

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL  
FACULDADE DE BIBLIOTECONOMIA E COMUNICAÇÃO  
DEPARTAMENTO DE CIÊNCIAS DA INFORMAÇÃO

Julio Cesar de Castro

**e-Science e Organização da Informação:** abordagens da Biblioteconomia e da  
Ciência da Informação para a curadoria de dados de pesquisa

Porto Alegre

2012

Julio Cesar de Castro

**e-Science e Organização da Informação:** abordagens da Biblioteconomia e da  
Ciência da Informação para a curadoria de dados de pesquisa

Trabalho de Conclusão de Curso submetido ao  
Curso de Biblioteconomia da Faculdade de  
Biblioteconomia e Comunicação da Universidade  
Federal do Rio Grande do Sul como requisito  
parcial para obtenção do título de Bacharel em  
Biblioteconomia.

Orientador: Me. Jackson da Silva Medeiros.

Coorientadora: Prof<sup>a</sup>. Dra. Sônia Elisa Caregnato.

Porto Alegre

2012

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

Reitor: Prof. Dr. Carlos Alexandre Netto

Vice-Reitor: Prof. Dr. Rui Vicente Oppermann

FACULDADE DE BIBLIOTECONOMIA E COMUNICAÇÃO

Diretora: Prof<sup>a</sup>. Dra. Regina Helena van der Laan

DEPARTAMENTO DE CIÊNCIAS DA INFORMAÇÃO

Chefe: Prof<sup>a</sup>. Dra. Ana Maria Mielniczuk de Moura

Chefe Substituta: Prof<sup>a</sup>. Dra. Sônia Elisa Caregnato

COMISSÃO DE GRADUAÇÃO DO CURSO DE BIBLIOTECONOMIA

Coordenadora: Prof<sup>a</sup>. Dra. Samile Andréa de Souza Vanz

Coordenadora Substituta: Prof<sup>a</sup>. Me. Glória Isabel Sattamini Ferreira

**C355e** Castro, Julio Cesar de

*e-Science* e Organização da Informação: abordagens da Biblioteconomia e da Ciência da Informação para a curadoria de dados de pesquisa/ Julio Cesar de Castro ; orientador Jackson da Silva Medeiros ; coorientadora Sônia Elisa Caregnato . – 2012.

48 f.

1. *e-Science*. 2. Organização da Informação. 3. Curadoria de dados. I. Medeiros, Jackson da Silva. II. Caregnato, Sônia Elisa. III. Título.

CDU 025.4

Departamento de Ciências da Informação  
Rua Ramiro Barcelos, 2705 Campus Saúde  
Bairro Santana  
Porto Alegre – RS  
CEP: 90035-007  
Telefone: (51) 3308-5146  
E-mail: [fabico@ufrgs.br](mailto:fabico@ufrgs.br)

Julio Cesar de Castro

**e-Science e Organização da Informação:** abordagens da Biblioteconomia e da  
Ciência da Informação para a curadoria de dados de pesquisa

Trabalho de Conclusão de Curso submetido ao  
Curso de Biblioteconomia da Faculdade de  
Biblioteconomia e Comunicação da Universidade  
Federal do Rio Grande do Sul como requisito  
parcial para obtenção do título de Bacharel em  
Biblioteconomia.

Data de aprovação: 11 de dezembro de 2012.

#### **BANCA EXAMINADORA**

---

Me. Jackson da Silva Medeiros – UFRGS (orientador)

---

Prof<sup>a</sup>. Dra. Sônia Elisa Caregnato - UFRGS (coorientadora)

---

Prof. Dr. Rafael Port da Rocha – UFRGS

---

Prof<sup>a</sup>. Me. Rita do Carmo Ferreira Laipelt – UFRGS

## **AGRADECIMENTOS**

Primeiramente, gostaria de agradecer à Universidade Federal do Rio Grande do Sul (UFRGS) e à Faculdade de Biblioteconomia e Comunicação (FABICO) por terem proporcionado tamanho desenvolvimento pessoal e intelectual por meio do convívio com profissionais conceituados durante minha graduação.

Ao orientador Jackson Medeiros e a coorientadora Sônia Caregnato, por terem orientado esse trabalho com ricas contribuições e amplos esclarecimentos.

Aos membros da banca, Rafael Port da Rocha e Rita Laipelt, por terem aceitado o convite e disponibilizado tempo para avaliar esse trabalho.

À Universidade Federal de Minas Gerais (UFMG) e à Escola de Ciência da Informação (ECI) por terem me recebido por intermédio da mobilidade acadêmica, o que possibilitou meu contato com profissionais referências na área da Ciência da Informação no Brasil.

À professora Ana Maria Dalla Zen, que me fez crescer não apenas intelectualmente, mas também como ser humano.

Aos meus pais, Augusto e Karla, pelo apoio desde a escolha do curso de Biblioteconomia até a realização dessa monografia.

À minha irmã Karine e à minha avó Maria da Penha, que, mesmo longe, me apoiaram ao longo desses anos com palavras de conforto e com boas energias.

Ao Henrique Casagrande, por ter me acompanhado durante minha primeira matrícula na faculdade, me apresentando à Fabico, além de se fazer presente em momentos bons e ruins de minha vida.

À minha amiga Janaína Bordignon, por me ajudar com a revisão do texto e com a tradução das citações em inglês, constante nos momentos desesperadores e munida de palavras reconfortantes e muito companheirismo.

Aos amigos Camila Bermúdez, Pedro Girardi, Mariane Brusque, Priscila Lopes, Ana Dedomenico, Andreli Dalbosco, Andressa Andrade e Patrícia Noms, que, ao longo desse trabalho, mandaram mensagens de apoio.

Por fim, agradeço a todos e a todas que, por questão de espaço, não puderam ser mencionados/as, porém sabem ter sido essenciais no processo de construção desse trabalho, contribuindo para meu crescimento emocional, psicológico e acadêmico.

*“As novas tecnologias nos condenaram a nos tornar inteligentes.”*

Michel Serres

## RESUMO

Esse estudo tem por objetivo investigar na literatura quais abordagens da Organização da Informação nas áreas da Biblioteconomia e da Ciência da Informação podem contribuir para a curadoria e para o compartilhamento de dados de pesquisa. Para tal, seu referencial teórico está dividido em duas grandes áreas, *e-Science* e Organização da Informação. A primeira foca na apresentação dos conceitos de compartilhamento e de curadoria de dados de pesquisa, ao passo que a segunda explora as abordagens de indexação, de classificação, de recuperação da informação e de inserção de metadados, bem como os instrumentos de vocabulário controlado, como cabeçalhos de assuntos, taxonomias, tesouros e ontologias. Conclui-se que os movimentos de compartilhamento e de curadoria de dados são importantes tanto para a comunidade científica quanto para a sociedade como um todo, demandando trabalho interdisciplinar entre profissionais da informação, profissionais da TI e pesquisadores responsáveis pela produção de dados. Notou-se, também, que a gama de abordagens provinda da Biblioteconomia e da Ciência da Informação poderá contribuir para a curadoria e o compartilhamento de dados na condição de que seja mediada por profissionais qualificados da área.

**Palavras-chave:** Organização da Informação. *e-Science*. Curadoria de dados. Compartilhamento de dados de pesquisa.

## ABSTRACT

This study has as an objective to investigate in the literature which Information Organization approaches in the areas of Librarianship and Information Science can contribute to the research data curation and sharing. In order to do so, the theoretical background is divided into two great areas, e-Science and Information Organization. The first focuses on the presentation of data curation and sharing concepts, while the latter explores the indexing, classification, information recovery and metadata insertion approaches, as well as the controlled vocabulary instruments, such as subject headings, taxonomies, thesaurus and ontologies. In conclusion, data curation and sharing are important not only to the scientific community, but also to society as a unit, demanding an interdisciplinary work among information professionals, IT professionals and researchers responsible for the data production. It can also be perceived that the range of approaches originally from the areas of Librarianship and Information Science will contribute to data sharing and curation, provided that it is mediated by qualified professionals.

**Keywords:** Information Organization. E-Science. Data curation. Data sharing.



## SUMÁRIO

<b>1 INTRODUÇÃO</b> .....	10
<b>2 OBJETIVOS</b> .....	13
2.1 OBJETIVO GERAL.....	13
2.2 OBJETIVOS ESPECÍFICOS.....	13
<b>3 METODOLOGIA</b> .....	14
<b>4 E-SCIENCE</b> .....	17
4.1 COMPARTILHAMENTO DE DADOS DE PESQUISA.....	19
4.2 CURADORIA DE DADOS.....	25
<b>5 ORGANIZAÇÃO DA INFORMAÇÃO</b> .....	30
5.1 ABORDAGENS DA BIBLIOTECONOMIA E CIÊNCIA DA INFORMAÇÃO PRESENTES NA ORGANIZAÇÃO DA INFORMAÇÃO.....	32
5.2 INSTRUMENTOS DE ORGANIZAÇÃO E REPRESENTAÇÃO DA INFORMAÇÃO .....	37
<b>6 CONSIDERAÇÕES FINAIS</b> .....	42
<b>REFERÊNCIAS</b> .....	45

## 1 INTRODUÇÃO

Com a inserção das tecnologias da informação e da comunicação no meio científico, a colaboração entre grupos de pesquisa deve se tornar mais ágil, uma vez que as barreiras físicas são quebradas por meio da utilização da internet. A colaboração em rede é uma das principais defesas da *e-Science*, assim como veremos no referencial teórico desse trabalho. Seu surgimento ocorreu no contexto das ciências naturais, uma vez que o volume de produção de dados nessa área é grandioso, porém, sua aplicação também tem atingido outras áreas da ciência.

Jim Gray (2009, p. xxx)<sup>1</sup> apresenta um plano ideal onde “o objetivo é ter um mundo no qual toda a literatura científica esteja online, todos os dados científicos estejam online e interoperem entre si”. Esse objetivo pode ser alcançado por intermédio de movimentos pertencentes à própria *e-Science*, denominados curadoria e compartilhamento de dados de pesquisa. Essas abordagens defendem o tratamento, a preservação e a disponibilização de dados, tanto para grupos de pesquisa quanto para usuários em geral, através de abordagens próprias para esse fim.

De forma a possibilitar o processo de curadoria e de compartilhamento de dados, é necessário um trabalho interdisciplinar<sup>2</sup>. Logo, conta-se com a colaboração entre profissionais da informação, os quais dominam conhecimentos sobre tratamento, organização e preservação, profissionais da TI, auxiliando na escolha e na manutenção de bases de dados, bem como no processamento de dados, e pesquisadores responsáveis pela produção de dados, que fornecem os próprios dados e as descrições sobre eles.

Estudar esses movimentos e a forma com que esses dados são armazenados para uma posterior disponibilização se faz relevante, uma vez que grande parte dessa produção é perdida. Considerando que dados preservados poderão servir para a produção de novas pesquisas, otimizando tempo e dinheiro de grupos de pesquisa, torna-se relevante a reflexão sobre possíveis formas de armazenamento e divulgação. Além disso, esses resultados podem ser comunicados para a população, demonstrando onde o dinheiro público está sendo aplicado.

---

<sup>1</sup> Texto original: The goal is to have a world in which all of the science literature is online, all of the science data is online, and they interoperate with each other. (GRAY, 2007, p. xxx)

<sup>2</sup> Significado: *dj m+f (inter+disciplina+ar<sup>3</sup>)* Comum a diversas disciplinas. (MICHAELIS, 2012, *online*)

A cada ano, cresce o número de pesquisas desenvolvidas em nosso país. Porém, parte dessa produção de dados não publicados acaba sendo perdida. Uma vez que nossos resultados de busca pelos temas que envolvem este trabalho revelam que a produção nacional sobre esse tema é baixa, este trabalho de pesquisa procura trazer o assunto ao nosso contexto, discutindo e difundindo questões atuais e emergentes.

