coletar Metadados	Tipo de repositórios coletados	Critérios para seleção dos repositórios a ser coletados;
	Esquemas de metadados	Lista dos esquemas de metadados coletáveis pela federação;
	Formas de adesão à federação	Fluxo de tarefas e formas de um repositório aderir à federação;
	Temporalidade da coleta	Cronograma de coleta da federação, mesmo que seja individualizada para cada repositório;
Indexar Metadados	Campos de metadados indexados	Lista dos campos de metadados indexados para gerar serviços;
	descarte	Critérios e formas de descarte de registros;
	vocabulários controlados	Indicação dos vocabulários controlados utilizados como apoio à indexação
	idiomas	Indicação dos idiomas que possuem facilidades de indexação
Apresentar Metadados	Formato de apresentação dos resultados	Indicação das formas de apresentação dos registros do acervo da federação;
	campos de metadados	Lista e formas de apresentação dos metadados
	refinamento e sugestões	Formas de refinamento das buscas e interação entre usuário e federação
	Opções de busca	Orientação para a ferramenta de busca
	formas de navegação pelo acervo	Orientações para arquitetura da informação

Fonte: o autor; 2015.

#### 6.1.2.4 MODELO DE FEDERAÇÃO

O modelo apresentado na seção 5.3.1 tem por base o levantamento dos processos das federações, apresentando uma compilação, no sentido em que foram analisadas várias federações. Da mesma forma propõe alguns processos para ofertar novas facilidades. Nesse ponto, a maior contribuição do presente estudo é a formalização e registro dos processos da federação, que são apresentados na seguinte forma:

##### ❖ A- 0 Diagrama Inicial

- **A0 - Federar Repositórios:** Processo geral em que se federaliza repositórios orientados por políticas, em forma de rede, mantendo a independência de cada um dos membros.
  - **A1 - Coletar Metadados:** Processo automático de se obter os metadados dos repositórios, de forma a constituir o acervo da federação.

- **A11 - Gerenciar Repositórios:** Processo de gerenciamento dos repositórios constituintes da federação.
  - **A111 - Prospectar Repositório:** Processo de prospectar por novos repositórios para fazer parte da federação, baseados nos critérios definidos nas políticas da federação.
  - **A112 - Atender Repositório:** Processo de atender os repositórios que entram em contato para participar da federação.
  - **A113 Catalogar Repositório:** Processo de registrar os repositórios, de forma a se tornarem membros da federação.
- **A12 - Executar Coleta:** Processo de se executar a coleta, de forma automatizada, alinhado com as políticas e características de cada repositório.
  - **A121 - Configurar Coleta:** Processo pelo qual se ajusta a coleta as políticas e características do repositório.
  - **A122 - Coletar:** Processo automático ou manual para execução da coleta de metadados.
  - **A123 - Verificar Coleta:** Processo de verificação da coleta de metadados, de forma a detectar erros ocorridos no processo coletador.
  - **A124 - Reportar Error:** Processo pelo qual a equipe da federação entra em contato com os repositórios para relatar os erros ocorridos no processo de coleta de metadados para correção.

- **A13 - Mapear Metadados:** Processo de ajuste nos metadados coletados.
  - **A131 - Desenvolver Mapeador:** Processo pelo qual se cria os mapeadores, baseados na análise de erros e conteúdos dos metadados.
  - **A132 - Aplicar Mapeador:** Processo pelo qual se aplica os mapeados nos metadados.
- **A2 - Indexar Metadados:** Processo pelo qual cria os índices para apoiar a recuperação das informações na federação.
  - **A21 - Processar Metadado:** Processo pelo qual se processa os metadados os preparado para a criação dos índices.
  - **A22 - Criar Índice:** Processo automático para a criação dos índices no sentido da Ciência da Computação.
- **A3 - Apresentar informações:** Processo executado nos portais, em que os dados são apresentados.
  - **A31 - Processar Busca:** Processo pelo qual as informações são recuperadas por meio de ferramentas de busca.
  - **A32 - Processar Navegação:** Processo pelo qual as informações são recuperadas por meio de geração automática de listas que possibilitam a navegação.
  - **A33 - Apresentar Metadados:** Processo pelo qual a informação é apresentada, podendo fer efetuda em vários formatos.

Nesse contexto, a utilização da ferramenta de registro IDEF0 para descrever os processos da federação revelou-se eficaz, na medida em que possibilitou refinar os processos a níveis mais atômicos. Com isso, revelar peculiaridades que ficam

despercebidas em processamentos automáticos, mas possuidores de importância na operacionalização da federação. Assim, do maior processo da federação "Federar Repositórios" aos mais refinados pode-se apresentar como uma federação opera, como as políticas e outras forças atuam e, principalmente, apresentando o fluxo de dados.

Ressalta-se a finalidade das federações no modelo, altamente embasado em facilitar a disseminação, com o objetivo de aumentar a visibilidade do documento, autor e instituição. Deve-se destacar, no entanto, que as federações possuem papel menor na cadeia da comunicação científica, visto que não possui finalidades de preservação, armazenamento ou publicação, papéis atribuídos aos repositórios. Todos os processos apresentados nos modelos direcionam-se à ofertar serviços que atendam a essa finalidade.

Pode-se constatar que a principal característica das federações é a contribuição, o compartilhamento de informações, em um ambiente altamente colaborativo. O modelo apresentado aporta essa característica, na medida em que adota várias das características observadas nas federações existentes. Da mesma forma que se baseia na interoperabilidade para a construção do acervo, apresentando os critérios para adesão de novos repositórios.

