



CONFERÊNCIA SOBRE TECNOLOGIA, CULTURA E MEMÓRIA

CURADORIA E DADOS ABERTOS

**ANAIS IV
CTCM 2017**

Májory Karoline Fernandes de Oliveira Miranda
Vildeane da Rocha Borba
Marcos Galindo Lima
(ORGs.)



CONFERÊNCIA SOBRE TECNOLOGIA, CULTURA E MEMÓRIA

CURADORIA E DADOS ABERTOS

RECIFE
NECTAR
2018

Catálogo na fonte
Bibliotecário Jonas Lucas Vieira, CRB4-1204

C748 Conferência sobre Tecnologia, Cultura e Memória (CTCM 2017):
curadoria e dados abertos/ Organização de Májory Karoline
Fernandes de Oliveira Miranda, Vildeane da Rocha Borba,
Marcos Galindo Lima. – Recife: NECTAR, 2018.
124 p. : il.

ISBN: 978-85-60323-59-3 (e-book)

1. Ciência da Informação. 2. Curadoria digital. 3. Dados abertos.
4. Dados de pesquisa. I. Miranda, Májory K. F. de Oliveira, Org. II.
Borba, Vildeane da Rocha, Org. III. Lima, Marcos Galindo, Org.

CDD 004.6

ORGANIZAÇÃO

Coordenação Geral

Prof. Dr. Marcos Galindo (Coordenador Geral)

(Programa de Pós-Graduação em Ciência da Informação/Depto de Ciência da Informação da Universidade Federal de Pernambuco / Laboratório Liber - Tecnologia para o Conhecimento.

Profa. Dra. Sandra de Siebra (Vice-Coordenadora)

(Programa de Pós-Graduação em Ciência da Informação/Depto de Ciência da Informação da Universidade Federal de Pernambuco / Laboratório Liber - Tecnologia para o Conhecimento.

Coordenação Científica

Profa. Dra. Májory Karoline F. de Oliveira Miranda

(Programa de Pós-Graduação em Ciência da Informação/Depto de Ciência da Informação da Universidade Federal de Pernambuco / Laboratório Liber - Tecnologia para o Conhecimento.

Profa. Dra. Sônia Aguiar Cruz Riascos

Depto de Ciência da Informação da Universidade Federal de Pernambuco / Laboratório Liber - Tecnologia para o Conhecimento.

Coordenação Técnica

Profa. Aureliana L. de Lacerda Tavares

Depto de Ciência da Informação da Universidade Federal de Pernambuco / Laboratório Liber - Tecnologia para o Conhecimento.

Profa. Vildeane da Rocha Borba

Depto de Ciência da Informação da Universidade Federal de Pernambuco / Laboratório Liber - Tecnologia para o Conhecimento.

Evaldo Rosa de Souza

Pesquisador/ Laboratório Liber - Tecnologia para o Conhecimento.



Comissão de Comunicação e Cultura

Celly Britto

Depto de Ciência da Informação da Universidade Federal de Pernambuco / Laboratório Liber - Tecnologia para o Conhecimento.

Helio Pajeu

Depto de Ciência da Informação da Universidade Federal de Pernambuco / Laboratório Liber - Tecnologia para o Conhecimento.

Comissão de Infraestrutura

Alexandre Valdevino

Laboratório Liber - Tecnologia para o Conhecimento. / Universidade Federal de Pernambuco.

Prof. Dr. Maurício Rocha de Carvalho

Departamento de Ciência da Informação / Universidade Federal de Pernambuco / Laboratório Liber - Tecnologia para o Conhecimento.

Comissão de Comunicação e Mídias Sociais

Bruno Flávio Espíndola Leite

Liber - Laboratório de Tecnologia para o Conhecimento / Universidade Federal de Pernambuco.

Designer Fundação Oswaldo Cruz Pernambuco.

Gabriel Ladislau Anceles da Rocha

Graduando em Gestão da Informação DCI / UFPE / Laboratório Liber - Tecnologia para o Conhecimento.

Jarluzia Herquita de Azevedo Afonso

Graduanda em Gestão da Informação DCI / UFPE / Laboratório Liber - Tecnologia para o Conhecimento.

Luiz Henrique Teixeira Bazilio

Graduando em Gestão da Informação DCI / UFPE / Laboratório Liber - Tecnologia para o Conhecimento.



Susimery Vila Nova

Mestre em Ciência da Informação DCI / UFPE

Bibliotecária Procit - Pró Reitoria de Comunicação, Informação e Tecnologia da Universidade Federal de Pernambuco.

Tony Bernardino de Macedo

Mestre em Ciência da Informação DCI / UFPE

Bibliotecário do Memorial Denis Bernardes (MDB) / BC / UFPE.

Comissão de Captação de Recursos

Maria Angela Oliveira

Laboratório Liber - Tecnologia para o Conhecimento. / Universidade Federal de Pernambuco.

Mário Gouveia

Laboratório Liber - Tecnologia para o Conhecimento / Universidade Federal de Pernambuco.

Sandra Maria Veríssimo Soares

Laboratório Liber - Tecnologia para o Conhecimento / Universidade Federal de Pernambuco / Articuladora da Rede Memorial.

Kleber Santos

Laboratório Liber - Tecnologia para o Conhecimento / Universidade Federal de Pernambuco.

Comissão de Apoio

Bruno Felipe de Melo Silva

Cynthia Barros

Danubia Souza Menezes

Elanna Beatriz

Eline Isobel

Faysa Oliveira e Silva

Guilherme Ulisses

Ismael Rodrigues Santos

Rafaella Arcoverde

Roman Ruiz Maranhão

Rosa de la Vega

Silvânia Cosmos Esmerindo



Comissão Avaliadora

Profa. Dra. Anna Elizabeth G. C. Correia.....	UFPE/BR
Profa. Dra. Celly de Brito Lima.....	UFPE/BR
Profa. Msc. Vildeane Borba.....	UFPE/BR
Prof. Dr. Diego Andres Salcedo.....	UFPE/BR
Prof. Dr. Hélio Márcio Pajeú.....	UFPE/BR
Prof. Dr. Marcos Galindo Lima.....	UFPE/BR
Prof. Dr. Maurício Rocha de Carvalho.....	UFPE/BR
Profa. Dra. Májory Karoline F. de Oliveira Miranda.....	UFPE/BR
Profa. Dra. Sandra de Albuquerque Siebra.....	UFPE/BR

Instituições Parceiras

Universidade Federal de Pernambuco (UFPE); Departamento de Ciência da Informação (DCI); Laboratório Liber - Tecnologia para o Conhecimento; Programa de Pós-Graduação em Ciência da Informação (PPGCI); Instituto Ricardo Brennand (IRB); Rede Memorial de Pernambuco.

SUMÁRIO

Apresentação.	08
Introdução.	11
Trabalhos.	12
1. Acesso e uso de dados de pesquisa: uma análise preliminar do Figshare. Autoria: Ronaldo Ferreira Araújo, Sarah de Oliveira Santos.	13
2. As práticas de indexação e suas interfaces com os repositórios digitais de acesso aberto: uma abordagem sobre o metadado assunto no Repositório Institucional da UFRN. Autoria: Raimunda Fernanda Santos.	23
3. Case Data Analysis and Open Data in the Course of Information Systems at UFRPE. Autoria: Roberta Macêdo Marques Gouveia, Maria da Conceição Moraes Batista, Lisandra Sousa da Cruz, Jonathan Henrique Andrade de Carvalho.	42
4. Classificação de documentos filatéticos em formato aberto para Web: o caso dos selos postais no Repositório Filatélico Brasileiro. Autoria: Kézia de Lira Feitosa, Diego Andres Salcedo.	53
5. Convergência entre repositórios digitais e periódicos. Autoria: Bruno Tenório Ávila, Milena Beatriz Lira Dias da Silva.	67
6. Dados abertos e curadoria digital no projeto Jugando a definir la Ciência: o papel do design da informação. Autoria: Laís Alpi Landim, Maria José Vicentini Jorente.	79
7. Design Thinking para a Curadoria Digital de Repositórios de Objetos Digitais de Aprendizagem: uma proposta inicial. Autoria: Natália Nakano, Maria José Vicentini Jorente.	94
8. Princípios e normas que fundamentam os dados abertos: um estudo de caso do software ICA-Atom. Autoria: Nandia Letícia Freitas Rodrigues, Lucineia da Silva Batista, Maria José Vicentini Jorente.	109

APRESENTAÇÃO

O Laboratório Liber e o Programa de Pós-Graduação em Ciência da Informação da Universidade Federal de Pernambuco organizam desde 2011 o evento científico bianual Conferência sobre Tecnologia, Cultura e Memória (CTCM). O evento reúne mentes inovadoras com o objetivo de promover o intercâmbio de ideias e difundir a pesquisa no campo de metodologias e tecnologias aplicadas a preservação da memória e do patrimônio cultural. Em sua quarta edição o evento já se consolidou como fórum de referência na reflexão de processos inovadores para gestão do patrimônio memorial.

O primeiro CTCM foi realizado em 2011 com o apoio da Embaixada do Reino dos Países Baixos e da Fundação de Amparo à Ciência e Tecnologia de Pernambuco (FACEPE). Reuniu no Instituto Ricardo Brennand cerca de 250 pessoas, entre palestrantes, convidados, profissionais e

estudantes. Um marco fundamental desta conferência foi a criação da Rede Memorial ou Rede Nacional das instituições comprometidas com políticas de digitalização dos acervos memoriais do Brasil.

A conferência de 2013 enfatizou o debate sobre as tendências contemporâneas e os problemas estratégicos que envolvem a tecnologia aplicada à gestão da memória no âmbito da Ciência da Informação e domínios afins. Foram discutidas as tendências universais para o tratamento, organização, difusão e acesso à informação registrada e o papel das networks no desenvolvimento de processos colaborativos. Nesse ano, o evento também sediou o Fórum da Rede Memorial em paralelo com sua programação.

Em 2015, o CTCM ampliou a discussão acerca de problemas estratégicos que envolvem a gestão, preservação e acesso da memória



registrada em meio digital, com ênfase no debate sobre as principais tendências contemporâneas em soluções colaborativas especialmente na curadoria de bens culturais em rede.

Considerando o potencial dos dados gerados pelas pesquisas e os desafios que a gestão deles impõe às pessoas e organizações, o CTCM em sua quarta edição (2017), a exemplo de outros eventos da área de Ciência da Informação, vem incentivar instituições de pesquisa, organizações, gestores, investigadores, bibliotecários e equipes de Tecnologia da Informação a explorar seus papéis e responsabilidades no processo de gestão de dados de pesquisa, do inglês, Research Data Management (RDM).

A gestão de dados de pesquisa diz respeito à organização dos dados, desde a sua entrada no ciclo de investigação até à divulgação e arquivamento de resultados. Trata-se de: criar dados e planejar seu uso; organizar, estruturar,

agregar valor e nomear dados; manter, armazenar, tornar seguro, fornecer acesso a longo prazo e possibilitar seu compartilhamento e reuso. Além disso, transversalmente, envolve o desenvolvimento e implementação de políticas que deem suporte a todo esse processo.

Neste sentido, a Conferência sobre Tecnologia, Cultura e Memória 2017 (CTCM), realizada no auditório Evaldo Coutinho do Centro de Artes e Comunicação (CAC) da UFPE entre os dias 16 a 18 de agosto, congregou estudantes, docentes e pesquisadores da área de Ciência da Informação e afins, nacionais e internacionais, para discutir temas caros a essas áreas, a exemplo das Tecnologias para Memória, Curadoria de Dados Abertos e Humanidades Digitais.

Com importantes contribuições de profissionais, estudantes e pesquisadores interessados em refletir sobre o futuro dos serviços de informação renovados pela ação



da tecnologia, repletu-se sobre as novas demandas de pesquisa e as tendências de inovação. Importou neste encontro questionamentos em direção a soluções sustentáveis e convergentes.

A temática da gestão de dados de pesquisa em todas as suas nuances, além do entendimento que a implantação de políticas públicas de cultura e preservação da memória, conscientizou os participantes do CTCM a buscarem oportunidades de processo de gestão de dados de pesquisa. O evento, assim, foi uma excelente oportunidade para reciclagem de conhecimentos, compartilhamento de experiências e reforço das redes de colaboração para o avanço dos processos de curadoria.

Como em outros anos, o CTCM 2017 abrigou o Fórum da Rede Memorial, que ocorreu em paralelo à programação do CTCM.

O Fórum da Rede Memorial já ocorre desde o primeiro CTCM, em 2011. O Fó-

rum reuniu representantes de instituições públicas e privadas envolvidas (ou desejosas de se envolverem) com projetos de digitalização de acervos para discutir caminhos práticos na contribuição dos processos em curso de valorização da cultura brasileira.

A parceria com a Fundação de Amparo à Ciência e Tecnologia do Estado de Pernambuco, FACEPE, tornou possível a realização da Conferência sobre Tecnologia, Cultura e Memória – 2017.

Májory Miranda

Julho 2018

INTRODUÇÃO

De acordo com Simon Hodson, do Joint Information Systems Committee (JISC), o processo científico é reforçado pela gestão e partilha de dados de pesquisa. Uma boa prática de gestão de dados de pesquisa permite tanto uma verificação confiável dos resultados, quanto a realização de pesquisas novas e inovadoras construídas sobre estoques de informação pré-existentes.

Nesta nova ambiência social, o gerenciamento de dados e a curadoria digital estão se consolidando como importantes tendências contemporâneas. No mundo inteiro, universidades e instituições governamentais estão incrementando sua infraestrutura para manter ativas ferramentas de gestão de dados. Pois, de fato, o gerenciamento de dados de pesquisa visa tornar o processo de investigação mais eficiente e atender às expectativas de universidades, de financiadores e às novas demandas por dados especia-

lizados. Isso vem sendo reconhecidos pelos principais conselhos de pesquisa do mundo desenvolvido que, a seu turno, vem sistematicamente adotando políticas públicas de gestão de dados científicos e incentivando a criação de planos de gestão de dados de pesquisa, além de incentivar o depósito dos dados pesquisados em centros de dados confiáveis, do ponto de vista da curadoria de dados.

Para refletir sobre estas temáticas, o CTCM 2017 lançou seu olhar sobre os novos campos do conhecimento e sobre as fronteiras metamorfolizadas sob o efeito gravitacional da TI. Interessou-se sobremaneira sobre a discussão de processos, ferramentas e modelos de gestão de dados de pesquisa, curadoria digital e curadoria de dados e o desenvolvimento e implementação de políticas públicas de cultura e preservação da memória, além de possibilidades de colaboração.

TRABALHOS

1. Acesso e uso de dados de pesquisa: uma análise preliminar do Figshare.

Autoria: Ronaldo Ferreira Araújo, Sarah de Oliveira Santos.

2. As práticas de indexação e suas interfaces com os repositórios digitais de acesso aberto: uma abordagem sobre o metadado assunto no Repositório Institucional da UFRN.

Autoria: Raimunda Fernanda Santos.

3. Case Data Analysis and Open Data in the Course of Information Systems at UFRPE.

Autoria: Roberta Macêdo Marques Gouveia, Maria da Conceição Moraes Batista, Lisandra Sousa da Cruz, Jonathan Henrique Andrade de Carvalho.

4. Classificação de documentos filatélicos em formato aberto para Web: o caso dos selos postais no Repositório Filatélico Brasileiro.

Autoria: Kézia de Lira Feitosa, Diego Andres Salcedo.

5. Convergência entre repositórios digitais e periódicos.

Autoria: Bruno Tenório Ávila, Milena Beatriz Lira Dias da Silva.

6. Dados abertos e curadoria digital no projeto Jugando a definir la Ciência: o papel do design da informação.

Autoria: Laís Alpi Landim, Maria José Vicentini Jorente.

7. Design Thinking para a Curadoria Digital de Repositórios de Objetos Digitais de Aprendizagem: uma proposta inicial.

Autoria: Natália Nakano, Maria José Vicentini Jorente.

8. Princípios e normas que fundamentam os dados abertos: um estudo de caso do software ICA-AtoM.

Autoria: Nandia Letícia Freitas Rodrigues, Lucineia da Silva Batista, Maria José Vicentini Jorente.



ACESSO E USO DE DADOS DE PESQUISA uma análise preliminar do *Figshare*

Ronaldo Ferreira Araujo

Doutor em Ciência da Informação - UFMG
Professor Adjunto do Curso de Biblioteconomia – UFAL
Laboratório de Estudos Métricos da Informação na Web (Lab-iMetrics)

Sarah Rúbia de Oliveira Santos

Bacharel em Biblioteconomia - UFAL
Laboratório de Estudos Métricos da Informação na Web (Lab-iMetrics)

RESUMO: Os resultados de pesquisa sempre foram os principais elementos de avaliação da produção científica, seja para análise de seus indicadores tradicionais ou mais recentes na compreensão de suas métricas alternativas. No âmbito de uma ciência 2.0 e das práticas mais abertas e colaborativas da comunicação científica os dados de pesquisa são cada vez mais importantes. Com o objetivo de investigar a visibilidade e impacto do acesso e (re)uso de dados de pesquisa, o presente trabalho, de caráter exploratório e abordagem quantitativa descritiva, analisa um conjunto de dados abertos de pesquisa disponibilizados no *Figshare*. O desempenho do indicador de visibilidade e acesso dos trabalhos somados chega a 136.799 visualizações e 67.301 downloads. As métricas de uso e impacto são mais tímidas com 1024 dados altmétricos e apenas 10 citações. Embora o compartilhamento de dados abertos de pesquisa seja uma prática relativamente nova essa análise preliminar do *Figshare* possibilitou observar que o alcance dos dados abertos é representativo, principalmente para os números de visualizações e downloads, o que pode potencializar seu (re)uso dentro e fora da comunidade científica.

Palavras-chave: Dados abertos de pesquisa, Ciência aberta, Reuso de dados, Altmétrie, *Figshare*.

ABSTRACT: *The research output has always been the main elements of evaluation of scientific production, whether for analysis of its traditional indicators or more recent in the understanding of altmetrics. In the science 2.0 context and the practices of scientific communication more open and collaborative, research data are increasingly important. To investigate the visibility and impact of access and reuse of research data, this exploratory and quantitative descriptive approach a set of open data research in Figshare were analyzed. The performance of the visibility and access indicator was 136,799 views and 67,301 downloads. The usage and impact metrics were lower with 1024 altmetric data and only 10 citations. Although open research data sharing is a relatively new practice, this analysis shows that the reach of open data is representative, especially for the numbers of views and downloads, which can enhance its reuse in the scientific community and the public.*

Keywords: *Open Data Research, Open Science, Data Reuse, Altmetrics, Figshare.*

1 INTRODUÇÃO

Os resultados de pesquisa sempre foram os principais elementos de avaliação da produção científica, seja para análise de seus indicadores tradicionais de impacto como os estudos bibliométricos e de citação ou mais recentes na compreensão de suas métricas alternativas e compreensão da atenção online que recebem e o interesse público que despertam ao circularem em ambientes digitais abertos como blogs e as mídias sociais.

Atualmente, para além dos resultados de pesquisa, no âmbito de uma ciência 2.0 ou uma ciência aberta (*open science*) tornam-se mais comuns as práticas de colaboração e compartilhamento de dados de pesquisa, tendo em vista a possibilidade de armazenamento destes em bases de dados e a publicização de informações que podem ser utilizadas, reutilizadas e adaptadas para diferentes realidades e objetivos distintos.

O presente trabalho faz uma breve incursão sobre a ciência aberta, seus benefícios para a visibilidade e impacto de dados de pesquisa, bem como suas características e indicadores que podem ser monitorados.

2 FUNDAMENTAÇÃO TEÓRICA

A ciência aberta possui diversas facetas e significados devido a sua interdisciplinaridade. Para Sayão e Sales (2014, p. 77) a face primordial da ciência aberta é aquela que “reconhece [...] que o conhecimento científico é um patrimônio da humanidade e, que, portanto, deve estar disponível livremente”. Tal ciência é essencialmente colaborativa, utilizando-se de recursos tecnológicos para promover a contribuição online entre pesquisadores, à distância e em tempo real (ALBAGLI; CLINIO; RAYTCHTOK, 2014, p. 436).

O ponto principal da ciência aberta é permitir a utilização e compartilhamento de dados científicos por qualquer pessoa, “tal movimento insere-se no contexto de uma ciência modo 2 por valorizar a produção colaborativa, a transparência e a transdisciplinaridade” (FAGUNDES, 2014, p. 491). Para Aventurier e Alencar (2016, p.3) ela “inclui um conjunto de práticas sobre a utilização da internet e das ferramentas da web social para mudar de maneira global a abordagem acadêmica, desde a formulação das questões e das hipóteses até a difusão dos resultados de pesquisas”.

Corroborando com essa ideia, Araújo (2014, p.32) entende que a ciência aberta sugere “novas práticas de cientistas que publicam resultados experimentais ou finais, matérias, novas teorias, reivindicações de descoberta e de projetos na web para que outros possam ver, compartilhar e comentar”. O consenso é de que a ciência aberta é um movimento transformador que possibilita a disponibilização de dados, a colaboração entre pesquisadores e a utilização dessas informações sem nenhuma restrição, geralmente utilizando-se de recursos tecnológicos para socialização deste conteúdo.

Assim, a internet e os recursos da web 2.0 são fundamentais quando se fala em ciência aberta, por se caracterizar como um espaço aberto que permite o diálogo, a colaboração e o compartilhamento de recursos diversos, como os textuais, hipertextuais e/ou audiovisuais. As plataformas digitais colaborativas são vistas como uma “oportunidade tecnológica em favor do conhecimento aberto e não proprietário” (ALBAGLI; CLINIO; RAYTCHTOK, 2014, p. 436).

Para Sayão e Sales (2014, p. 78), estas tecnologias “têm transformado de maneira vertiginosa a forma como os dados de pesquisa podem ser produzidos, disseminados, gerenciados, compartilhados e usados”. Segundo a Comissão Europeia de Pesquisa e Inovação¹, as características e indicadores da ciência aberta podem ser monitoradas por três facetas principais: a) os dados abertos de pesquisa, b) o acesso livre às publicações e c) a comunicação científica aberta.

A primeira compreende o uso de repositórios de dados de pesquisa, de políticas de financiamento no compartilhamento de dados e das atitudes dos pesquisadores relativas ao compartilhamento desses dados como o *Figshare* <<https://figshare.com>> e o GenBank <<https://www.ncbi.nlm.nih.gov/genbank/>> que possibilitam o compartilhamento de forma aberta de dados e resultados de pesquisa, facilitando o trabalho de outros pesquisadores que podem fazer uso e reuso dos dados disponibilizados. Para Sayão e Sales (2014, p. 78) esse processo “encurta o ciclo clássico de comunicação científica e abre novas formas de interlocução e de socialização no mundo científico, além de contribuir para a racionalização dos recursos financeiros públicos aplicados na pesquisa científica”.

A segunda baseia-se em estratégias para o acesso aberto, uma delas por meio do autoarquivamento, “pelos autores ou seus representantes, dos artigos

¹ <http://ec.europa.eu/>

publicados nas revistas científicas em repositórios, disciplinares ou institucionais” (via verde) e a outra através da publicação em revistas de acesso livre “que não restringem o acesso e o uso do material que publicam e não cobram assinatura nem taxas de acesso” (BAPTISTA et al., 2007, p. 5).

A terceira, comunicação científica aberta, culmina das duas anteriores, promovendo a colaboração, acessibilidade e transparência através da disseminação de pesquisas por meios formais e informais. De acordo com Albagli, Clinio e Raytchtok (2014, p. 435) esta faceta é contrária aos modos atuais de produção e de comunicação científica que estão “submetidos a mecanismos que criam obstáculos artificiais de várias ordens, especialmente legais e econômicos, à sua livre circulação e, logo, a seu avanço e difusão”.

3 MATERIAL E MÉTODO

Trata-se de uma pesquisa exploratória de abordagem quantitativa descritiva sobre o acesso e uso de dados abertos de pesquisa. Universo de pesquisa é composto por dados de pesquisa disponibilizados no *Figshare*, um repositório de dados interdisciplinares na web, de iniciativa privada (Digital Science) mantido em parceria com universidades (AVENTURIER; ALENCAR, 2016).

Entre as suas funcionalidades, estão o espaço limitado para uso privado e espaço ilimitado para o compartilhamento de materiais públicos, permitindo que os usuários publiquem figuras, vídeos, áudios, posters, papers, teses, códigos, apresentações, *datasets* e *filesets*; o alocamento dos materiais em 20 categorias com diversas subcategorias; e a identificação dos materiais com o *Digital Object Identifier* (DOI). O Figshare ainda mantém armazenados dados de uso e apresenta estatísticas referentes a cada item em suas páginas (KRAKER et al., 2015).