Além disso, outro fator motivador para a realização desse estudo é a possibilidade de implementação de um repositório na Universidade Federal do Rio Grande do Sul, que visará o armazenamento de dados de pesquisa produzidos pela própria Instituição, necessitando de estudos sobre a própria essência da *e-Science* e seus processos de curadoria e de compartilhamento de dados. Como veremos no referencial teórico desse trabalho, abordagens da Biblioteconomia e da Ciência da Informação referentes à Organização da Informação podem contribuir para a realização de projetos como esse.

Assim, esse trabalho pretende responder a seguinte questão: como o processo de Organização da Informação, utilizando abordagens da Biblioteconomia e da Ciência da Informação, pode contribuir para a curadoria e o compartilhamento de dados de pesquisa?

Para tal, o referencial teórico está dividido em duas grandes áreas, sendo elas a *e-Science* e a Organização da Informação. A primeira é composta por textos dos idiomas inglês e espanhol. Alguns dos autores citados foram: Hey e Hey (2006); Gray (2009); Martinez-Uribe e Macdonald (2008); Yekel (2007); Borgman (2012) e Torres-Salinas, Robinson-Gracia e Cabezas-Clavijo (2012). A segunda permitiu utilizar literatura nacional ou traduções, uma vez que foram abordados conceitos já consolidados da área, sendo que alguns dos autores consultados foram: Robredo e Cunha (1986); Lancaster (1993); Alvarenga (2001; 2003; 2006); Bräsher e Café (2008) e Medeiros (2012).

Em relação à metodologia do trabalho, foi realizada uma pesquisa bibliográfica de cunho qualitativo e de natureza exploratória, a fim de compreender o que os autores que defendem a curadoria e o compartilhamento de dados têm publicado, além de levantar abordagens da Biblioteconomia e da Ciência da Informação presentes no processo de Organização da Informação. Desse modo, traz-se também à literatura científica nacional um tópico amplamente debatido internacionalmente.

Por fim, tendo introduzido os aspectos abordados nesse trabalho, justificando e apresentando o problema de pesquisa, a próxima seção apresenta os objetivos geral e específicos desse estudo.

## 2 OBJETIVOS

Esta seção restringe-se a apresentar os objetivos do estudo, dividindo-os em objetivo geral e objetivos específicos, como segue.

### 2.1 OBJETIVO GERAL

Investigar abordagens da Organização da Informação na Biblioteconomia e Ciência da Informação que podem contribuir para a curadoria e o compartilhamento de dados de pesquisa.

### 2.2 OBJETIVOS ESPECÍFICOS

- a) Verificar como se apresentam as questões sobre curadoria e compartilhamento de dados de pesquisa na literatura;
- b) Relacionar abordagens sobre Organização da Informação existentes na Biblioteconomia e Ciência da Informação;
- c) Apresentar formas de Organização da Informação baseadas na literatura da Biblioteconomia e na Ciência da Informação que podem contribuir para a curadoria e compartilhamento de dados de pesquisa.

Para cumprir esses objetivos, a pesquisa foi delineada com base na metodologia exposta na próxima seção.

### 3 METODOLOGIA

Nessa seção são apresentados os processos metodológicos utilizados durante a realização desse trabalho. Em relação à abordagem, pode-se considerá-la uma pesquisa qualitativa, defendido por van Maanen<sup>3</sup> (1979, p. 520 apud NEVES, 1996, p. 1) como um estudo que

[...] procura reduzir a distância entre a teoria e os dados, entre o contexto e a ação, usando a lógica da análise fenomenológica, isto é, da compreensão dos fenômenos pela sua descrição e interpretação. (VAN MAANEN, 1979, p. 520 apud NEVES, 1996, p. 1)

Ademais, essa pesquisa é de natureza exploratória, pois visa “[...] proporcionar maior familiaridade com o problema, com vistas a torná-lo explícito ou a construir hipóteses.” (GIL, 2002, p. 41). Foi realizada, também, uma pesquisa bibliográfica, a qual consiste no “[...] levantamento de toda a bibliografia publicada em forma de livros, publicações avulsas e imprensa escrita.” (LAKATOS; MARCONI, 1992, p. 43).

O referencial teórico está separado em duas grandes áreas: *e-Science* e Organização da Informação. A primeira trata de materiais bibliográficos, mais especificamente de artigos de periódicos científicos em idiomas inglês e espanhol, sendo que o período de publicação não foi delimitado, uma vez que os estudos referentes ao tema são muito recentes. Por outro lado, a segunda seção teve como fontes livros, artigos de periódicos, anais de eventos, dissertações e teses de língua portuguesa. Nesse caso, também não foi delimitado um período de publicação dos materiais, pois as abordagens relacionadas à Biblioteconomia e à Ciência da Informação, ambas presentes no processo de Organização da Informação, exigem conceitos e abordagens tanto atuais quanto consolidados na área.

A coleta de dados é definida por Lakatos e Marconi (2003, p. 165) como a “[...] etapa da pesquisa em que se inicia a aplicação dos instrumentos elaborados e das técnicas selecionadas, a fim de se efetuar a coleta dos dados previstos.” Os instrumentos e as técnicas utilizados nesse trabalho são descritos abaixo, divididos de acordo com as duas grandes áreas mencionadas anteriormente.

---

<sup>3</sup> VAN MAANEN, J. Reclaiming Qualitative Methods for Organizational Research: A Preface. **Administrative Science Quarterly**, Ithaca, v. 24, n. 4, dec. 1979.

Ao tratar da *e-Science*, os artigos foram coletados em bases de dados nacionais e internacionais de conteúdos multidisciplinar ou especializado. O critério utilizado para a escolha das bases foi a constatação de referência de artigos da área de Ciência da Informação. As bases multidisciplinares utilizadas foram: Web of Science; Scopus e Google Acadêmico. Além disso, as bases especializadas foram: BRAPCI, LISA e LISTA. Uma vez que essas bases não são de acesso livre (exceto o Google Acadêmico e BRAPCI), o acesso foi possível por meio do Portal de Periódicos da Capes.

A busca foi realizada entre as datas 16/09/2012 e 24/09/2012. No que se refere à *e-Science*, a busca foi realizada a partir dos termos E-Science, eScience e cyberinfrastructure, retornando 25924 resultados. Para o compartilhamento de dados, foram recuperados, a partir dos termos, data sharing, sharing research data, reusing scientific data e compartilhamento de dados, resultando em 15964. Os materiais sobre curadoria de dados, por sua vez, foram buscados a partir dos termos data curation, curadoria digital e curadoria de dados, respondendo por 429 artigos.

Em cada base foram utilizados diferentes campos de busca: no caso da Web of Science, utilizou-se o campo “assunto”; na Scopus, os campos “título, resumo e palavra-chave”; no Google Acadêmico, o campo “título” e nas bases especializadas BRAPCI, LISA e ISTA, fez-se uso de “todos os campos”. Para filtrar os resultados nas bases de dados multidisciplinares, os resultados foram restringidos para a área de Ciência da Informação no contexto da Web of Science. Na Scopus a área selecionada foi a das Ciências Sociais Aplicadas, uma vez que só era possível restringir a pesquisa pelas grandes áreas do conhecimento. No Google Acadêmico não foi possível filtrar por área, portanto foram verificadas as primeiras vinte e uma páginas de resultados (quando tal número era excedido).

A partir dos resultados gerados a seleção dos artigos foi feita por intermédio da análise do título e dos resumos. Em seguida foi realizada uma leitura técnica<sup>4</sup> a fim de selecionar os artigos que seriam lidos na íntegra para, finalmente, realizar uma leitura integral e os fichamentos, com o objetivo de separar o conteúdo que seria utilizado como referencial teórico dessa pesquisa.

Na seção da Organização da Informação, o material foi coletado a partir do repositório institucional LUME (dissertações e teses) e da base de dados BRAPCI

---

<sup>4</sup> O processo de leitura técnica foi realizado através da leitura do resumo, introdução, título dos capítulos e considerações finais.

(artigos e anais). O registro dos livros foi recuperado por meio do Sabi - Catálogo *online* das bibliotecas da UFRGS, para coletar posteriormente os livros físicos na Biblioteca da Faculdade de Biblioteconomia e Comunicação. Os termos utilizados para realizar a busca foram: organização da informação; representação da informação; organização do conhecimento; representação do conhecimento; indexação; e documentação. Após a seleção processo similar ao anterior ocorreu, no qual foi realizada primeiramente uma leitura técnica para então uma apreciação das partes pertinentes e um fichamento do conteúdo.

Essa seção dedicou-se à descrição dos métodos utilizados para a composição desse trabalho. Na seção seguinte, serão apresentados os resultados dessas coletas enquanto construção do referencial teórico sobre o tema.



## 4 E-SCIENCE

Antes de abordar a curadoria e o compartilhamento de dados de pesquisa, discorreremos brevemente sobre a grande área a qual esses movimentos estão inseridos, denominada *e-Science*. Segundo Jankowski (2007, p. 551), o termo foi utilizado pela primeira vez em 1999 por John Taylor, na época diretor geral do *Office of Science and Technology* no Reino Unido. O autor ainda lembra que tal terminologia foi concebida para o processamento de grandes volumes de dados, além de ter como foco as Ciências Naturais e Biológicas.

Em definição do próprio Taylor<sup>5</sup> (2001 apud HEY; HEY, 2006, p. 517), “*e-Science* trata da colaboração global entre áreas-chave da ciência e da próxima geração de infraestrutura que irá disponibilizá-la”<sup>6</sup>. Essa colaboração se empodera no meio científico, uma vez que as barreiras físicas são quebradas e grupos de qualquer parte do mundo podem compartilhar dados por meio da *internet*. Porém é relevante estabelecer que a *e-Science* não é uma nova disciplina científica; é, sim, um atalho entre as ferramentas tecnológicas e a colaboração científica em rede. Dessa forma, cientistas são possibilitados a realizar pesquisas de uma forma distinta, melhor e mais ágil. (HEY; HEY, 2006, p. 517).

Apesar da origem do termo *e-Science* remeter a estudos realizados na área das Ciências Biológicas, suas contribuições não estão limitadas às ciências naturais, onde a *e-Science* tornou-se, em alguns países e disciplinas, o *modus operandi*, mas também está penetrando os domínios das Ciências Sociais e Humanas”<sup>7</sup> (JANKOWSKI, 2007, p. 549). Contudo, cada campo científico produz diferentes tipos de dados, conforme exemplificam abaixo os autores Martínez-Uribe e Macdonald (2008, p. 274)

<sup>5</sup> TAYLOR, J.M. (2001), Talk given at UK e-Science Town Meeting, July.

<sup>6</sup> Texto original: e-Science is about global collaboration in key areas of science, and the next generation of infrastructure that will enable it. (TAYLOR, 2001 apud HEY; HEY, 2006, p. 517)

<sup>7</sup> Texto original: This transformation is not limited to the natural sciences, where e-science has become, in some countries and disciplines, the *modus operandi*, but is also penetrating the domains of the social sciences and humanities. (JANKOWSKI, 2007, p. 549)

[...] nas Ciências Naturais, são medições de precipitação de chuvas, observações astronômicas, bases de dados de modelos genéticos ou estruturas cristalinas. Nas Ciências Sociais, os dados são gerados através de pesquisas de opinião ou mapas com censos de informação georeferenciada. Em Humanidades, podem incluir fotografias de antigas escrituras em pedra, e, na Medicina, neuroimagens que captam a atividade cerebral.<sup>8</sup> (MARTÍNEZ-URIBE E MACDONALD, 2008, p. 274)

Para Gray (2009, p. xxx), essas transformações ocorridas durante os últimos anos indicam que um novo paradigma científico pode estar surgindo. Ademais, ele instrui que

[...] quase tudo sobre a ciência está mudando em função do impacto da tecnologia de informação. As Ciências Experimental, Teórica e Computacional estão todas sendo afetadas pelo dilúvio de dados, e um quarto paradigma científico “intensificado por dados” está emergindo. O objetivo é ter um mundo no qual toda a literatura científica esteja online, todos os dados científicos estejam online, e eles interoperem entre eles.<sup>9</sup> (GRAY, 2009, p. xxx)

Esse objetivo pode ser alcançado por meio dos movimentos científicos denominados como compartilhamento de dados de pesquisa (*data sharing*) e curadoria de dados (*data curation*). Por defender o compartilhamento, a preservação e o acesso aos dados de pesquisa via internet, permitindo melhor colaboração entre pesquisadores e grupos de pesquisa, tais movimentos adquirem um papel crucial dentro da *e-Science*.