O modelo também contribuiu com uma proposta de política, apresentada na forma completa em 5.3.1.1 e mais resumidamente em 6.1.2.3, sendo uma das maiores contribuições do presente estudo, visto que se torna base para o desenvolvimento de federações. As políticas englobam todos os processos descritos no modelo, orientando cada etapa.

O registro dos processos da federação pode servir de apoio a outras pesquisas ou para propostas de atualizações, ao passo que os descreve com refinamento. Com isso, tem-se a descrição do processo de federalização de repositórios científicos, que podem ser adaptado para outras finalidades e outros tipos de repositórios, apoiando a disseminação das federações.

#### **6.1.2.5 FEDERAÇÕES E AS TENDÊNCIAS TECNOLÓGICAS E DA CIÊNCIA**

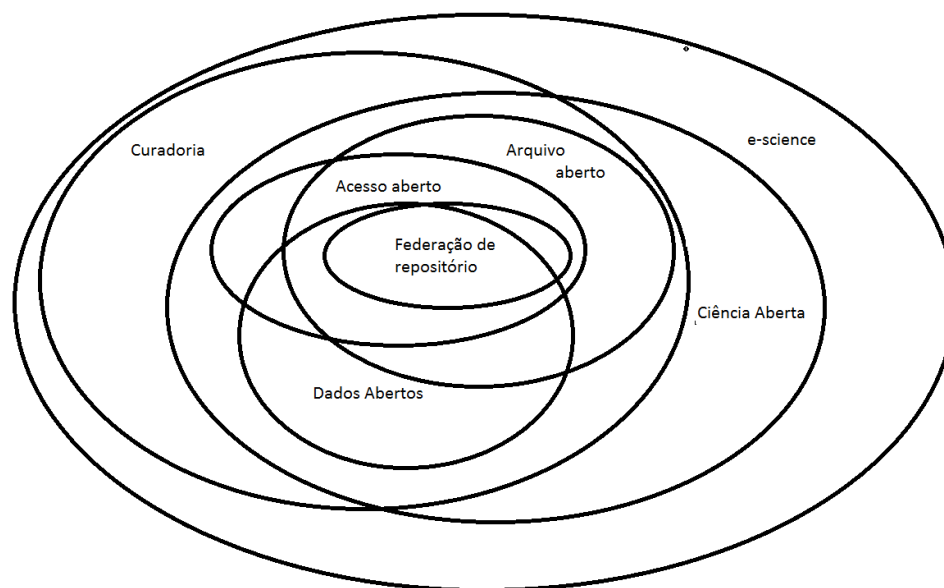
A adaptação do modelo de federação para atender às tendências tecnológicas propõe novos serviços ou nova infraestrutura, de forma a atender as necessidades futuras dos usuários. Com isso, atua na sustentabilidade tecnologia da federação, no sentido que a federação como um serviço disponível na internet requer estar atualizado com as suas tendências.

Mesmo que as tendências tecnológicas apresentadas como “nexus das forças” estejam mais alinhadas com serviços voltados aos negócios, o estudo revelou a compatibilidade das federações com essas tendências. Assim, a contribuição do estudo nesse ponto ampara na questão relacionada à utilização de orientações para melhoria de sistemas voltada a negócios para sistemas acadêmicos. Com isso, revela não haver disparidades entre essas duas orientações, no que diz respeito às federações. Assim, apresenta as oportunidades e desafios para que federações existentes possam aderir às novas tecnologias, de forma a ofertar outros serviços, além do apoio à disseminação da informação científica.

As federações, também, se revelaram como uma ferramenta apropriada para apoiar as novas tendências da ciência, ao passo que podem facilitar o acesso à informação. Entretanto, revelou-se da mesma forma que os repositórios possuem um papel mais importante nesse novo cenário. Assim, as federações se apresentam como uma ferramenta coadjuvante perante as novas tendências da ciência.

Com isso, ratificou-se o diagrama de relação entre a federação de repositórios e as tendências da ciência, em que a federação se torna apenas uma das ferramentas disponíveis aos novos preceitos (figura 51). Revela-se que as novas tendências englobam conceitos e ferramentas de outras tendências como em um processo evolutivo, apresentando novas formas de fazer ciência.

FIGURA 51 – RELAÇÃO ENTRE AS FEDERAÇÕES E AS TENDÊNCIAS DA CIÊNCIA



Fonte: o autor; 2014.

Ressaltou-se que o e-science é mais abrangente, enlobando conceitos todas as outras tendências. Da mesma forma que a federação de repositórios é uma ferramenta que adere a todas as tendências, entre tantas outras que podem ser aplicadas as necessidades específicas de cada tendência.

## 6.2 ESTUDOS FUTUROS

A federação de repositórios institucionais se apresenta como um tema atual de estudo, tendo sido pouco explorado. Em parte, possui um aspecto mais aplicado, mas revela um desafio, visto a contemporaneidade do tema, suas relações interdisciplinares, principalmente com a Ciência da Computação. Assim, apresenta desafios e oportunidades de estudos futuros, revelando um tema fértil a pesquisa.

As federações estudadas baseiam-se na interoperabilidade proposta pelos preceitos dos arquivos abertos, por meio do uso do protocolo OAI-PMH ou OAI-ORE. Esses dois protocolos operam em sistema assíncrono, ou seja, coletam automaticamente os metadados dos repositórios em datas determinadas. Essa questão cria uma defasagem



entre os repositórios e as federações, principalmente para os repositórios muito produtivos e dinâmicos. Assim, requer-se estudos para implementação de sistema sincrônicos de coletas de metadados, baseados em outras tecnologias.