A coleta foi realizada na segunda quinzena de março de 2017, com consulta manual no campo de busca da plataforma com a expressão: “*open data*”. Para garantir a análise de critérios de visibilidade e impacto considerou-se uma amostra intencional para a análise dos 20 primeiros trabalhos ordenados por “popularidade” <sort by: *Popular*>, pelo tipo “conjunto de dados” <type: *data set*>. Os trabalhos são descritos pelo tipo de autoria, data de publicação e forma de licença, analisados segundo orientação temática (título, palavras-chave e categoria) e na compreensão

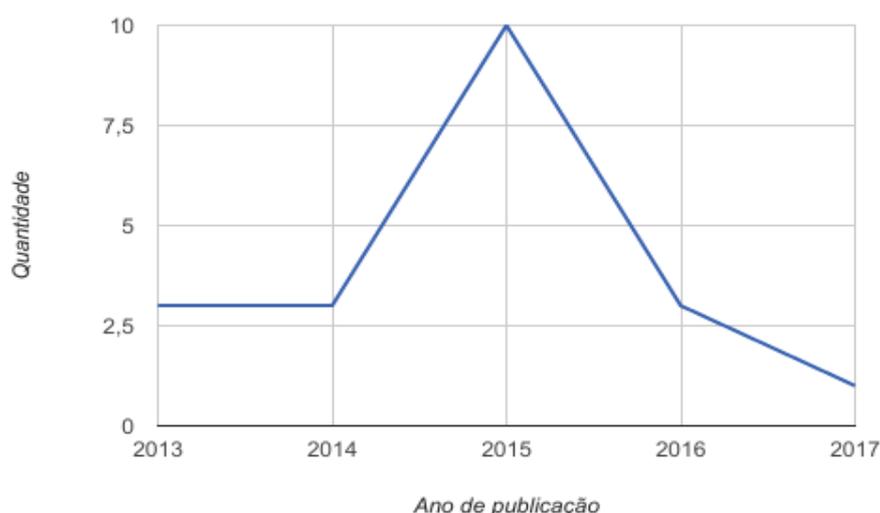
de seus indicadores de visibilidade e acesso (visualizações e downloads) e uso e impacto (citações e menções).

4 RESULTADOS

O compartilhamento de dados abertos de pesquisa, quando acompanhados da identificação da publicação no qual são vinculados podem servir de elementos para revisão pós-publicação. Metade dos 20 conjunto de dados possuem autoria única, sendo complementado por 45% de projetos com 2 a 5 autores e apenas uma minoria de 5% com mais de 6 autores. No total foram contabilizados 43 autores, apenas 5 destes tiveram mais de um trabalho compartilhado, sendo Ellery Wulczyn autora de 3 projetos, seguida pelo Digital Science, Lucas Dixon, Nithum Thain e Randy Olson com 2 projetos cada.

A distribuição dos trabalhos por ano de publicação pode ser visualizada no Gráfico 1. No Figshare é possível adicionar novas versões de *datasets* na página criada para o projeto, desta forma, ao acessar a página, abaixo do título do projeto, consta a data da última atualização feita. Apesar disso, analisamos a data de criação de cada *dataset* a partir da data da primeira versão adicionada.

Gráfico 1. Publicações por ano



Fonte: dados da pesquisa (2017)

No gráfico, observa-se um pico no ano de 2015 referente à criação de *datasets*, nos dois anos antecedentes, a criação de *datasets* permaneceu na média

O Figshare possui 20 categorias com diversas subcategorias definidas para identificação dos documentos enviados para o site. As categorias que possuíam mais *datasets* dentro do nosso escopo foram *Applied Computer Science* empatado com *Library and Information Studies* (6,8%), seguidos de *Biochemistry*, *Biophysics*, *Natural Language Processing*, *Science Policy* e *Sociology* (4,5%), todas as outras categorias foram utilizadas apenas uma vez.

As opções de licença disponíveis para os usuários são: CC-BY, CC-0, MIT, GLP (-2.0 e -3.0) e Apache-2.0, para instituições mais alguns tipos de licença são disponibilizados como CC BY-SA, CC BY-ND entre outras. Todos os *datasets* disponibilizados estavam sob licença *Creative Commons*.

Dos 20 projetos, 75% estavam em CC-BY, licença que permite que terceiros distribuam, remixem, adaptem e criem a partir desse trabalho, mesmo que para fins comerciais, desde que se atribua o devido crédito pela criação original. Os outros 25%, estavam sob CC-0, que permite que os licenciados renunciem a todos os direitos e coloquem um trabalho no domínio público². A Tabela 1 apresenta a distribuição do conjunto de dados com sua data de criação, tipo de licença e indicadores de visibilidade e acesso e uso e impacto.

Tabela 1. Indicadores de visibilidade e acesso e uso e impacto

Dataset	Projeto		Visibilidade e acesso		Uso e impacto	
	Data	Licença	Views	Downloads	Altmetrics	Citações
D1	2015	CC-BY	8526	<u>21673</u>	57	0
D2	2015	CC-0	<u>13067</u>	<u>19762</u>	<u>129</u>	<u>1</u>
D3	2014	CC-BY	<u>35191</u>	1405	18	0
D4	2017	CC-BY	20	<u>6221</u>	0	0
D5	2014	CC-0	<u>13175</u>	1435	<u>439</u>	<u>3</u>
D6	2015	CC-BY	<u>21496</u>	488	0	0
D7	2015	CC-BY	270	<u>3674</u>	0	0
D8	2015	CC-BY	269	3163	0	0
D9	2014	CC-BY	6130	1121	<u>159</u>	<u>6</u>
D10	2015	CC-BY	6577	371	0	0
D11	2015	CC-BY	3482	1241	12	0
D12	2015	CC-BY	5266	482	53	0
D13	2015	CC-0	3743	642	4	0
D14	2016	CC-0	3505	718	16	0

² <https://creativecommons.org/>

D15	2016	CC-BY	94	1510	0	0
D16	2013	CC-BY	2605	922	6	0
D17	2015	CC-BY	4652	260	0	0
D18	2016	CC-0	1976	1065	<u>128</u>	0
D19	2013	CC-BY	2965	722	3	0
D20	2013	CC-BY	3790	426	0	0
Total			136799	67301	1024	10

Fonte: dados da pesquisa (2017)

O desempenho do indicador de visibilidade e acesso dos trabalhos somados chega a 136.799 visualizações e 67.301 downloads. Quanto às métricas de uso e impacto obteve-se 1.024 dados altmétricos e 10 citações. Não parece haver uma correlação direta entre um indicador e outro. O único conjunto de dados com bom desempenho em praticamente todos os indicadores é o D2, publicado em 2015 com licença CC-0, 13.067 visualizações, 19.762 downloads, 129 dados altmétricos e 1 citação.

Mediante os indicadores de visibilidade e acesso é possível observar que o alcance dos dados abertos é representativo, principalmente para os números de visualizações e downloads, o que pode potencializar seu (re)uso, tendo em vista que além de dados científicos, os dados abertos, referem-se à categoria de dados de ordem “cultural, financeira, estatística, climática, ambiental, viária e geográfica” (ALBAGLI; CLINIO; RAYTCHTOK, 2014, p.440), ou seja, dados que poderiam ser (re)utilizados por qualquer pessoa para outros fins além de novas publicações científicas.

Os dados de uso e impacto ainda são tímidos. E pelas datas de criação dos conjuntos de dados serem recentes (entre 2013 e 2017) os valores altmétricos (1.024) superam significativamente os de citação (apenas 10). A principal fonte dos 1.024 dados altmétricos é o Twitter (951), seguido dos Blogs (25), do Google+ (22), do Mendeley (18), do Facebook (7) e do Reddit (1). O posicionamento do Twitter como mídia mais expressiva em análises altmétricas na comunicação científica aberta é recorrente (ALPERIN, 2014; ARAÚJO, 2015; ARAÚJO; FURNIVAL, 2016) e parece se reproduzir quando se analisa dados de pesquisa da ciência aberta.

5 CONSIDERAÇÕES FINAIS

O *Figshare* é um dos parceiros da Comissão Europeia de Pesquisa e Inovação no desenvolvimento do Monitor da Ciência Aberta e por isso o mesmo compreende as três facetas apresentadas no modelo adotado por esta comissão: (a) é um repositório de dados de pesquisa; (b) possibilita o autoarquivamento de projetos pelos próprios autores e disponibilização de acesso livre para aquele que tiver interesse; e (c) promove a comunicação científica aberta por meio da colaboração, acessibilidade e transparência através da disseminação tanto de resultados quanto de dados de pesquisa.

Embora o compartilhamento de dados abertos de pesquisa seja uma prática relativamente nova essa análise preliminar do *Figshare* possibilitou observar que o alcance dos dados abertos é representativo, principalmente para os números de visualizações e downloads, o que pode potencializar seu acesso e (re)uso dentro e fora da comunidade científica.

REFERÊNCIAS

- ALBAGLI, S.; CLINIO, A.; RAYCHTOCK, S.. Ciência Aberta: correntes interpretativas e tipos de ação. **Pesquisa Brasileira em Ciência da Informação e Biblioteconomia**, v. 10, n. 1, 2015. Disponível em: <<http://revista.ibict.br/liinc/article/view/3593/3072>>. Acesso em: 02 maio 2017.
- ALPERIN, J. P. Ask not what altmetrics can do for you, but what altmetrics can do for developing countries. **Bulletin of the American Society for Information Science and Technology**, v. 39, n. 4, p. 18-21, 15 abr 2014.
- ARAUJO, R. F.. Marketing científico digital e métricas alternativas para periódicos: da visibilidade ao engajamento. **Perspectivas em Ciência da Informação**, v. 20, p. 67-84, 2015
- ARAUJO, R. F. Ciência 2.0 e a Presença Online de Pesquisadores: visibilidade e impacto. **Ciência da Informação em Revista**, v. 1, n. 3, p. 32-40, 2014. Disponível em: <<http://www.seer.ufal.br/index.php/cir/article/view/1608>>. Acesso em: 02 maio 2017.
- AVENTURIER, P.; ALENCAR, M. C. F.. Os desafios dos dados de pesquisas abertos. **RECIIS - Revista de Comunicação Informação & Inovação em Saúde**, v. 10, n. 3, p. 19 p., 2016. Disponível em: <<https://hal.archives-ouvertes.fr/hal-01410525/document>>. Acesso em: 02 maio 2017.

BAPTISTA, A. A. et al. Comunicação científica: o papel da Open Archives Initiative no contexto do Acesso Livre. 2007. Disponível em: <<https://periodicos.ufsc.br/index.php/eb/article/viewFile/1518-2924.2007v12nesp1p1/435>>. Acesso em: 21 maio 2017.

FAGUNDES, V. O.. Ciência aberta e bactérias extraterrestres: Transparência e colaboração na produção do conhecimento. **Liinc em Revista**, v. 10, n. 2, 2014. Disponível : <<http://revista.ibict.br/liinc/article/view/3574/3062>>. Acesso em: 02 maio 2017.

KRAKER, P. et al. Research Data Explored II: the Anatomy and Reception of figshare. **ArXiv**. Disponível em: <<https://arxiv.org/abs/1503.01298>>. Acesso em: 21 maio 2017.

SAYÃO, L. F.; SALES, L. F. Dados abertos de pesquisa: ampliando o conceito de acesso livre. **Revista Eletrônica de Comunicação, Informação & Inovação em Saúde**, [S.l.], v. 8, n. 2, jun. 2014. ISSN 1981-6278. Disponível em: <<http://www.reciis.icict.fiocruz.br/index.php/reciis/article/view/611/1252>>. Acesso em: 02 maio 2017.

SINGH, J. FigShare. **Journal of Pharmacology and Pharmacotherapeutics**, n. 2, v.2, abr./jun. 2011. Disponível em: <www.jpharmacol.com>. Acesso em: 21 maio 2017. DOI: 10.4103/0976-500X.81919.



AS PRÁTICAS DE INDEXAÇÃO E AS SUAS INTERFACES COM OS REPOSITÓRIOS DIGITAIS DE ACESSO ABERTO: uma abordagem sobre o metadado assunto no Repositório Institucional da UFRN

Raimunda Fernanda Santos

RESUMO: Estuda aspectos relativos à representação temática da informação no contexto dos Repositórios Digitais. Objetiva analisar a representação da informação em tais ambientes. Especificamente, visa analisar as implicações da indexação por extração no processo de representação do conteúdo dos documentos no Repositório Institucional da Universidade Federal do Rio Grande do Norte (RI/UFRN); analisar os descritores utilizados na indexação das produções científicas no metadado assunto da Repositório em questão. Utiliza como metodologia a pesquisa bibliográfica pautada em revisão de literatura sobre indexação e representação da informação em repositórios digitais e pesquisa exploratória com abordagem qualitativa mediante a análise da indexação realizada no metadado assunto do RI/UFRN. Destaca que os termos atribuídos em linguagem natural viabilizam implicações negativas no processo de representação temática dos documentos em Repositórios Digitais. Identifica a presença de algumas implicações mediante a análise qualitativa do metadado "assunto" do Repositório Digital, dentre elas: emprego de frases, termos com polissemia, presença de descritores abrangentes e com multiplicidade de sentidos, abreviações nos termos, uso de siglas e entre outras. Conclui enfatizando a importância dos gestores dos Repositórios Institucionais de acesso aberto elaborarem uma política de indexação que vise o aperfeiçoamento das práticas já utilizadas nesses ambientes incluindo o auxílio de linguagens documentárias e vocabulários controlados que possam elevar o nível da qualidade de descrição desses documentos visando a precisão no processo de busca e recuperação da informação.

Palavras-chave: Representação Temática da Informação. Indexação por extração. Repositórios Institucionais.

THE INDEXING PRACTICES AND THEIR INTERFACES WITH THE OPEN ACCESS DIGITAL REPOSITORIES: an approach on the metadata subject in the Institutional Repository of UFRN

ABSTRACT: *It studies aspects related to thematic representation of information in the context of Digital Repositories. It aims to analyze the representation of information in such environments. Specifically, it aims to analyze the implications of indexation by extraction in the process of representation of the content of the documents in the Institutional Repository of the Federal University of Rio Grande do Norte (RI / UFRN); Analyze the descriptors used in the indexation of scientific productions in the subject metadata of the Repository in question. It uses as a methodology the bibliographic research based on literature review on indexing and representation of information in digital repositories and exploratory research with qualitative approach through the analysis of the indexation carried out in the subject metadata of the RI / UFRN. It emphasizes that the terms assigned in natural language enable negative implications in the process of thematic representation of documents in Digital Repositories. It identifies the presence of some implications through the qualitative analysis of the "subject" metadata of the Digital Repository, among them: use of terms with polysemy, presence of comprehensive descriptors and multiplicity of meanings, abbreviations in the terms, use of acronyms and others. It concludes by emphasizing the importance of the managers of the Institutional Repositories to elaborate an indexation policy that aims to improve the practices already used in these environments including the aid of documentary languages and controlled vocabularies that can raise the level of the quality of description of these documents aiming the accuracy in the process of Search and retrieval of information*

Keywords: *Thematic Representation of Information. Indexing by extraction. Institutional repositories.*

1 INTRODUÇÃO

A explosão da informação e o advento das novas formas de registrá-la culminaram na complexidade do trabalho de identificação e acesso à informação. Nesse entendimento, tornar as informações acessíveis para os usuários que delas necessitam constituiu-se uma das preocupações centrais da Ciência da Informação. Assim sendo, para que os registros documentais sejam recuperados, é necessária a realização de um conjunto de procedimentos que visam facilitar o seu acesso pela atual e futura geração.

Nessa perspectiva, a indexação consiste na descrição do conteúdo de um documento, de forma concisa e condensada, por meio do emprego de termos - também denominados como palavras-chave ou descritores - que exercem a função de pontos de acesso mediante os quais um documento pode ser identificado e recuperado. Destarte, para a realização dessa atividade existem as linguagens documentárias/vocabulários controlados os quais são instrumentos tradicionais de representação da informação que objetivam facilitar a comunicação por meio da padronização de termos para a descrição dos conteúdos dos documentos.

Contudo, diante do crescente aumento na produção de documentos e do desenvolvimento de produtos e serviços que beneficiam a participação dos usuários na coordenação de ações de produção, organização, representação e disseminação da informação na *Web*¹, é evidenciada uma nova forma de organização e representação das informações no âmbito dos Repositórios Digitais em que não se adotam regras de indexação envolvendo controle de vocabulário na descrição dos recursos.

Sendo assim, em linhas gerais, observa-se que na atividade de indexação das produções científicas em Repositórios Digitais Institucionais é permitida a utilização dos mesmos termos empregados pelo autor para a apresentação das suas ideias no texto (ou seja, linguagem natural) para a indexação do seu trabalho no Repositório - selecionando-se assim as palavras-chave empregadas no resumo do documento

¹ A *World Wide Web* (ou simplesmente *Web*) está relacionada à palavra inglesa cujo significado é “rede de alcance mundial”. É um sistema de informações interligado e executado na Internet.

para representá-lo, configurando-se como uma indexação por extração de acordo com Lancaster (2004).

Dessa forma, o uso da indexação por extração e as suas implicações, como demais questões que daí decorrem, oferecem inúmeras possibilidades de discussões, uma delas refere-se à sua utilização no contexto dos Repositórios Digitais como o Repositório Institucional da Universidade Federal do Rio Grande do Norte (RI/UFRN).

O RI/UFRN objetiva promover o acesso e uso a produção intelectual da comunidade universitária (docentes, técnicos e discente de pós-graduação) da Universidade, bem como divulga e preserva parte da memória acadêmica da instituição, uma vez que a sua missão consiste em armazenar, preservar e disponibilizar na internet, textos completos de acesso livre. Sob esse viés, esse Repositório se insere internacionalmente no crescente movimento de constituição de Repositórios Digitais de acesso aberto nos moldes da *Open Access Initiative* (Iniciativa de Acesso Aberto). Para tanto, a sua política de indexação permite a utilização dos termos livres informados pelos autores das produções científicas, cujos descritores podem ser as palavras-chave extraídas dos resumos dos seus trabalhos para representar o conteúdo das produções científicas - configurando-se como uma indexação por extração mediante o emprego de termos livres para representar a informação.

Sob esse viés, a presente pesquisa tem como objetivo geral analisar a representação da informação em Repositórios Digitais. Especificamente, visa analisar as implicações da indexação por extração no processo de representação do conteúdo dos documentos no Repositório Institucional da UFRN; avaliar os descritores utilizados na representação do conteúdo das produções científicas no metadado² assunto do Repositório Institucional em questão visando verificar como os autores ((docentes, técnicos e discente de pós-graduação) estão descrevendo o conteúdo de suas produções intelectuais nos Repositórios Digitais.

A importância dessa investigação decorre, em âmbito geral, das tendências de representação da informação em Repositórios Digitais e uma das principais justificativas para a realização desta pesquisa é o fato da representação do assunto,

² Dado que “descreve informações de um recurso informacional em meio digital ou não digital” (ROSETTO, 2003, p.9).

nessas ferramentas de acesso aberto como o Repositório Institucional da UFRN, ocorrer de forma livre e sem padronização.

Nessa concepção, a escolha dessa fonte de informação é relativa à importância que ela representa para a comunidade acadêmica da instituição, haja vista que é por meio dessa ferramenta depositária que é armazenada, preservada e divulgada parte da memória acadêmica da Universidade Federal do Rio Grande do Norte. Outrossim, a seleção desse universo está atrelada à sua política de indexação cuja atividade de indexação orientada foi mencionada anteriormente.

Torna-se relevante também pesquisar aspectos relativos à representação da informação no Repositório Digital supracitado levando em conta que ela consiste em uma ferramenta dinâmica que subsidia alunos, professores e pesquisadores a encontrar as produções acadêmicas construídas no âmbito da instituição, possibilitando também a preservação histórica dos documentos nesse contexto.

2 AS PRÁTICAS DE INDEXAÇÃO E AS SUAS INTERFACES COM OS REPOSITÓRIOS DIGITAIS DE ACESSO ABERTO

Uma das preocupações principais na Ciência da Informação consiste em tornar a informação acessível para aqueles que dela necessitam haja vista que os itens informacionais devem passar por um conjunto de tarefas de armazenamento, organização e representação para que seja possível a sua recuperação em tais sistemas (acesso).

Sob esse viés, Guimarães (2009, p. 1) afirma que no âmbito da organização da informação dois universos se descortinam: "o primeiro, ligado ao acesso aos documentos e o segundo, de natureza mais complexa, voltado para o acesso ao conteúdo informacional, genericamente denominado de tratamento temático da informação" ou representação temática da informação, como é designada na literatura da área de organização da informação.

Nessa perspectiva, verifica-se que o documento deve ser tratado em seu todo levando em consideração tanto as operações que contemplam a descrição física (tratamento descritivo ou representação descritiva), como a descrição do conteúdo temático (tratamento temático ou representação temática) haja vista que a diferenciação entre a forma e o conteúdo "[...] reside na busca do *o que*

(materialização) e do *sobre o que* (teor) que convivem no âmbito do documento" (GUIMARÃES, 2009, p. 1, grifo nosso).

Nesse entendimento, a representação temática da informação centra-se nas questões relativas "à análise, descrição e representação do conteúdo dos documentos, bem como suas inevitáveis interfaces com as teorias e sistemas de armazenamento e recuperação da informação" (BARITÉ, 1999, p. 124). A indexação, por sua vez, consiste em uma atividade integrante da representação temática da informação e que apresenta etapas de caráter integrativo.

Sob esse viés, a indexação consiste em uma atividade integrante do tratamento temático da informação, ou seja, uma "combinação metodológica altamente estratégica entre o tratamento do conteúdo do documento e a sua recuperação por um usuário", demonstrando uma relação direta entre o processo e a finalidade da indexação, visando a disponibilização de informações pertinentes às necessidades dos usuários (FUJITA, 2003, p. 61).

Mediante tais considerações, de acordo com Dias e Naves (2007), a indexação consiste na ação de descrever um recurso em relação ao seu conteúdo por meio de termos (também denominados palavras-chave ou descritores). Nesse entendimento, em um Repositório Digital, os termos são pontos de acesso para os documentos, haja vista que é possível recuperar esses objetos informacionais sob qualquer uma das suas palavras-chave (que foram empregadas no sistema no ato da sua representação) ou até mesmo por meio da combinação delas. Em outras palavras, esses termos são atribuídos pelo indexador (pessoa responsável por realizar a atividade de indexação) e exercem a função de pontos de acesso mediante os quais um documento é identificado e recuperado.

Apesar dos avanços tecnológicos, observa-se a necessidade da presença de um indexador humano nessa tarefa levando em conta que a indexação consiste em uma atividade intelectual considerada como o limite da máquina, tendo em vista que a máquina ainda não pode realizar o trabalho humano de abstrair, distinguir, compreender e assimilar o que está presente em um texto. Portanto, todas as fases do processo de indexação sofrem influências de fatores linguísticos, cognitivos e lógicos (simultaneamente), os quais são designados por Dias (2001) como o tripé de sustentação do processo de análise de assunto - gerando assim um caráter interdisciplinar.

Lancaster (2004) realiza a distinção entre *indexação por extração* (também conhecida como *indexação derivada*) e *indexação por atribuição*. Para esse autor, na *indexação por extração* as palavras ou expressões empregadas pelo autor no documento analisado são selecionadas pelo indexador para representar o seu conteúdo. Em outras palavras, nesse tipo de indexação empregam-se os mesmos termos utilizados pelo autor para a apresentação das suas ideias no texto (ou seja, uma linguagem natural), selecionando-se assim palavras encontradas no título, nas palavras-chave ou na própria estrutura textual do documento. Já a *indexação por atribuição* compreende a atribuição de descritores ao documento utilizando-se uma linguagem originária de uma fonte que não é o próprio documento (LANCASTER, 2004).

Sendo assim, são determinados os cabeçalhos a serem empregados objetivando-se controlar sinônimos, diferenciar homógrafos³ e ligar termos por meio de relações hierárquicas, associativas e de equivalência (ou seja, utiliza-se uma linguagem artificial/linguagem controlada). De acordo com Lancaster (2004, p.19), "a indexação por atribuição envolve o esforço de representar a substância da análise conceitual mediante o emprego de termos extraídos de alguma forma de vocabulário controlado".

Diante do exposto, para esse autor, a linguagem natural configura-se como discurso comum, ou seja, é uma linguagem usualmente empregada na escrita e na fala, diferentemente das linguagens documentárias ou vocabulários controlados que consistem em linguagens construídas de acordo com as necessidades dos usuários. Analogicamente, no âmbito dos Repositórios Digitais, os termos presentes no título, resumo e palavras-chave dos documentos apresentam linguagem natural.

Destarte, Lopes (2002, p.48) apresenta em sua pesquisa as vantagens e desvantagens da linguagem natural, as quais são elencadas no Quadro 1 a seguir.

Quadro 1- Linguagem natural: vantagens e desvantagens

VANTAGENS	DESVANTAGENS
Permite o imediato registro da informação em um repositório digital, sem necessidade de consulta a uma linguagem de controle.	Os usuários da informação, no processo de busca, precisam fazer um esforço intelectual maior para identificar os sinônimos, as grafias alternativas, os homônimos etc.
O processo de busca é facilitado com a ausência de treinamentos específicos no uso de uma linguagem de controle.	Haverá alta incidência de respostas negativas ou de relações incorretas entre os termos

³ Termos que coincidem na sua forma, porém representam conceitos diferentes.

	usados na busca (por ausência de padronização).
Os termos de entrada de dados são extraídos diretamente dos documentos que vão constituir o Repositório Digital.	Os custos de acesso tendem a aumentar com a entrada de termos de busca aleatórios.
Os temas específicos citados nos documentos podem ser encontrados.	Uma estratégia de busca que arrole todos os principais conceitos e seus sinônimos deve ser elaborada para ferramenta (ex.: nomes comerciais de substâncias químicas não ocorrem no <i>Chemical Abstracts</i>).
Elimina os conflitos de comunicação entre os indexadores e os usuários, pois ambos terão acesso aos mesmos termos.	Perda de confiança do usuário em uma possível resposta negativa.

Fonte: Lopes (2002, p.48).

Nesse sentido, o emprego de termos livres na indexação de documentos no âmbito dos Repositórios Digitais pode incitar algumas implicações negativas no processo de representação da informação, levando em consideração que as informações disponíveis em tais ambientes são descritas com palavras provenientes da linguagem natural sem o auxílio de um instrumento de representação da informação. Para tanto, os usuários representam conteúdos com termos que usam e/ou acreditam que entendem, sendo assim alguns termos podem fazer sentido apenas a quem os elencou e nem sempre para a comunidade ou grupo de pessoas que utilizam o sistema.