Após apresentarmos o conceito da *e-Science*, na próxima seção serão abordadas as características do compartilhamento de dados de pesquisa, além de suas dificuldades e benefícios para a comunicação científica.

---

<sup>8</sup> Texto original: en ciencias de la naturaleza son las mediciones de precipitación de lluvias, las observaciones astronómicas, bases de datos de modelos genéticos o de estructuras cristalográficas. En ciencias sociales, los datos son generados a través de encuestas de opinión o mapas con información de censos geo-referenciados. En humanidades pueden incluir fotografías de antiguas escrituras en piedra, y en medicina neuroimágenes que captan la actividad del cerebro. (MARTÍNEZ-URIBE E MACDONALD, 2008, p. 274)

<sup>9</sup> Texto original: [. . .] almost everything about science is changing because of the impact of information technology. Experimental, theoretical, and computational science are all being affected by the data deluge, and a fourth, “data-intensive” science paradigm is emerging. The goal is to have a world in which all of the science literature is online, all of the science data is online, and they interoperate with each other. (GRAY, 2007, p. xxx)

#### 4.1 COMPARTILHAMENTO DE DADOS DE PESQUISA

Recentemente o compartilhamento de dados de pesquisa se tornou foco de discussões para além de revistas da área científica. Periódicos científicos como a *Nature* em 2008 e 2009 e a *Science* em 2011 e revistas como *Wired* em 2008 e *The Economist* em 2010 publicaram sobre o tema. (BORGMAN, 2012, p. 1059).

De forma geral, pode-se afirmar que o “compartilhamento de dados é a liberação de dados de pesquisa para o uso de outros”<sup>10</sup> (BORGMAN, 2012, p. 1060). Torres-Salinas, Robinson-Garcia e Cabezas-Clavijo (2012, p. 173) complementam essa ideia ao dizerem que o compartilhamento de dados de pesquisa “consiste em compartilhar os dados finais de pesquisa entre os cientistas, com o objetivo de maximizar esforços e recursos.”<sup>11</sup> Cragin et al (2010, p. 4024) enfatizam seu caráter colaborativo e afirmam que “compartilhamento é a base do sucesso, em vista que a coleta, o armazenamento e o uso de dados podem apenas ocorrer após a existência de meios para o compartilhamento”.<sup>12</sup>

Ao tratar de compartilhamento de dados, torna-se relevante a definição do termo dados. Conforme o National Research Council<sup>13</sup> (1999 p. 19 apud BORGMAN, 2012, p. 1061), “dados são fatos, números, letras e símbolos que descrevem objetos, ideias, condições, situações ou outros fatores”<sup>14</sup>. Devido a suas especificidades, cada campo científico produzirá tipos distintos de dados, ilustrados por Borgman<sup>15</sup> (2007, 2009 apud BORGMAN, 2012, p. 1061)

<sup>10</sup> Texto original: Data sharing is the release of research data for use by others. (BORGMAN, 2012, p. 1060).

<sup>11</sup> Texto original: consistente en compartir los datos finales de investigación entre los científicos, con el objetivo de maximizar esfuerzos y recursos.” (TORRES-SALINAS, ROBINSON-GARCIA E CABEZAS-CLAVIJO, 2011, p. 173)

<sup>12</sup> Texto original: sharing is at the heart of success, as collecting, storing and making use of data can only come after the means for sharing are in place. (CRAGIN et al, 2010, p. 4024)

<sup>13</sup> National Research Council. (1999). A question of balance: Private rights and the public interest in scientific and technical databases. Washington, DC: National Academies Press.

<sup>14</sup> Texto original: “Data are facts, numbers, letters, and symbols that describe an object, idea, condition, situation, or other factors” (NATIONAL RESEARCH COUNCIL, 1999, p. 15 apud BORGMAN, 2012, p. 1061)

<sup>15</sup> Borgman, C.L. (2007). *Scholarship in the digital age: Information, infrastructure, and the Internet*. Cambridge, MA: MIT Press. / Borgman, C.L. (2009). *The digital future is now: A call to action for the humanities*. *Digital Humanities Quarterly*, 3(4), 233.

Nas ciências biológicas e físicas, a maioria dos dados são reunidos ou produzidos por pesquisadores, por meio de observações, experimentos ou modelos. Nas ciências sociais, pesquisadores podem reunir ou produzir seus próprios dados, ou eles podem obter dados em outras fontes, tais quais registros públicos ou atividade econômica.<sup>16</sup> (BORGMAN, 2007; 2009 apud BORGMAN, 2012, p. 1061) :

Devido ao caráter abrangente, existem diferentes categorias de dados. Os dados de pesquisa, foco desse trabalho, são descritos pela OECD - Organization for Economic Co-operation and Development<sup>17</sup> (2007 apud TORRES-SALINAS, ROBINSON-GARCIA E CABEZAS-CLAVIJO, 2012, p. 175) como

[...] todo aquele material que tenha sido registrado durante a pesquisa, reconhecido pela comunidade científica e que serve para certificar os resultados da pesquisa que se realiza. Além dessas características, também se especifica que deve provir de uma fonte única e devem ser difíceis ou impossíveis de se obter novamente, por ser de um momento ou circunstâncias que não se repetiriam de forma exatamente igual. Todos os dados que atendem as características acima são conhecidos como dados finais de pesquisa.<sup>18</sup> (OECD, 2007 apud TORRES-SALINAS, ROBINSON-GARCIA E CABEZAS-CLAVIJO, 2012, p. 175)

Observa-se também que a diversidade entre tipos de dados de pesquisa influi diretamente no que se refere ao seu compartilhamento. Borgman (2012, p. 1066) acredita que os dados são a alma da pesquisa, independentemente do seu campo científico. A complexidade e as dificuldades de compartilhá-los dependerá da variedade de tipos e origens de dados, da forma como os pesquisadores desejam compartilhá-los (integralmente, parcialmente ou temporariamente) e as circunstâncias locais de sua coleta. A autora, porém, aponta que os critérios para identificação e compartilhamento de dados ainda não são bem compreendidos. (BORGMAN, 2012, p. 1062)

---

<sup>16</sup> Texto original: In the physical and life sciences, most data are gathered or produced by researchers, such as by observations, experiments, or models. In the social sciences, researchers may gather or produce their own data, or they may obtain data from other sources such as public records of economic activity. (BORGMAN, 2007, 2009 apud BORGMAN, 2012, p. 1061) .

<sup>17</sup> OECD (Organisation for Economic Co-operation and Development). *OECD Principles and guidelines for access to research data from public funding*. Paris: OECD Publications, 2007.

<sup>18</sup> Texto original: es la que considera datos de investigación todo aquel material que ha sido registrado durante la investigación, reconocido por la comunidad científica y que sirve para certificar los resultados de la investigación que se realiza. Además de estas características también se especifica que debe provenir de una fuente única y deben ser difíciles o imposibles de obtener de nuevo por ser propios de un momento o circunstancias irrepetibles de una forma exactamente igual. Todos los datos que cumplen las características anteriores se conocen como datos finales de investigación. (OECD, 2007 apud TORRES-SALINAS, ROBINSON-GARCIA E CABEZAS-CLAVIJO, 2012, p. 175)

Torres-Salina (2010, *online*) esclarecem que “publicar dados não é o mesmo que publicar ppts [arquivos Microsoft PowerPoint] ou compartilhar links, já que nos deparamos com informações muito mais complexas, com especificidades de cada especialidade”.<sup>19</sup> Em outras palavras, compartilhar não é uma tarefa simples, uma vez que demanda o envolvimento de várias pessoas, a disponibilidade de tempo, de dinheiro, etc. Por tal motivo, ainda não há uma prática entre pesquisadores de compartilhar seus dados, em vista que, “tradicionalmente cientistas tinham boas desculpas para não salvar e documentar tudo para sempre, era não econômico ou inviável” (GRAY et al, 2002, p. 2).<sup>20</sup>

Borgman (2012, p. 1059) menciona outras razões para a dificuldade enfrentada no compartilhamento de dados, como falta de recursos financeiros, ausência de incentivos para compartilhá-los ou até questões éticas e/ou epistemológicas. Outro motivo exposto por Cragin et al (2010, p. 4033) é o medo que os pesquisadores têm de que seus dados sejam utilizados de forma indevida. Borgman (2012, p. 1070) reforça essa afirmação ao mencionar que alguns preferem compartilhá-los com pesquisadores da mesma área, mas não com o público em geral, justamente pelo receio de que seus dados sejam também mal interpretados devido à falta de conhecimento.

Dessa forma, alguns pesquisadores optam por compartilhar seus dados informalmente. Segundo Martínez-Urbe e Macdonald (2008, p. 279),

[...] em muitos casos, a comunidade científica prefere métodos mais informais para compartilhar recursos. Esses procedimentos permitem a pessoa que compartilha avaliar o uso que se quer dar a seus dados para decidir em cada caso, dar ou não acesso a eles.<sup>21</sup> (MARTÍNEZ-URIBE E MACDONALD, 2008, p. 279)

O problema desse tipo de compartilhamento é que resulta no impedimento da reutilização dos dados de forma permanente, como afirmam Torres-Salinas, Robinson-Garcia e Cabezas-Clavijo (2012, p. 178):

<sup>19</sup>Texto original: colgar datos no es igual que colgar ppts o compartir enlaces; nos encontramos con información mucho más compleja, con especificidades propias de cada especialidad (TORRES-SALINA, 2010)

<sup>20</sup>Texto original: “Traditionally scientists have had good excuses for not saving and documenting everything forever, it was uneconomic or infeasible.” (GRAY et al, 2002, p. 2).”

<sup>21</sup>Texto original: en muchos casos la comunidad científica prefiere métodos más informales para compartir recursos. Estos procedimientos permiten a la persona que comparte evaluar el uso que se quiere dar de sus datos para decidir en cada caso si se da acceso a ellos. (MARTÍNEZ-URIBE E MACDONALD, 2008, p. 279)

[...] é evidente que um grave obstáculo impedirá a reutilização permanente dos dados, referente ao formato em que se apresentam e a sua obsolescência pode complicar a reutilização permanente dos dados e complicar a conservação e a disponibilidade futura dos dados. Por outro lado, os dados compartilhados de maneira informal cumprem unicamente o critério de normalização do autor, que por sua vez se adéqua a sua forma de trabalho, impedindo em muitos casos que eles sejam reutilizados com vantagem.<sup>22</sup> (TORRES-SALINAS, ROBINSON-GARCIA E CABEZAS-CLAVIJO, 2012, p. 178)

Pode-se entender que, portanto, grande parte dos dados produzidos por estudos acaba sendo inutilizada e perdida. Segundo Gray (2009, p. xvii), os resultados publicados na literatura científica representam apenas a ponta do *iceberg* dos dados existentes. Essa situação torna-se preocupante ao considerarmos a quantidade de dados que poderia fazer parte da construção de novos conhecimentos.