A indexação de registros multitemáticos é outro ponto de estudo. As federações armazenam grandes quantidades de metadados de muitas disciplinas, proveniente de repositórios, gerando problemas para o processo de indexação mais semântico. Assim, estudos voltados à indexação automática com utilização de ontologias torna-se um desafio, principalmente pela extensão de temas mantidos na federação, pois necessitaria de utilização de múltiplas ontologias, de forma a desambiguar dos termos, que muitas vezes possuem significados diferentes em disciplinas distintas.

Por fim, estudos mais profundos sobre a sustentabilidade de iniciativas de acesso aberto. Atualmente, novos modelos de publicação e disseminação da informação estão sendo estudados, testados, validados de forma a criar formas de sustentabilidade dessas iniciativas, visto a necessidade de continuidade de oferta do serviço, mas sem a cobrança para o acesso. As federações dependem da sobrevivência dos repositórios para, pois até certo ponto, podem ser consideradas como fontes terciárias, no caso do artigo científico, por exemplo, a fonte primária é a revista, a secundária o repositório. Assim, estudos de sustentabilidade de toda essa cadeia tornam-se essencial

## REFERÊNCIAS

---

AGUILLO, I. F. , et all Indicators for a webometric ranking of open access repositories. *Scientometrics*, v. 82. n 3, p. 477-486, 2010.

ASSIS, T. B. de. Análise das políticas de autoarquivamento nos repositórios institucionais brasileiros e portugueses. *InCID: Revista de Ciência da Informação e Documentação, Brasil*, v. 4, n. 2, p. 212-227, dez. 2013. ISSN 2178-2075. Disponível em: <<http://www.revistas.usp.br/incid/article/view/69329>>. Acesso em: 04 Jan. 2015. doi:<http://dx.doi.org/10.11606/issn.2178-2075.v4i2p212-227>

BALARINE, O. F. O. Tecnologia da informação como vantagem competitiva. *RAE-eletrônica*, Volume 1, Número 1, jan-jun/2002

BAUN, S. *Cloud Computing: web-based dynamic IT services*, Springer, Berlin, 2011

BAWDEN, D.; ROBINSON, L. The dark side of information: overload, anxiety and other paradoxes and pathologies. *Journal of Information Science*, v. 35, n. 2, 2009.

BECKER, S. *Developing Quality Complex Database Systems: Practices, Techniques and Technologies*. Idea Group Inc (IGI), Hershey PA.2000.

BENANTAR, M. *Access Control Systems: Security, Identity Management and Trust Models*. Springer, 2006

BERNERS-LEE, T. Creating a Science of the Web, *Science*, 313, (5788), 769-771. 2006.

BERNERS-LEE, T. Linked data: design issues. 2006.

BEZERRA, B. Gêneros digitais: Apresentando livros na Internet. *Revista Signos*, Número Especial Monográfico N° 1, PP. 45-61, 2010.

BICKERTON, P.; BICKERTON, M. & SIMPSON-HOLLEY, K. *Cyberstrategy*. Oxford, Butterworth-Heinemann, 1999.

BIZER, C.; HEATH, T.; BERNERS-LEE, T. Linked data: the story so far. *International Journal on Semantic Web and Information Systems*, v.5, n.3, p.1-22, 2009.

BJÖRK, B. C. Scientific communication life-cycle model. 2005. disponível em: <<http://oacs.shh.fi/publications/Model35explanation2.pdf>>. Acesso em: Dez. 2007.

BONGIOVANI, P. C.; NAKANO, S. Acceso Abierto en Argentina: La experiencia de articulación y coordinación institucional de los repositorios digitales en ciencia y tecnología. Volumen 1 N. 2 Julio – Diciembre 2011 ISSN: 2027 - 7415 e-colabora Pág. 163.

BORGES, M. E. N. A informação como recurso gerencial das organizações na sociedade do conhecimento. *Ciência da Informação* - V. 24, n. 2, 1995

BORGMAN, C. L. *Scholarship in the digital age: Information, Infrastructure, and the Internet*. Cambridge: MIT Press, 2007.

BRETERNITZ, V. J.; SILVA, L. A. Big data: um novo conceito gerando oportunidades e desafios. *Revista RETC – Edição 13<sup>a</sup>*, 2013.

BUSH, V. As we may think. *The Atlantic Monthly*, Boston, v. 176, n. 1, p. 101-108, July 1945

BUYYA, R.; BROBERG, J.; GOSCINSKI, A. M. *Cloud Computing: Principles and Paradigms*. John Wiley & Sons, Nova Iorque, 2010

CAFÉ, L. et al. Repositórios institucionais: nova estratégia para publicação científica na Rede. In: CONGRESSO BRASILEIRO DE CIÊNCIAS DA COMUNICAÇÃO, 26; Belo Horizonte: INTERCOM – Sociedade Brasileira de Estudos Interdisciplinares da Comunicação 2003 e ENDOCOM, 13; Belo Horizonte, 2003.