Sob esse viés, depreende-se que com a liberdade da atribuição dos descritores um mesmo termo pode ter diferentes significados para aqueles que atribuíram as etiquetas/descriptores para fins de representação do conteúdo.

Noruzi (2007) em sua pesquisa menciona algumas implicações negativas dessas práticas no processo de representação da informação em ambientes digitais, dentre as quais: **a) ocorrência de plurais** - ausência de padronização em relação aos termos com flexões de número (singular ou plural): palavras com flexões gramaticais e variações linguísticas de gêneros nos descritores utilizados, por exemplo; **b) existência de polissemia**: palavra que representa mais de um significado; **c) emprego de sinonímia**: palavras distintas que têm o mesmo significado; **d) profundidade/ especificidade da descrição**: refere-se à especificidade dos descritores elencados pelo usuário no processo de representar o conteúdo do objeto informacional em tags. Tais descritores (em sua maioria) podem não representar o objeto em sua totalidade, sendo elencados apenas para interesse particular do usuário que o atribuiu.

Além das implicações supracitadas, Amstel (2007), Gouvêa; Loh (2007) e Catarino; Baptista (2009); Santos (2013) enfatizam que em ambientes que utilizam a representação colaborativa da informação é comum a **presença de descritores homógrafos** - os quais são termos que coincidem na sua forma, porém representam conceitos diferentes; **erros ortográficos**; **expressões regionais** (que são empregadas como *tags* de acordo com o contexto social/regional em que o usuário estar inserido); **etiquetas em diferentes idiomas e com letras maiúsculas e minúsculas** (sem padronização na descrição) e **etiquetas assimétricas** (*tags* que não se relacionam entre si) e abreviações.

Outrossim, alguns termos atribuídos são dotados de **ambiguidade morfológica**, ou seja, "pertencem a mais de uma categoria gramatical, como proposta, que pode ser substantivo ou adjetivo ou verbo" (BRASCHER, 2002, p.4) impossibilitando classificá-los pelo fato de ter formas mínimas portadoras de significado concorrentes para uma mesma palavra.

Nessa concepção, para fins de indexação existem as linguagens documentárias/vocabulários controlados os quais são instrumentos que visam facilitar a comunicação mediante a padronização de termos para a descrição de conteúdos. Diante disso, verifica-se que tais instrumentos objetivam controlar as dispersões lexicais existentes nas linguagens naturais orientando também o indexador no que concerne à escolha dos termos para representar o assunto dos documentos, bem como indicar os melhores termos para que ele possa realizar a indexação do texto, atuando também como instrumento de comunicação entre a linguagem dos usuários e a linguagem do sistema.

Lancaster (2004) enfatiza que o tipo de vocabulário controlado (esquema de classificação, cabeçalhos de assuntos, tesouro) não é o aspecto mais importante a influenciar na etapa de tradução da indexação. Muito mais importante é a sua *abrangência* (alcance) e a sua *especificidade*. Porém, de acordo com Café e Sales (2009), no que concerne à representação temática da informação por meio de um controle terminológico, os tesouros e as ontologias⁴ são considerados como os principais instrumentos utilizados e estudados no âmbito da Ciência da Informação.

⁴ Ontologias são instrumentos de indexação que apresentam uma estrutura de termos e as relações entre eles na perspectiva do sujeito e da linguagem de um determinado domínio. Apresentam-se como alternativas viáveis para organizar o conhecimento em ambiente Web na forma de uma teia de relações permitindo a ligação entre os conceitos.

Tais instrumentos possibilitam a realização da representação temática intimamente relacionada com a ideia do autor do documento a ser indexado; a ampliação dos métodos de busca, bem como o controle dos termos que apresentam polissemia e sinonímia; a proximidade entre a linguagem natural (do usuário) e a linguagem adotada pelo sistema de recuperação de informação.

A partir de tais considerações, observa-se que o tratamento temático da informação, mediante o uso dos processos e instrumentos desenvolvidos em sistemas tradicionais, consiste em uma tarefa indispensável para a recuperação da informação em ambientes digitais. Sob esse viés, se faz necessário atentar-se para as práticas de indexação realizadas no âmbito dos Repositórios Digitais Institucionais como o Repositório Institucional da UFRN.

Nesse segmento, levando em consideração os aspectos teóricos enfatizados anteriormente e objetivando revisitar tais questões na prática, será apresentada a seguir os resultados da análise do metadado assunto do Repositório Institucional da Universidade Federal do Rio Grande do Norte, cuja política de indexação adotada foi mencionada anteriormente.

3 METODOLOGIA

Quanto aos objetivos a serem alcançados para este estudo, foi empreendida uma pesquisa do tipo exploratória. Quanto aos meios, os procedimentos metodológicos foram empreendidos a partir de uma pesquisa bibliográfica, baseada em revisão de literatura sobre as implicações do emprego de termos provenientes de linguagem natural no processo de representação da informação. Foram utilizadas na concepção do trabalho fontes impressas como livros, e em meio digital como artigos de periódicos científicos eletrônicos da área de Ciência da Informação encontradas por meio de bases de dados, mecanismos de busca e bibliotecas digitais na Web.

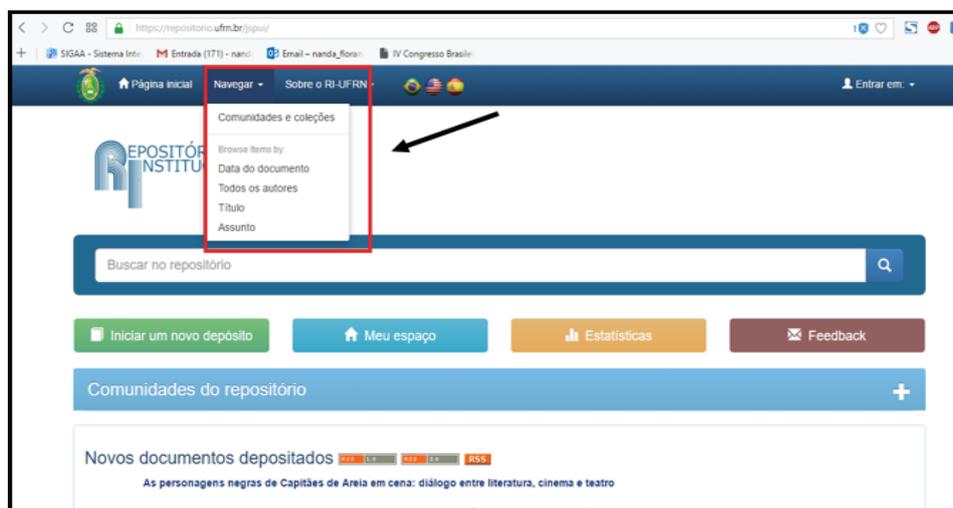
Além disso, o presente estudo possui caráter qualitativo tendo em vista que foi feita uma investigação pela literatura para seu melhor entendimento, bem como foram analisados, essencialmente, alguns aspectos qualitativos referentes à representação da informação no metadado assunto das produções científicas depositadas nas comunidades do RI/UFRN. Para tanto, a análise qualitativa aleatória do metadado em questão foi realizada mediante a verificação das palavras-chave empregadas pelos autores das produções científicas armazenadas no Repositório em questão.

Até o dia 15 de maio de 2017 o Repositório Institucional da UFRN apresentou 10.769 (dez mil, setecentos e sessenta e nove) produções científicas depositadas no sistema.

4 ANÁLISE DO METADADO ASSUNTO DO REPOSITÓRIO INSTITUCIONAL DA UFRN (RI/UFRN)

Após a análise do repositório em evidência verificou-se que o processo de busca e recuperação das monografias neste ambiente ocorre de maneira clara e simples, haja vista que são disponibilizados alguns mecanismos de acesso em sua página principal, tais como: busca por comunidades e coleções, data do documento, nome dos autores, título e assunto - conforme pode ser visualizado na Figura 1.

Figura 1 - Página principal do RI/UFRN



Fonte: < <https://repositorio.ufrn.br/jspui/> >

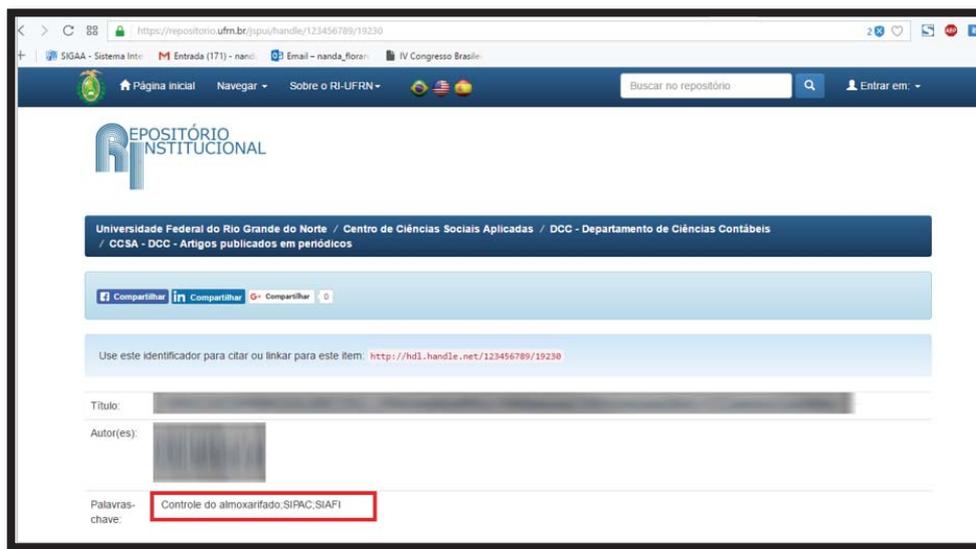
Ao clicar na opção de navegação “Comunidades e coleções” é possível visualizar a lista de comunidades, subcomunidades e coleções presentes no RI/UFRN. Em tais comunidades são armazenados os seguintes documentos: a) Artigos publicados em periódicos científicos; b) Trabalhos apresentados em eventos; c) Dissertações e Teses defendidas em outras instituições; d) livros eletrônicos; e) Capítulos de livros; f) BDTD (Biblioteca Digital de Teses e Dissertações).

Através do interesse em compreender e visualizar aspectos práticos relativos às implicações da participação dos usuários na representação da informação, considerou-se importante realizar uma análise qualitativa de representação da informação no metadado “assunto” do RI/UFRN por meio da análise das palavras-chave empregadas pelos autores das produções científicas incluídas no Repositório.

Nesse segmento, esta tarefa possibilitou pesquisar também como os autores estão representando os conteúdos de suas produções intelectuais através da Folksonomia⁵/ livre representação da informação.

É importante ressaltar que nas figuras abaixo foram resguardados dados como os nomes dos autores e os respectivos orientadores das suas monografias visando uma análise ética da representação de conteúdos nesse contexto.

Figura 3 – Ocorrência de siglas em palavras-chave no RI/UFRN



Fonte: < <https://repositorio.ufrn.br/jspui/community-list>>.

Conforme é possível observar na Figura 3, o autor do documento em questão empregou três termos para representar o conteúdo da produção intelectual depositada no âmbito do RI/UFRN, cujos descritores foram: “Controle do almoxarifado”; “SIPAC” e “SIAFI”. Portanto, verificou-se o emprego de abreviações de palavras (por meio das siglas “SIPAC” e SIAFI”), as quais dizem respeito às seguintes expressões - respectivamente: Sistema Integrado de Patrimônio, Administração e Contratos; e Sistema Integrado de Administração Financeira. Destarte, a aplicação de siglas e abreviações em descritores podem implicar no processo de recuperação dos documentos no âmbito dos Repositórios Institucionais, uma vez que elas são

⁵ De acordo com Santos e Corrêa (2015, p.7) a Folksonomia consiste no resultado do processo de etiquetagem livre (atribuição de etiquetas, palavras-chave) realizada pelos usuários mediante o emprego de termos provenientes de linguagem natural - dispensando o auxílio de vocabulários controlados- em ambientes digitais colaborativos visando indexar recursos informacionais compartilhados em qualquer formato (textos, imagens, áudio, vídeo etc.) para fins de sua representação.

originalmente empregadas para reduzir palavras/termos – sem que os mesmos sejam descritos na íntegra.

Os metadados são considerados elementos que visam proporcionar o desenvolvimento de sistemas de recuperação de informação, permitindo a interoperabilidade entre eles e o acesso mais amplo às informações disponibilizadas em diversos ambientes. Nesse entendimento, os metadados auxiliam na descrição, identificação, localização e acesso dos documentos. O metadado assunto, em especial, é destinado à representação do conteúdo dos recursos digitais possibilitando a inclusão de palavras-chave (termos, descritores) para fins de indexação e recuperação da informação. Entretanto, mediante as análises realizadas, foi identificado o emprego de frase e de sigla no metadado assunto do RI/UFRN, conforme pode ser visualizado na Figura 4 a seguir:

Figura 4 – Ocorrência de frases e sigla em palavras-chave no RI/UFRN



Fonte: < <https://repositorio.ufrn.br/jspui/community-list>>.

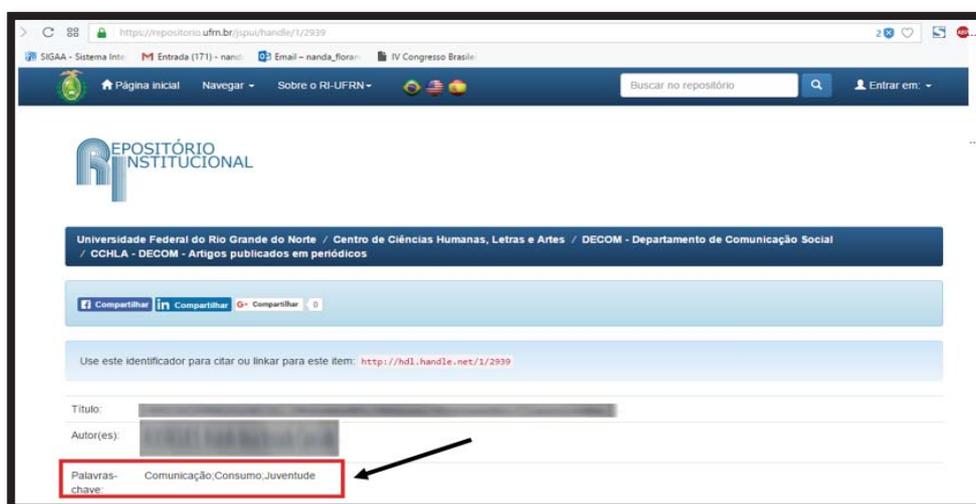
Como é perceptível na Figura 4, o autor do trabalho em questão elencou a frase “Convenção nº 102 da OIT⁶ sobre normas mínimas de seguridade social” para representar a sua produção intelectual, bem como os termos “Seguridade Social” e “Direito Comparado”. Nesse segmento, verifica-se que o emprego de frases e sigla no metadado assunto dos sistemas de recuperação da informação pode implicar em problemas de revocação e precisão, as quais são medidas relativas à probabilidade dos documentos recuperados serem compatíveis com aquilo que os usuários estão

⁶ Sigla destinada a abreviação do termo “Organização Internacional do Trabalho” - (OIT).

buscando em um sistema de informação. Dessa forma, a qualidade das palavras-chave atribuídas, bem como a terminologia adotada são fundamentais para o sucesso de qualquer sistema, independentemente de sua forma e estrutura.

De outro modo disposto, verificou-se o emprego de termos generalistas no âmbito do metadado assunto do RI/UFRN, conforme é apresentado na Figura 5 a seguir.

Figura 5 – Ocorrência de termos generalistas em palavras-chave no RI/UFRN

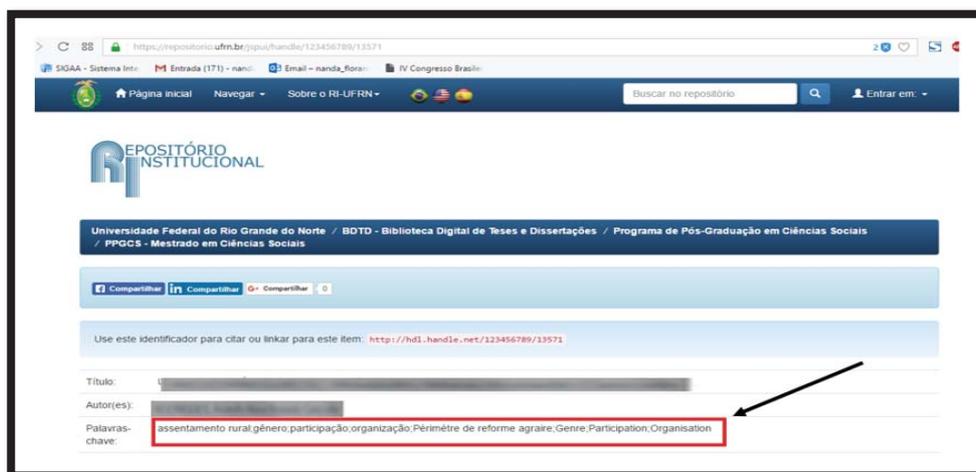


Fonte: < <https://repositorio.ufrn.br/jspui/community-list>>.

Nesse entendimento, outro elemento vinculado à linguagem natural (de acordo com o que foi abordado nos aspectos teóricos deste estudo) é o emprego de termos generalistas ou com multiplicidade de sentidos. Na figura anterior verificou-se a ocorrência desse fenômeno nos termos “Comunicação”, “Consumo” “e”, “Juventude”. Assim, percebe-se que estes termos poderiam estar inseridos em diversos contextos sob enfoques e filtros diversos, podendo resultar em implicações negativas no processo de busca no Repositório Institucional.

Mediante as análises realizadas, verificou-se também a ocorrência de polissemia em palavras-chave no RI/UFRN:

Figura 6 – Ocorrência de polissemia em palavras-chave no RI/UFRN



Fonte: < <https://repositorio.ufrn.br/jspui/community-list>>.

Como é perceptível na Figura 6, o autor do trabalho monográfico apresentado anteriormente elencou 8 (oito) palavras-chave/descriptores para representarem a sua produção intelectual, dentre elas palavras em língua estrangeira e o descritor “Organização” que consiste em uma polissemia – tendo em vista que esse termo está vinculado a, pelos menos, dois significados - são eles:

- **Organização** → Ordem = Ordenação/Arrumação
Associação = Entidade/Instituição

Nessa concepção, infere-se que a representação do conteúdo de um documento por uma palavra polissêmica (como apresentado na Figura 6) e a ausência de controle desse fenômeno pode resultar em problemas na sua recuperação. Assim, ao fazer uma busca com o termo “Organização”, o Repositório Digital pode recuperar para o usuário documentos sobre entidade/instituição, como também outros trabalhos monográficos que abordem aspectos relativos à ordenação/arrumação de algo.

Nessa perspectiva, o desenvolvimento das tecnologias exige do profissional da informação além de conhecimentos específicos na sua área, outros conhecimentos e habilidades relativas ao tratamento da informação em suportes e ambiências diversificadas como os Repositórios Digitais esboçando assim novos campos de competência profissional.

Sendo assim, no momento da representação do conteúdo do documento, se o autor deixar de aplicar um termo mais específico para representar o assunto, empregando dessa forma um termo generalista ao conteúdo temático, provavelmente ocorrerão falhas na recuperação da informação.

Sob esse entendimento, a aplicação de descritores genéricos na representação da informação é uma questão que precisa ser observada/analísada, uma vez que estes termos também servirão de base para a comunidade da área de conhecimento em questão (que por sua vez possui terminologia específica) para fins de recuperação da informação e, possivelmente, auxílio para o desenvolvimento de outras pesquisas.

Por essa razão, a ausência de controle dos fenômenos elencados anteriormente pode resultar em problemas na recuperação das produções científicas em sistemas de recuperação de informação. Dessa forma, as palavras-chave atribuídas, bem como a terminologia adotada são fundamentais para o sucesso de qualquer sistema, independentemente de sua forma e estrutura.

5 CONSIDERAÇÕES FINAIS

Mediante tais considerações, a partir dos estudos realizados, observou-se nesta pesquisa que os termos atribuídos em linguagem natural viabilizam implicações no processo de representação e recuperação dos documentos em ambiente digital. Destarte, na análise qualitativa do metadado “assunto” do Repositório Institucional da Universidade Federal do Rio Grande do Norte verificou-se a presença de algumas implicações, dentre elas: emprego de frases, termos com polissemia, ocorrência de erros ortográficos, presença de descritores abrangentes e com multiplicidade de sentidos, abreviações nos termos, uso de siglas e entre outras.

Excluir os termos provenientes da coleta automática das palavras-chave empregadas na indexação das produções científicas do Repositório Institucional para adicionar outra forma de representação temática pode significar um risco para esse sistema. Dessa forma, observa-se a importância de manter os mesmos descritores informados pelos autores nas palavras-chave dos resumos dos seus documentos no metadado "assunto" do Repositório. Contudo, se faz necessário que o usuário administrador (ou indexador do sistema) analise e corrija eventuais erros ortográficos existentes nesses termos, bem como adicione remissivas para eles (com descritores semelhantes àqueles já sugeridos pelos usuários ou explicitamente identificados como equivalentes disponíveis no instrumento de controle terminológico a ser associado ao sistema). Além disso, com essas adequações propostas o sistema pode

permitir a correção automática de erros ortográficos dos descritores coletados ou atribuídos para a indexação de uma produção científica.

Mediante tais considerações, infere-se a importância dos gestores do RI/UFRN (e dos demais Repositórios Digitais Institucionais que adotam práticas semelhantes no que concerne à indexação dos seus documentos) elaborarem uma política de indexação que vise o aperfeiçoamento das práticas já utilizadas neste ambiente incluindo o auxílio de linguagens documentárias e vocabulários controlados que podem elevar o nível da qualidade de descrição do conteúdo desses documentos visando, conseqüentemente, a precisão no processo de busca e recuperação da informação neste ambiente.

Nessa perspectiva, a política de indexação deve ser considerada como um elemento de grande importância não somente para designar as práticas de representação da informação, como também para servir de auxílio nas decisões administrativas dos Repositórios Digitais Institucionais. Desse modo, diante das implicações elencadas anteriormente, torna-se relevante que os administradores do Repositório Digital em questão (bem como os administradores de repositórios digitais que adotam políticas semelhantes) repensem o processo de indexação dos documentos nessas ferramentas de acesso aberto.

Tais ações podem ser voltadas para a preocupação na construção de modelos colaborativos de indexação social, os quais são descritos por Santos (2016)⁷ como modelos que delineiam atividades colaborativas de representação de conteúdo a serem executadas pelos usuários - como estratégias de indexação dos conteúdos dos recursos em um sistema - e servem de referência para o desenvolvimento de sistemas colaborativos de indexação, podendo ser adaptados ou aperfeiçoados a diversos contextos. Para a autora citada anteriormente, a alternativa do emprego de um modelo colaborativo de indexação é controlar o nível de liberdade do usuário autor do documento ao atribuir termos na indexação de conteúdos em Repositórios Digitais para que o processo de iteratividade possa resultar na reconstrução do conhecimento de forma coletiva e moderada - permitindo uma nova estrutura de organização e

⁷ Mais informações acerca dos modelos colaborativos de indexação social ler a Dissertação Referenciada a seguir: SANTOS, Raimunda Fernanda dos. **Modelos colaborativos de indexação social e a sua aplicabilidade na Base de Dados Referencial de Artigos de Periódicos em Ciência da Informação (BRAPCI)**. 184f. Dissertação (Mestrado em Ciência da Informação) – Universidade Federal de Pernambuco, Recife, 2016.

representação de informação mais consistente para o domínio em que o sistema está inserido. Outrossim, essas propostas permitem que os termos/descriptores tenham uma relação maior no que concerne ao conteúdo do objeto depositado, principalmente dentro do contexto em que elas estão sendo empregadas.

Nessa perspectiva, verifica-se que a percepção da atuação dos usuários enquanto mentes tradutoras e idealizadoras de novos arranjos e categorizações informacionais, desafia os profissionais da informação a repensarem a construção de sistemas e metodologias frente à representação da informação em ambiente digital, sobretudo no âmbito dos Repositórios Digitais que se constituem como ferramentas depositárias e disseminadoras de informação.

Nesse entendimento, necessita-se que novos processos, metodologias e instrumentos venham a ser desenvolvidos no âmbito dos Repositórios Digitais - seja aplicando técnicas para a ampliação ou aperfeiçoamento dos termos fornecidos pelos autores no metadado assunto dessas ferramentas, seja adicionando os termos oriundos das linguagens documentárias. Sendo assim, tais reflexões se constituem como objetivos de estudo dos trabalhos atuais e futuros da pesquisadora.

Em linhas gerais, acredita-se que o presente trabalho é relevante no contexto da Ciência da Informação uma vez que aborda aspectos representação da informação no âmbito dos Repositórios Digitais e a relaciona com os desafios do profissional da informação frente às novas possibilidades de representação da informação, tendo em vista que é possível aperfeiçoar as técnicas já utilizadas, além de adaptá-las às necessidades da sociedade atual.

REFERÊNCIAS

AMSTEL, Van Frederick. **Folksonomia**: vocabulário descontrolado na arquitetura da informação ou samba do crioulo doido. 2007. Disponível em: <http://www.guilhermo.com/ai_biblioteca/referencia.asp?referencia=317> . Acesso em: 20 maio 2015.

BARITÉ, M. **Formación de recursos humanos en el área de información en el Mercosur**: compatibilización curricular y competencias del profesional de la información en el Mercosur. Santiago, Chile: Universidad Tecnológica Metropolitana, 1999, p.121-128.

BOCCATO, V. R. C. O Contexto sociocognitivo do indexador no processo de representação temática da informação. **Encontros Bibli**: Revista Eletrônica de Biblioteconomia e Ciência da Informação, Florianópolis, v.17, n.esp., 2012.