Borgman (2012, p. 1059) relaciona quatro razões para compartilhar os dados de pesquisa:

- a) reproduzir ou verificar uma pesquisa;
- b) tornar os resultados de pesquisa com financiamento público disponível para o público;
- c) permitir que outros possam realizar novos questionamentos sobre dados existentes;
- d) avançar o estado da pesquisa e inovação.<sup>23</sup>

Percebe-se, assim, que os benefícios de compartilhar dados de pesquisa não são apenas para a comunidade científica, mas também para a sociedade como um todo. Torres-Salina (2010, *online*) defende que principalmente os dados de pesquisa financiados com dinheiro público devem ser compartilhados:

<sup>22</sup>Texto original: se deduce un grave impedimento que obstaculiza la reutilización permanente de los datos, y es el referente al formato en que se presentan ya que su obsolescencia puede complicar la conservación y disponibilidad de los datos en el futuro. Por otro lado, los datos compartidos de manera informal cumplen únicamente el criterio de normalización del autor, que a su vez se adecua a su forma de trabajo, impidiendo en muchos casos que puedan ser empleados con provecho. (TORRES-SALINAS, ROBINSON-GARCIA E CABEZAS-CLAVIJO, 2012, p. 178)

<sup>23</sup>Texto original: Four rationales for sharing data are examined, drawing examples from the sciences, social sciences, and humanities: (1) to reproduce or to verify research, (2) to make results of publicly funded research available to the public, (3) to enable others to ask new questions of extant data, and (4) to advance the state of research and innovation. (BORGMAN, 2012, p. 1059)

Elas também [administração pública] teriam que reclamar que os dados gerados deveriam ser utilizados por outros cientistas do mesmo sistema público. Assim, se reciclaríamos todos os projetos que não alcançaram os resultados esperados, mas cujos dados podem ser de interesse e, em última instância, poderíamos evitar duplicação de pesquisa, poupando algum dinheiro.<sup>24</sup> (TORRES-SALINA, 2010, *online*)

Borgman (2012, p. 1069) também aprofunda sua argumentação em relação às pesquisas financiadas com dinheiro público ao determinar que

o sentimento público pelo compartilhamento de dados de pesquisa é baseado largamente na lógica de que o dinheiro dos impostos deveria ser alavancado para servir o bem público. Nesse ponto de vista, dados produzidos com financiamento público deveriam estar disponíveis para uso e não ser acumulados por pesquisadores.<sup>25</sup> (BORGMAN, 2012, p. 1069).

Assim, defende-se que principalmente os dados financiados com dinheiro público deveriam ser compartilhados. Torres-Salina (2010, *online*) vai além e argumenta a favor da obrigatoriedade de disponibilização de dados ao considerar que “se pode concluir que os pesquisadores não se animam em compartilhar de forma natural, o melhor seria atuar com políticas de decreto [...]”<sup>26</sup>

Objetivando impor que os pesquisadores deem mais importância a essas questões sobre o compartilhamento de dados, é sugerida a criação de políticas que beneficiem aqueles que adotarem a prática. Considerando o contexto profissional de um pesquisador, Borgman (2012, p. 1072) esclarece que

<sup>24</sup> Texto original: Ellas también tendrían que reclamar los datos generados que deberían poder ser utilizados por otros científicos del mismo sistema público. Asimismo se reciclarían todos esos proyectos que no alcanzaron los resultados esperados pero cuyos datos sí pueden ser de interés y, en última instancia, se podrían evitar investigaciones duplicadas ahorrándose un poco de dinero. (TORRES-SALINA, 2010, *online*)

<sup>25</sup> Texto original: Public sentiment for sharing research data is based largely on the rationale that tax monies should be leveraged to serve the public good. In this view, data produced with public funds should be available for use and should not be hoarded by researchers. (BORGMAN, 2012, p. 1069)

<sup>26</sup> Texto original: se puede concluir que los investigadores no se animan a compartir de forma natural, lo mejor es actuar con políticas de decretazo [ . . . ] (TORRES-SALINA, 2010, *online*)

Essa é uma afirmação simples de interesse próprio. Pesquisadores colaboram, mas eles também competem por bolsas, empregos, locais para publicação e alunos. Eles devem escolher cuidadosamente onde gastar seu tempo e recursos. Tempo e dinheiro gastos na documentação de dados para o uso por outros são recursos não gastos em coleta e análise de dados, equipamentos, taxas de publicação, conferências, escrita de artigos e propostas, ou outras necessidades de pesquisa. (BORGMAN, 2012, p. 1073)<sup>27</sup>

Iniciativas para a conscientização sobre a importância do compartilhamento de dados científicos têm partido de alguns periódicos científicos. Publicações renomadas como *Science* e *Nature* têm exigido em suas políticas de publicação que os pesquisadores compartilhem os dados de pesquisa dos estudos a ser publicados. A *Science*, por exemplo, oferece aos autores a possibilidade de publicar os dados em materiais adicionais que ficarão disponíveis no site da revista, juntamente ao artigo. (TORRES-SALINAS, ROBINSON-GARCIA E CABEZAS-CLAVIJO, 2012, p. 180).

Considerando a disponibilização de dados, os autores recomendam que ela seja de acesso livre. Torres-Salinas, Robinson-Gracia e Cabezas-Clavijo (2012, p. 178) exploram a lógica de que “os dados que provem de projetos de financiados com dinheiro público também são resultados da pesquisa e, da mesma forma que os artigos, devem ser públicos.”<sup>28</sup> Também a favor do acesso livre, Borgman (2012, p. 1071) enfatiza que, “quando os dados são compartilhados de forma rápida e aberta, os pesquisadores podem recorrer aos dados uns dos outros de forma mais fácil.”<sup>29</sup>

Entretanto, independentemente ao tipo de financiamento, uma avaliação prévia à disponibilização de dados deve ser realizada. Por exemplo: dados de cunho comercial, confidenciais ou que possuam direitos de propriedade intelectual, não podem fazer parte de uma política de acesso livre. (MARTÍNEZ-URIBE; MACDONALD, 2008, p. 278).

Dessa forma, a partir das considerações apresentadas nessa seção, é possível concluir que existem várias questões ainda não resolvidas referentes ao

<sup>27</sup> Texto original: This is a simple statement of self-interest. Researchers collaborate, but they also compete for grants, for jobs, for publication venues, and for students. They must choose carefully where to spend their time and resources. Time and money spent on documenting data for use by others are resources not spent in data collection, analysis, equipment, publication fees, conference travel, writing papers and proposals, or other research necessities. (BORGMAN, 2012, p. 1073)

<sup>28</sup> Texto original: “Los datos que provienen de proyectos financiados con dinero público también son resultado de la investigación y por tanto, al igual que los artículos, deben hacerse públicos.” (TORRES-SALINAS, ROBINSON-GRACIA E CABEZAS-CLAVIJO, 2012, P. 178)

<sup>29</sup> Texto original: When data are shared quickly and openly, researchers can draw upon each other’s data more readily. (BORGMAN, 2012, p. 1071)



compartilhamento de dados de pesquisa. Borgman (2012, p. 1059) pondera que “compartilhar dados de pesquisa é então um problema difícil e complexo – em outras palavras, um enigma”<sup>30</sup>, relacionado também com o tratamento necessário para disponibilizá-los. Além disso, é o foco de outro movimento, denominado curadoria de dados, que objetiva selecionar, tratar e preservar os dados de pesquisa para futuro compartilhamento.

Durante esse processo, profissionais da informação, como bibliotecários, arquivistas e museólogos, podem e devem se inserir. Habilidades de indexação e seleção, atividades presentes no dia a dia desses profissionais, são fundamentais para a execução da tarefa de curadoria.

Segundo Gray et al (2002, p. 5), os dados tornam-se incompreensíveis sem uma descrição detalhada a respeito do contexto do qual foram retirados. Assim, destaca-se a relevância da descrição detalhada, que pode ser entendida como o tratamento da informação, presente na curadoria para disponibilizar o acesso e preservar os dados de pesquisa. Na próxima seção, a curadoria de dados e sua relação com as práticas de Biblioteconomia e Ciência da Informação são exploradas.

## 4.2 CURADORIA DE DADOS

Matínez-Uribe e Macdonald (2008, p. 274) definem curadoria de dados como “uma gestão associada tanto à organização e preservação quanto a outras tarefas que podem agregar valor aos dados”.<sup>31</sup> De forma que estes possam ser descobertos e reutilizados, a curadoria compõe-se de atividades como a preservação digital, que, conforme Yakel (2007, p. 338), causa intenso interesse por representar uma necessidade atual. Em concordância com os autores acima, Shreeves e Cragin (2008, p. 93) adicionam que curadoria de dados consiste no

---

<sup>30</sup> Texto original: “Sharing research data is thus an intricate and difficult problem—in other words, a conundrum” (BORGMAN, 2012, p. 1059)

<sup>31</sup> Texto original: “una gestión asociada tanto a la organización y preservación como a otras tareas que puedan añadir valor a los datos.” (MATÍNEZ-URIBE; MACDONALD, 2008, p. 274)

[...] gerenciamento ativo e em andamento de dados por meio de seu ciclo de vida de interesse e utilidade para o conhecimento, a ciência e a educação, o que inclui apreciação e seleção, representação e organização desses dados para o uso e acesso com o tempo.<sup>32</sup> (SHREEVES E CRAGIN, 2008, p. 93)

Por sua vez, Yakel (2007, p. 338) utiliza a definição disponível no relatório *e-Science Curation*<sup>33</sup>, a qual segue abaixo:

A atividade de gerenciamento e promoção do uso de dados a partir de seu ponto de criação, de forma a garantir sua adequação para propósitos contemporâneos e sua disponibilidade para descobertas e reutilizações. Para conjuntos de dados dinâmicos, isso pode significar o enriquecimento e a atualização contínuos para manter sua adequação ao propósito. (YAKEL, 2007, p. 338)<sup>34</sup>

Nota-se que a curadoria se propõe a tratar dados para que sejam preservados e posteriormente disponibilizados, já existindo, inclusive, iniciativas dedicadas a estudar formas de depósitos de dados, possibilitando sua posterior recuperação.

O projeto “DISC-UK *DataShare*”<sup>35</sup> é uma dessas iniciativas, tendo como objetivo principal a elaboração de um modelo de depósito em repositórios institucionais (MARTÍNEZ-URIBE; MACDONALD, 2008, p. 277). Outra iniciativa é o projeto “Data Curation Profile”<sup>36</sup>, desenvolvido por pesquisadores das Universidade de Purdue e da Universidade de Illinois. Objetivando identificar as necessidades de pesquisadores em relação aos seus dados de pesquisa, foram realizadas entrevistas, pesquisas e observações de forma a elaborar perfis dos pesquisadores. Esses perfis apresentaram informações relacionadas às necessidades, ao acesso, ao uso e à reutilização dos dados de pesquisa por esses pesquisadores. (WITT et al, 2009).

Além de projetos, há também cursos de especialização em curadoria de dados nas Universidades de Illinois e da Carolina Norte. O primeiro é voltado à preservação dos dados científicos, ao passo que o último apresenta ênfase em

<sup>32</sup> Texto original: [ . . . ] the active and ongoing management of data through its lifecycle of interest and usefulness to scholarship, science and education which includes appraisal and selection, representation and organization of these data for access and use over time. (SHREEVES E CRAGIN, 2008, p. 93)

<sup>33</sup> Disponível em: <[www.jisc.ac.uk/publications/publications/pub\\_escience.aspx](http://www.jisc.ac.uk/publications/publications/pub_escience.aspx)>

<sup>34</sup> Texto original: The activity of managing and promoting the use of data from its point of creation, to ensure it is fit for contemporary purpose, and available for discovery and re-use. For dynamic datasets this may mean continuous enrichment or updating to keep it fit for purpose. (YAKEL, 2007, p. 338)

<sup>35</sup> Página do projeto: <http://www.disc-uk.org/datashare.html>

<sup>36</sup> Página do projeto: <http://datacurationprofiles.org/>

*cyberinfrastructure* e tecnologia. (TORRES-SALINAS, ROBINSON-GARCIA E CABEZAS-CLAVIJO, 2012, p. 181).