CAMARGO, L. S. de A. ; VIDOTTI, S. A. B. G. . Arquitetura da informação para repositórios digitais. In: Luis Sayão; Lídia Brandão Toutain; Flavia Garcia Rosa; Carlos Henrique Marcondes. (Org.). *Implantação e gestão de repositórios institucionais : políticas, memória, livre acesso e preservação*. Salvador: EDUFBA, 2010, v. 1, p. 55-82

CAPLAN, P. *Metadata Fundamentals for All Librarians*. American Library Association, Chicago. 2003

CARVALHO, J.; MOREIRA, J.M.; RODRIGUES, E.; SARAIVA, R.. Repositórios institucionais Democratizando o acesso ao conhecimento IN: Flavia Rosa EDUFBA

CARVALHO, K. Disseminação da informação e informação de inteligência organizacional. DataGramaZero - Revista de Ciência da Informação - v.2 n.3 jun/01.

CARVALHO, M. C. R.; SILVA, C. H.; GUIMARÃES, M. C. S.. Repositório institucional da saúde: a experiência da Fundação Oswaldo Cruz. Informação & Sociedade: Estudos, João Pessoa, v.22, n.1, p.97-103, jan./abr. 2012.

CASTELLS, M. A sociedade em rede, Editora Paz e Terra, São Paulo, 2002.

CASTRO, R. L.; PELUFÊ, M. S.; ARRUDA, R. G. Repositórios digitais da Embrapa: acesso livre a produção técnico-científica. In: XXV Congresso Brasileiro de Biblioteconomia, Documentação e Ciência da Informação, ANAIS, Florianópolis, SC, 2013.

CHIAVENATO, I. Administração Nos Novos Tempos. Elsevier Brasil, 2005.

COHEN, W. A. *Uma aula com Drucker: As lições do maior mestre de administração*. Trad. Afonso Celso da Cunha Serra. Rio de Janeiro: Campos/Elsevier, 2008

COMBER, T. Building usable web pages: an HCI perspective. In: A Ellis & R Debreceny (eds), AusWeb95, innovation and diversity : the World Wide Web in Australia, proceedings of AusWeb95, the first Australian World Wide Web Conference, Ballina, NSW, Norsesearch Ltd, Lismore, NSW, pp. 119-124.

COSTA, S. M. S. Abordagens, estratégias e ferramentas para o acesso aberto via periódicos e repositórios institucionais em instituições acadêmicas brasileiras. Liinc em Revista, 2008, v. 4, n. 2 disponível em <<http://revista.ibict.br/liinc/index.php/liinc/article/view/281>> Acesso em: abr/2009.

COSTA, Sely M. S. ; LEITE, Fernando César Lima. Repositórios institucionais: potencial para maximizar o acesso e o impacto da pesquisa em universidades. In: CONFERÊNCIA IBEROAMERICANA DE PUBLICAÇÕES ELETRÔNICAS NO CONTEXTO DA COMUNICAÇÃO CIENTÍFICA, 1., 2006, Brasília. Anais... Brasília: Universidade de Brasília, 2006.

CRESWELL, J. W. *Projeto de pesquisa: método qualitativo, quantitativo e misto*.

Tradução de Luciana de Oliveira da Rocha. 2. ed. Porto Alegre: Artmed, 2007.

CROW, R. The Case for Institutional Repositories: A SPARC Position Paper. Washington DC: Scholarly Publishing & Academic Resources Coalition (SPARC). 2002.

CURRY, E. FREITAS, A. O'RIÁIN, S. The role of community-driven data curation for enterprises. In. WOOD, D. (Eds). Linking Enterprise Data. Springer Science & Business Media, Boston. 2010.

D. Nurmi et al., The Eucalyptus Open-source Cloud Computing System, 9th IEEE/ACM International Symposium on Cluster Computing and the Grid, 2009

DAUN, B. Modelagem de Objetos de Negócio com Xml. Elsevier, 2003.

DENZIN, N. K.; LINCOLN, Y. S. Planejamento da pesquisa qualitativa – teorias e abordagens. Porto Alegre: Artmed, 2006, p.367-388.

DEVEDŽIC, V. Semantic Web and Education. Berlin: Springer, 2006;

DI FOGGI, R. A.; FURNIVAL, A. C. M.. Mapeamento e Análise da Percepção das Mudanças Associadas ao Acesso Aberto à Literatura Científica com Bibliotecários e Profissionais de Informação de Universidades Públicas Federais e Estaduais do Brasil. InCID: Revista de Ciência da Informação e Documentação, Brasil, v. 4, n. 2, p. 75-94, dez. 2013. ISSN 2178-2075. Disponível em: <<http://www.revistas.usp.br/incid/article/view/69274>>. Acesso em: 04 Jan. 2015. doi:<http://dx.doi.org/10.11606/issn.2178-2075.v4i2p75-94>.

DODEBEI, Vera . Novos meios de memória: livros e leitura na época dos weblogs. Encontros Bibli (Online), v. 1, p. 129-143, 2009.

ESPIRITO SANTO, A. Delineamentos de Metodologia Científica. São Paulo: Edições Loyola, 1992. 176 p.

EVANS, D. Social media marketing: an hour a day. Indianápolis: John Wiley & Sons, 2012.

FIPS PUBS (Federal Information Processing Standards Publications - US), INTEGRATION DEFINITION FOR FUNCTION MODELING (IDEF0): Draft Federal Information Processing Standards Publication: National Institute of Standards and Technology 183, 1993. Patrocinado pela Secretary of Commerce pursuant to Section 111(d) of the Federal Property and Administrative Services Act of 1949 as amended by the Computer Security Act of 1987, Public Law 100-235.