BRANDT, Mariana Baptista ; MEDEIROS, Marisa Brascher Basílio. Folksonomia: esquema de representação do conhecimento? **TransInformação**, Campinas, v.22, n.2, p.111-121, maio/ago., 2010. Disponível em: < <http://periodicos.puc-campinas.edu.br/seer/index.php/transinfo/article/view/489>>. Acesso em: 13 out. 2013.

BRASCHER, M. A ambiguidade na recuperação da informação. **DataGramZero - Revista de Ciência da Informação**, Rio de Janeiro, v.3, n.1, fev. 2002.

CAFÉ, L.; SALES, R. de. Tesouros e ontologias sob o olhar da teoria comunicativa da terminologia. In: CONGRESS ISKO-SPAIN, 9., 2009, Valência. **Anais...** Valência, 2009.

CATARINO, M.; BAPTISTA, A. Folksonomias: características das etiquetas na descrição de recursos da web. **Inf. Inf.**, Londrina, v.14, n. esp., p.46-67, 2009.

DIAS, E. W. Contexto digital e tratamento da informação. **DataGramZero**, Rio de Janeiro, v. 2, n. 5, out. 2001. Disponível em: < http://www.dgz.org.br/out01/Art_01.htm>. Acesso em: 25 set. 2014.

DIAS, E. W.; NAVES, M. M. L. **Análise de assunto**: teoria e prática. Brasília: Thesaurus, 2007.

FUJITA, M. S. L. A identificação de conceitos no processo de análise de assunto para indexação. **Revista Digital de Biblioteconomia & Ciência da Informação**, Campinas, v. 1, n. 1, p. 60-90, jul./dez. 2003. Disponível em: <http://server01.bc.unicamp.br/revbib/artigos/art_5.pdf>. Acesso em: 26 jun. 2016.

GOUVÊA, C.; LOH, S. Folksonomias: identificações de padrões na seleção de tags para descrever conteúdos. **Revista Eletrônica de Sistemas de Informação**, [S.l.], v. 11, n.2, 2007.

GUIMARÃES, J. A. C. **Abordagens teóricas de Tratamento Temática da Informação (TTI)**: catalogação de assunto, indexação e análise documental. Ibersid, 2009, p.105-117.

LANCASTER, F. W. **Indexação e resumos**: teoria e prática. 2.ed. Brasília: Briquet de Lemos, 2004.

LOPES, I. L. Uso das linguagens controlada e natural em base de dados: revisão de literatura. **Ci. Inf.**, Brasília, v.31, n.1, p.41-52, jan./abr. 2002. Disponível em: < http://eprints.rclis.org/14449/1/linguagens_controladas.pdf>. Acesso em: 11 set. 2014.

NORUZI, A. Folksonomies: Why do we need controlled vocabulary? **Webology**, v.4, n.2, 2007. Disponível em: < <http://www.webology.org/2007/v4n2/editorial12.html>>. Acesso em: 20 ago. 2013.

REPOSITÓRIO INSTITUCIONAL DA UFRN. 2017. Disponível em: < <https://repositorio.ufrn.br/jspui/>>. Acesso em: 15 maio 2017.

ROSETTO, M. Metadados e recuperação da informação: padrões para bibliotecas digitais. In: CIBERNÉTICA: SIMPÓSIO INTERNACIONAL DE PROPRIEDADE INTELECTUAL, INFORMAÇÃO E ÉTICA, 2., Florianópolis, **Anais...** Florianópolis: [s.n.], 2003.

SANTOS, H. P. Etiquetagem e Folksonomia: o usuário e sua motivação para organizar e compartilhar informação na web 2.0. **Perspectivas em Ciência da Informação**, Belo Horizonte, v.1, n.2, p.91-104, abr./jun. 2013.

SANTOS, Raimunda Fernanda dos. **Modelos colaborativos de indexação social e a sua aplicabilidade na Base de Dados Referencial de Artigos de Periódicos em Ciência da Informação (BRAPCI)**. 184f. Dissertação (Mestrado em Ciência da Informação) – Universidade Federal de Pernambuco, Recife, 2016.

SANTOS, Raimunda Fernanda dos; CORRÊA, Renato Fernandes. Análise e síntese dos diversos usos do termo "Folksonomia" no âmbito da ciência da informação. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 16., 2015. **Anais...** João Pessoa: Associação Nacional de Pesquisa e Pós-Graduação em Ciência da Informação, 2015.



Case Data Analysis and Open Data in the Course of Information Systems at UFRPE

Roberta Macêdo Marques Gouveia
Maria da Conceição Moraes Batista
Lisandra Sousa da Cruz
Jonathan Henrique Andrade de Carvalho

Departamento de Estatística e Informática – Universidade Federal Rural de Pernambuco (UFRPE)
Rua Dom Manuel de Medeiros, s/n – Dois Irmãos – CEP 52.171-900 – Porto Alegre – RS – Brazil

{robertammg, cecamoraes, lisansouza, hcarvalho.jon}@gmail.com

Abstract. This article describes the data analysis and open data cases obtained from 2015 to 2017 in scientific projects and several works of conclusion of course by students of the Bachelor's degree in Information Systems at the Federal Rural University of Pernambuco. It briefly describe the work, results, difficulties and conclusions obtained. The works were oriented and/or co-oriented by the teachers Roberta Macêdo and Maria da Conceição.

Resumo. Este artigo descreve os casos de análises de dados e dados abertos obtidos de 2015 a 2017 em projetos científicos e diversos trabalhos de conclusão de curso de estudantes do Bacharelado em Sistemas de Informação da Universidade Federal Rural de Pernambuco. Relata-se brevemente os trabalhos, resultados, dificuldades e conclusões obtidas. Os trabalhos foram orientados e/ou coorientados pelas professoras Roberta Macêdo e Maria da Conceição.

1. Introdução

A maioria dos trabalhos aqui apresentados, fundamentaram-se em traçar e analisar o perfil de dados educacionais (estudantes, docentes, cursos e instituições de ensino), visando obter os índices de evasão e retenção/reprovação, bem como investigar o rendimento escolar, os indicadores de qualidade dos cursos e infraestrutura das instituições de ensino fundamental, médio e superior do Brasil. Na tentativa de melhor entendimento do cenário educacional brasileiro, esses trabalhos empenham-se em investigar e detectar o perfil socioeconômico dos estudantes e as metodologias de ensino-aprendizagem, e para alcançar este objetivo, fez-se uso do processo de descoberta de conhecimento em bases de dados, com ênfase em *Data Mining* – Mineração de Dados. A análise de grande volume de dados permitiu que se observassem tendências, que se detectassem regiões e/ou cursos cuja evasão é mais frequente; a revelação de possíveis deficiências no processo de ensino e aprendizagem dos estudantes; a constatação de categorias de estudantes propícios à conclusão do ensino médio e superior, entre outros diagnósticos.

A motivação das pesquisas surgiu do interesse em investigar e diagnosticar as possíveis deficiências e obstáculos enfrentados pelos gestores das instituições, docentes e discentes. Como resultados dos estudos de caso realizados na pesquisa, pretende-se obter respaldo científico necessário para detectar padrões e descobrir regras

significativas na tentativa de melhor compreender o cenário atual da educação no ensino fundamental, médio e universidades brasileiras.

O processo *Knowledge Discovery in Database* (KDD) se propõe em encontrar e interpretar padrões através das análises nas fontes de dados. De acordo com Fayyad et al. (1996), o processo KDD é uma área de pesquisa que tem como principal objetivo o desenvolvimento de métodos do aprendizado de máquina, visando explorar conjuntos de dados coletados de diversos ambientes operacionais. O objetivo é extrair de grandes bases de dados, sem nenhuma formulação prévia de hipóteses, as informações desconhecidas, válidas e acionáveis, que poderão ser úteis para a tomada de decisão. O processo de KDD foi proposto para determinar as etapas que produzem conhecimentos a partir dos dados e, principalmente, definir a etapa de *Data Mining*, que é a fase que transforma dados em conhecimento. Cada fase da execução do processo KDD possui uma interseção com as demais, deste modo, os resultados produzidos em uma fase podem ser utilizados para melhorar os resultados das próximas fases. Este cenário revela um processo iterativo, que busca sempre aprimorar os resultados a cada iteração.

Conforme Inmon (2005), o termo *Data Warehouse* (DW) é definido como “Um depósito de dados orientado por assunto, integrado, não volátil, variável com o tempo, para apoiar as decisões da gerência”. Onde não volátil significa que, uma vez inseridos, os dados não podem ser alterados, embora possam ser excluídos. O conceito de armazém de dados surgiu por duas razões: primeiro, pela necessidade de fornecer uma origem de dados única, limpa e consistente para fins de apoio à decisão; segundo, pela necessidade de fazê-lo sem causar impacto sobre os sistemas operacionais.

As ferramentas Online Analytical Processing (OLAP) são utilizadas para a criação de relatórios gerenciais, facilitando a formatação multidimensional e análise dos dados. Enquanto que, as ferramentas de *Data Mining* utilizam técnicas da aprendizagem de máquina e estatísticas com a finalidade de encontrar correlações e padrões entre os dados que ajudem nas decisões estratégicas. Para construção do DW normalmente utiliza-se a modelagem multidimensional, que é uma técnica de projeto lógico definida sobre dois pilares: tabelas fato e tabelas dimensão. Tabela fato é a tabela primária em um modelo dimensional, onde as medidas de desempenho numéricas do negócio são armazenadas. As tabelas de dimensão são complementos integrais para uma tabela fato. As tabelas dimensão contém os descritores textuais do negócio (KIMBALL e ROSS 2011). Os DW foram implementados como parte do processo de descoberta de conhecimento na base de dados abertas educacionais. Os ambientes de DW organizaram e disponibilizaram os cubos de dados, visando facilitar as análises por meio de ferramentas OLAP e dos algoritmos de *Data Mining*. O uso das tecnologias OLAP proporcionara agregações e sumarizações dos dados contidos no DW, gerando informações úteis ao processo decisório e oferecendo uma análise mais detalhada da educação no Brasil.

Os dados contidos nos DW, que se encontravam íntegros e consistentes, também foram utilizados pelos algoritmos de Aprendizagem de Máquina. Foram utilizados algoritmos de Mineração de Dados do Aprendizado Indutivo Supervisionado e do Aprendizado Indutivo Não Supervisionado para o processo de descoberta de conhecimento a partir da seleção de atributos relevantes. O propósito da mineração de dados é apontar relacionamentos entre os dados e possíveis medidas corretivas e

preventivas para minimizar, entre outros problemas, a evasão das instituições de educação básica e superior.

A aplicação do *Data Mining* visa encontrar o perfil do estudante, das instituições de ensino e detectar possíveis ineficiências da educação básica e superior. Os resultados obtidos com o *Data Mining* serão utilizados a fim de detectar padrões, descobrir regras significativas e estabelecer relações entre os índices de evasão e retenção, o perfil socioeconômico dos estudantes, a situação de infraestrutura das instituições, e demais características inerentes da educação no Brasil. Ao constatar tais relacionamentos e pontos fracos, ações poderão ser tomadas, por parte das instituições, a fim de mitigar as deficiências, buscando reduzir os altos índices de evasão e retenção.

2. Cases de Análise de Dados e Open Data

Esta seção descreve *cases* referentes à análise de dados e *open data* resultantes de projetos de pesquisa do Departamento de Estatística e Informática (DEINFO) e trabalhos de conclusão de curso do Bacharelado em Sistemas de Informação da Universidade Federal Rural de Pernambuco, coordenados, orientados e/ou coorientados pelas professoras Roberta Macêdo Marques Gouveia e Maria da Conceição Moraes Batista, ambas do DEINFO/UFRPE.

O projeto de pesquisa intitulado “Processo de Descoberta de Conhecimento em Ambientes Virtuais de Aprendizagem da Educação a Distância (FACEPE/CNPq)”, coordenado pela professora Roberta Macêdo, propõe a aplicação de tecnologias contemporâneas de banco de dados e inteligência artificial para análise de dados do ensino superior brasileiro no cenário da Educação a Distância (EaD). O referido projeto deu origem a dois novos projetos científicos: “Data Mining em Ambientes Virtuais de Aprendizagem para Educação a Distância (PIBITI/CNPq), realizado entre 2014 e 2015, e “Mineração de Dados Educacionais em Ambientes *B-learning* de Instituições Federais de Ensino Superior (PIBIC/PIC/UFRPE)”, realizado entre 2015 e 2016.

O estudo realizado por Freitas (2015), confirma que a EaD no Brasil está em constante expansão, e isso deve-se a um aumento considerável de instituições de ensino superior adotando a modalidade a distância em seus cursos de graduação e especialização, bem como o incentivo do governo, por meio do projeto Universidade Aberta do Brasil (UAB), e a crescente adesão de estudantes. Segundo o Censo da EaD, dentre as instituições educacionais com ofertas de cursos na modalidade a distância, 64% estão sediadas nas capitais brasileiras e 36% nos interiores, com concentração de 42% no Sudeste, Sul (21%), Nordeste (18%), Centro-Oeste (13%) e Norte (6%). No Brasil, os cursos são ofertados em todos os níveis e áreas de conhecimento, com destaque para as áreas de Ciências Sociais Aplicadas, com 608 ofertas de cursos regulamentados totalmente a distância, e entre os cursos semipresenciais a preferência é pelas Ciências Humanas. As instituições contam, em média, com 1.000 a 5.000 estudantes (ABED, 2016).

A maioria das matrículas em cursos a distância e cursos semipresenciais (*b-learning*) encontra-se nas licenciaturas, totalizando aproximadamente 700 mil estudantes. O Censo da EaD 2015/2016 contabilizou mais de 5 (cinco) milhões de estudantes, sendo 1,1 milhão em cursos regulamentados totalmente a distância e semipresenciais, e aproximadamente 4 (quatro) milhões em cursos livres. No Censo atual, 2015/2016, tem-se 1,2 milhão de estudantes a mais do que no Censo 2014/2015,

ou seja, houve um aumento de 23,4% de estudantes matriculados em cursos a distância no Brasil. Em se tratando das taxas de evasão reportadas nos cursos a distância, ABED (2016) registra uma evasão de 26% a 50% – alertando que a desistência dos estudantes da EaD é maior que nos cursos presenciais. As instituições apontam o fator “tempo” como o mais influente no fenômeno da evasão, seguido do fator “finanças”.

Para dar suporte ao processo de apoio à decisão foi projetado um Data Warehouse, tendo como base a modelagem multidimensional constelação de fatos, aplicação de tecnologias OLAP e mineração de dados da aprendizagem de máquina supervisionada. Os dados de alta granularidade foram obtidos dos sistemas e-MEC e Universidade Aberta do Brasil (UAB), por meio de dados abertos. Também foram obtidos dados dos sistemas de gestão acadêmica (SIG@ e SIGAA) de duas IES: UFRPE e UFPB. Portanto, o trabalho fundamenta-se no desenvolvimento de um ambiente computacional analítico – *Decision Support Systems* (DSS), visando traçar perfil de estudantes e cursos da EaD (PE, PB e Nacional), a fim de investigar obstáculos enfrentados pelos docentes, discentes e instituições que ofertam cursos na modalidade a distância.

O projeto de pesquisa intitulado “Algoritmos do Aprendizado de Máquina Aplicados na Mineração de Dados Educacionais Abertos do INEP”, coordenado pela professora Roberta Macêdo, tem como objetivo a descoberta de padrões e novos conhecimento (processo KDD) em grandes volumes de dados abertos educacionais. Para isso são utilizadas técnicas de *Machine Learning*, mais especificamente algoritmos de classificação por meio de Árvore de Decisão e classificação Bayesiana, com foco na análise do perfil de estudantes e instituições brasileiras de ensino fundamental, médio e superior. As bases de dados utilizadas nesse projeto referem-se ao Censo Escolar e Censo da Educação Superior de 2014 e 2015, sendo disponibilizadas abertamente pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP). O referido projeto deu origem a dois novos projetos: “Algoritmos do Aprendizado de Máquina Aplicados na Mineração de Dados Educacionais do INEP (PIBIC/CNPq)” e “Técnicas de Classificação e Associação Aplicadas em Dados Abertos dos Censos da Educação Básica e Superior (PIBITI/CNPq), ambos iniciados em 2017.

2.1. Descoberta de Conhecimento Utilizando Mineração de Dados Educacionais Abertos

A adoção de técnicas de mineração de dados tornou-se uma realidade para varejistas, bancos, fabricantes, seguradoras, dentre outros, proporcionando-lhes a descoberta de relações entre padrões não visíveis em grandes volumes de dados. A mineração de dados tem ganho destaque também no cenário educacional, oferecendo aos educadores e gestores educacionais subsídios para a tomada de decisão, além de perspectivas para mitigar desafios tradicionais do processo de ensino-aprendizagem.

Neste sentido, o trabalho de Gomes (2015) versa sobre a aplicação de técnicas e métodos de mineração de dados a fim de descobrir padrões e regras de associação em dados estatísticos educacionais oriundos do Exame Nacional do Ensino Médio (ENEM) nos anos de 2013 e 2014 no âmbito da região Nordeste.

Assim, com este trabalho, espera-se contribuir, por meio de trabalhos futuros que serão embasados no trabalho atual, na oferta de subsídios para a criação e/ou fortalecimento de iniciativas que visem melhorias na diminuição da evasão e retenção dos estudantes universitários, avaliação e diagnóstico do corpo discente, bem como

auxiliar, por exemplo, os gestores públicos na criação/consolidação de ações afirmativas dentro da universidade, que vão desde o estabelecimento de cotas, programas de nivelamento a programas de bolsas de estudo e de apoio acadêmico.

2.2. Comparação de Algoritmos do Aprendizado de Máquina Aplicados na Mineração de Dados Educacionais

A mineração de dados é uma técnica comumente usada para a descoberta de conhecimento em grandes bases de dados relacionadas a qualquer tipo de contexto. Para que esta descoberta de conhecimento seja alcançada, é necessário atender às etapas do processo de Descoberta de Conhecimento em Base de Dados. Estas etapas vão desde a seleção, pré-processamento e transformação dos dados, até a própria implantação do conhecimento recém descoberto.

Um exemplo de aplicação da mineração de dados é a Mineração de Dados Educacionais (MDE). Trata-se de uma área de pesquisa interdisciplinar que lida com o desenvolvimento de métodos para explorar dados oriundos de contextos educacionais. Uma aplicação da mineração de dados neste contexto significa coletar informações e conhecimentos de alunos com o objetivo de entender como eles estão explorando o ambiente de estudo e o que pode ser feito para tornar o espaço de estudo ainda mais atrativo para eles.

Na mineração de dados, diversos algoritmos podem ser executados para a descoberta de conhecimento. A escolha de quais algoritmos executar é uma tarefa não trivial que deve ser feita considerando fatores, tais como, o volume de dados a serem processados, o tipo de aprendizado a ser obtido e as características dos dados. Para contribuir nesse sentido, o trabalho de Mirela Souza (2015) empregou o processo KDD e executou algoritmos do aprendizado de máquina supervisionado e não-supervisionado na base de dados do ENEM 2013. Com os resultados gerados a partir dos algoritmos, foi realizado um estudo comparativo a partir da análise de métricas pré-estabelecidas, como o tempo de execução, estatística kappa e a taxa de acurácia (matriz de confusão) de cada algoritmo. Desta forma, pode-se tomar este estudo comparativo baseado nas métricas como apoio para a escolha de que algoritmos aplicar.

De acordo com o objetivo de estudo deste trabalho não foi aqui apresentado o melhor algoritmo dos sete estudados, foram apresentadas características de cada algoritmo, a performance de cada um deles em diferentes configurações e quantidade de instâncias mostrando o desempenho individualmente. As performances apresentadas ao longo do estudo indicaram diferenças consideráveis e algumas sutis como por exemplo o tempo, alguns tempos de execução foram mais elevados e outros não tão elevados, a quantidade de tuplas executadas pelos algoritmos foram diferentes, assim como a quantidade de regras geradas pelos algoritmos de associação.

2.3. Implementação de SIG e Mapas de Kernel visando Acessibilidade na Educação Superior

Identificar a acessibilidade oferecida por uma instituição de ensino superior é um desafio. Para vencê-lo, é necessário verificar as normas técnicas e legislação vigente para identificar os itens de acessibilidade ofertados. Portanto, o objetivo do trabalho de Fernanda Souza (2015) é implementar um Sistema de Informações Geográficas que possibilite o armazenamento e consultas espaciais implementando Banco de Dados Geográficos. Através da identificação dos recursos ofertados, é feita uma análise destes

através de mapas de Kernel que possibilitam identificar as áreas nas quais devem ser feitas ações para melhoria da acessibilidade ofertada.

O estudo fornece subsídios para melhorias de propostas para acessibilidade de pessoas com mobilidade reduzida na Universidade Federal Rural de Pernambuco (UFRPE). Como resultado do sistema de análise dos itens de acessibilidade ofertados pela UFRPE desenvolvido ao longo desse trabalho, utilizando SIG, Bancos de Dados Geográficos e Mapas de Kernel, foi possível obter uma visão dos pontos que demandam uma maior atenção. Além do fornecimento de uma visão analítica para a solução da falta de acessibilidade, o presente trabalho visa o uso de Bancos de Dados Geográficos como tecnologia assistiva na área de análise através da produção de mapas de Kernel.

2.4. Análise de Modelos de Dados Não Relacionais e Multidimensionais no Contexto de *Big Data*

Atualmente são geradas grandes quantidades de dados por dia. Esses dados vêm de diversos lugares, tais como registros de transações financeiras, internet banking, call centers, caixas automáticos, sensores, etc. E esse número só vem aumentando, pois 90% desses dados foram produzidos somente nos dois últimos anos. Esse fenômeno é chamado de *Big Data*: quantidades gigantescas de dados de diversos tipos vindos de vários lugares. Porém, muitas vezes tem-se todos esses dados, mas não é possível retirar informação útil deles, por isso faz-se necessária a utilização de ferramentas e tecnologias de diversas áreas, tais como banco de dados, Sistemas Distribuídos e Estatística, que auxiliem na coleta, no armazenamento, na análise e na transformação destes dados.

Um banco de dados relacional tradicional talvez não seja capaz de suportar essa quantidade de dados, por isso a utilização de outros modelos de dados vem sendo explorada na busca de tecnologias que possam solucionar questões relacionadas a *Big Data*. O trabalho de Lira (2016) apresenta uma análise comparativa entre dois modelos de dados, um modelo de dados multidimensional (*Data Warehouse*) e um modelo de dados não relacional (NoSQL), com a finalidade de mostrar qual se adapta melhor entre eles para se trabalhar neste contexto utilizando dados de compra e venda de produtos na região metropolitana de Recife.

A comparação foi entre um DW multidimensional - utilizando as ferramentas MySQL para armazenamento dos dados, o Pentaho para fazer as transformações dos dados, e o Saiku como ferramenta OLAP – e o HBase, um banco de dados NoSQL orientado à coluna, que funciona sobre o sistema distribuído Hadoop, e para fazer consultas ao banco, foi utilizado o Hive.

Após a realização dos estudos de caso, o resultado deste trabalho mostrou que em se tratando com dados de mercado, que foram transformados de XML para tabelas do modelo relacional, o modelo multidimensional teve um desempenho superior ao modelo não relacional. Isso se dá ao fato de que nesse contexto é requerida uma grande quantidade de consultas complexas e análise de dados de rotina, características que não fazem parte da essência dos modelos não relacionais. O Data Warehouse utilizou apenas 36% do tempo gasto pelo HBase para realizar as consultas.

Por fim, observou-se que mesmo o HBase não tendo um desempenho tão bom em relação às consultas, ele foi capaz de realizar as mesmas funções que um SGBD relacional realizaria, e ainda obteve um tempo menor na importação das tabelas. O fato

é que a presença da tabela fato no modelo multidimensional, onde os dados quantitativos já estão presentes, deixa as consultas mais rápidas. No modelo não relacional orientado à coluna, além da necessidade de sempre fazer junções nas duas grandes tabelas, em cada consulta que busca os dados valor total de vendas e unidades vendidas é necessário fazer uma operação de soma para cada uma delas, diminuindo assim a performance da consulta. Além disso, outro fator que comprometeu a performance do HBase foram os recursos de hardware limitados. Como ele é um elemento do Hadoop, que é um sistema distribuído, possivelmente sua performance seria melhor se existisse um outro nó dividindo a carga com a única máquina utilizada.

2.5. Análise de Desempenho de Banco de Dados Não Relacionais no Cenário de Dados Abertos Educacionais

A necessidade de tecnologias de banco de dados e sistemas distribuídos para armazenamento e manipulação de grande quantidade de dados está cada vez mais comum. Os bancos de dados relacionais apresentam certas limitações para lidar com grandes volumes de dados, devido ao fato de não apresentarem muita flexibilidade para sua estrutura mais rígida de dados. Para suprir essa necessidade têm-se os bancos de dados não relacionais que apresentam estruturas mais flexíveis, e possuem um bom escalonamento horizontal e vertical, sendo bancos de dados mais indicados para a manipulação de dados no cenário de *big data*.

O trabalho desenvolvido por Melo (2016) tem como objetivo analisar o desempenho de banco de dados não relacionais sobre o cenário de dados abertos educacionais. O objetivo foi verificar qual destes modelos apresentam os melhores desempenhos para lidar com os dados do censo escolar. Esta análise foi desenvolvida sobre as operações de criação, inserção/carga, remoção, alteração e consulta aos dados educacionais. Os modelos de banco de dados não relacionais que são avaliados neste trabalho, com os seus respectivos sistemas de banco de dados são: modelo orientado à coluna (Cassandra), modelo orientado a documentos (MongoDB) e modelo orientado à chave-valor (DynamoDB).