Em relação aos passos necessários para a realização de uma curadoria de dados, Yakel (2007, p. 338) nos apresenta cinco atividades fundamentais:

- a) Gestão do ciclo de vida / continuidade dos materiais talvez até chegar de volta para a criação do sistema de manutenção de registros;
- b) Envolvimento ativo, ao longo do tempo, de criadores de registros e potenciais curadores digitais;
- c) Avaliação e seleção de materiais;
- d) Desenvolvimento e fornecimento de acesso;
- e) Garantia de preservação (usabilidade e acessibilidade) de objetos.<sup>37</sup>

Essas etapas podem ser relacionadas a práticas presentes em bibliotecas e centros de informação, como a “Avaliação e seleção de materiais”, a “Garantia de preservação” e o “Desenvolvimento e fornecimento de acesso”. Alvaro et al (2011, *online*) afirmam que “bibliotecários podem ser importantes membros da equipe em projetos *e-Science*, porém as responsabilidades e as habilidades desejáveis de um bibliotecário *e-science* ainda estão abertas ao debate.”<sup>38</sup> Da mesma forma, não existe um parecer final sobre o papel de bibliotecários e profissionais da informação no auxílio à execução dessas atividades.

Considerando as funções e os papéis de profissionais da informação, pesquisadores e profissionais da TI, Lyon<sup>39</sup> (2007 apud MARTÍNEZ-URIBE; MACDONALD, 2008, p. 276) defende um trabalho conjunto:

<sup>37</sup> Texto original: (1) Life cycle/continuum management of the materials perhaps even reaching back to the creation of the record keeping system; (2) Active involvement over time of both the records creators and potentially digital curators ; (3) Appraisal and selection of materials; (4) Development and provision of access; (5) Ensuring preservation (usability and accessibility) of the objects. (YAKEL, 2007, p. 338)

<sup>38</sup> Texto original: Librarians can be important team members on e-Science projects, but both the responsibilities and desirable skills of an e-science librarian are still up for debate.” (ALVARO et al, 2011, online)

<sup>39</sup> Lyon, Liz. “Dealing with data—roles, rights, responsibilities and relationships” En: Informe de Ukoln para the Joint Information Systems Committee, 2007.

Para conservar os dados científicos para as gerações futuras é fundamental que as comunidades de pesquisa acadêmica (usuários e produtores de dados), os serviços de informática (que sabem gerir a tecnologia em organizações) e as bibliotecas (com suas habilidades em preservação e sua experiência em repositórios) trabalhem juntos.<sup>40</sup> (LYON, 2007 apud MARTÍNEZ-URIBE; MACDONALD, 2008, p. 276)

Compartilhando da mesma ideia, os autores Hey e Hey (2006, p. 515) determinam que

Para analisar, centralizar e extrair esses dados, será necessária colaboração entre cientistas e cientistas computacionais; para organizar, curar e preservar esses dados, esses dados requererão colaboração entre cientistas e bibliotecários.<sup>41</sup> (HEY; HEY, 2006, p. 515)

Buscando uma visão integrada, pode-se estabelecer que os pesquisadores, com seu conhecimento especializado, contribuem com dados/informações essenciais para a indexação, indicando os melhores termos a serem usados, e para a contextualização da origem dos dados. Já os profissionais da TI auxiliam no desenvolvimento, escolha e/ou manutenção de softwares, ao passo que os profissionais da informação contribuem com suas habilidades de indexação, preservação e disseminação.

Contudo, a curadoria ainda possui limitações. Dentre elas, Borgman (2012, p. 1073) menciona seu alto custo ao afirmar que

a liberação de dados é mais efetiva se tais dados são curados de maneiras que os tornem úteis a outros durante algum período longo de tempo. Da mesma forma, a curadoria de dados é muito cara e improvável de ser justificável para todas as formas de dados. Questões de seleção e de apreciação de forma a determinar quais dados valem a curadoria e por quanto tempo são assuntos urgentes que requerem muita mais atenção.<sup>42</sup> (BORGMAN, 2012, p. 1073)

---

<sup>40</sup>Texto original: Para conservar los datos científicos para las generaciones futuras es fundamental que las comunidades de investigación académica (usuarios y productores de datos), los servicios informáticos (que saben gestionar la tecnología en organizaciones) y las bibliotecas (con sus habilidades en preservación y su experiencia en repositorios) trabajen juntos. (LYON, 2007 apud MARTÍNEZ-URIBE; MACDONALD, 2008, p. 276)

<sup>41</sup>Texto original: To analyze, federate and mine this data will require collaboration between scientists and computer scientists; to organize, curate and preserve this data will require collaboration between scientists and librarians. (HEY; HEY, 2006, p. 515)

<sup>42</sup>Texto original: Data release is more effective if those data are curated in ways that make them useful to others over some long period of time. Data curation likewise is very expensive, and unlikely to be justifiable for all forms of data. Issues of selection and appraisal to determine which data are worth curating and for how long are urgent matters that require much more attention. (BORGMAN, 2012, p. 1073)

Portanto um dos grandes desafios de pesquisadores e profissionais de informação é elaborar alternativas que tornem a curadoria de dados uma atividade acessível aos grupos de pesquisa, além de realizável. Tanto a curadoria quanto o compartilhamento dos dados de pesquisa são atividades essenciais e emergenciais, e Lord e Macdonald<sup>43</sup> (2004 apud MARTÍNEZ-URIBE MACDONALD 2008, p. 273) advertem que “[...] se não forem tomadas medidas apropriadas de preservação, a velocidade vertiginosa com que a tecnologia avança põe em risco os dados digitais produzidos por atividades científicas.”<sup>44</sup> Nessa problemática, os bibliotecários e profissionais da informação adquirem um papel essencial.

Ao longo dessa seção, foi possível perceber e elencar algumas atividades fundamentais para o processo de curadoria de dados e relacioná-las com as áreas de Biblioteconomia e Ciência da Informação. Especificamente, partir das cinco etapas fundamentais no processo de curadoria de Yakei (2007, p. 338), destaca-se as atividades de “seleção e avaliação de materiais”, também inclusas dentro das atividades de Desenvolvimento de Coleções de bibliotecas e centros de informação e, ainda, as atividades de “desenvolvimento e fornecimento de acesso”, as quais estão relacionadas com o processo de Organização da Informação, composto por atividades como indexação e recuperação de objetos informacionais.

Dessa forma, no contexto do compartilhamento e da curadoria de dados, buscaram-se processos e instrumentos de organização da informação, a partir de abordagens da Biblioteconomia e da Ciência da Informação, a serem utilizadas para subsidiar o processo de curadoria de dados, o que é meta da seção a seguir.

---

<sup>43</sup>Lord, Philip; Macdonald, Alison. “Digital data—a new knowledge based research”. En: Informe del Joint Information Systems Committee, 2004.

<sup>44</sup>Texto original: “[...] si no se toman medidas apropiadas de preservación, la velocidad de vértigo con la que avanza la tecnología pone en riesgo los datos digitales producidos por actividades científicas”. (LORD E MACDONALD, 2004 apud MARTÍNEZ-URIBE E MACDONALD 2008, p. 273)

## 5 ORGANIZAÇÃO DA INFORMAÇÃO

Anterior à inserção das tecnologias da informação e da comunicação no contexto da Biblioteconomia e da Ciência da Informação, o tratamento e a organização de bibliotecas eram centralizados no documento, “objeto que fornece um dado ou uma informação e pode ser diferenciado entre outros documentos, de acordo com suas características físicas ou intelectuais.” (FEITOSA, 2006, p. 17). Porém a utilização dessa terminologia tornou-se restritiva, devido aos novos suportes, tanto no meio físico quanto no digital, disponibilizados por tais tecnologias. Considerando essa nova demanda, Victorino e Bräscher (2009, grifo do autor) apontam que, “no intuito de usar uma expressão mais ampla que ‘documento’ para referenciar estes suportes e respectivos conteúdos, emprega-se a denominação *objetos informacionais*”.

Em concordância, Bräscher e Café definem organização da informação como

[...] um processo que envolve a descrição física e de conteúdo dos objetos informacionais. O produto desse processo descritivo é a representação da informação, entendida como um conjunto de elementos descritivos que representam os atributos de um objeto informacional específico. (BRASCHER; CAFÉ, 2008, p. 5)

Além disso, cabe acrescentar que, conforme Alvarenga (2003, p. 29), a organização da informação é estabelecida com o objetivo de oportunizar aos usuários uma recuperação eficaz, ou seja, uma recuperação de qualidade, que traga resultados pertinentes ao que o usuário precisa. Logo, cabe discutir possíveis formas pelas quais os profissionais da informação podem otimizá-la. Sobre isso, a autora sugere que são necessários

[...] sistemas representacionais que estabeleçam a confluência entre a organização cognitiva imposta ao conhecimento pelo seu produtor (representação primária) e a organização conceitual imposta ao documento pelo especialista da informação (representação secundária). A confluência entre essas duas estruturas de representação constitui-se no componente central do trabalho dos arquivistas, bibliotecários e demais profissionais da informação, desde que não seja esquecida a representação que ocorre em nível dos usuários dos sistemas de informação. (ALVARENGA, 2003, p. 29).

Victorino e Brächer (2009, *online*) vão além ao afirmar que outro “objetivo da organização da informação é dar suporte ao fluxo de tratamento e recuperação dos objetos informacionais estruturados, semi-estruturados e não estruturados.” A diferenciação entre tipos de objetos informacionais está relacionada com a sua estrutura de armazenamento. Dessa forma, Dittrich e Domenig<sup>45</sup> (1999 apud VICTORINO, BRÄSCHER, 2009) esclarecem este contexto ao apontar que

[...] os objetos informacionais estruturados possuem uma estrutura rígida de armazenamento (planilhas eletrônicas, tabelas relacionais, etc.), os semi-estruturados possuem estrutura de armazenamento, porém esta não é rígida (e-mail, páginas HTML, etc.) e os não estruturados não possuem estrutura de armazenamento definida além de sequências de bytes ou caracteres (imagens, textos, etc.). (DITTRICH E DOMENIG, 1999 apud VICTORINO, BRÄSCHER, 2009)

Considerando a geração de informação por diferentes áreas científicas, cada vez mais dinâmica e abrangente, questionam-se os efeitos dessa nas formas utilizadas até então para o tratamento da informação. Segundo Vickery<sup>46</sup> (1986 apud ALVARENGA, 2003, p. 35), um dos resultados é a multidisciplinaridade de áreas e dos consequentes profissionais responsáveis pelo tratamento, expandindo a equipe de trabalho para além dos campos da Biblioteconomia e da Ciência da Informação. Alvarenga ressalta que o objetivo em comum dos diferentes profissionais deve ser de “resolver problemas na área da formação e representação de conceitos”. (ALVARENGA, 2001, *online*).

O advento de novas tecnologias, no entanto, permitiu que novas possibilidades de trabalhar com a organização da informação. Para Alvarenga, o mundo digital

[...] ocasionou novas mudanças no trabalho de autores e profissionais da informação, fazendo com que estes se envolvessem com novas possibilidades tecnológicas, diretamente incidentes nos processos de produção, armazenagem, representação e recuperação de documentos e informações, alterando seus processos de trabalho e produtos finais. (ALVARENGA, 2003, p. 34).

---

<sup>45</sup> DITTRICH, K., DOMENIG, R. Towards **Exploitation of the Data Universe**. In: 3RD INTERNATIONAL CONFERENCE ON BUSINESS INFORMATION SYSTEM, abril 1999.

<sup>46</sup> VICKERY, D.C. **Knowledge representation**: a brief review. *Journal of documentation*, v. 42, n. 3, set. 1986. p. 145-59.

Nessa seção exploraram-se os conceitos de diferentes autores sobre a organização da informação. É sabido que abordagens presentes na Biblioteconomia e Ciência da Informação tratam dos processos de produção, armazenagem, representação e de recuperação, os quais são apresentados na próxima seção.

### 5.1 ABORDAGENS DA BIBLIOTECONOMIA E CIÊNCIA DA INFORMAÇÃO PRESENTES NA ORGANIZAÇÃO DA INFORMAÇÃO

De forma a descrever objetos informacionais, é necessária a aplicação de abordagens da Biblioteconomia e da Ciência da Informação. Esta seção discorre sobre elas, definindo-as e apresentando seus papéis em meio ao processo da organização da informação.