FOX, E. et al. NDLTD: Encouraging International Collaboration in the Academy. DESIDOC Bulletin of Information Technology, Vol. 17, No. 6, November 1997, pp. 45-56.

FURHT, B. Cloud Computing Fundamentals. In: FURHT, B.; ESCALANTE, A. Handbook of Cloud Computing. Springer, Nova Iorque, 2010.

GARDNER, H.; MANDUCHI, G. Design Patterns for e-Science. Springer Science & Business Media, 03/05/2007

GARTNER. The nexus of forces: social, mobile, cloud and information. 2013. Special Report. Disponível em: <<http://www.gartner.com/technology/research/nexus-of-forces/>>. Acesso em: 28 nov. 2013.

GARVEY, D. W.; GRIFFITH, B. C. Communication and information processing withinscientific disciplines: empirical findings for ppsychology. In: \_\_\_\_\_Communication: TheEssence of Science. Oxford: Pergamon Press, 1979. p. 127-147.

GIARETTA, D.. Advanced Digital Preservation. New York: Springer, 2011.

GIL, A. C. Como elaborar um projeto de pesquisa. 4. ed. São Paulo: Atlas, 2006.

GILMAN, I. Library Scholarly Communication Programs: Legal and Ethical Considerations. Woodhead Pub. Oxford, U.K. 2013.

GILMAN, I. Library Scholarly Communication Programs: Legal and Ethical Considerations. Woodhead Pub. Oxford, U.K. 2013.

GILMAN, I. Library Scholarly Communication Programs: Legal and Ethical Considerations. Woodhead Pub. Oxford, U.K. 2013.

GRAY, J. e-Science: a transformed scientific method. Transcrição de palestra ministrada por Jim Gray no Conselho Nacional de Pesquisa (EUA), 11 Jan. 2007. In: HEY, T.; TANSLEY, S.; TOLLE, K (Ed.). The fourth paradigm: data-intensive scientific discovery. Redmond: Microsoft Research, 2009.

GUTIERREZ, C. Modeling the Web of Data (Introductory Overview). Reasoning Web 2011: 416-444

HALFORD,S.; POPE, C.; CARR, L. A Manifesto for Web Science?. In: Proceedings of the WebSci10: Extending the Frontiers of Society On-Line, April 26-27th, 2010, Raleigh, NC: US.

HARNAD, S. et al. The access/impact problem and the green and gold roads to open access. 2004. disponível em: <http://eprints.soton.ac.uk/265852/2/serev-revised.pdf>, acessado em: mar. 2013.

HENDLER , J. , SHADBOLT et all. ( 2008 ). Web science: An interdisciplinary approach to understanding the Web . Commun ACM , 51 ( 7 ), 60 – 69 . [CrossRef], [Web of Science ®]

HEY, T.; HEY, J. E-science and its implications for the library community. Library Hi Tech, v. 24, n. 4, p. 515-528, 2006.

HOPKINS, B. Top Technology Trends To Watch: 2014 To 2016. Forrester Research, Inc. 2013.

HOUCADE, V. Processos abertos na prática científica. Cienc. Cult., São Paulo, v. 65, n. 4, 2013 . Disponível em

HURD, J. The transformation of scientific communication: A model for 2020. Journal of theamerican society for information science, New York, v.51, n. 14, p. 1279-1283, 2000.

HUSSAIN, S. et al. <http://www.dlib.org/dlib/september01/suleman/09suleman-pt1.html>

IKEMATU, R. S. Gestão de Metadados: Sua Evolução na Tecnologia da Informação. DataGramZero - Revista de Ciência da Informação - v.2 n.6 dez/01

JAILET, A. How the introduction of new technology into higher education is revealing a need for change In: Kindt, M; Dittler, U.; Kahler, H. (Eds); E-Learning in Europe - Learning Europe. Waxmann Verlag. Munique, 2005.

- JANSA, K. Cloud Computing. Jones & Bartlett Publishers, Burlington 2013
- JEWEL, D. et al. Performance and Capacity Implications for Big Data. IBM Red Books, 2014.
- JOHNSON, S. Emergência: a dinâmica de redes em formigas, cérebros, cidades e softwares, Rio de Janeiro: Jorge Zahar, 2003.
- KAKUTA, S. M.; RIBEIRO, J. Livro dá dicas a empreendedor. Artigo Sebrae, p. 1. 2011.
- KIM, J. Faculty self-archiving behavior: Factors affecting the decision to self-archive. Proceedings of the American Society of Information Science and Technology, 44: 1-5. 2007. doi: 10.1002/meet.1450440353.
- KUHN, T. S. *A estrutura das revoluções científicas*. Tradução de Beatriz Viana Boeira e Nelson Boeira. 9. ed. São Paulo: Perspectiva, 2006.
- LAGOZE, C. et al. The Open Archives Initiative Protocol for Metadata Harvesting. <http://www.openarchives.org/OAI/openarchivesprotocol.html>. Acesso em 17 de fevereiro de 2003
- LARA, M. L. G. de; CONTI, V. L.. Disseminação da informação e usuários. São Paulo Perspec., São Paulo, v. 17, n. 3-4, Dec. 2003. Available from <[http://www.scielo.br/scielo.php?script=sci\\_arttext&pid=S0102-88392003000300004&lng=en&nrm=iso](http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0102-88392003000300004&lng=en&nrm=iso)>. access on 04 Jan. 2015. <http://dx.doi.org/10.1590/S0102-88392003000300004>.
- LEITE, F. C. L.; COSTA, S. M. S. Repositórios institucionais como ferramentas de gestão do conhecimento científico no ambiente acadêmico. Perspectivas em Ciência da Informação, v.11, n.2, p. 206-2119, maio/ago. 2006
- LIEVROUW, L. A. Communication, representation, and scientific knowledge: a conceptual framework and case study. Knowledge and Policy, New Brunswick, v. 5, n. 1, p. 6-28, spring 1992.
- LIMA, J. A. L.; LIMA, V. A. O. Do open source ao open education: novos conceitos para educomunicação. In Ecom (Anais), Campina Grande, 2013
- LINDE, P. et al, How can libraries and other academic institutes engage in making data open. in: P. Polydoratos, M. Dobrevá (eds) Let's Put Data to Use: Digital Scholarship for the Next Generation: Proceedings of the 18th International Conference on Electronic Publishing
- LINHARES, B. F.; ALVES, D. S. Metodologia de ensino em pesquisa social quantitativo. Pensamento Plural, n. 14 v. 7, Santa Maria, RS, 2014.
- LIU, B. Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data. Chicago: Springer, 2011.
- LOPES, A. V. S. As transformações do ensino universitário de graduação com a utilização da internet. Estudo do caso PUC-Rio. Dissertação de mestrado. FGV, 2002