Com base nos resultados obtidos nos estudos de caso, constatou-se que o banco de dados orientado a documento MongoDB foi mais eficiente na execução das operações de inserção, consulta, atualização e remoção. Os experimentos realizados no trabalho corroboram com esta colocação, já que o MongoDB foi 18 vezes mais rápido na inserção dos dados se comparado ao Cassandra, e 353 vezes mais rápido se comparado ao DynamoDB. O MongoDB foi 6 vezes mais rápido na consulta dos dados se comparado ao Cassandra, e 43 vezes mais rápido se comparado ao DynamoDB. O MongoDB foi 22 vezes mais rápido na atualização dos dados se comparado ao Cassandra, e 627 vezes mais rápido se comparado ao DynamoDB. O MongoDB foi 24 vezes mais rápido na remoção dos dados se comparado ao Cassandra, e 1062 vezes mais rápido se comparado ao DynamoDB.

Então, como o banco de dados orientado à chave-valor, DynamoDB, apresentou os piores resultados entre os três bancos de dados analisados, ele é o menos indicado para utilização no cenário de dados abertos do censo escolar. Já o banco de dados do modelo orientado à documento, MongoDB, apresentou os melhores resultados em todas as operações analisadas no trabalho, e por isso, é o mais indicado para ser utilizado com a base de dados do censo escolar 2014.

2.6. Mineração de Dados Educacionais e Visualização de Informações Geográficas Utilizando Mapas de Calor

Os centros educacionais acumulam uma gama de dados, que podem se referir às instituições de ensino (localização, infraestrutura, dependências), aos alunos (desempenho escolar, perfil socioeconômico), aos profissionais da educação, (nível de instrução), entre outras informações. Estes dados, se vistos individualmente, não fornecem informações relevantes, são apenas dados brutos. No entanto, se estes grandes volumes dados forem analisados de forma mais minuciosa, utilizando técnicas computacionais que os associem entre si, estará sendo aplicada a descoberta de conhecimento, mais especificamente a mineração de dados. A Descoberta de Conhecimento em Banco de Dados, do inglês *Knowledge Discovery in Databases* (KDD), é um processo que vem sendo utilizado com o objetivo de descobrir padrões ocultos e relevantes para o cenário educacional.

Neste trabalho, Nascimento (2016) realizou a mineração de dados da educação básica, envolvendo ensino infantil, fundamental e médio. É com a relevância dos resultados oriundos deste processo que o presente trabalho visa aplicar técnicas de mineração de dados na base de dados educacional do Censo Escolar dos anos de 2012 a 2015 no cenário do estado de Pernambuco, através de algoritmos de classificação, como o J48 e o Naive Bayes e de associação, como o Apriori. Após a obtenção dos resultados, é desenvolvido um sistema de representação geográfica, transformando as informações em mapas de calor. Os resultados obtidos ao fim deste processo promovem a identificação de fatores, tendências e problemas que podem ajudar na tomada de decisão, na reformulação de procedimentos e de políticas internas ou públicas na área educacional.

Ao fim, pode-se destacar que, em relação à infraestrutura básica e à acessibilidade, as escolas urbanas possuem melhores resultados se comparadas às escolas rurais. Este cenário não apresenta diferenças se as escolas forem analisadas quanto à presença de atividades complementares, como esportes, leitura e inclusão digital, com destaque para as escolas municipais do ambiente rural, que apresentam maior carência neste cenário. Sendo assim, os investimentos na área de educação devem ser melhor aplicados à estes cenários que possuem necessidades de melhorias, visto que houveram mudanças mínimas no intervalo de tempo analisado (entre 2012 e 2015), para assim, melhor atender alunos, professores, diretores e outros profissionais relacionados a área de educação do estado de Pernambuco.

2.7. Um Framework Multiplataforma para Análise e Monitoramento de Dados Governamentais

Dados são a base para a descoberta de informações, já as informações geram conhecimentos para o desenvolvimento de inovações de alcance mundial. Logo, dados mudam o mundo. Visando a descoberta de conhecimentos implícitos nos dados, governos em todo o mundo passaram a disponibilizá-los com o intuito de oferecer transparência, inovação e novos negócios.

Cruz Júnior (2016) desenvolveu um *framework* para análise e monitoramento de dados abertos governamentais embasado em algoritmos de inferências estatísticas, correlações não lineares, *outliers* e análises de sentimento. As informações geradas são disponibilizadas através de uma suíte visual de gráficos, tabelas e mapas interativos.

A arquitetura desenvolvida é orientada a serviços, facilitando a disponibilização das suas funcionalidades. A aplicação base é desenvolvida em padrões *web*, possibilitando seu funcionamento e disponibilização em múltiplas plataformas. A interface da ferramenta segue padrões de usabilidade e do *flat design*, implementando preceitos de sistemas colaborativos.

O *framework* foi utilizado em distintos estudos de caso, com bases de dados do Portal Brasileiro de Dados Abertos, do Portal de Convênios e do Portal de Dados Abertos da Prefeitura de Recife, juntamente com dados minerados da rede social *Twitter* para a descoberta de informações relativas à saúde, infraestrutura, corrupção e comportamentos sociais. Durante os estudos de caso pôde-se constatar, na base de convênios federais, que cerca de 12% dos repasses foram identificados como *outliers*, alguns destes apontados pela mídia como possíveis casos de corrupção. Também foi identificado um coeficiente de correlação de 0,536, entre a quantidade de vítimas de acidentes de trânsito e a média de temperatura, supõe-se que quanto maior a temperatura na cidade do Recife, maior a quantidade de acidentados.

Com a utilização do *framework* foi constatado que, de maneira geral, a cidade do Recife é a que mais recebe incentivos do PAC em Pernambuco, que o bairro do Ibura, região da periferia da cidade do Recife, é o que mais sofre com casos de deslizamentos de terra e o de Boa Viagem o que mais acontece acidentes de trânsito. Também foi possível identificar casos de possível corrupção, já que 11,89% dos repasses feitos pelo governo federal em forma de convênios representam 93,14% de todo o dinheiro investido, além disso, identificou-se que os dois maiores pagamentos, que somados resultam em 2,2 trilhões de reais, foram para um município de 6.328 habitantes.

3. Conclusões

Sabe-se desde sempre, que dados é um bem difícil de manter com integridade, consistência e correte. Ao longo de anos, um sem número de tecnologias foram criadas e aperfeiçoadas para tratar, manter, processar e transmitir dados. Contudo, ter permissão para acessar dados para fins de pesquisas para benefício da sociedade ou resolução de um problema, é difícil. As empresas não cedem seus dados com receio de ter a confidencialidade violada, mesmo quando oferecemos assinar termos de confidencialidade para garantir a privacidade de dados sensíveis e dando garantia à privacidade.

Aconteceu conosco, tivemos autorização de acesso aos dados de um consultório de nutrição especializado em pessoas que passaram por cirurgias de redução de estômago, tanto da rede de saúde pública, como privada. A ideia era minerar dados de cerca de 15 anos de consultas, exames e acompanhamento de pacientes e encontrar correlações, e até pessoas em situação de risco, sempre preservando dados sensíveis e pessoais dos pacientes. Foi constatado que a mineração iria auxiliar bastante os profissionais do consultório. Entretanto, a empresa de TI que desenvolveu o software que atende ao consultório, também mantém os dados dos pacientes em seus servidores. Essa empresa se recusou a nos fornecer o acesso aos dados. Infelizmente, o consultório - real "dono" dos dados - não se dispôs a rejeitar este comportamento por questões de relacionamentos pessoais entre pessoas do consultório e da empresa de TI.

Portanto, obter acesso a dados privados, mesmo que para fins de pesquisas, é uma situação complicada. Assim, usar dados abertos é a opção mais indicada para tocar

pesquisas relevantes. Entretanto, os portais de dados abertos ainda não disponibilizam uma variedade de dados muito grande. Tem-se ainda muito o que melhorar em relação à quantidade, qualidade e variedade dos portais de dados abertos. Mas, continuamos investindo nesta modalidade de dados - os abertos - e reivindicando, que os governos cada vez mais disponibilizem e tornem públicos mais conjuntos de dados.

Referências

- ABED – Associação Brasileira de Educação a Distância. (2016) Censo EAD.BR: Relatório Analítico da Aprendizagem a Distância no Brasil. Curitiba: InterSaber, 2016.
- CRUZ JÚNIOR, Geraldo Gomes. (2016) Um Framework Multiplataforma para Análise e Monitoramento de Dados Governamentais. Trabalho de Conclusão de Curso, Bacharelado em Sistemas de Informação - Universidade Federal Rural de Pernambuco, Recife, PE.
- FAYYAD, U.M.; PIATETSKY-SHAPIRO, G.; SMYTH, P. (1996) From Data Mining to Knowledge Discovery. *AI Magazine*, p. 37-54.
- FREITAS, C. N.; GOUVEIA, R. M. M.; SILVA, A. (2015) Online Analytical Processing em Ambientes Virtuais de Aprendizagem da Educação a Distância. In Anais do 4º DesafIE - Workshop de Desafios da Computação Aplicada à Educação – XXXV Congresso da Sociedade Brasileira de Computação.
- GOMES, Tancicleide Carina Simões. (2015) Descoberta de Conhecimento Utilizando Mineração de Dados Educacionais Abertos. Trabalho de Conclusão de Curso, Bacharelado em Sistemas de Informação - Universidade Federal Rural de Pernambuco, Recife, PE.
- INMON, William H. (2005) *Building the Data Warehouse: Getting Started*. 4a Edição. Editora: Wiley Publishing, Inc.
- KIMBALL, Ralph; ROSS, Margy. (2011) *The data warehouse toolkit: the complete guide to dimensional modeling*. John Wiley & Sons.
- LIRA, Maria Camila Soares. (2016) Análise de Modelos de Dados Não Relacionais e Multidimensionais no Contexto de Big Data. Trabalho de Conclusão de Curso, Bacharelado em Sistemas de Informação - Universidade Federal Rural de Pernambuco, Recife, PE.
- MELO, Felipe Fernandes de Lima Melo. (2016) Análise de Desempenho de Banco de Dados Não Relacionais no Cenário de Dados Abertos Educacionais. Trabalho de Conclusão de Curso, Bacharelado em Sistemas de Informação - Universidade Federal Rural de Pernambuco, Recife, PE.
- NASCIMENTO, Rafaella Leandra Souza. (2016) Mineração de Dados Educacionais e Visualização de Informações Geográficas Utilizando Mapas de Calor. Trabalho de Conclusão de Curso, Bacharelado em Sistemas de Informação - Universidade Federal Rural de Pernambuco, Recife, PE.
- SOUZA, Fernanda Correia Monteiro. (2015) Implementação de SIG e Mapas de Kernel visando Acessibilidade na Educação Superior. Trabalho de Conclusão de Curso, Bacharelado em Sistemas de Informação - Universidade Federal Rural de Pernambuco, Recife, PE.

SOUZA, Mirela Natali Vieira. (2015) Comparação de Algoritmos do Aprendizado de Máquina Aplicados na Mineração de Dados Educacionais. Trabalho de Conclusão de Curso, Bacharelado em Sistemas de Informação - Universidade Federal Rural de Pernambuco, Recife, PE.



CLASSIFICAÇÃO DE DOCUMENTOS FILATÉLICOS EM FORMATO ABERTO PARA WEB: o caso dos selos postais no Repositório Filatélico Brasileiro¹

Kézia de Lira Feitosa
Diego Andres Salcedo

RESUMO: Indica os resultados parciais da pesquisa em desenvolvimento no ambiente do Grupo Imago, vinculado ao Departamento de Ciência da Informação da Universidade Federal de Pernambuco. Tem como objetivo classificar a documentação filatélica do Repositório Filatélico Brasileiro para fins de armazenamento e disseminação em formato aberto na WEB. De natureza exploratória e descritiva, utiliza como procedimento metodológico a revisão da literatura especializada nos campos da Biblioteconomia, Ciência da Informação, Filatelia, Coleccionismo, Memória e Computação. Descreve o procedimento de elaboração do modelo de classificação. Ilustra dois exemplos de classificação de selos postais. Conclui que a classificação de selos postais demanda tanto um olhar crítico que respeite as suas especificidades documentais, quanto um modelo de classificação que complemente àqueles utilizados para classificar documentos similares.

PALAVRAS- CHAVE: Biblioteconomia. Filatelia. Repositório. Classificação. Imagem.

ABSTRACT: *Indicates the partial results of the research under development in the environment of the Imago Group, linked to the Department of Information Science of the Federal University of Pernambuco. It aims to classify the philatelic documentation of the Brazilian Philatelic Repository for storage and dissemination purposes in an open format on the WEB. Of exploratory and descriptive nature, it uses as a methodological procedure the revision of the specialized literature in the fields of Librarianship, Information Science, Philately, Collecting, Memory and Computing. Describes the procedure for drawing up the classification model. It illustrates three examples of classification of brazilian postage stamps. It concludes that the classification of postal stamps requires both a critical look that respects its documentary specificities and a classification model that complements those used to classify similar documents.*

KEYWORDS: *Library Studies. Philately. Repository. Classification. Image.*

1 INTRODUÇÃO

O projeto de desenvolvimento do Repositório Filatélico Brasileiro (REFIBRA), devidamente homologado na Universidade Federal de Pernambuco (UFPE) e no Diretório de Grupos de Pesquisa do CNPq, articula atividades de pesquisa (na graduação e pós-graduação), ensino, extensão e cultura. A gestão ocorre no âmbito do Grupo IMAGO, vinculado ao Departamento de Ciência da Informação, da UFPE.

¹ Pesquisa com apoio financeiro para bolsa de iniciação científica PROPESQ/UFPE/CNPQ.

O REFIBRA contribui com um conjunto de princípios que norteiam os programas estratégicos de uma rede, nacional e internacional, de instituições comprometidas com políticas de preservação, conservação e acesso de dados abertos na WEB de acervos memoriais e patrimônios materiais e imateriais (SALCEDO et al, 2015).

O projeto corrobora com as ações de preservação da memória digital recomendadas pela Organização das Nações Unidas para a Educação, a Ciência e Cultura (UNESCO), na sua Declaração de Vancouver (2012). Logo, insere o Brasil no cenário internacional neste tipo de ação no campo da Filatelia e da História Postal. No âmbito nacional contribui, de forma inédita ao ser considerado o campo da Filatelia, para a preservação, conservação e difusão da memória, da cultura e da identidade nacional em consonância com o Plano Nacional de Cultura – PNC, do Ministério da Cultura do Brasil (2010).

Esse patrimônio documental e memorial resulta de mais de 170 anos de atividades filatélicas no país e é constituído por diversos tipos de documentos: selos, inteiros, blocos, editais, catálogos, livros, periódicos, boletins, folhetos, atas, ofícios, cartas, fotografias etc. Reúne materiais que englobam saberes, celebrações e formas de expressão de grande relevância para memória nacional.

São registros que permitem a leitura, interpretação e reconstrução das relações sociais, culturais, políticas e econômicas. Constitui um patrimônio que pode servir de subsídio para pesquisas, ações de extensão, atividades pedagógicas, práticas culturais e projetos de economia criativa, com enfoque local, regional, nacional e internacional.

Foi nesse cenário que nasceu um projeto de pesquisa que propôs como objetivo geral classificar a documentação filatélica do Repositório Filatélico Brasileiro para fins de armazenamento e disseminação em formato aberto na WEB. Assim, neste artigo será feita uma apresentação das ações que já foram realizadas cobrindo aspectos relacionados ao referencial teórico, aos procedimentos metodológicos utilizados e às exemplificações de aplicação do modelo de classificação dos selos postais que constituem o REFIBRA.

2 FUNDAMENTAÇÃO TEÓRICA

O ato de classificar é um fenômeno social que sempre fez parte da vida

humana. No decorrer das rotinas diárias, nós atribuímos valores, juízos, títulos, nomenclaturas e estereótipos às pessoas e aos objetos. Além dessas classificações serem feitas por indivíduos de certo grupo social ou comunidade, elas também são parte concreta da vida social e podem ser intituladas, segundo (COSTA, 1997/1998, p. 65) de “classificações sociais”.

Para além das classificações sociais são reconhecidas diversas maneiras de classificar o conhecimento produzido pelos humanos. Por exemplo, no período da Idade Moderna podemos observar um esforço maior na organização de forma sistemática do conhecimento humano. Burke (2003) revela que neste momento histórico de transição este conhecimento era classificado por grupos diferentes de maneiras diferentes também. Estas classificações divergem entre duas abordagens distintas e complementares: a filosófica e a voltada para a organização e recuperação de documentos, denominada “classificação bibliográfica”.

Segundo Buckland (1991, p. 4),

seres humanos fazem coisas *com* a informação ou *para* a informação. Examinam, descrevem e categorizam. Compreendem, interpretam bem ou mal, resumem ou refutam. Podem até falsificá-la, alterá-la ou destruí-la. A essência da evidência é precisamente o que a percepção das pessoas pode alterar aquilo que acreditam que sabe.

Um dos papéis predominantes do profissional da informação é o de emissor-mediador entre a memória registrada pelas gerações passadas e a sociedade atual e futura. Fazer com que este material esteja disponível em sua completude, por muito mais tempo, é de uma responsabilidade grande, visto que é através da memória humana, representada nos patrimônios históricos, que podemos voltar no tempo e tomar partido do conhecimento gerado.

Na área da Biblioteconomia o conceito dado ao ato de classificar por (PIEDADE, 1977, p. 09) mostra a essência desse processo, intrínseco às culturas humanas, à medida que separamos em partes um conjunto de objetos a partir das similitudes e distinções, considerando as qualidades atribuídas pela própria sociedade e muitas vezes pelo senso comum.

A classificação bibliográfica busca aperfeiçoar o acesso aos documentos requeridos de forma a tornar o processo de busca mais eficaz e efetivo. A denominação de palavras-chave que representam o conteúdo de um documento

funciona como remissiva que oferece pistas sobre o conteúdo no documento original, traços esses que podem possibilitar uma noção ampla ou com maior grau de especificidade. O estudo do conceito de documento tratado demasiadamente nas áreas de documentação, biblioteconomia e Ciência da Informação permite observar que o uso do conceito é usado tanto de forma especializada quanto de forma generalista, de forma a delimitá-lo, por vezes, a alguma fonte de informação física, ou não, em um meio específico.

Conforme indica Salcedo (2010, p. 124), “um selo postal pode ser considerado um documento, quando olhado atentamente por um historiador e que, para manter a memória social, dialoga com outros textos”. Dessa forma, sua classificação deve ser pensada enquanto representação de imagem em sua totalidade, o plano de expressão, a interface e a superfície material do selo postal. Para o campo da semiótica, toda linguagem possui dois planos dos quais resultam o sentido: o plano de expressão e o plano de conteúdo. O sentido, sob esta perspectiva, caracteriza-se segundo Pietroforte (2004, p. 12-13) por uma “rede de relações”, o que quer dizer que os elementos do conteúdo só adquirem sentido por meio das relações estabelecidas entre eles.

Floch (2001, p.9) conceitua o plano de expressão enquanto “terreno onde as qualidades sensíveis que possui uma linguagem para se manifestar são selecionadas e articuladas por variações diferenciais”. Entende-se que é tudo aquilo perceptível ao olhar, a veiculação do conteúdo. Trata-se da materialidade do texto, um recorte de elementos de uma determinada linguagem que serão organizados para significar um conjunto de ideias e é no plano de conteúdo que as significações nascem de acordo com as ordenações e encadeamentos sob a visão de cada cultura.

Existe um acentuado debate teórico e a contínua experimentação de procedimentos metodológicos, dos quais à área de representação temática, específica da Biblioteconomia, pode ser favorecida. Ainda, é possível encontrar na literatura científica da Ciência da Informação certo debate sobre as metodologias utilizadas na representação temática dos documentos pictóricos. O Quadro 1, a seguir, indica algumas tendências metodológicas de análise imagético-textual utilizadas desde o século XX.

Quadro 1 - Autores e aparatos teórico-metodológicos para análise de

Autor/es	Metodologias
Panofsky (1939)	Método Iconológico: descrição pré-iconográfica, análise iconográfica e iconologia.
Eco (1968)	Sistema de códigos visuais: verbal e visual (icônico)
Barthes (1970)	Abordagem Semiológica: denotação, conotação, ancoragem, revezamento. Retórica da Imagem.
Floch (1981)	Semiótica Estrutural. Teoria Gerativa de Sentido.
Joly (1994)	Descrição da Imagem. Reprodução do Texto. Separação e Análise da Imagem (plástica, icônica e linguística)
Prosser (1998)	Antropologia e Sociologia Visual. Etnografia. Evidência fotográfica. Iconografia e Iconologia, Mitologias. Análise de gênero e técnica, forma, estilo e semiótica. Estruturalismo, reconstrução, contexto físico. Hermenêutica.
Emmison e Smith (2000)	Análise de Evidência em duas dimensões: quantitativa e qualitativa
Leeuwen e Jewitt (2001)	Análise do conteúdo. Estudos Culturais. Semiótica. Iconografia. Perspectiva terapêutica. Sociosemiótica. Etnometodologias.
Rose (2001)	Interpretação Compositiva. Análise de Conteúdo. Semiologia, Psicanálise, Análise do Discurso.
Emery (2002)	Abordagem de 7 marcos interpretativos: formalista, desconstrucionista, gênero, cultural, semiótico, psicanalítico e social realista.
Walker e Champlin (2002)	Forma e conteúdo. Análise do Conteúdo. Iconografia e Iconologia, Mitologias. Análise de gênero e técnica, forma, estilo e semiótica. Estruturalismo, reconstrução, contexto físico. Hermenêutica.

imagens.

Fonte: Salcedo (2010, p. 155)

A análise de uma imagem diz respeito à tradução verbal e interpretação do aspecto visual do documento, como fotografias, selos postais, pinturas etc. Há um trabalho de significação dos conteúdos implícitos e explícitos. Desse conjunto de documentos, o selo postal do tipo comemorativo, tem características verbovisuais que transcendem a função administrativo-postal e por isso interpelam, para além das esferas do mercado filatélico, todos que com ele se deparam.

Diferentemente de outros tipos de selos postais, os comemorativos têm sua tiragem e seu período de validade e circulação pré-determinados pelos Atos Normativos e Editais. Por convenção três elementos verbovisuais devem constar impressos conforme um padrão normativo internacional: motivo da emissão, valor facial e nome do país ou da instituição emissora.

Os elementos contidos neste tipo de documento modificaram-se progressivamente. Sob uma perspectiva didática pode-se separar essas mudanças em dois momentos descritos por Salcedo (2010): um primeiro momento em que predominam as representações de grandes personalidades, geralmente ligadas ao sistema político, assim como certa alusão a alguns eventos específicos e um

segundo momento em que começam a surgir as recorrências temáticas na representação, como por exemplo os esportes individuais, fauna e flora.

Esses elementos verbovisuais constituem o plano de expressão, a interface, a superfície material de um selo postal. Ao ser analisado sob esse enfoque, o documento estará sendo interpretado com base nos assuntos que estão ali representados, caracterizando, assim, o seu tratamento informacional e documental. Essa abordagem considera as suas estruturas significantes, que por sua vez possibilita construção de sentidos, além de supor processos de decodificação e interpretação. Na fase de decodificação dos elementos do objeto estudado é importante ter noção das delimitações da sua macroestrutura.

No caso do selo postal comemorativo, a macroestrutura alcança obrigatoriamente o motivo da emissão, o valor facial e o nome do país ou entidade emissora, além da gravura e termos que funcionam como legenda da história que pretende-se representar e divulgar através do selo postal comemorativo. Parte desse processo remete à leitura de seus códigos icônicos. Surge o significado ao leitor a partir da interpretação dos conteúdos, do contato com a interface imagética, e da disposição técnica do suporte.

Dentro de suas ramificações, a Ciência da Informação tem um segmento específico que analisa as etapas e procedimentos adotados para a organização e representação dos dados documentais. Estes estudos indicam metodologias que seguem padrões lógicos a fim de que o produto final seja compreendido massivamente e que as informações agrupadas sejam claras e atendam as necessidades potenciais. Estes estudos indicam um trabalho minucioso no que diz respeito à tradução dos elementos documentais até a elaboração de metadados descritivos. Os metadados podem possuir diferentes funções, tais quais: identificar, descrever, facilitar a busca e recuperação, proporcionar o acesso e a preservação, dentre outras funções (MÉNDEZ RODRÍGUEZ, 2002). Em relação a este conceito, Gilliland Swetland (2000, p. 1) adota uma definição ampla, conceituando metadado enquanto “a soma total do que pode ser dito sobre algum objeto informacional em algum nível de agregação”.

Os metadados descritivos são aqueles que descrevem um objeto informacional com o objetivo de identificar e remeter às informações deste mesmo

objeto, como por exemplo os registros de catalogação, palavras-chave e índices especializados. Este olhar sobre o objeto informacional interpreta-o como um item ou grupo de itens que pode ser manipulado, independente de tipo ou formato, como um objeto singular, com suas individualidades, por ser humano ou por um sistema computacional. Enfim, expostos alguns importantes referenciais teóricos que norteiam a pesquisa em andamento, a seguir são apontados os procedimentos metodológicos utilizados.

3 METODOLOGIA

A pesquisa em desenvolvimento é de natureza exploratória e descritiva. De acordo com seus procedimentos, convém dizer que é uma pesquisa de caráter bibliográfico ao utilizar a literatura científica da Biblioteconomia e Ciência da Informação, bem como das seguintes obras de referências filatélicas: Catálogo de Selos do Brasil (MEYER, 2015) e Guia dos Editais, Envelopes de Primeiro Dia de Circulação e Máximos Postais (MELLO, 2004).