Primeiramente, deve-se atentar à etapa de tratamento intelectual da informação, que Dias (2006, p. 67) define como “a função de descrever os documentos, tanto do ponto de vista físico (características físicas dos documentos) quanto do ponto de vista temático (ou de descrição do conteúdo)”. Essa descrição de conteúdo é realizada por intermédio da técnica de indexação, definida por Robredo e Cunha (1986, p. 244) como um processo intelectual cujo objetivo é “[...] indicar o conteúdo temático de uma unidade de informação, mediante a atribuição de um ou mais termos (ou códigos) ao documento, de forma a caracterizá-lo de forma unívoca.” van der Laan (2002, p. 11) adiciona que se trata de um “[...] processo de identificação e organização dos itens necessários a posterior recuperação das informações contidas em um documento, observando-se que os principais pontos de acesso para a recuperação de informação são autor, título e assunto”.

Em relação ao propósito de um serviço de indexação, Carneiro (1985, p. 221) afirma que é “[...] assegurar da forma mais eficiente e econômica possível, que qualquer documento ou informação seja fornecido ao usuário no momento preciso”. Porém, para que haja tal garantia é relevante compreender os requisitos desse tipo de serviço, sobre os quais Robredo e Cunha (1986, p. 249) explanam abaixo.

A indexação é uma operação que requer tempo e exige conhecimentos adequados por parte do indexador. Consequentemente, é uma operação cara. Convém observar, entretanto, que o tempo gasto na indexação é, normalmente, tempo ganho no momento da busca e da recuperação. (ROBREDO ; CUNHA, 1986, p. 249)



As atividades inerentes a esse processo podem ser divididas, de acordo com Robredo e Cunha (1986, p. 244-245) em:

- a) análise conceitual do conteúdo significativo do documento, ou seja, a identificação do assunto;
- b) expressão dessa análise por meio de um conjunto de palavras, frases ou códigos que representam o assunto;
- c) tradução das descrições dos assuntos relevantes para a linguagem de indexação; e,
- d) organização das descrições padronizadas dos assuntos de acordo com a sintaxe da linguagem de indexação.

Todas essas atividades devem ser gerenciadas por uma visão comum, isto é, a busca pela satisfação das necessidades de usuários potenciais.

Por outro lado, Lancaster divide o processo de indexação em duas etapas. A primeira, denominada “análise conceitual”, a qual implica em “[...] decidir do que trata um documento, isto é, qual é o seu assunto” (1993, p. 8). Ou seja, é feita a identificação de conteúdos tratados em determinada obra, documento ou objeto. A segunda, intitulada “tradução”, é definida como a atividade que “[...] envolve a conversão da análise conceitual de um documento num determinado conjunto de termos de indexação.” (1993, p. 13). Nessa etapa, a indexação pode ser feita de duas formas: por extração ou por atribuição. Na primeira, “[...] palavras ou expressões que realmente ocorrem num documento são selecionadas para representar seu conteúdo temático.” (LANCASTER, 1993, p. 13). A segunda envolve “[...] a atribuição de termos a um documento a partir de uma fonte que não é o documento.” (LANCASTER, 1993, p. 14). No último caso o indexador poderá selecionar termos considerados relevantes, mesmo que esses não estejam explicitados no documento através de seu conhecimento sobre área ou utilizar um vocabulário controlado.

Esses termos escolhidos para representar o documento na indexação podem ser compreendidos, segundo Feitosa (2006, p. 21), como “[...] uma espécie de mini-resumo, uma vez que reúne os termos em uma lista ou índice”. Sendo assim, as abordagens de indexação e de elaboração de resumos estão conectadas entre si. Lancaster (1993, p. 5) reforça essa interdependência ao assegurar que “[...] a

indexação de assuntos e a redação de resumos são atividades intimamente relacionadas, uma vez que ambas implicam a preparação de uma representação do conteúdo temático dos documentos.”

Logo, a indexação resumirá um objeto informacional utilizando termos relevantes e representantes de seu conteúdo. De forma paralela, o resumo representará o conteúdo do objeto por meio de um texto. Devido à existência de diferentes tipos de resumos, a escolha por um dependerá da informação que se pretende fornecer ao usuário final. De acordo com a ABNT NBR 6028 (2003, p. 1) há três tipos de resumos, esses podendo ser (i) críticos, (ii) indicativos ou (iii) informativos. O primeiro é um resumo feito por especialista que realiza uma análise crítica sobre um documento. O segundo indica apenas os pontos principais de um documento, sem dispensar a consulta ao original. E o terceiro informa a finalidade, a metodologia, o resultado e as conclusões do documento, dispensando até a consulta ao original. Contudo, indiferentemente de sua tipologia, o resumo tem por finalidade, conforme Lancaster (1993, p. 91), facilitar o processo de seleção de um objeto, além de auxiliar o leitor a decidir se um item específico atende às suas necessidades.

Outra abordagem pertencente ao processo de organização da informação é a de classificação. Robredo e Cunha (1986, p. 204) a definem como o “[...] processo mental de designar e ordenar o universo do conhecimento, para determinar, com base em suas relações o lugar preciso e exato das coisas, num esquema organizado.” Em análise mais ampla, Langridge acrescenta que

[...] sem classificação não poderia haver nenhum pensamento humano, ação e organização que conhecemos. A classificação transforma impressões sensoriais, isoladas e incoerentes em objetos reconhecíveis e padrões recorríveis. (LANGRIDGE, 1977, p. 11)

Sendo a classificação tão relevante para o pensamento humano, o contexto de sua utilização por profissionais em unidades de informação, constituído pelo seu armazenamento e pela sua recuperação, não ocorreria de forma diferenciada. Lancaster (2003, p. 17) afirma que a classificação de assuntos compreende na atividade de “[...] formar classes de objetos com base em seu conteúdo temático.” Essa atividade nada mais é do que a indexação de assuntos, tópico essencial para que seja possibilitada a recuperação de objetos informacionais a partir de seu assunto.

Em relação à recuperação da informação, a abordagem possui a função orgânica de “[...] permitir que o usuário localize o maior número possível de itens relevantes.” (FEITOSA, 2006, p. 28). A relação entre o total de itens úteis para um usuário e o total de itens recuperados é denominada coeficiente de precisão. Esse coeficiente foca na capacidade do sistema de relacionar informações que serão consideradas relevantes pelo usuário. De forma contrária, denomina-se coeficiente de revocação aquele que relaciona o total de itens úteis existentes na base de dados e os itens úteis recuperados pelo usuário em determinada busca. Dessa forma, Feitosa (2006, p. 28) conclui que

o índice de revocação e o coeficiente de precisão são grandezas inversamente proporcionais. Portanto, quanto menor for o índice de revocação, mais preciso será o resultado da pesquisa, isto é, maior será a quantidade de itens úteis recuperados. (FEITOSA, 2006, p. 28)

Um dos fatores determinantes para um menor índice de revocação é a especificidade escolhida no momento da indexação (CARNEIRO, 1985, p. 234). Portanto, conforme menção anterior é preferível um maior investimento de tempo durante o processo de indexação para que a recuperação da informação ocorra de forma mais eficiente. Além disso, Feitosa (2006, p. 28) defende que resultados mais satisfatórios dependerão da “[...] política de indexação utilizada, das regras usadas para a redação do resumo, da qualidade do vocabulário controlado, da quantidade das estratégias de busca, entre outros fatores.”

Por fim, outra abordagem dentro do processo de organização e recuperação de objetos digitais é a inserção dos metadados, uma vez que todos os processos descritos até agora dependem, essencialmente, da criação de descritores dos conteúdos como uma maneira de formalização. Segundo Alvarenga, metadado significa, etimologicamente, “dado sobre dado” e descreve “[...] a essência, atributos e contexto de emergência de um recurso (documento, fonte, etc.) e caracteriza suas relações, visando-se ao acesso e ao uso potencial.” (ALVARENGA, 2003, p. 36). Para Feitosa (2006, p. 50-51), metadado é “[...] todo dado físico, isto é, contido em algum software ou qualquer tipo de mídia, e que fornece informações sobre a informação.” Percebe-se que estes são essenciais no contexto de dados digitais, uma vez que “[...] podem ser indicados diretamente nos próprios objetos,

constituindo-se esses em chaves de acesso a documentos e conteúdos.” (ALVARENGA, 2006, p. 80).

Relacionando-os à área da Biblioteconomia, Feitosa (2006, p. 52) considera que os metadados são como um “[...] dado estruturado, que compartilha diversas características similares para a catalogação, que descreve as características de um determinado recurso informacional.” Dessa forma, os metadados têm papel fundamental na descrição e facilitam a recuperação de objetos informacionais. Em conformidade, Alvarenga (2003, p. 36) compara o metadado com o “*ponto de acesso*, termo da área da catalogação bibliográfica, e parece ter sido cunhado em contextos externos à ciência da informação.” Alguns exemplos de esquemas de metadados na área da Ciência da Informação são: Dublin Core, AACR2 (Código de Catalogação Anglo-Americano) e GILS (Government Information Locator Service). (FEITOSA, 2006, p. 53).

Além de ponto de acesso, os padrões de metadados (XML, *Dublin Core* etc.) podem ser considerados linguagens de marca, que, segundo Bax (2001, p. 32-33),

[...] identificam, de forma descritiva, cada ‘entidade informacional’ digna de significado presente nos documentos, como por exemplo parágrafos, títulos, tabelas, gráficos. A partir dessas descrições os programas de computador podem melhor tratar ou processar a informação contida em documentos eletrônicos. (BAX, 2001, p. 32-33)

A inserção dessas linguagens de marca, de acordo com Alvarenga (2003, p. 37), pode ser atribuída *a priori*, ou seja, durante a etapa de criação do documento pelos próprios autores, ou *a posteriori*, após a inclusão do objeto no espaço digital. A inserção de metadados na etapa de criação de objetos informacionais poderá agilizar a indexação desse material posteriormente, caso seja indexado em uma base de dados ou em algum tipo de repositório. Para a autora, essa é uma “importante na discussão sobre a representação e recuperação de documentos constantes na web.” Os elementos constantes nos metadados estão associados a propriedades como “nome, domínio, tipo, tamanho, formato, responsabilidade e, no caso de sistemas informatizados, origem.” (FEITOSA, 2006, p. 52). Ao finalizar, é relevante mencionar Alvarenga (2001, *online*) ao advertir que

[...] os metadados não se restringem a dados temáticos, constantes no texto mas, podem também incluir outros tipos de pontos de acesso, relacionados à descrição física e ao contexto de produção, dependentes do tipo de objeto em questão. (ALVARENGA, 2001, *online*)

Considerando as reflexões apresentadas nessa seção, compreende-se que a organização da informação é um processo que se utiliza de uma série de abordagens presentes nas áreas da Biblioteconomia e da Ciência da Informação, como: indexação, elaboração de resumos, recuperação da informação, classificação e utilização de metadados. Apresentaremos, na próxima seção, alguns instrumentos utilizados por estas áreas e que se utilizam das abordagens tratadas acima, como cabeçalhos de assunto, taxonomias, tesouros e ontologias.

## 5.2 INSTRUMENTOS DE ORGANIZAÇÃO E REPRESENTAÇÃO DA INFORMAÇÃO

Nessa seção serão apresentados os instrumentos conhecidos como vocabulários controlados. Além de serem utilizados no processo de organização e de representação da informação, também são fundamentais nas abordagens de indexação e de recuperação de objetos informacionais. Segundo Currás (1995, p. 78), os vocabulários controlados podem ser compreendidos como uma “[...] linguagem na qual os elementos (palavras) que a compõem são controlados de acordo com regras preestabelecidas”.