LOPES, A. V. S. As transformações do ensino universitário de graduação com a utilização da internet: estudo do caso PUC-Rio. dissertação, FVG, 2002.

MACEDO, D. J. et al. Theses and dissertations digital library: Ten years of Open Access and Open Archives in Brazil. In: Anais do Open Repositories 2014, Helsink, 2014a.

MARCONDES, C. H.; SAYÃO, L. F. Integração e interoperabilidade no acesso a recursos informacionais eletrônicos em C&T: a proposta da Biblioteca Digital Brasileira. *Ciência da Informação*, v. 30, n. 3, p. 24-33, set./dez., 2001. Disponível em: <<http://revista.ibict.br/index.php/ciinf/article/viewFile/190/167>>. Acesso em: 01 mar. 2010.

MARCONDES, Carlos Henrique. “Linked data” – dados interligados - e interoperabilidade entre arquivos, bibliotecas e museus na web. *Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação*, [S.l.], v. 17, n. 34, p. 171-192, ago. 2012. ISSN 1518-2924. Disponível em: <<https://periodicos.ufsc.br/index.php/eb/article/view/1518-2924.2012v17n34p171/22782>>. Acesso em: 16 Fev. 2015. doi:<http://dx.doi.org/10.5007/1518-2924.2012v17n34p171>

MASIERO, Paulo Cesar et al . A Biblioteca Digital de Teses e Dissertações da Universidade de São Paulo. *Ci. Inf.*, Brasília , v. 30, n. 3, Dec. 2001 . Available from <[http://www.scielo.br/scielo.php?script=sci\\_arttext&pid=S0100-19652001000300005&lng=en&nrm=iso](http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0100-19652001000300005&lng=en&nrm=iso)>. access on 30 Dec. 2014. <http://dx.doi.org/10.1590/S0100-19652001000300005>.

MEADOWS, A. J. A comunicação científica. Tradução de Briquet de Lemos e A.A. Briquet de Lemos. Brasília: Livros, 1999.

MEDEIROS, J. S.; CAREGNATO, S. E. Compartilhamento de dados e e-Science: explorando um novo conceito para a comunicação científica. *Liinc em Revista*, v.8, n.2, setembro, 2012, Rio de Janeiro, p. 311-322 - <http://www.ibict.br/liinc>

MELL, P.; GRANCE, T. The NIST Definition of Cloud Computing: Recommendations of the National Institute of Standards and Technology. National Institute of Standards and Technology. Special Publication 800-145. Gaithersburg. 2011.

MIRANDA, Roberto Campos da Rocha. O uso da informação na formulação de ações estratégicas pelas empresas. *Ciência da Informação*, Brasília, v. 28, n. 3, p. 286-292, 1999.

MONK, S. Tecnologia da Informação para Gestão: Em Busca de um Melhor Desempenho Estratégico e Operacional (8ed). Bookman Editora. São Paulo, 2013.

MURAKAMI, Tiago Rodrigo Marçal; FAUSTO, Sibebe. Panorama atual dos Repositórios Institucionais das Instituições de Ensino Superior no Brasil. *INCID: Revista de Ciência da Informação e Documentação*, Brasil, v. 4, n. 2, p. 185-201, dez. 2013. ISSN 2178-2075. Disponível em: <<http://www.revistas.usp.br/incid/article/view/69327>>. Acesso em: 19 Out. 2014. doi:<http://dx.doi.org/10.11606/issn.2178-2075.v4i2p185-201>.



NAVEGA, S. Princípios essenciais do data mining. In: INFOIMAGEM. 2002. Anais... Cenadem, nov. 2002. Disponível em: <[www.intelliwise.com/snavega](http://www.intelliwise.com/snavega)>. Acesso em: 4 mai. 2014.

NELSON, R. R. The Market economy and the scientific commons. In: ANDERSEN, B. Intellectual Property Rights: Innovation, Governance and the Institutional Environment. Edward Elgar Publishing, 2006.

NEWSON, I.; HOUGHTON, D.; PATTEN, J. Blogging and other social media: exploiting the technology and protecting the enterprise. Surrey: Gower, 2009.

NIELSEN, M. Reinventing Discovery: The New Era of Networked Science. Princeton University Press, 2011

OSBURN, C. B.. The Changing System of Scholarly Communication. Washington D. C.: Association of Research Library, 1986.