Sendo o selo postal um objeto de pesquisa imagético, de caráter interpretável, torna-se ele objeto de estudo da Ciência da Informação como fonte informacional a ser analisada e disponibilizada como qualquer outro documento. Manini (2002) propõe um esquema para representar o conteúdo informacional de imagens baseada no sistema de análise proposto por Smit (1996) reunindo distintas categorias informacionais:

Quadro 2: esquema de indexação de imagens

Categoria	Conteúdo Informacional		Dimensão Expressiva	
	DE			SOBRE
	Genérico	Específico		
Quem / O Que				
Onde				
Quando				
Como				

Fonte: adaptado por Manini (2002)

Este esquema tem o objetivo de sistematizar as informações contidas no documento visual, a fim de levantar termos descritores que se baseiam tanto na imagem, quanto das inferências de quem o descreve, a partir de conhecimentos

prévios. O fato é que mesmo com a existência de instrumentos facilitadores, produzidos e estudados na Ciência da Informação, a extração de todo conteúdo de um documento imagético, especificamente o selo postal comemorativo, ainda é um desafio.

Baseados nos métodos e ferramentas já citados e levando em consideração o tipo documental em questão, um método pioneiro para a classificação de selos postais comemorativos é proposto e esquematizado a seguir:

Quadro 3 - Protótipo de modelo para classificar selos postais

ELEMENTOS PICTÓRICOS		ELEMENTOS VERBAIS	
DESCRIÇÃO DA IMAGEM-MOTIVO:		LEGENDA:	
INDIVÍDUO REPRESENTADO:		VALOR FACIAL:	
EVENTO REPRESENTADO:		DATA DE EMISSÃO:	
INSTITUIÇÃO REPRESENTADA:		EMISSOR:	
COLORAÇÃO:		Observações :	
RESUMO DE CONTEÚDO:			
PALAVRAS-CHAVE:			

Fonte: os autores (2017)

Considerando que a documentação filatélica brasileira é dividida em várias dezenas de tipos distintos de documentos e que este artigo traz os resultados parciais de uma pesquisa em andamento, analisaremos a classificação de 2 dos 150 selos postais do tipo comemorativo emitidos pelos Correios de 1900 até 1940. Nessa amostra, de um população de mais de 5.000 selos postais, estão incluídos selos separados dos blocos comemorativos emitidos no período, bem como selos de uma mesma série temática, mas que têm especificidades em relação a possibilidade de classificação.

4 RESULTADOS E DISCUSSÕES

Nesta pesquisa, o primeiro contato com o objeto de estudo foi feito por meio do Catálogo de Selos do Brasil RHM (2013). Esta é a principal obra de referência nacional, com publicação anual, utilizada pelos colecionadores, comerciantes e pesquisadores da documentação filatélica no Brasil.

Nele estão catalogadas os metadados de toda a documentação filatélica Brasileira desde seu surgimento em 1843 até os dias atuais. O detalhe interessante é que nesta obra foi elaborado um sistema de classificação dos autores que conta com uma codificação própria e o uso de termos descritores que identificam cada documento.

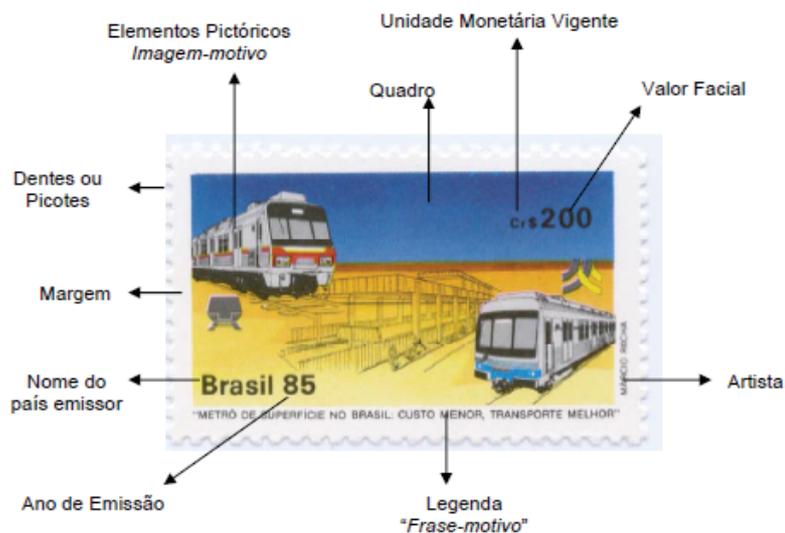
Inicialmente, a utilização desta obra de referência se fez necessária para identificar os selos que fariam parte do corpus da pesquisa. Contudo, após análise do sistema utilizado para descrever os documentos surgiu a necessidade de enriquecer as informações existentes no catálogo, de forma que seja feito o uso adequado dos termos descritores com um olhar atento sob o grau de precisão na representação desses documento, encarando as atribuições do catálogo como uma análise preliminar.

Desta forma, foi feita a leitura da seção de selos comemorativos emitidos entre 1900 até 1940 do catálogo. Foram consideradas as informações impressas nos selos postais comemorativos e também as atribuições dos editores de forma a complementar a análise descritiva para posterior atribuição de conceitos.

Diante disso foi pensada a importância que deve ser dada aos termos descritores do selo postal comemorativo levando em consideração a sua natureza e elementos intrínsecos a este documento, que por sua vez exigem maior grau de exaustividade, que (LANCASTER, 2004, p. 27) conceitua como o “emprego de termos em número suficiente para abranger o conteúdo temático principal do documento, ou seja, conseguir o maior número de termos indexadores para compreender a temática primeira do documento. Quanto mais termos forem utilizados para indexar um documento, mais acessível ele se tornará e, provavelmente, mais vezes será recuperado.

O selo postal comemorativo deve ser analisado em sua totalidade no processo de representação temática. Dito isso, na Figura 1, a seguir, tem-se os descritores que podem ser considerados na análise desse tipo de documento.

Figura 1 - Componentes do selo postal comemorativo



Fonte: Salcedo (2008, p. 116)

Diante da exposição dos elementos que devem ser levados em consideração ao analisar o selo postal comemorativo, bem como já exposto o método de classificação elaborado e utilizado, sigamos então para o enfoque prático da pesquisa. Seleccionamos inicialmente dois selos postais, retirados de uma coleção particular de um filatelista, para que o processo de representação seja analisado, a fim de ilustrar, na prática, a metodologia abordada na pesquisa. Os documentos e suas respectivas análises estão dispostos de acordo com a ordem crescente do ano de emissão de cada selo postal dentro do escopo temporal delimitado na pesquisa.

Figura 2 - 4º Centenário da Independência do Brasil



Fonte: os autores (2017)

ELEMENTOS PICTÓRICOS		ELEMENTOS VERBAIS	
DESCRIÇÃO DA IMAGEM-MOTIVO:	Representação de Dom Pedro 2º e sua guarda de honra. Dom Pedro empunhando uma espada, vestido com uniforme militar e montado em um cavalo às margens do rio Ipiranga.	LEGENDA:	Independência ou morte: 7 de setembro de 1822
INDIVÍDUO REPRESENTADO:	Dom Pedro de Alcântara; dragões da independência.	VALOR FACIAL:	200 réis
EVENTO REPRESENTADO:	O grito da independência proclamado por Dom Pedro 2º às margens do Rio Ipiranga.	DATA DE EMISSÃO:	1900
INSTITUIÇÃO REPRESENTADA:	Não aplicável.	EMISSOR:	Correio do Brasil
COLORAÇÃO:	Policromático; verde e amarelo com variações de tom.	Observações:	
RESUMO DE CONTEÚDO:	Comemoração do 4º centenário da independência do Brasil. O selo reproduz o momento histórico em que Dom Pedro 2º dá o grito da independência, evento que representa o momento em que o príncipe foi proclamado imperador e o Brasil é nomeado Império.		
PALAVRAS-CHAVE:	Independência do Brasil. Dom Pedro de Alcântara. Grito da independência. 1822.		

Figura 3 - Centenário da Revolução Republicana de Pernambuco



Fonte: os autores (2017)

ELEMENTOS PICTÓRICOS		ELEMENTOS VERBAIS	
DESCRIÇÃO DA IMAGEM-MOTIVO:	Representação da Bandeira de Pernambuco toda em azul entre duas colunas jônicas que sustentam o nome BRAZIL. No canto superior esquerdo encontra-se o brasão pernambucano.	LEGENDA:	Centenário da Revolução Republicana de Pernambuco. Bandeira da república 6 de março de 1817.
INDIVÍDUO REPRESENTADO:	Não aplicável	VALOR FACIAL:	100 réis
EVENTO REPRESENTADO:	Revolução pernambucana de 1817	EMISSOR:	Correio do Brasil
INSTITUIÇÃO REPRESENTADA:	República de pernambuco	DATA DE EMISSÃO:	1917
COLORAÇÃO:	Monocromático	Observações:	
RESUMO DE CONTEÚDO:	Comemoração do centenário da revolução pernambucana de 1817. O selo aborda em destaque a bandeira de Pernambuco criada em 1817 no movimento separatista, mas que só foi oficializada no seu centenário, em 1917, pelo então presidente da província Manuel Antônio Pereira Borda.		
PALAVRAS-CHAVE:	Revolução Pernambucana. Bandeira da república. 1817.		

O grau de especificidade da classificação documental irá depender do quão exaustivo será o trabalho do classificador. Os documentos variam de acordo com a sua tipologia, entre eles existem os documentos não-gráficos que englobam os audiovisuais, os pictóricos e os sonoros. No caso que interessa ao estudo os selos postais enquadram-se na qualidade de documentos verbovisuais (pictóricos).

Observa-se neste estudo preliminar o quanto os graus de especificidade e exaustividade da representação temática alteram-se de acordo com cada imagem a ser analisada, para além de todas as especificidades comuns aos selos postais comemorativos. Importante salientar que a pesquisa em outras fontes para além do documento a ser analisado, bem como a compreensão do assunto tratado por parte de quem classifica, incluindo os conhecimentos inerentes à pessoa, são indispensáveis para o enriquecimento deste serviço.

6 CONSIDERAÇÕES FINAIS

Sob um olhar ainda inacabado, o debate gerado desde a ferramenta metodológica, elaborada para a classificação do corpus da pesquisa, bem como a sua aplicação atende ao propósito da pesquisa e funcionam de forma eficaz e efetiva, considerando o quantitativo que já foi trabalhado sob caráter experimental.

Diante dos resultados já alcançados pretende-se daqui em diante dar continuidade ao tratamento dos documentos postais diante da perspectiva adotada, tendo em vista que a partir das especificidades que serão encontradas diante da análise dos documentos restantes o método aplicado deve ser mutável e recorrentemente discutível. Cabe ainda indicar que os produtos finais deste trabalho fomentarão o banco de dados do REFIBRA contribuindo com a difusão livre da memória e da identidade nacional.

A disponibilização desse tipo de patrimônio documental em formato aberto na WEB significa criar a chance para que a população preserve e também renove o olhar sobre a sua própria identidade, bem como possibilita o amplo acesso aos dados e informação filatélica nacional alcançando todo público-alvo em potencial, para que tenha independência relativa à sua pesquisa e que possam fazer livre uso do material encontrado.

REFERÊNCIAS

BRASIL. Ministério da Cultura. Lei nº 12.343, de 2 de dezembro de 2010. Plano Nacional de Cultura – PNC. **Diário Oficial da União** - Seção 1, de 03/12/2010. Disponível em: <<http://goo.gl/Eewttx>> Acesso em: 20 abr. 2017.

BUCKLAND, M. K. Information as thing. **Journal of the American Society for Information Science (JASIS)**, v. 45, n.5, p.351-360, 1991.

BURKE, Peter. **Uma história social do conhecimento: de Gutenberg a Diderot**. Rio de Janeiro: Zahar, 2003.

COSTA, Antônio Firmino. Classificações sociais. **Leitura**. Lisboa, v. 3, n. 2, p. 65-75, out. 1997/abr. 1998.

FLOCH, Jean-Marie. **Alguns fundamentais em Semiótica Geral**. Documentos de estudo do Centro de Pesquisas Sociosemióticas, 2001.

GILLILAND-SWETLAND, A. J. Introduction to metadata: Setting the stage. 2000. Disponível em: <https://goo.gl/CH9Ebk> Acesso em: 19 abr. 2017.

LANCASTER, L. W. **Indexação e resumos: teoria e prática**. 2. ed. Brasília [DF]: Briquet de Lemos, 2004.

MANINI, Miriam P. **Análise documentária de fotografias: um referencial de leitura de imagens fotográficas para fins documentários**. São Paulo, 2002. Tese (Doutorado) – Escola de Comunicações e Artes, USP.

MELLO, Eduardo Cavalcanti de. **Guia dos Editais, Envelopes de Primeiro Dia de Circulação e Máximos Postais emitidos pelos Correios do Brasil: 1965-2003**. João Pessoa: Novo Mundo, 2004.

MÉNDEZ RODRÍGUEZ, E. Metadados y recuperación de información: estándares, problemas y aplicabilidad en bibliotecas digitales. Gijón: Trea, 2002. 429 p.

MEYER, Peter. **Catálogo de Selos do Brasil**: 2015. São Paulo: RHM, 2015.

ORGANIZAÇÃO DAS NAÇÕES UNIDAS PARA A EDUCAÇÃO, A CIÊNCIA E A CULTURA. A memória do mundo na era digital: digitalização e preservação. Vancouver, [BC]: UNESCO, 2012. Disponível em: <<http://goo.gl/yBN70e>> Acesso em: 20 abr. 2017.

PIEIDADE, Maria. **Introdução a Teoria da Classificação**. 2. ed. Rio de Janeiro: Interciência, 1983

PIETROFORTE, Antônio Vicente. **Semiótica visual**: os percursos do olhar. São Paulo: Contexto, 2004.

RAMALHO E OLIVEIRA, Sandra. **Imagem também se lê**. São Paulo: Rosari, 2005.

SALCEDO, Diego Andres. **A ciência nos selos postais comemorativos brasileiros**: 1900-2000. Recife: EDUFPE, 2010.

_____. A imagem científica nos selos postais brasileiros. IN: SALCEDO, Diego Andres; OLIVEIRA, Maria Cristina G.; OTERO, Maria Mercedes D. F. (orgs.). **Construção, prática e identidades da Ciência da Informação**. Recife: NECTAR/UFPE, 2008.

_____. ÁVILA, Bruno; FEITOSA, Kézia. Repositório Filatélico Brasileiro: ações e primeiros passos. CONFERÊNCIA SOBRE TECNOLOGIA, CULTURA E MEMÓRIA. Recife, 2015. **Anais...** Recife: LIBER, 2015. Disponível em: <<https://goo.gl/jZluPW>> Acesso em: 20 abr. 2017.

SMITH, Johanna W. A representação da imagem. **Informare**, Rio de Janeiro, v.2, n.2. p.28-36. jul.-dez. 1996.



CONVERGÊNCIA ENTRE REPOSITÓRIOS DIGITAIS E PERIÓDICOS

Bruno Tenório Ávila
Milena Beatriz Lira Dias da Silva

RESUMO: A comunidade científica sofre problemas em sua comunicação, principalmente por causa da demora da publicação de pesquisas como também o escurecido processo de revisão por pares. Este trabalho possui como objetivo a proposta de uma discussão acerca de possíveis soluções para estes problemas. Para isso, através de uma pesquisa exploratória, aborda brevemente sobre a história do periódico científico e como tomou a forma que possui hoje o percurso do Movimento de Acesso Aberto e quais influências ele teve sobre a comunicação científica. Define o papel do repositório digital na comunidade científica como conclusões, afirma a existência da necessidade de um processo de publicação científica mais democrático, público e rápido. Para isso, propõe a criação de um produto que seja o resultado dos processos de qualidade e velocidade de numa publicação científica, tendo o repositório digital como local de publicação.

Palavras-chave: Comunicação científica. Publicação periódica. Repositório digital. Revisão por pares.

ABSTRACT: The scientific community suffers from problems in its communication, mainly due to the delay in the publication of researches as well as the obscure process of peer reviews. This paper has the goal to propose a discussion on the possible solutions to those problems. In order to do that, through an exploratory research it briefly touches on the history of the scientific journal and how it took the form that it has today and the course of the Open Access Movement and what influences it had on scientific communication. It defines the role of the digital repository in the scientific community as conclusions; it asserts the existence of the need for a faster, more democratic and public process of scientific publication. To this end, it proposes the creation of a product that is the result of the processes of quality and speed of a scientific publication, having the digital repository as a place of publication.

Keywords: Scientific communication. Periodical publication. The digital repository. Peer review.

1 INTRODUÇÃO

O desenvolvimento e progresso da ciência dependem da comunicação das novas ideias e avanços entre cientistas. É através da comunicação científica que publicações são viabilizadas, para que pesquisadores possam acessá-las e então, produzir mais conhecimentos científicos. É dever da comunicação científica o estímulo à compreensão de novos saberes e de fornecer retorno para os pesquisadores.

A comunicação científica pode ser definida desde o momento que o cientista concebe a ideia até o momento de sua publicação; dentro de todas as etapas da

pesquisa, escrita, veiculação em um documento formal até sua disseminação, que muitas vezes, é restrito. “Como também os periódicos pagos limitam a divulgação do conhecimento a um número restrito de especialistas que podem arcar com os custos de acesso” (MORENO; LEITE; 2006, p. 83).

O processo de produção científica é longo; o pesquisador pode demorar anos para que seu artigo esteja apto a ser submetido para publicação e passar pela revisão dos pares. O atual processo de revisão de artigos científicos, que é fechado e lento, deixa brechas à abusos e avaliações equivocadas pelos pares, além de ser um meio passível de corrupção. Price e Flach (2017) também reconhecem que a revisão por pares é um processo obscuro e, indo mais longe, afirmam que sua configuração no mundo editorial já está enraizada e difícil de mudança. Portanto, é preciso um novo ambiente suscetível a mudanças, como os repositórios digitais.

Os repositórios digitais são fontes de informação em plataforma online de acesso livre, que permite o armazenamento de conhecimentos científicos e sua recuperação. Existem os repositórios digitais a) institucional, que é composto pela produção científica de um instituto de pesquisa ou universidade e b) temático, que é composto pela produção intelectual de determinada área temática (INSTITUTO BRASILEIRO DE INFORMAÇÃO EM CIÊNCIA E TECNOLOGIA, 2017).

Kling (2004 apud WEITZEL, 2006) acredita que o aumento da velocidade e transparência da literatura científica, acontecerá fora do sistema de periódicos. Van de Sompel e Lagoze afirmam que os repositórios de e-prints possibilitam uma forma mais igualitária e eficiente de divulgação de pesquisas.

Os repositórios, atualmente, oferecem alternativas para alguns dos problemas mencionados. Deixar produções acadêmicas institucionais disponíveis simplifica o acesso à ciência, portanto, contribuindo para a comunicação e divulgação científica.

Este é um indício de uma lenta convergência dos cientistas às formas mais democráticas de acesso à ciência, sinalizando uma necessidade de produção científica mais democrática, mantendo a qualidade. Neste sentido, os repositórios digitais e periódicos são mutuamente complementares: um pode suprir os pontos que o outro deixa a desejar.

Assim, é percebido um ponto de convergência entre esses dois véis da comunicação científica. Enfim, este trabalho tem como objetivo, através de uma pesquisa exploratória, propor uma discussão acerca de soluções para os problemas de qualidade e velocidade das publicações científicas.

2 COMUNICAÇÃO CIENTÍFICA E O MOVIMENTO DE ACESSO ABERTO

A informação científica é a base para o desenvolvimento científico e tecnológico que, através do seu compartilhamento, possibilita a geração de novos conhecimentos. “Portanto, a comunicação das pesquisas científicas tem a função de validar e consolidar os avanços da ciência e tecnologia, bem como permitir o registro e a recuperação das informações” (BOMFÁ, 2009, p.35). De fato, a comunicação se localiza no “coração” da ciência (MEADOWS, 1999) de tal forma que o avanço do conhecimento científico depende da forma que ela é estabelecida na comunidade.

A origem da comunicação científica foi a troca de ideias informais entre pesquisadores, cuja necessidade de comunicar resultados de pesquisa, como também de solicitação de revisões do que estavam produzindo, fossem realizados, até o surgimento do primeiro periódico.

O primeiro periódico surgiu no século 17 na Europa como forma de simples anúncio em folhetos acerca dos andamentos e/ou resultados de pesquisas científica e foi se modificando até tomar a forma atual, com a formação (introdução, metodologia, revisão teórica, resultados e conclusão) como também aos tipo de artigos publicados, como de revisão ou original (MIRANDA; PEREIRA, 1996).



Figura 1 – Fluxo da editoração de publicações periódicas científicas
Fonte: Bomfá, 2009

Como demonstra a Figura 1, a comunicação científica abrange, segundo Garvey (1979 apud Miranda e Pereira, 1996, p. 375), "o conjunto de atividades associadas à produção, disseminação e uso da informação". O processo de comunicação se dá desde o momento que o cientista concebe a ideia até o momento de sua publicação.

“As discussões em torno dos resultados das pesquisas científicas remontam da Grécia Clássica, através da atividade dos grandes cientistas” (BOMFÁ, 2009, p.35). Portanto, entende-se que a comunicação científica se consolida através da publicação periódica, onde a exposição de pesquisas é realizada para amplo acesso à comunidade científica.

Com base nos conceitos apresentados pelos autores entende-se que o periódico científico, ou técnico-científico, é uma publicação seriada, com periodicidade definida, cujo objetivo é oferecer ao autor registro, disseminação e recuperação das pesquisas mediante serviços de indexação e mecanismos de busca, bem como o reconhecimento e prestígio profissional. O caráter científico se dá mediante a aprovação e o reconhecimento dos pares (BOMFÁ, 2009, p.47).

Com a popularização do uso da Web na década de 1990, os periódicos migraram pouco a pouco para o formato eletrônico. Desta forma, eles puderam inovar no que tange à comunicação científica, com a inserção de recursos integrados que não podiam ser usados no formato impresso de publicação, como possibilidade de processamento de dados, indexação automática, recuperação e busca em bases de dados, como também a utilização de recursos gráficos visuais (BOMFÁ, 2009; LE COADIC, 1996).

Entretanto, a inegável transformação do monopólio das editoras quanto ao conhecimento científico, com muita espera para realização da publicação por partes dos autores como também as cobranças de altos custos para acesso às publicações, geraram frustrações para a comunidade de pesquisadores. Dessa forma, como consequência, surgiu o Movimento de Acesso Livre, que, é resultado de:

[...] (1) de uma reação dos pesquisadores ao modelo de negócios de editoras comerciais de revistas científicas (e seus preços cada vez mais altos preços de assinatura); e da (2) crescente conscientização do aumento de impacto provocado pela disponibilização de documentos científicos livres de barreiras ao acesso. O mote do movimento mundial em favor do Acesso Livre a resultados de pesquisa, portanto, é a disseminação ampla e irrestrita dos resultados de pesquisas financiadas com recursos públicos (BAPTISTA *et. al.*, 2007, p.2).

O precursor do movimento acesso aberto pode ser considerado o Projeto Gutenberg – primeiro agregador de livros integrados num computador em rede – lançado por Michael Hart em 1971. Posteriormente, ocorre em 1989 o lançamento da *Psychology*, de Steven Harnad, a primeira revista online de consulta pública e

livre, que veio a ser revisada pelos pares em 1990. Em 2002, ocorreu a declaração em Budapeste, conhecida como Budapest Open Access (BOAI) e em 2003, a Declaração de Berlim sobre Acesso ao Conhecimento nas Ciências e Humanidades¹ e a Declaração de Bethesda². Finalmente, em 2005, o Brasil aderiu oficialmente ao movimento através do Manifesto Brasileiro de Apoio ao Acesso Aberto, pelo Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT)³.

A Declaração de Budapeste (2003) trás duas formas de acesso aberto *I*. A via verde, onde o autor realiza o depósito de sua obra num repositório de acesso aberto após sua publicação. Para isso, o autor necessita da permissão da editora; e *II*. A via dourada, onde o autor publica em revistas de acesso aberto. Estas revistas não cobram taxas por suas assinaturas e deixam suas edições para livre acesso na internet, entretanto, as revistas podem redirecionar seus gastos de publicação aos autores.

A cartilha de manifesto conhecida como um marco principal ao movimento de acesso aberto, BOAI, afirmou que a comunicação científica aberta precisava ser alavancada para algo além de “pequenas porções da literatura de periódicos” (2002, p.1, tradução nossa).

Nos anos 2000 a Open Archives Initiative (OAI) é consolidada dentro da iniciativa de acesso aberto, se definindo como um sistema de arquivos abertos, onde os objetivos científicos localizados nele podem ser acessados livres de custo e interoperáveis. Dentro das recomendações abordadas no modelo, o ponto principal é a criação de um repositório digital institucional de acesso aberto, onde a memória científica da instituição seria preservada em longo prazo e disseminada sem fins lucrativos.

Na literatura, o conceito dos repositórios digitais está bastante atrelado à preservação dos documentos digitais. Segundo a Research Libraries Group (RLG) e Online Computer Center Library (OCLC) (2002), que fornecem um modelo para a certificação de um repositório digital, está entre os principais pontos a serem seguidos é a preservação dos documentos digitais a longo prazo e então, dar acesso à eles. Para um RT ou RI ser considerado confiável e de qualidade, ele precisa ter critérios de preservação, tendo a segurança de que os documentos

¹ <https://repositorium.sdum.uminho.pt/about/DeclaracaoBerlim.htm>

² <http://legacy.earlham.edu/~peters/fos/bethesda.htm>

³ <http://livroaberto.ibict.br/Manifesto.pdf>

depositados ali não vão se perder, seja por uma perda de dados ou obsolescência tecnológica.