Essas regras são instauradas para que haja um controle não somente na utilização de sinônimos e de palavras homógrafas, mas também no agrupamento de termos afins (LANCASTER, 1993, p. 15). Dessa forma, o resultado dessas atividades interferirá diretamente na precisão da recuperação da informação, em vista que, em sistemas automatizados, “[...] qualquer diferença na grafia das palavras significa, para o computador, palavras diferentes.” (ROBREDO E CUNHA, 1986, p. 223)

Para atender necessidades específicas, existem diferentes tipos de vocabulários controlados. Primeiramente, há listas de cabeçalho de assunto, as quais, para Medeiros (2012, p. 21), consistem em

[...] uma lista de termos em linguagem controlada, referentes a um determinado domínio do conhecimento, onde esses termos estão relacionados de forma a demonstrar equivalência e onde o termo não-preferido contém uma remissiva que indica o termo preferido, além de ter como objetivo a indexação de assuntos existentes em documentos.” (MEDEIROS, 2012, p. 21)

Porém, devido a sua baixa complexidade, existem algumas limitações em sua utilização, exemplificadas por Hodge (2000, p. 6 apud MEDEIROS, 2012, p. 22) ao mencionar casos nos quais as listas de cabeçalhos de assuntos “[...] são geralmente muito superficiais, com uma estrutura hierárquica limitada”. Em concordância, Batista (2004, p. 7) complementa que essas listas “[...] não atingem o nível de especificidade desejável na representação do conteúdo temático de documentos [...]”. Assim, para algumas instituições e alguns sistemas mais especializados, é recomendada a utilização de outros tipos de vocabulário controlado, tais quais as taxonomias.

Martinez et al. (2004, p. 106) explicam que,

[...] em um sentido amplo, é a criação da estrutura (ordem) e dos rótulos (nomes) que ajudam a localizar a informação relevante. Em um sentido mais específico, é o ordenamento e rotulação de metadados, que permite organizar sistematicamente a informação primária. (MARTINEZ et al., 2004, p.106)

Além disso, Campos e Gomes (2007, *online*) acrescentam que as taxonomias são “estruturas classificatórias que têm por finalidade servir de instrumento para a organização e recuperação de informação em empresas e instituições”. Logo, apresentam grande relevância em sistemas automatizados e colaborativos por possibilitar “[...] ao usuário fazer parte do processo de construção, apresentando uma estrutura flexível, respondendo as necessidades de informação de ambientes específicos, com vistas à recuperação eficaz.” (VITAL, 2012, p. 38)

Além das listas de cabeçalhos de assunto e das taxonomias, os tesauros constituem um outro tipo de vocabulário controlado. Mais complexo que os anteriores, são descritos por Motta (1987, p. 24) como um

[...] sistema de vocabulário baseado em conceitos, incluindo termos preferidos (descritores), termos não preferidos (não descritores) e suas inter-relações, que se aplica a um determinado ramo do conhecimento e que se destina a controlar a terminologia utilizada para a indexação/recuperação de documentos. (MOTTA, 1987, p. 24).

É possível notar similaridades entre os tipos de vocabulários mencionados acima, como, por exemplo, o objetivo principal, que é a recuperação da informação. No entanto, os cabeçalhos de assunto possuem uma estrutura hierárquica imperfeita e não distinguem claramente as relações entre termos. Por tal motivo, em contextos de instituições e sistemas mais especializados, a utilização de tesouros pode ser uma alternativa mais adequada e eficaz. Medeiros (2012, p. 31) explana que a função deles é “[...] controlar terminologicamente a tradução de conceitos da linguagem natural para uma linguagem artificial, permitindo a representação e a recuperação da informação de uma determinada área do conhecimento humano.” De forma complementar, a UNESCO, através de um programa denominado UNISIST, define tesouro como

Um dispositivo de controle terminológico usado para traduzir a linguagem natural dos documentos, dos indexadores ou dos usuários numa mais restrita *linguagem do sistema* (linguagem documentária, linguagem de informação), bem como para traduzir a linguagem do sistema de volta em linguagem natural. No que tange à sua estrutura, o tesouro é um vocabulário controlado e dinâmico de termos semântica e genericamente relacionados que cobrem um domínio específico do conhecimento. (UNESCO, 1973 apud MEDEIROS, p. 28, grifo do autor)

Em plano de superior complexidade aos tesouros, encontram-se as ontologias. Etimologicamente, o “[...] termo está ligado ao verbo grego *ontos* (ser), como prefixo, e a *logos* (estudo), como sufixo, podendo ser traduzida como o *estudo da existência do ser*” (MEDEIROS, 2012, p. 51, grifo do autor). Ademais, o autor estabelece que é o “[...] estudo da existência das coisas a partir de suas categorias, podendo ser utilizada na metafísica para tentar responder a seguinte questão: que coisas existem?”.

Em descrição mais concisa, Guarino (1998, *online*) explora a ontologia como uma

[..] teoria lógica correspondente ao significado pretendido de um vocabulário formal, ou seja, seu compromisso ontológico com uma conceitualização particular de mundo. Os modelos pretendidos de uma linguagem lógica usada como vocabulário são restritos pelo seu compromisso ontológico. Uma ontologia indiretamente reflete esse compromisso (e a conceitualização subjacente) pela aproximação desses modelos pretendidos. (GUARINO, 1998, *online*)<sup>47</sup>

As razões para a elaboração de ontologias ultrapassam as características dos vocabulários controlados. Noy e MacGuinness<sup>48</sup> (2004 apud FEITOSA, 2006, p. 73) as listam como: (a) compartilhar um entendimento comum da estrutura da informação entre humanos e agentes inteligentes; (b) permitir o reuso do domínio do conhecimento; (c) tornar explícitos os pressupostos de um domínio; (d) separar o conhecimento sobre o domínio do conhecimento operacional; e (e) analisar o conhecimento sobre o domínio.

Em relação a utilização, Feitosa (2006, p. 73) alerta que

Uma ontologia não deve ser concebida apenas como um vocabulário informal, ou mesmo como uma linguagem de termos estruturados – como um tesouro, por exemplo -, mas requer uma possibilidade de interpretação algorítmica dos seus significados e, por conseguinte, uma representação em uma linguagem formal, cujo procedimento dos significados pode ser realizado por máquinas. (FEITOSA, 2006, p. 73)

A partir das considerações apresentadas nessa seção, é possível atentar à existência de instrumentos essenciais para o processo de organização da informação. Apesar de possuírem características próprias e níveis distintos de complexidade, o seu principal objetivo é controlar a utilização de termos na indexação, possibilitando uma melhor recuperação da informação. Porém, as ontologias vão além dessas questões, com o objetivo de permitir que o processamento de significados seja realizado por computadores.

No próximo capítulo serão apresentadas as considerações finais do trabalho, relacionando os movimentos da *e-Science* de compartilhamento e de curadoria de

---

<sup>47</sup> Texto original: logical theory accounting for the intended meaning of a formal vocabulary, i.e. its ontological commitment to a particular conceptualization of the world. The intended models of a logical language using such a vocabulary are constrained by its ontological commitment. An ontology indirectly reflects this commitment (and the underlying conceptualization) by approximating these intended models. (GUARINO, 1998, *online*)

<sup>48</sup> NOY, F. N.; GUINNESS, D.L. *Ontology development 101: a guide to create your first ontology*. 2001.



dados de pesquisa com as abordagens da Biblioteconomia e da Ciência da Informação, ambas presentes no processo de Organização da Informação.

## 6 CONSIDERAÇÕES FINAIS

Após a busca e a leitura do referencial a respeito da *e-Science*, especificamente sobre a curadoria e o compartilhamento de dados de pesquisa, é possível constatar a importância desses movimentos, não apenas para a comunidade científica, mas também para a sociedade como um todo. É preocupante o número de dados que poderiam ser aproveitados para o desenvolvimento de novas pesquisas que acabam inutilizados e perdidos. Frente a tal problemática, tanto periódicos científicos quanto revistas de circulação comercial têm publicado sobre o tema com o intuito de pensar em caminhos e soluções.

A curadoria e o compartilhamento de dados de pesquisa são abordagens interdependentes. Enquanto a primeira foca no tratamento e na preservação, a segunda se concentra no compartilhamento e na disponibilização desses dados. Porém, não há sentido em tratar e preservar se não será dado acesso a esses dados. Dessa forma, para que se consiga compartilhar é necessário o estabelecimento de meios de descrição desses objetos, acarretando em uma melhor recuperação. Conforme explorado no referencial teórico, há pesquisadores que compartilham seus dados de modo informal, sem qualquer tipo de tratamento e padronização. No entanto, em casos assim, o reaproveitamento posterior será muito restrito ou até inacessível por grande parte de possíveis interessados.

Ao pensarmos em uma Unidade de Informação, seja bibliotecas, arquivos ou museus, seu objetivo central é selecionar, tratar e preservar objetos, disponibilizando-os para um público. Concomitantemente, a curadoria e o compartilhamento de dados de pesquisa trabalham com os mesmos objetivos.

Anteriormente, o tratamento da informação na Biblioteconomia era voltado ao suporte “documento”, de forma que, em uma perspectiva atual, com a inserção das tecnologias da informação e comunicação, a Biblioteconomia e a Ciência da Informação trabalham com objetos em diferentes formatos. Assim, autores contemporâneos destas áreas defendem a utilização do termo “objeto informacional”, enquadrando uma ampla variedade de suportes, físicos ou digitais. Dessa forma, os dados produzidos em pesquisas podem ser tratados no contexto da Biblioteconomia e da Ciência da Informação como um objeto informacional.

A partir da indexação, é possível fazer uma descrição desses dados, com o auxílio de metadados, que irão descrever informações como contexto de criação,

local, período, etc. Com a possibilidade de realizar inserções no momento da criação dos dados ou posteriormente, os metadados exercem papel fundamental no contexto de objetos digitais, permitindo que seja estabelecida uma descrição detalhada desses objetos, provendo uma melhor recuperação. Percebe-se, assim, a importância do envolvimento de profissionais da informação nesse processo, estabelecendo políticas que atenderão às necessidades de acordo com seu contexto.

A indexação tem como caráter resumir as características de um objeto a partir da inserção de termos significativos. Sendo assim, para que haja uma melhor recuperação desses dados, é fundamental que, durante o processo de indexação, sejam utilizados instrumentos de vocabulário controlado. Para que a inserção de termos seja padronizada, pode-se utilizar listas de cabeçalhos de assunto, taxonomias e tesouros ou até, em casos que exijam grande processamento automático de dados e alta complexidade, a utilização de ontologias, sendo que a escolha de um tipo específico de vocabulário controlado dependerá do contexto no qual está sendo trabalhado. Por mais que existam níveis de complexidade entre esses instrumentos, nem sempre o mais complexo será a melhor opção a ser utilizada, pois o sistema de recuperação da informação pode requerer funções mais simples.

Conclui-se que existe uma gama de abordagens provindas da Biblioteconomia e da Ciência da Informação que pode contribuir com a curadoria e com o compartilhamento de dados. Contudo, na condição de que sejam mediados por profissionais qualificados da área para que sejam definidos instrumentos adequados para cada especificação. Além disso, projetos de curadoria e posterior compartilhamento devem contar com o trabalho conjunto entre profissionais da informação, profissionais da ciência da computação, pesquisadores e produtores dos dados a serem tratados.

Como apontado pelos autores estudados, ainda existem diversos problemas que não foram resolvidos e sobre os quais não se tem certeza sobre como proceder, requerendo estudos mais aprofundados. A curadoria e todas as atividades que envolvem o seu processo demandam tempo e alto custo para os grupos de pesquisa. Além de envolver profissionais de áreas diferentes, em alguns casos, existem questões éticas e de direito autoral que impedem a disponibilização. Essas limitações, porém, não podem ser fatores para desmotivação, visto o número de

benefícios a serem agregados a partir do compartilhamento e da curadoria de dados. A solução é que novos estudos continuem sendo feitos e que cada vez mais os profissionais da informação se insiram nessas pesquisas. Como consequência dessas atitudes, a Ciência da Informação e a Biblioteconomia, a partir das abordagens presentes na Organização da Informação, têm muito a contribuir para a curadoria e para o compartilhamento dos dados de pesquisa. É fundamental que, além disso, estudos nacionais que se aprofundem sobre o tema.