OTERO, A. P. S.; GUTIÉRREZ, M. M. B. Biblioteca Digital Colombiana (bdcol): ¿unportal de acceso abierto a la información?. Revista de Derecho, comunicaciones y nuevas tecnologías. n° 5. Bogota, 2011.

PAVÃO, C. P. et al. Repositório digital : acesso livre à informação na Universidade Federal do Rio Grande do Sul . In: Conferência Ibero-Americana de Publicações Eletrônicas no Contexto da Comunicação Científica (2. : 2008 : Rio de Janeiro). Anais. Rio de Janeiro : [s.n.], 2008. Científica (2. : 2008 : Rio de Janeiro). Anais. Rio de Janeiro : [s.n.], 2008.

PEGGY, J. Fundamentals of Collection Development and Management. 2nd ed. Chicago, IL: American Library Association. 2009

PERRENOUD, P. Ciência da Computação: Uma Visão Abrangente(11ed). Bookman Editora: São Paulo, 2013.

PERUZZO, C. M. K. Internet e democracia comunicacional: entre os entraves, utopias e o direito à comunicação. In: MARQUES de MELO, J. STHLER, L. Direitos à comunicação na sociedade da informação. São Bernardo do Campo, SP: UMESP. 2005.

PILLER, R. (2014). Open data. In K. Harvey (Ed.), Encyclopedia of social media and politics. (Vol. 16, pp. 919-920). Thousand Oaks, CA: SAGE Publications, Inc.

PINHEIRO, L.. Do Acesso Livre à Ciência Aberta: conceitos e implicações na comunicação científica. Revista Eletrônica de Comunicação, Informação & Inovação em Saúde, Brasil, 8, jun. 2014. Disponível em: <http://www.reciis.cict.fiocruz.br/index.php/reciis/article/view/946/1806>. Acesso em: 04 Jan. 2015.

PORTER, M. Competitive Advantage: creating and sustaining superior performance. New York, Free Press, 1985.

PREECE, J. ROGERS, Y. SHARP, H. Design de Interacao, Bookman, 2005

REINO UNIDO, Open Data White Paper: Unleashing the Potential. The Stationery Office, 28/06/2012.

RIBEIRO PINTO. O acesso aberto à investigação em ciência da informação em Portugal: alcance e impacto <http://repositorio-aberto.up.pt/bitstream/10216/57297/2/63193.pdf> 2009 Páginas a&b : arquivos e bibliotecas

ROSA, F.; MEIRELLES, R. F.; PALACIO, M. Repositório Institucional da Universidade Federal da Bahia: implantação e acompanhamento. Informação e Sociedade: Estudos. v. 12, n. 1, João Pessoa, 2011.

RUFF, J., 2002. Information overload: Causes, symptoms and solutions. Disponível em: <[http://lila.pz.harvard.edu/\\_upload/lib/InfoOverloadBrief.pdf](http://lila.pz.harvard.edu/_upload/lib/InfoOverloadBrief.pdf)>. Acessado em: 02/11/2013.

SANTOS NETO, A. L. dos et al . Tecnologias de dados abertos para interligar bibliotecas, arquivos e museus: um caso machadiano. Transinformação, Campinas , v. 25, n. 1, Apr. 2013 . Available from <[http://www.scielo.br/scielo.php?script=sci\\_arttext&pid=S0103-37862013000100008&lng=en&nrm=iso](http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0103-37862013000100008&lng=en&nrm=iso)>. access on 16 Feb. 2015. <http://dx.doi.org/10.1590/S0103-37862013000100008>.

SAYÃO LF, SALES LF. Dados de pesquisa: contribuição para o estabelecimento de um modelo de curadoria digital para o país. TPBCI. 2013; 6(1):1-26.

SAYÃO, SALES, Dados abertos de pesquisa: ampliando o conceito de acesso livre. RECIIS – Rev. Eletron. de Comun. Inf. Inov. Saúde. 2014 jun.; 8(2) – p.76-92

SAYÃO, L. F.; MARCONDES, C. H. Softwares livres para repositórios institucionais: alguns subsídios para a seleção. In: SAYÃO, L. et. al. (Org.). Implantação e gestão de repositórios institucionais: políticas, memória, livre acesso e preservação. Salvador: EDUFBA, 2009. p. 23-54

SERRA, P. A. Credibilidade da informação na web. 2006. Disponível em: <<http://www.bocc.ubi.pt/pag/serra-paulo-credibilidade-web.pdf>>. Acesso: 23 mar. 2014.

SHERMAN, C.; PRICE, G. The invisible web: uncovering information sources: search engines can't see. Medford: Cyberage Books, 2001.

SHINTAKU, M.; DUQUE, C. G.; SUAIDEN, E. J. Análise sobre o uso das tendências tecnológicas nos repositórios brasileiros. Pesquisa Brasileira em Ciência da Informação e Bibliotecnologia. v. 9, n. 2, João Pessoa, 2014.

SILVA, A. P. L. da; ALENCAR, L. L.; de SOUZA, J. S. Iniciativas de acesso aberto à informação científica: uma abordagem sobre o portal brasileiro de acesso aberto à informação científica - Oasisbr. In: XVIII Seminário Nacional de Bibliotecas Universitárias (SNBU), Belo Horizonte, 2014. disponível em: [https://www.bu.ufmg.br/snbu2014/trabalhos/index.php/sn\\_20\\_bu\\_14/sn\\_20\\_bu\\_14/paper/view/305](https://www.bu.ufmg.br/snbu2014/trabalhos/index.php/sn_20_bu_14/sn_20_bu_14/paper/view/305), acessado em 23/11/2014.