A comunidade científica sentiu o repositório digital como um divisor de águas, por ter quebrado algumas barreiras, como diminuição do custo e tempo para o acesso à pesquisa, como também o armazenamento de objetos científicos; também retirando as publicações científicas do monopólio das grandes editoras e periódicos (SANTAREM SEGUNDO, 2010). Assim como forma de preservação em longo prazo dos documentos institucionais, atribuindo também aos RIs o papel de guarda de memória institucional ou no caso dos RTs, depósito de *eprints*, preservação e disponibilização de forma digital de diversos documentos nos RTs.

Um caso diferenciado do conceito popular de preservação e disponibilização de artigos científicos aplicado aos repositórios é o *Arxiv* (<http://arxiv.org/>), que permite a realização de depósitos de trabalhos inéditos é o repositório da Universidade de Cornell nos Estados Unidos que iniciou o serviço em 1991, ou seja, curiosamente antes das iniciativas de acesso livre. Em janeiro de 2015, registrou-se a marca de um milhão de *e-prints*. Além do serviço de auto arquivamento, o *Arxiv* é visitado por usuários em busca dos trabalhos científicos mais recentes, que são publicados diariamente, em várias áreas de atuação.

O *Arxiv* demonstra um possível uso de repositório para além de memória e guarda e compartilhamento de trabalhos e artigos já aprovados pelos pares; um local onde seja possível a inclusão de obras inéditas de forma rápida (dentro de 24 horas a publicação pode ser aprovada ou não pela moderação do repositório). Portanto, já é confirmada a possibilidade de usar repositórios digitais para a realização de publicações científicas de forma rápida, que também estará preservada e disponibilizada para a comunidade científica.

É percebida uma convergência dos periódicos para os repositórios, como a disponibilização livre de seu conteúdo para acesso, como também dos repositórios para os periódicos, através da possibilidade de publicação. Entretanto, se é possível unir o conceito de publicação rápida com publicação de qualidade, será discutido na seção 3.

3 CONVERGÊNCIA ENTRE REPOSITÓRIOS DIGITAIS E PERIÓDICOS

Byrnes *et. al.* (2014) propõem quatro pilares em que a comunidade científica precisa se apoiar visando um futuro mais aberto para as publicações acadêmicas. São elas: 1) uma abrangência maior do que é definido por produto acadêmico, 2) acesso aberto e imediato a esses produtos desde o início de sua revisão até a versão final, 3) abrir uma revisão pública para os produtos acadêmicos e 4) melhoria da capacidade de avaliação de produtos acadêmicos e as contribuições gerais dos estudiosos.

Os pilares propostos por Byrnes e colegas se apoiam, principalmente, na abertura do processo de revisão no aumento da capacidade de reconhecimento de qualidade tanto de trabalhos, quanto de revisores e pesquisadores. Entretanto, é reconhecido que apesar de mudanças substanciais necessitam serem realizadas para uma progressão a este cenário, está havendo uma série de avanços, identificados em revistas periódicas como Atmospheric Chemistry & Physics (<http://www.atmospheric-chemistry-and-physics.net/>), PLOS One, PeerJ, Push (<http://push.cwcon.org/about/>), Biology Direct, entre outros, que liberam seus artigos ao público através da Web e possibilitam uma publicação mais rápida de seus artigos (no caso da PLOS One).

Os periódicos já estão com suas bases firmadas e dominam o mercado de publicações científicas, juntamente com os livros. O mercado consumidor está consolidado, juntamente com a existência de uma infraestrutura de acesso já e existência de grupos e instituições que legitimam os periódicos (MIRANDA; PEREIRA, 1996). Ainda, como também os problemas de tempo de publicação, custo e o compartilhamento do conhecimento científico estejam diretamente conectados com as editoras e periódicos.

Um dos problemas trazidos diante de tantos periódicos no atual mercado consumidor pode ser considerado oriundo das avaliações de qualidade para periódicos, que é exigida de forma comumente pela comunidade. São eles: *i.* o fator de impacto, que é a medição de prestígio que aquele periódico tem e é medido pela quantidade de vezes que aquele periódico foi citado em estudos (PINTO; ANDRADE, 1999); *ii.* a regularidade de publicação e indexação em bases secundárias.

Pressupõe-se que cientistas experientes e com boa qualificação tenham um melhor desempenho em pesquisas, assim, ele é citado. Citações aumentam o fator de impacto do periódico, tornando-o referencial na área. Dessa forma, acaba por ser

criada uma tendência aos periódicos aceitarem publicações com mais facilidade de cientistas mais famosos, tendo por consequência o chamado “Efeito Mateus”, onde os cientistas famosos ficam mais famosos e os desconhecidos permanecem desconhecidos.

Também, a concorrência para a publicação de artigos em periódicos de prestígio faz com que seja mais difícil a publicação.

As altas taxas de rejeição de artigos dos periódicos científicos de maior prestígio fomentaram um ciclo vicioso onde somente poucos conseguem publicar. Outras questões pertinentes aos critérios de avaliação e ao tempo que um artigo submetido leva para ser avaliado também contribuíram para restringir, sobretudo, a ação dos benefícios dos periódicos científicos para o desenvolvimento da ciência (WEITZEL, 2006, p. 149).

Os problemas de comunicação ainda permeiam a comunidade científica, apesar dos inúmeros repositórios digitais e periódicos de acesso aberto existentes. Mesmo havendo a divulgação do conhecimento científico de forma mais democrática, a 1) impossibilidade de uma publicação rápida com qualidade (no caso dos repositórios que permitem depósitos inéditos), como também o 2) obscurecido e demorado processo de revisão por pares não foram sanados.

Buck, Flagan e Coles (apud WEITZEL, 2006) acreditam, indo mais longe, que o sistema de avaliação dos periódicos científicos está sofrendo uma disjunção, ou como os autores chamam, “desacloppamento” (*sic*) do processo editorial dos periódicos científicos e está migrando para novas alternativas em ambientes digitais. Para isso, este trabalho sugere os repositórios digitais como local para um estabelecimento de um sistema de revisão por pares.

Como a popularização e uso frequente dos repositórios possuindo pouco mais que uma década (NICHOLAS *et. al.*, 2012), suas definições e objetivos ainda estão aptos a se formarem e transformarem; a ideia do repositório se tornar uma ferramenta capaz de solucionar os problemas da comunicação científica pode ser viabilizada.

Os RDs como estão configurados atualmente, possuem como diferencial o fácil acesso a produção intelectual de instituições ou de determinados temas, divulgando os avanços científicos de instituições. Por outro lado, trazem produções vinculadas às suas instituições que muitas vezes são públicas; e estas produções muitas vezes se limitam a monografias, trabalhos de conclusão de curso, dissertações e teses da instituição, ou artigos já publicados.

O repositório digital hoje é, resumidamente, caracterizado apenas pelo depósito de um objeto e não por sua publicação; todos os artigos contidos na plataforma já foram finalizados, revisados e de qualidade atestada e estão aptos a fazerem parte da literatura científica. Portanto, no fluxo de editoração de publicações periódicas demonstrado na Figura 1, o repositório se encontra na disponibilização e divulgação do trabalho científico, divulgando as pesquisas científicas para outros pesquisadores; para que assim eles possam produzir.

Contudo, um exemplo que foge ao estabelecido é o repositório Arxiv. Nele, é possível a realização de publicações inéditas. O serviço oferecido pelo Arxiv é a possibilidade de publicação dentro de 24 horas, cujo artigo submetido passa por uma comissão (moderação) que faz uma revisão; porém, não relativa a conteúdo (como o processo de revisão pelos pares), e sim à forma (verificação de comentários inadequados, metadados, tema, se o que está sendo submetido se enquadra na política do repositório).

Apesar de ser um exemplo diferenciado do que é entendido por ser repositório (RLC; OCLC, 2002), ele é a confirmação da possibilidade de outro uso para repositório além de preservação de publicações já realizadas ou compartilhamento de conhecimentos científicos: o uso de repositórios para publicações rápidas e inéditas. Um caso em destaque é publicação de artigos para a solução da conjectura de Poincaré, um importante problema matemático, apenas no Arxiv (PERELMAN, 2002, 2003a, 2003b). Entre julho de 1991 e agosto de 2015, o número de depósitos mensais cresceu a uma taxa de 28,48 documentos⁴ atingindo, em 2015, uma média de 8.272 depósitos por mês. Em janeiro de 2015, registrou-se a marca de um milhão de e-prints. Além do serviço de auto arquivamento, o Arxiv é visitado por usuários em busca dos trabalhos científicos mais recentes, que são publicados diariamente, em várias áreas de atuação.

Entretanto, a qualidade do que foi publicado de forma inédita num repositório, como se dá hoje, não passa por um processo de análise; o processo de produção pode se tornar informal quando outros cientistas acessam a pesquisa e contatam o autor através de meios fora da plataforma digital, como por exemplo, e-mail (FERREIRA; MODESTO; WEITZEL, 2003).

⁴ Valor obtido por interpolação linear dos dados fornecidos em http://arxiv.org/stats/monthly_submissions resultando em $d(x) = 28,48x - 472,96$ com $R^2 = 0,96$, onde $d(x)$ é o número de depósitos no mês x .

Este é um indício de uma necessidade que a comunidade científica tem em se comunicar, não apenas da maneira formal como publicações em periódicos (FERREIRA; MODESTO; WEITZEL, 2003); mas sim de um sistema que possibilite interações humanas, dando suporte a pesquisa e, principalmente, a realização de publicações de forma mais veloz.

Para um repositório digital ser um meio possível para publicação, é necessário uma política bem delineada, um sistema com ferramentas e aplicações que atendam à necessidade de interação entre cientistas e também a necessidade de acesso à pesquisas, como também um sistema de revisão eficiente e eficaz. Com estes processos sendo viabilizados num repositório digital, ele se tornaria uma fonte primária de comunicação científica, assim como os periódicos; desde que o que esteja sendo publicado seja um artigo original (CUNHA, 2001).

Portanto, o ponto de convergência entre periódicos e repositórios é o produto resultante entre os processos de qualidade e velocidade de publicação. Com isso, os autores deste trabalho sugerem a formação de um sistema de publicação dentro de um repositório digital, de caráter livre, de qualidade, não anônimo e não monetarizado.

4 CONSIDERAÇÕES FINAIS

Os problemas de comunicação ainda permeiam a comunidade científica, apesar dos inúmeros repositórios digitais e periódicos de acesso aberto existentes. Pois, é notável que mesmo havendo a divulgação do conhecimento científico de forma mais democrática, a impossibilidade de uma publicação rápida com qualidade como também o obscurecido processo de revisão por pares não foi sanado. A comunidade científica concorda que o acesso ao conhecimento precisa ser mais democrático, e que a revisão pelos pares precisa ser feita de forma aberta e ainda, demonstra a necessidade de reconhecimento destes revisores.

Publicações sem qualidade constam no currículo, mas não acrescentam em nada à comunidade científica e o processo de pesquisa. Portanto, a discussão aqui realizada propõe a criação de um modelo de publicação por revisão pelos pares dentro de um repositório digital, de forma que consiga atender aos problemas da comunidade científica, que integre tanto os recursos dos periódicos como os do

repositório. Entretanto, deixa em aberto a formação, processos e uso deste produto, para ser confeccionado nos próximos trabalhos.

REFERÊNCIAS

BAPTISTA, A.; COSTA, S.; KURAMOTO, H.; RODRIGUES, E. Comunicação científica: o papel da Open Archives Initiative no contexto do Acesso Livre. **Encontros Bibli**: revista eletrônica de biblioteconomia e ciência da Informação, n. esp., 2007.

BUCK, A. M.; FLAGAN, R. C.; COLES, B. Scholars' forum: a new model for scholarly communication, [S.L.:s.n.], 2002.

BOMFÁ, C. R. Z. **Modelo de gestão de periódicos científicos eletrônicos com foco na promoção da visibilidade**. 2009. 238f. Tese (Doutorado em Engenharia de Produção) – Universidade Federal de Santa Catarina, Programa de pós-graduação em Engenharia de Produção, 2009.

BYRNES, J. E. K.; BASKERVILLE, E. B.; CARON, B.; NEYLON, C. ; TENOPIR, C.; SCHIDHAUER, M.; BUDDEN, A.; AARSSSEN, L. W.; LORTIE, C. J. The four pillars of scholarly publishing: the future and a foundation. **Ideas in Ecology and Evolution**, [s.l.], v. 7, p. 27-33, 2014. Disponível em: <<http://ojs.library.queensu.ca/index.php/IEE/article/view/4628/5123>>. Acessado em: 4 março 2017.

CUNHA, M. B. **Para saber mais: fontes de informação em ciência e tecnologia**. Brasília: Briquet de Lemos, 2001. 168p.

FERREIRA, S.M.; MODESTO, F.; WEITZEL, S.R. Comunicação científica e o protocolo OAI: uma proposta na área de ciências da comunicação. In: INTERCOM - CONGRESSO BRASILEIRO DE CIÊNCIAS DA COMUNICAÇÃO, 26., 2003, Belo Horizonte. **Anais...** Belo Horizonte: [s.n.], 2003.

GARVEY, W.D. **Communication**: the essence of science. Oxford: Pergamon, 1979.

INSTITUTO BRASILEIRO DE INFORMAÇÃO EM CIÊNCIA E TECNOLOGIA. **Sobre Repositórios Digitais**. Disponível em: <<http://www.ibict.br/informacao-para-ciencia-tecnologia-e-inovacao%20/repositorios-digitais>>. Acesso em: 22 maio 2017.

KLING, R. The internet and unrefereed scholarly publishing. In: CRONIN, Blaise (ed.). **Annual Review of Information Science and Technology**, Medford, NJ, v. 38, p. 591-631, 2004.

LE COADIC, Y. F. **A ciência da informação**. Brasília: Briquet de Lemos/Livros, 1996. 119 p.

MEADOWS, A. J. **A comunicação científica**. Brasília: Briquet de Lemos/Livros, 1999. 268p.

MIRANDA, D.B.; PEREIRA, M.N.F. O periódico científico como veículo de comunicação: uma revisão de literatura. **Ciência da Informação**, Brasília, v. 25, n. 3, p. 375-382, set./dez. 1996.

MORENO, F. P.; LEITE, F. C. L.; ARELLANO, M. A. M. Acesso livre a publicações e repositórios digitais em ciência da informação no Brasil. **Perspectivas em Ciência da Informação**, Belo Horizonte, v.11 n.1, p. 82-94, 2006.

PINTO, A. C.; ANDRADE, J.B. fator de impacto de revistas científicas: qual o significado deste parâmetro? **Química nova**, São Paulo, v. 22, n. 3, 1999. Disponível em: <<http://www.scielo.br/pdf/qn/v22n3/1101.pdf>>. Acessado em: 21 maio 2017.

PERELMAN, G. **The entropy formula for the Ricci flow and its geometric applications**. 2002. Disponível em: <<http://arxiv.org/abs/math/0211159>>.

PERELMAN, G. **Ricci flow with surgery on three-manifolds**. 2003a. Disponível em: <<http://arxiv.org/abs/math/0303109>>.

PERELMAN, G. **Finite extinction time for the solutions to the Ricci flow on certain three-manifolds**. 2003b. Disponível em: <<http://arxiv.org/abs/math/0307245>>.

PRINCE, S.; FLACH, P. A. Computational support for academic peer review: a perspective from artificial intelligence. **Communications of the ACM**, [s.l.], v. 6, n. 3, p. 70-79, 2017. Disponível em: <<http://cacm.acm.org/magazines/2017/3/213825-computational-support-for-academic-peer-review/fulltext>>. Acessado em: 4 março 2017.

RESEARCH LIBRARIES GROUP. **Trusted Digital Repositories: Attributes and Responsibilities**. RLG, 2002.

SANTAREM SEGUNDO, J. E. **Representação iterativa: um modelo para repositórios digitais**. 2010. 224 f. Tese (Doutorado) - Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, São Paulo, 2010. Disponível em: <<http://hdl.handle.net/11449/103346>>. Acessado em: 3 março 2017.

VAN DE SOMPEL, H.; LAGOZE, C. The Santa Fe Convention of the open archives initiative. **D-Lib Magazine**, [s.l.], v.6, n.3, 2000. Disponível em: <<http://www.dlib.org/dlib/february00/vandesompel-oai/02vandesompel-oai.html>>. Acessado em: 22 maio 2017.

WEITZEL, S.R. **Os repositórios de e-prints como nova forma de organização da publicação científica: o caso da área de ciências da comunicação do Brasil**. 2006. 361f. Tese (Doutorado em Ciência da Informação) – Escola de Comunicação e Artes, Universidade de São Paulo, 2006. Disponível em: <<http://www.teses.usp.br/teses/disponiveis/27/27151/tde-14052009-133509/pt-br.php>>. Acessado em: 22 maio 2017.



DADOS ABERTOS E CURADORIA DIGITAL NO PROJETO *JUGANDO A DEFINIR LA CIENCIA*: o papel do *Design* da Informação

Laís Alpi Landim

Maria José Vicentini Jorente

RESUMO: O acesso facilitado às tecnologias possibilitou um aumento no potencial de compartilhamento de dados, cuja consequência foi o surgimento do movimento pela ciência aberta, práticas científicas em que os dados coletados são compartilhados, além dos resultados (acesso aberto), e que permitem a participação de sujeitos não especializados (ciência cidadã). Esse novo patamar da ciência exige cuidados específicos de gerenciamento dos dados, o que pode ser encontrado nas pesquisas em curadoria digital, considerada aqui como uma abordagem holística de gerenciamento de dados digitais em todo o seu ciclo de vida, o que envolve um planejamento de processos que vão do momento de sua criação até a disponibilização e acesso. Neste trabalho, abordamos o *Microscopi*, ambiente digital de compartilhamento de dados do projeto *Jugando a Definir la Ciencia*, sob a perspectiva da ciência aberta, a fim de refletir sobre a necessidade de uma abordagem de curadoria digital e *design* da informação para o gerenciamento e apresentação desses dados. A pesquisa caracteriza-se como descritivo-exploratória. O projeto pode ser considerado uma iniciativa valiosa de ciência cidadã, que pode inspirar a realização de projetos similares, e os dados de pesquisa se beneficiariam de uma abordagem de curadoria digital e de *design* da informação.

Palavras-chave: Informação e tecnologia. Curadoria digital. Design da Informação. Dados abertos. Ciência cidadã.

ABSTRACT: Easier access to technologies has enabled an increase in the potential for data sharing and collaborative work, which has resulted in the emergence of the open science movement, scientific practices in which data is shared and accessible, as well as results (open access), and that allow the participation of subjects that are not in the traditional research institutions (citizen science). These new science practices require specific procedures of data management, which can be found in digital curation research (considered here as a holistic approach to digital data management throughout its life cycle, which involves process planning from the moment of its creation until the availability and access). In this work, we approach the *Microscopi*, a digital data sharing environment of the project *Jugando a Definir la Ciencia*, from the perspective of open science, in order to reflect on the need for a digital curation and information design approach to the management and presentation of this data. The research is characterized as descriptive-exploratory. The project can be considered a valuable citizen science initiative that can inspire similar projects, and the research data would benefit from a digital curation and information design approach.

Keywords: Information and technology. Digital curation. Information Design. Open data. Citizen science.

1 Introdução

O desenvolvimento das tecnologias de informação e comunicação (TIC), observado a partir da década de 1980, ocasionou mudanças profundas em nosso cotidiano, principalmente com a popularização da internet, a partir da década de 1990, o que resultou em possibilidades crescentes de acesso às informações por parte da população em geral – abriu-se um grande acervo bibliográfico a partir de um computador em sua própria casa ou cidade. Nesse cenário, as facilidades de acesso e disponibilização de uma quantidade cada vez maior de dados trouxeram o potencial de um maior compartilhamento de dados de pesquisa, entre pesquisadores e não pesquisadores, e da construção da chamada ciência aberta. Porém, esse potencial traz novos problemas para as áreas científicas e especialmente para a Ciência da Informação, visto que é necessário pensar em formas satisfatórias de coleta, armazenamento, organização, preservação, disponibilização, disseminação e acesso a esses dados.

Algumas perspectivas encontradas na literatura relacionam ciência aberta aos dados abertos, visto que pressupõem que se trata de uma prática científica em que não se divulgam apenas os resultados das pesquisas em forma de artigos ou teses, mas também os dados coletados, para compartilhamento continuado e reinterpretação por outros pesquisadores, entre outras possibilidades. Porém, ciência aberta pode se referir também a práticas de construção de conhecimento científico com a participação colaborativa de sujeitos ligados ou não a instituições formais de investigação científica.

A crescente quantidade de dados de acesso aberto publicados representa um grande desafio quanto à forma de coleta, armazenamento, organização, representação, acesso e disseminação. Seria inútil apoiar um movimento de ciência aberta sem realizar o gerenciamento adequado desses dados, sem o qual seu potencial de gerar conhecimentos de forma compartilhada não se realizaria. Por isso, uma curadoria digital adequada é desejável para que os dados abertos sejam satisfatoriamente preservados, representados e acessados.

A Curadoria digital - apesar de ser um conceito em construção - pode ser definida como uma abordagem holística de gerenciamento de dados digitais em todo o seu ciclo de vida, o que envolve um planejamento de atribuição de metadados de representação e de preservação desde o momento de sua criação,

passando pelas fases de armazenamento, preservação, avaliação, disponibilização e acesso. A abordagem da curadoria digital, por ser uma abordagem de processos sistematizados e interligados, pode evitar problemas de preservação dos dados digitais ocasionados principalmente pela obsolescência tecnológica. Por interligar as etapas de representação e atribuição de metadados de preservação, a curadoria digital possibilita a previsão de problemas e, conseqüentemente, ações tomadas previamente com o intuito de evitá-los. Além disso, as fases de representação, disponibilização e apresentação da informação também se beneficiam dessa visão integrada dos processos e dos preceitos do *design* da informação, que auxiliam com elementos a serem observados na escolha dos elementos gráficos e de interação na apresentação da informação.

Neste trabalho, abordaremos o *Microscopi*¹, ambiente digital de compartilhamento de dados do projeto *Jugando a Definir la Ciencia*², sob a perspectiva da ciência aberta, a fim de refletir sobre a necessidade de uma abordagem de curadoria digital para o gerenciamento desses dados em seu ciclo de vida e o papel do *design* da informação nas fases de representação, disponibilização e apresentação da informação.

O projeto *Jugando a Definir la Ciencia* é um projeto de construção colaborativa de dicionários com definições elaboradas por crianças para crianças, definições essas verbais e pictóricas de termos básicos da ciência, por meio de um ambiente digital.

Estopà e seus colaboradores, da Universidade Pompeu Fabra, em Barcelona, na fase diagnóstica do projeto *Jugando a Definir la Ciencia*, concluíram que a falta de conhecimento de conceitos científicos básicos nos primeiros anos de escola afetavam negativamente o aprendizado posterior de conteúdos científicos mais complexos. Por isso realizaram, com 500 crianças de escolas catalãs, a primeira fase do projeto, que consistiu na confecção de um dicionário de termos científicos a partir da captação de definições, exemplos de uso e desenhos cunhados por crianças da primeira e segunda série do ensino básico espanhol (entre 06 e 08 anos de idade). De acordo com Estopà, o conhecimento científico e tecnológico é representado e transmitido por meio de palavras que possuem

¹ <http://defciencia.iula.upf.edu/Microscopi>.

² <http://defciencia.iula.upf.edu/ClubLexicografs>.

significado específico, preciso e conciso e acesso ao conhecimento especializado possibilita um uso adequado e preciso da terminologia. Para ela, o processo de especialização em uma área é progressivo e se inicia nos primeiros anos de vida de uma pessoa. Por isso,

Trabalhar a linguagem juntamente com o conhecimento científico desde o início é crucial para qualquer profissão científica futura, uma vez que não há ciência sem linguagem para descrevê-la, referi-la ou comunicá-la. O conhecimento holístico fortalece interdependência entre as palavras (língua) e o conhecimento (ciência). No projeto "*Jugant a definir la ciència*", partimos do pressuposto de que as bases do conhecimento especializado começam a ser adquiridas nos primeiros anos de vida de uma pessoa. Assim, mesmo antes de começar o aprendizado escolar básico, já temos concepções significativas que influenciarão na construção do conhecimento científico. (ESTOPÀ, 2013, p. 199, tradução nossa)

O objetivo do projeto, de acordo com Lemos (2016), foi criar um dicionário apropriado à faixa etária das crianças, para que elas fossem apresentadas de maneira adequada às definições de termos científicos básicos, sem as quais haveria um entrave para o desenvolvimento do seu conhecimento posterior em ciências. Na metodologia colaborativa, utilizada pelo projeto realizado entre os anos de 2009 e 2012, os próprios sujeitos constroem conhecimentos sobre sua realidade, o que os torna mais adequados às suas necessidades informacionais.

Na primeira etapa, as crianças dedicavam uma hora semanal a desenhar e explicar uma palavra. Na etapa seguinte, as fichas produzidas eram analisadas e os fragmentos mais adequados eram selecionados, a partir dos quais era construída uma única explicação atrelada a um desenho. O produto tradicional resultante do *design* do projeto foi um dicionário colaborativo com 100 termos relacionados com a ciência no âmbito do ensino infantil: o *Petit Diccionari de Ciència*. (ESTOPÀ; GAYA, 2013)

O número de crianças envolvidas no projeto (500) foi percebido como uma limitação, de acordo com Estopà e Gaya (2013), uma vez que ele fora realizado manualmente com fichas de papel, além de ter sido feito somente na língua catalã. Somadas ao fato de que as crianças envolvidas no projeto puderam trabalhar com diversas competências, como a categorização, a expressão escrita, compreensão leitora, etc. e de que universitários tiveram acesso a um material que lhes permitiu desenvolver trabalhos de diversos tipos.