Por fim, espera-se que com esse trabalho, novos estudos focados na *e-Science* sejam realizados, especificamente exploradores da curadoria e do compartilhamento de dados, principalmente a partir de uma perspectiva nacional e de profissionais da informação, os quais podem contribuir significativamente em equipes de projetos de curadoria de dados.

## REFERÊNCIAS

- ALVARENGA, L. A Teoria do Conceito revisada em conexão com ontologias e metadados no contexto das bibliotecas tradicionais e digitais. **DataGramZero**, Rio de Janeiro, v. 2, n. 6, dez. 2001. Disponível em: <[http://dgz.org.br/dez01/Art\\_05.htm](http://dgz.org.br/dez01/Art_05.htm)>. Acesso em: 2 nov. 2012
- \_\_\_\_\_. Representação do conhecimento na perspectiva da Ciência da Informação em tempo e espaço digitais. **Encontros Bibli**, Florianópolis, v. 8, n. 15, 2003. Disponível em: <<http://www.periodicos.ufsc.br/index.php/eb/article/view/1518-2924.2003v8n15p18>>. Acesso em: 2 nov. 2012.
- \_\_\_\_\_. Organização da informação nas bibliotecas digitais. In: KURAMOTO, H.; NAVES, M. M. L. (Orgs.). **Organização da Informação: princípios e tendências**. Brasília, DF: Bricquet de Lemos, 2006.
- ALVARO, E. et al. E-Science Librarianship: Field Undefined. **Issues in Science and Technology Librarianship**, v. ?, n. 63, 2011. Disponível em: <<http://www.istl.org/11-summer/article1.html>>. Acesso em: 28 out. 2012.
- ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. NBR 6028: Informação e documentação: Resumo. Rio de Janeiro, 2003.
- BATISTA, G. H. R.. Redes de Conceitos. **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 9, n. 1, p. 6-17, jan./jun. 2004. Disponível em: <<http://www.brapci.ufpr.br/documento.php?dd0=0000006658&dd1=5a879>>. Acesso em: 02 nov. 2012
- BAX, M. Introdução às linguagens de marca. **Ciência da Informação**, Belo Horizonte, v. 30, n. 1, p. 32-38, jan./abr. 2001. Disponível em: <<http://www.scielo.br/pdf/ci/v30n1/a05v30n1.pdf>>. Acesso em: 28 out 2012.
- BORGMAN, C. L. The Conundrum of Sharing Research Data. **Journal of the American Society for Information Science and Technology**, v. 63, n. 6, p. 1059-1078, jun. 2012. Disponível em: <<http://onlinelibrary.wiley.com.ez45.periodicos.capes.gov.br/doi/10.1002/asi.22634/pdf>>. Acesso em: 17 jun. 2012.
- BRÄSCHER, M.; CAFÉ, L. Organização da Informação ou Organização do Conhecimento? In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 9., 2008, São Paulo. **Anais...** São Paulo: Ancib, 2008. Disponível em: <<http://www.ancib.org.br/media/dissertacao/1835.pdf>>. Acesso em: 1 nov. 2012.
- CAMPOS, M. L. A.; GOMES, H. E. Taxonomia e Classificação: a categorização como princípio. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 8., 2007, Salvador. **Anais...** Salvador: Ancib, 2007. Disponível em: <<http://www.enancib.ppgci.ufba.br/artigos/GT2--101.pdf>>. Acesso em: 02 nov. 2012.
- CARNEIRO, M. V. Diretrizes para uma política de Indexação. **Revista da Escola de Biblioteconomia da UFMG**, Belo Horizonte, v. 14, n. 2, p. 221-241, set. 1985.

Disponível em:

<<http://www.brapci.ufpr.br/documento.php?dd0=0000002649&dd1=5dba2>> . Acesso em: 28 out. 2012

CRAGIN, M. et al. Data sharing, small science and institutional repositories. **Philosophical Transactions of the Royal Society**, Londres, v. 368, n. 1928, p. 4023-4038, out. 2010.

CURRÁS, E. **Tesauros**: linguagens terminológicas. Brasília, DF: IBICT, 1995.

DIAS, E. W. Organização do Conhecimento no contexto de bibliotecas tradicionais e digitais. In: KURAMOTO, H.; NAVES, M. M. L. (Orgs.). **Organização da Informação**: princípios e tendências. Brasília, DF: Briquet de Lemos, 2006.

FEITOSA, A. **Organização da Informação na Web: das tags à web semântica**. Brasília, DF: Thesaurus, 2006.

GIL, A. C. **Como elaborar projetos de pesquisa**. 4. ed. São Paulo: Atlas, 2002.

GRAY, J. Jim Gray on eScience: a transformed scientific method. HEY, T.; TANSLEY, S.; TOLLE, K. (ed.). **The fourth paradigm**: data-intensive scientific discovery. Redmont: Microsoft Research, 2009. Disponível em: <[http://research.microsoft.com/en-us/collaboration/fourthparadigm/4th\\_paradigm\\_book\\_jim\\_gray\\_transcript.pdf](http://research.microsoft.com/en-us/collaboration/fourthparadigm/4th_paradigm_book_jim_gray_transcript.pdf)> . Acesso em: 20 out. 2012.

GRAY, J.; SZALAY, A. S.; THAKAR, A. R.; STOUGHTON, C.; VANDENBERG, J. **Online Scientific Data Curation, Publication, and Archiving**. Redmond: Microsoft Research, 2002. Disponível em: <<http://research.microsoft.com/pubs/64568/tr-2002-74.pdf>>. Acesso em: 16 out. 2012.

GUARINO, N. Formal Ontology and Information Systems. In: FORMAL ONTOLOGY IN INFORMATION SYSTEMS, 1998, Trento. **Anais...** Amsterdam: IOS Press, 1998. Disponível em: <<http://www.loa.istc.cnr.it/Papers/FOIS98.pdf>>. Acesso em: 02 nov. 2012

HEY, T.; HEY, J. e-Science and its implications for the library community. **Library Hi Tech**, v. 24, n. 4, p. 515-528, 2006. Disponível em: <<http://www.emeraldinsight.com/journals.htm?issn=0737-8831>> . Acesso em: 19 out. 2012.

LAKATOS, E. M. ; MARCONI, M. A. **Metodologia do trabalho científico**. 4. ed. rev. ampl. São Paulo: Atlas, 1992.

LAKATOS, E. M. ; MARCONI, M. A. **Fundamentos da metodologia científica**. 5. ed. São Paulo: Atlas, 2003.

JANKOWSKI, N. W. Exploring E-Science: An Introduction. **Journal of Computer-Mediated Communication**, v. 12, n. 2, p. 549-562, 2007. Disponível em: <<http://jcmc.indiana.edu/vol12/issue2/jankowski.html>>. Acesso em: 19 out. 2012.

LANCASTER, F. W. **Indexação e resumos: teoria e prática.** Brasília, DF: Briquet de Lemos, 1993.

LANGRIDGE, D. W. **Classificação: abordagem para estudantes de Biblioteconomia.** Rio de Janeiro: Interciência, 1977.

MARTINEZ, A. et al. Las categorías o facetas fundamentales: una metodología para el diseño de taxonomías corporativas de sitios Web argentinos. **Ciência da Informação**, Brasília, DF, v. 33, n. 2, p. 106-111, maio/ago. 2004. Disponível em: <<http://revista.ibict.br/ciinf/index.php/ciinf/article/view/523/475>>. Acesso em: 30 out. 2012.

MARTÍNEZ-URIBE, U.; MACDONALD, S. Un Nuevo cometido para los bibliotecarios académicos: data curation. **El profesional de la información**, Barcelona, v. 17, n. 3, mayo-jun. 2008.

MEDEIROS, J. S. **Tesauros conceituais e ontologias de fundamentação: abordagem comparativa entre modelos conceituais.** São Paulo: Ixtlan, 2012.

MICHAELIS. **Moderno Dicionário da Língua Portuguesa.** São Paulo: Melhoramentos, 2012. Disponível em: <<http://michaelis.uol.com.br/moderno/portugues/index.php>> . Acesso em: 02 nov. 2012.

MOTTA, D. F. **Método relacional como nova abordagem para a construção de tesauros.** Rio de Janeiro, SENAI, 1987.

MULLINS, J. L.; Bringing Librarianship to E-Science. **College & Research Libraries**, v. 70, n. 3, p. 212-213, maio 2009. Disponível em: <<http://crl.acrl.org/content/70/3/212.short>>. Acesso em: 21 out. 2012

NEVES, J. L. Pesquisa Qualitativa: características, usos e possibilidades. **Caderno de Pesquisas em Administração**, São Paulo, v. 1, n. 3, 1996. Disponível em: <<http://www.ead.fea.usp.br/cad-pesq/arquivos/C03-art06.pdf>>. Acesso em: 17 nov. 2012

ROBREDO, J.; CUNHA, M. B. **Documentação de hoje e de amanhã: uma abordagem informatizada da Biblioteconomia e dos sistemas de informação.** 2. ed. São Paulo: Global, 1986.

SHREEVES, S. L.; CRAGIN, M. H. Institutional Repositories: current state and future. **Library Trends**, Baltimore, v. 57, n. 2, p. 89-97. Fall 2008.

STEWART, C. A. et. al. What is Cyberinfrastructure?. Norfolk, Proceedings of the 38th annual fall conference on SIGUCCS, ACM Digital Library, p. 37-44. Disponível em: <<http://dl.acm.org/citation.cfm?id=1878347>>. Acesso em: 21 out. 2012.

TORRES-SALINAS, D.; ROBINSON-GARCIA, N.; CABEZAS-CLAVIJO, A. Compartir los datos de investigación en ciencia: introducción al data sharing. **El profesional de la información**, Barcelona, v. 17, n. 3, mar-abr. 2012.

TORRES-SALINAS, D. **Compartir datos (data sharing) en ciencia: el contexto de una oportunidad**. Anuario ThinkEPI, 2010. Disponível em: <http://www.thinkepi.net/compartir-datos-data-sharing-en-ciencia-el-contexto-de-una-oportunidad> Acesso em: 15 out. 2012.

VAN DER LAAN, R. H. **Tesauro e terminologia**: uma inter-relação lógica. 2002. 196 f. Tese (doutorado) – Departamento do Programa de Pós-graduação em Letras, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2002. Disponível em: <<http://www.lume.ufrgs.br/handle/10183/3487>>. Acesso em: 2 nov. 2012.

VICTORINO, M.; BRÄSCHER, M. Organização da Informação e do Conhecimento, Engenharia de Software e arquitetura orientada a serviços: uma abordagem holística para o desenvolvimento de sistemas de informação computadorizados. **DataGramZero**, Rio de Janeiro, v. 10, n. 3, jun. 2009. Disponível em: <[http://www.dgz.org.br/jun09/Art\\_03.htm](http://www.dgz.org.br/jun09/Art_03.htm)>. Acesso em: 2 nov. 2012.

VITAL, L. P. Taxonomia como elemento estruturante em portais corporativos. **Revista Digital de Biblioteconomia & Ciência da Informação**, Campinas, v. 9, n. 2, p. 25-40, jan./jun. 2012.

WITT, M.; CARLSON, J.; BRANDT, D. S.; CRAGIN, M. Constructing Data Curation Profiles. **The International Journal of Digital Curation**, Bath, v. 4, n. 3, p. 93-103, dez. 2009. Disponível em: <<http://www.ijdc.net/index.php/ijdc/article/view/137/165>>. Acesso em: 16 out. 2012.

YAKEL, E. Digital Curation, **OCLC Systems & Services**, v. 23, n. 4, p. 335-340, 2007. Disponível em: <[dx.doi.org/10.1108/10650750710831466](http://dx.doi.org/10.1108/10650750710831466)>. Acesso em: 23 out. 2012