SOUTHWICK, S. B. The Brazilian electronic theses and dissertations digital library: providing open access for scholarly information. *Ciência da Informação*, V. 35, n. 2, 2006.

SUBER, P. *Open Access*. MIT Press, 2012.

SULEMAN, H., et al, *Networked Digital Library of Theses and Dissertations: Bridging the Gaps for Global Access – Part 1 and 2*, in *D-Lib Magazine*, Vol. 7, No. 9, September 2001. Available <http://www.dlib.org/dlib/september01/suleman/09suleman-pt1.html> and <http://www.dlib.org/dlib/september01/suleman/09suleman-pt2.html>

SWIFT, R. *CRM: customer relationship management: o revolucionário marketing de relacionamento com o cliente*. Elsevier Brasil, Rio de Janeiro. 2001.

TAPIA, R. V. *La Red Mexicana de Repositorios Institucionales – REMERI: Un proyecto nacional para compartir y dar visibilidad al conocimiento*. PKP Scholarly Publishing Conference 2013

TARANTINO, A., CERNAUSKAS, D. *Risk Management in Finance: Six Sigma and other Next Generation Techniques*. Hoboken, Nova Jersey: John Wiley and Sons, 2009. 320 p.

TARAPANOFF, K.; ARAÚJO JÚNIOR, R. H. de; CORMIER, P. M. J. *Sociedade da informação e inteligência em unidades de informação*. *Ciência da Informação*, Brasília, v. 29, n. 3, p. 91-100, set./dez. 2000

TAURION, C. *Big Data*. Brasport. Rio de Janeiro, 2013.

TEIXEIRA, L. G. *Elogio da pedagogia (O)*. Edições Loyola, 2001.

TOBIN, J.; BRAZIEL, L. *Social media is a cocktail party: why you already know the rules of social media marketing*. Cary, NC: Ignite social media, 2008.

TONINI, G. *Apresentação do Repositório de Conteúdo Digital da UFSC*, 2011. disponível em: <http://repositorio.ufsc.br/handle/123456789/119>, Acesso em 01/01/2015.

TRISKA, R., CAFÉ, L. *Arquivos abertos: subprojeto da Biblioteca Digital Brasileira*. *Ciencia da Informacao*, Brasília, v.30, n.3, p.92-96, set./dez. 2001.

TURBAN, E. et al. *Tecnologia da Informação para Gestão: Transformando os Negócios na Economia Digital*. Bookmann, 2008.

VAN de SIMPEL, H., LAGOZE, C. (2000) *The Santa Fe Convention of the Open Archives Initiative*. *D-Lib Magazine*, 6(2). doi:10.1045/february2000-vandesompel-oai ; <http://dx.doi.org/10.1045/february2000-vandesompel-oai>

van WESTRIENEN, G., LYNCH, C. *Academic Institutional Repositories*. *D-Lib Magazine*. v. 11, n. 9, 2005, disponível em: [http://www.dlib.org/dlib/september05/westrienen/09westrienen.html?utm\\_source=dbpia&utm\\_medium=article\\_detail&utm\\_campaign=reference](http://www.dlib.org/dlib/september05/westrienen/09westrienen.html?utm_source=dbpia&utm_medium=article_detail&utm_campaign=reference). Acesso em 05/01/2015.

VENKADESAN. s. (2009) 7th International Convention on Automation of Libraries in Education and Research <http://www.inflibnet.ac.in/caliber2009/CaliberPDF/48.pdf>

VIANA, C.L.M.; MÁRDERO ARELLANO, M.A.; SHINTAKU, M. Repositórios institucionais em ciência e tecnologia: uma experiência de customização do DSpace. In: PROCEEDINGS SIMPOSIO INTERNACIONAL DE BIBLIOTECAS DIGITAIS, 3., 2005, São Paulo. Anais... São Paulo.

VICARI, R. FEB – Federação de Repositórios Educa Brasil, Disponível em [http://www.rnp.br/\\_arquivo/gt/2008/Proposta\\_GT-FEB.pdf](http://www.rnp.br/_arquivo/gt/2008/Proposta_GT-FEB.pdf). Acessado em: fev. 2013.

VIEIRA, M. R. et al. Banco de dados NoSQL: conceitos, ferramentas, linguagens e estudos de casos no contexto do Big Data. In: SBBD, São Paulo, 2012.

XEXEO, G. Modelagem de sistemas de informação: da análise de requisitos ao modelo de interface. online, 2007. disponível em: <http://pt.scribd.com/doc/5290929/Modelagem-de-Sistemas-de-Informacao#scribd>. acessado em: 04/01/2015.

YAMAMOTO, Y.; TANAKA, K. Findings comparative facts and aspects for judging the credibility of uncertain[ facts. In: VOSSSEN, G.;LONG, D. D. E.; YU, J. X. (Eds). Web Information Systems Engineering-WISE. Poznań, Polônia: Springer, 2009.

YANG, X.; WANG, L.; Jie, W. Guide to e-Science: Next Generation Scientific Research and Discovery. Springer Science & Business Media, 26/05/2011.

ZIMAN, J. M. An Introduction of science studies: The Philosophical and Social Aspects of Science and Technology. Cambridge: Cambridge University Press, 1984.