A limitação levou Estopà a criar, posteriormente, uma plataforma digital que permitisse a outras escolas realizar um trabalho similar, em diversas outras línguas e com ferramentas úteis de acesso livre a escolas e acadêmicos. Neste segundo projeto, segundo Estopà e Gaya (2013, p. 03, tradução nossa), os objetivos foram “construir uma interface digital para trabalhar o vocabulário acadêmico através de definições e desenhos” com o intuito de “confeccionar dicionários individuais de cada aluno e dicionários colaborativos com todas as informações de meninos e meninas de diversas escolas” e “construir uma base de dados com todo o material proveniente da escola para que os universitários possam explorá-lo, analisá-lo e realizar trabalhos com dados reais”.

Para isso, o projeto foi realizado em três fases: a primeira, a elaboração de uma base de dados para armazenamento da informação com facilitado acesso para análise posterior, que foi chamada de *DicCiencia*; a segunda, uma ferramenta online de acesso aberto, atrativa para crianças, para incentivar sua colaboração e trabalho com o vocabulário especializado, chamado de *Club Léxic*; e a terceira, uma ferramenta online, de acesso aberto, para apresentar ao mundo acadêmico a informação compilada a fim de facilitar o trabalho de análise dos dados, o *Microscopi*. (ESTOPÀ; GAYA, 2013)

A plataforma *Microscopi*, representada abaixo, foi criada para apresentar à comunidade universitária as informações coletadas no Club Léxic, com o intuito de facilitar a elaboração de dicionários colaborativos ou realizar estudos sobre as ideias das crianças. Ela funciona como ferramenta de busca do material existente na base de dados e facilita sua análise. (ESTOPÀ; GAYA, 2013) Os dados coletados, definições verbais, pictóricas e exemplos de uso dos termos científicos, são disponibilizados no *Microscopi* a fim de fornecer material para pesquisadores de diversas áreas do conhecimento.

Figura 1 - Plataforma *Microscopi*



Fonte: <http://defciencia.iula.upf.edu/Microscopi>.

Nossa pesquisa centra-se na análise do ambiente digital em que situa-se o *Microscopi* e caracteriza-se como descritivo-exploratória; encontra-se em andamento em virtude de mestrado em curso. Foram realizados levantamentos bibliográficos em torno dos termos “ciência aberta”, “dados abertos”, “*design* da informação” e “curadoria digital” na base de dados da Biblioteca da UNESP³, no *Google Scholar*⁴ e na Base de Dados de Periódicos em Ciência da Informação (BRAPCI)⁵. Os ambientes digitais do projeto *Jugando a Definir la Ciencia* foram explorados, além da bibliografia publicada pelos realizadores do projeto.

2 Ciência aberta, dados abertos e ciência cidadã

O movimento pela ciência aberta, de acordo com Albagli (2015, p. 14), encontra-se num contexto de conflito entre “novas formas de produção colaborativa, interativa e compartilhada da informação, do conhecimento e da cultura” e “mecanismos de captura e privatização desse conhecimento que é coletiva e

³ http://www.parthenon.biblioteca.unesp.br/primo_library/libweb/action/search.do.

⁴ <https://scholar.google.com.br/>.

⁵ <http://www.brapci.ufpr.br/brapci/>.

socialmente produzido”. Para a autora, sua presença em nível internacional indica a inadequação dos modos dominantes de produção e comunicação científica, que atuam sob mecanismos que impedem, de diversas maneiras, sua livre circulação, colaboração, avanço e difusão.

O conceito de ciência aberta não é definido de forma consensual pelos estudiosos das diversas áreas pelas quais é discutido. O portal do projeto FOSTER (*Facilitate Open Science Training for European Research*)⁶ o define como “o movimento para tornar a pesquisa, os dados e a disseminação científica acessível a todos os níveis de uma sociedade que pesquisa.”⁷ Já Albagli, Clinio e Raychok (2014, p. 435) o consideram um termo “guarda-chuva”, uma vez que engloba diversos significados, práticas e iniciativas sob diferentes perspectivas, pressupostos e implicações que incluem “desde a disponibilização gratuita dos resultados da pesquisa (acesso aberto), até a valorização e a participação direta de não cientistas e não especialistas no fazer ciência, tais como ‘leigos’ e ‘amadores’ (ciência cidadã).”

Para Albagli (2015, p. 15), há duas grandes vertentes que relacionam, de formas diferentes, a ciência aberta ao que se chama ciência cidadã: uma delas, chamada “pragmática”, ou “instrumental”, não pressupõe necessariamente a abertura de dados nem a influência dos voluntários nos desdobramentos da pesquisa e abarca práticas de mobilizações voluntárias de diversos tipos “em esforços de pesquisa, por parte de não cientistas, incluindo desde compartilhamento de recursos computacionais até coleta de informações de relevância científica, no que já se chama de *crowdsourcing science*.”; a outra, chamada “democrática”, abarca “iniciativas orientadas para maior participação, intervenção e empoderamento de cidadãos não só nas formas de produção e uso, mas nos próprios rumos da pesquisa”.

Além da ciência cidadã, outro conceito importante no que se refere à ciência aberta é o de dados abertos (*open data*). Ele pode ser compreendido a partir do conceito de dados e conteúdos abertos, definido pela *Open Knowledge International (OKI)* como aqueles que “podem ser usados, modificados e compartilhados

⁶ <https://www.fosteropenscience.eu/about>. Acesso em: 25 abr 2017.

⁷ “Open science is the movement to make scientific research, data and dissemination accessible to all levels of an inquiring society.” Disponível em: <<https://www.fosteropenscience.eu/foster-taxonomy/open-science>>. Acesso em: 25 abr 2017.

livremente por qualquer pessoa para qualquer propósito”⁸. A OKI⁹ define-se como uma organização global sem fins lucrativos, com foco na compreensão do valor dos dados abertos para a sociedade, ajudando grupos da sociedade civil no acesso e uso dos dados para agir perante problemas sociais.

O movimento pela ciência aberta tem angariado inúmeros apoiadores ao redor do mundo. Pesquisadores e instituições que o advogam apresentam argumentos que indicam vantagens que vão desde um maior potencial de inovação a partir da construção colaborativa de conhecimento até a relação direta existente entre acesso aberto a dados de pesquisa e o desenvolvimento de um país. Para Machado (2015, p. 207), são vários os benefícios da abertura dos dados científicos para a humanidade: o direito de acesso à informação é considerado um direito humano fundamental, e o aumento das iniciativas abertas em ciência leva à emergência de novos paradigmas na produção e distribuição de conhecimento. Além disso, o autor considera o acesso ao conhecimento e à informação “fundamentais para o desenvolvimento humano, social e econômico”, como expresso em documentos como a Declaração dos Direitos Humanos (UDHR, 2014).

Para que o potencial da ciência aberta se realize adequadamente, é necessário pensar formas de gerenciamento e apresentação satisfatória dos dados gerados nestes contextos. Nesta perspectiva cabem reflexões sobre o que se chama curadoria digital.

3 O *Design* da Informação na Curadoria Digital de dados abertos

No contexto digital, emergem novos problemas em relação à coleta, preservação, recuperação, representação, apresentação e acesso à informação, devido à grande quantidade de dados disponíveis, à atribuição inadequada de metadados e, principalmente, à baixa expectativa de vida dos suportes de armazenamento digitais e à obsolescência tecnológica. No paradigma da ciência aberta e dos dados abertos, os questionamentos relacionam-se a quais seriam as formas mais adequadas de armazenar, organizar, preservar e disseminar satisfatoriamente os dados de pesquisa produzidos nessa perspectiva. As

⁸ “Open data and content can be freely used, modified, and shared by anyone for any purpose”. Disponível em: <<http://opendefinition.org/>>. Acesso em: 25 abr 2017.

⁹ <https://okfn.org/about/>

discussões em torno da curadoria digital podem trazer contribuições esclarecedoras e esse debate.

O conceito de curadoria digital também se encontra em construção. Aqui, considera-se sua aplicação no contexto da Ciência da Informação, na qual se observam três abordagens: a Curadoria de Conteúdo, ou de Informações, a Curadoria Digital, e a curadoria de dados de pesquisa. De acordo com Siebra, Borba e Miranda (2016, p. 07):

(...) a curadoria digital para a área de Ciência da Informação vem sendo usada em especial com foco em três abordagens: na filtragem e seleção de dados na Web (Curadoria de Conteúdo ou de Informações), na agregação de valor, gestão ativa e preservação de dados digitais (Curadoria Digital) e na curadoria de dados de pesquisa (e-science).

Grande parte das pesquisas em curadoria digital têm sido desenvolvida por pesquisadores do *Digital Curation Center* (DCC)¹⁰, uma instituição britânica internacionalmente reconhecida como um centro de excelência em curadoria digital, que tem como foco o desenvolvimento de soluções para o gerenciamento de dados de pesquisa.

Para Daisy Abbott, do DCC, curadoria digital se refere à forma de “gerenciamento e preservação dos dados digitais em longo prazo”, em que estão envolvidas todas as atividades de gerenciamento de dados, isto é, “o planejamento de sua produção e boas práticas na digitalização e documentação, assegurando sua disponibilidade e sustentabilidade para descoberta e reuso futuro”, além de incluir o “gerenciamento de vastos conjuntos de dados para uso diário, assegurando, por exemplo, que eles possam ser buscados e continuar legíveis”. (ABBOTT, 2008, p. 01, tradução nossa) Segundo Abbott, a curadoria digital é aplicada a diversos contextos em que a informação necessita ser tratada do início ao fim de seu ciclo de vida, como a digitalização, a criação de metadados e a criação de repositórios, por exemplo.

De acordo com Maureen Pennok, também ligada ao DCC, a curadoria digital é o “gerenciamento ativo e a avaliação da informação digital durante todo seu ciclo de vida” e consiste em “manter e agregar valor a um conjunto confiável de informação digital para uso atual e futuro.” (PENNOK, 2007, p. 01, tradução nossa)

¹⁰ <http://www.dcc.ac.uk/>.

Pennok afirma a importância da abordagem do ciclo de vida da curadoria digital, que é necessária por fatores como: a fragilidade e suscetibilidade dos materiais digitais às mudanças advindas dos avanços tecnológicos durante seu ciclo de vida; o impacto da presença ou ausência de atividades em cada etapa do ciclo de vida na capacidade de gerenciar e preservar materiais digitais em etapas subsequentes; o fato de que só é possível o reuso confiável de materiais digitais se eles forem curados de forma a manter sua autenticidade e integridade.

A curadoria digital, no contexto dos dados de pesquisa, apresenta desafios próprios. De acordo com Sayão e Salles (2012), o trabalho de pesquisa científica é uma das atividades humanas em que mais são gerados e manipulados materiais digitais, o que gera grande fluxo de dados. Gera-se, assim, a necessidade de práticas que possibilitem tornar esses dados mais visíveis e acessíveis, sem prejudicar sua integridade, fidedignidade e característica de evidência. Nesse contexto discute-se a chamada curadoria de dados.

No contexto da ciência aberta, a curadoria de dados, de acordo com Santos (2014), necessita envolver atividades de gestão necessárias para manter dados de pesquisa disponíveis a longo prazo, de forma a favorecer seu reuso, a colaboração entre pesquisadores, o avanço da ciência e a preservação do conhecimento científico. Essas atividades exigem dos profissionais da informação, de acordo com Siebra, Borba e Miranda (2016, p. 13), a criação de estratégias de curadoria desde o momento de recebimento até as etapas de uso e preservação. Por serem os dados de pesquisa fontes primárias, os autores recomendam que “a curadoria de dados digitais de pesquisa deve ser realizada em repositório arquivístico digital confiável, atendendo aos procedimentos arquivísticos em suas diferentes fases e aos requisitos de um repositório digital confiável (...)”

O *Design* da informação (DI) pode ser considerado como uma área do *design* que tem como objetivo “equacionar os aspectos sintáticos, semânticos e pragmáticos que envolvem os sistemas de informação por meio da contextualização, planejamento e produção de interface gráfica de informação para sua audiência”. (PORTUGAL, 2013, n.p.) Os estudos em DI fornecem uma metodologia própria para a construção adequada de ambientes informacionais. De acordo com Jorente (2015, p. 11), o DI

trata de estudar a percepção e cognição humana para definir e criar modelos visando a melhoria dos trânsitos de conteúdos informacionais em diversos meios e contextos; trata, por outro lado, da representação da informação, de suas estruturas e codificação. Ao atuar nas formas de recepção e de produção de informação, cria meios para facilitar o processo de aquisição da informação e do conhecimento, que se efetivam a partir dos sistemas de comunicação, sejam estes analógicos ou digitais.

Para isso, o DI, como metodologia, apresenta importantes aspectos que devem ser levados em conta no momento de criação de espaços informacionais, relacionados à experiência de usuário, *design* de interação e interface gráfica, entre outros. No processo de curadoria digital dos dados de pesquisa, o *Design* da informação tem papel crucial na última fase do ciclo de vida dos objetos digitais, isto é, a representação, disponibilização e apresentação da informação em ambientes informacionais digitais para acesso, interação e reinterpretação. Essa apresentação deve ser adequada de modo que proporcione o trânsito dos conteúdos informacionais em direção aos sujeitos que os acessam, com o intuito de favorecer a aquisição de informação e a construção de conhecimento. Sem isso, o potencial de gerar conhecimento a partir da disponibilização aberta de dados de pesquisa não poderia ser realizado.

4 Dados abertos e curadoria digital no projeto *Jugando a Definir la Ciencia*

Os realizadores do projeto *Jugando a Definir la Ciencia* consideram que as plataformas *Club Léxic* e *Microscopi* cumprem com as expectativas propostas. O primeiro permite o trabalho com vocabulário científico de forma colaborativa e sistemática pelas escolas. Por contar com a colaboração de sujeitos não especializados na construção de conhecimento, pode ser considerada uma iniciativa de ciência cidadã. O segundo possibilita o acesso a dados reais por pesquisadores e a elaboração de dicionários colaborativos, podendo ser considerado uma iniciativa caracterizada pela abertura de dados. Para Estopá e Gaya (2013), o *Microscopi* representa uma rica fonte de conhecimento aberto para a realização tanto de trabalhos de pesquisa quanto para observação ou análise do pensamento infantil, podendo ser útil para pesquisadores de diversas áreas do conhecimento, como educação, psicologia, línguas, linguística, entre outras.

Uma vez que a plataforma *Microscopi* apresenta acesso aberto aos dados de pesquisa coletados durante a execução do projeto, pode-se considerar o projeto *Jugando a Definir la Ciencia* uma iniciativa de ciência aberta e, como tal, seus dados de pesquisa devem contar com uma abordagem de curadoria digital adequada para garantir o acesso, a preservação e a manutenção da integridade de seus dados para acesso atual e futuro. Para que cumpra seu potencial de continuar fomentando o acesso, a interação e a reinterpretação dos dados de pesquisa para a construção colaborativa de conhecimento, o *Microscopi* seria ricamente favorecido pela metodologia do *design* da informação em sua curadoria digital.

5 Considerações

As facilidades de acesso e comunicação, proporcionadas pelo desenvolvimento das TIC, tornaram possível o surgimento de um número cada vez maior de iniciativas de construção de conhecimento colaborativo e com acesso aberto a dados de pesquisa. Essas iniciativas resultaram no movimento pela ciência aberta, que angaria apoiadores pelo mundo todo e que abarca. Considerado um termo guarda-chuva, a ciência cidadã relaciona-se também ao que se chama ciência cidadã e aos dados abertos.

O projeto *Jugando a Definir la Ciencia*, por representar uma iniciativa de construção colaborativa de conhecimento, pode ser considerado uma prática de ciência cidadã, e o fato de compartilhar de forma aberta os dados coletados durante sua realização, constitui-se numa iniciativa de abertura de dados, uma vez que os dados coletados em seu ambiente digital são disponibilizados para acesso, interação e reinterpretação no ambiente *Microscopi*, favorecendo a construção de conhecimento de forma colaborativa.

Trata-se de uma iniciativa bem sucedida que pode inspirar a realização de projetos similares, o que se somaria às iniciativas existentes de ciência aberta. Como tal, seus dados devem contar com uma abordagem adequada de gerenciamento que permitam sua preservação, disseminação e acesso para reuso atual e futuro, isto é, uma abordagem de ciclo de vida de curadoria digital. E, na fase de representação, disponibilização e apresentação da informação na curadoria digital dos dados, é crucial uma abordagem de acordo com os preceitos do *design*

da informação para que seja possibilitada a interação com os conteúdos pelos sujeitos que os acessam.

Considerando este cenário, entendemos que uma abordagem de curadoria digital pode ser de grande contribuição para o gerenciamento dos dados do projeto aqui descrito e de outros com características similares. Especialmente na fase de representação, disponibilização e apresentação da informação nesta curadoria, o *Design* da informação como metodologia de construção de ambientes informacionais digitais fornece os elementos essenciais para realizá-la de maneira eficiente e eficaz. Sugerimos, a partir deste diagnóstico inicial, uma análise mais aprofundada do ambiente a partir de instrumentos automatizados de análise e também de observação participante no ambiente. Os resultados dessas ações deverão guiar um rol sugestões para a apreciação das autoras do projeto. Farão parte também de uma extensão da presente apresentação.

REFERÊNCIAS

ABBOT, D. **What is digital curation?** Digital Curation Centre, 2008. Disponível em: <<https://www.era.lib.ed.ac.uk/bitstream/handle/1842/3362/Abbott%20What%20is%20digital%20curation%20%20Digital%20Curation%20Centre.pdf?sequence=2>>.

Acesso em: 18 maio 2016.

ALBAGLI, S. Ciência aberta em questão. In: ALBAGLI, S.; MACIEL, M. L.; ABDO, A. H. (Org.). **Ciência aberta, questões abertas**. Brasília: Ibict; Rio de Janeiro: Unirio. p. 09–26. Disponível em: <<http://livroaberto.ibict.br/handle/1/1060>>. Acesso em: 25 abr. 2017.

ALBAGLI, S.; CLINIO, A.; RAYCHTOCK, S. Ciência aberta: correntes interpretativas e tipos de ação. **Liinc em Revista**, Rio de Janeiro, v. 10, n. 2, nov. 2014. Disponível em: <<http://revista.ibict.br/liinc/article/view/3593>>. Acesso em: 25 abr. 2017.

ESTOPÀ, R. Juguem a definir la ciència: recursos per treballar el lèxic especialitzat a l'escola. **La terminologia en les ciències de la vida, en la química i en el món educatiu**, n. 4, p. 199–209, 2013.

ESTOPÀ, R.; GAYA, M. C.. El Club Lexic y el Microscopio, plataformas en línea para construir diccionarios científicos colaborativos en un proyecto universidad-escuela. In: IV Congreso Internacional UNIVEST, 2013, Girona. **Anais eletrônicos...** Girona: UdG, 2013. Disponível em: <<http://dugi-doc.udg.edu/handle/10256/8140>>. Acesso em: 20 mar. 2016.

JORENTE, M. J. V. **Tecnologia e Design da Informação: interdisciplinaridades e novas perspectivas para a Ciência da Informação**. Canal 6: São Paulo, 2015.

LEMOS, J. G. **Perspectivas Transdisciplinares de Aproximação com a Ciência Sob o Olhar da Ciência da Informação: Uma Metodologia Bottom-up para a TV Unesp**. 188 f. Tese (Doutorado) - Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília, 2016.

MACHADO, J. Dados abertos e ciência aberta. In: ALBAGLI, S.; MACIEL, M. L.; ABDO, A. H. (Org.). **Ciência aberta, questões abertas**. Brasília: Ibict; Rio de Janeiro: Unirio. p. 201-228. Disponível em: <<http://livroaberto.ibict.br/handle/1/1060>>. Acesso em: 25 abr. 2017.

PENNOCK, M.. Digital curation: A lifecycle approach to managing and preserving usable digital information. **Library and Archives Journal**, n. 1, v.18, 2007. Disponível em: <http://www.ukoln.ac.uk/ukoln/staff/m.pennock/publications/docs/lib-arch_curation.pdf>. Acesso em: 20 maio 2017.

PORTUGAL, C. **Design, Educação e Tecnologia**. Rio Books: Rio de Janeiro, 2013.

SANTOS, T. N. C. **Curadoria digital: o conceito no período de 2000 a 2013**. 2014. 165 f. Dissertação (Mestrado em Ciência da Informação) - Programa de Pós-Graduação em Ciência da Informação, Universidade de Brasília, Brasília, 2014.

SAYÃO, L. F.; SALES, L. F. Curadoria Digital: um novo patamar para preservação de dados digitais de pesquisa. **Inf. & Soc.:Est.**, João Pessoa, v. 22, n. 3, p. 179-191, set./dez. 2012.

SIEBRA, S. A.; BORBA, V. R.; MIRANDA, M. J. K. F. O. Curadoria digital: um termo interdisciplinar. **Encontro Nacional de Pesquisa em Ciência da Informação**, v. 17, 2016. Disponível em:
<<http://www.ufpb.br/evento/lti/ocs/index.php/enancib2016/enancib2016/paper/view/4107>>. Acesso em: 01 maio 2017.

UDHR - **Universal Declaration of Human Rights**. 10 December 1948. Disponível em: <<http://www.un.org/en/documents/udhr/>>. Acesso em: 25 abr. 2017.



***Design Thinking* para a Curadoria Digital de Repositórios de Objetos Digitais de Aprendizagem: uma proposta inicial**

Natália Nakano
Maria José Vicentini Jorente

Resumo: *Repositórios de objetos digitais de aprendizagem são sistemas de informação que permitem o armazenamento para aproveitamento e reutilização desses objetos, que podem consistir em animações, imagens, vídeos, mapas, software, entre outros, que constituam um acervo para subsidiar práticas pedagógicas em diferentes níveis educacionais. Assim, embora os repositórios convencionais sejam de importância essencial para os pesquisadores, docentes ensinando em ambientes digitais de aprendizagem desejam desenvolver materiais que possam ser reutilizados em outros cursos. Nesse contexto, é fundamental que as instituições públicas de ensino superior disponibilizando cursos em ambientes digitais disponibilizem também um sistema de gerenciamento para armazenagem, manutenção, disseminação e reuso dos objetos digitais de aprendizagem produzidos com investimento público, nos princípios do movimento do acesso livre. A proposta deste trabalho é sugerir que o modelo de ciclo de vida para objetos digitais proposto pelo Digital Curation Center seja utilizado para objetos digitais de aprendizagem em repositórios específicos para esses objetos, e que a metodologia do Design Thinking seja utilizado na curadoria dos repositórios de objetos digitais de aprendizagem. A metodologia do DT pode ser utilizada em todas e em cada fase do ciclo de vida da curadoria digital, quando a equipe de curadoria do repositório se deparar com um problema a ser resolvido.*

Palavras-chave: Informação e Tecnologia, Design da Informação, Curadoria Digital, Educação a Distância

Abstract: Digital Learning Object Repositories are information systems that allow the storage for use and reuse of these objects, which may consist of animations, images, videos, maps, software, among others, that constitute a collection to subsidize pedagogical practices at different educational levels. Thus, while conventional repositories are of key importance to researchers, teachers working in digital learning environments want to develop materials that can be reused in other courses. In this context, it is fundamental that public higher education institutions providing courses in digital environments also provide a management system for the storage, maintenance, dissemination and reuse of digital learning objects produced with public investment, in the principles of free access movement. The purpose of this paper is to suggest that the lifecycle model for digital objects proposed by the Digital Curation Center be used for digital learning objects in specific repositories for those objects and that the methodology of Design Thinking be used in the curation of digital learning object repositories. The DT methodology can be used in each and every phase of the digital curator's lifecycle, when the repository's curator team encounters a problem to be solved.

Keywords: Information and Technology, Information Design, Digital Curation, Distance Education

1 Introdução

Devido à sua origem, missão e objetivos, os repositórios institucionais estão, na maioria dos casos, focados em materiais tradicionais como dissertações, teses e trabalhos de conclusão de cursos. No entanto, ao longo dos anos, novos recursos foram se adaptando aos princípios do movimento Acesso Aberto, entre eles os recursos educacionais.

Repositórios de objetos digitais de aprendizagem são sistemas de informação que permitem o armazenamento para aproveitamento e reutilização desses objetos, que consistem em animações, imagens, vídeos, mapas, *software*, entre outros, que constituam um acervo para subsidiar práticas pedagógicas em diferentes níveis educacionais. (AFONSO et.al., 2011)

Assim, embora as teses, dissertações e artigos produzidos nas instituições sejam de essencial importância para o desenvolvimento da ciência no país, de acordo com Cervone (2012), docentes ensinando em ambientes digitais de aprendizagem desejam desenvolver materiais que possam ser reutilizados em outros cursos.

Nesse contexto, é fundamental que as instituições públicas de ensino superior disponibilizando cursos em ambientes digitais disponibilizem também um sistema de gerenciamento para armazenagem, manutenção, disseminação e reuso dos objetos digitais de aprendizagem (ODA) produzidos com investimento público, nos princípios do movimento do acesso aberto.

De acordo com Santos (2014), “o conceito de Acesso Aberto tornou-se um elemento importante na educação, especialmente na Educação a Distância (EaD), onde seus efeitos são especialmente sentidos na implantação de repositórios voltados para objetos de aprendizagem criados para os cursos desta modalidade de ensino”. (p. 2)

Considerando que o formato dos objetos educacionais produzidos atualmente tenderem a utilizar multimídia - como por exemplo áudio, vídeo e animações - a disseminação deste tipo de material em formato impresso se torna impossível, tornando a criação de repositórios de objetos educacionais de aprendizagem imprescindível.

Da mesma maneira que os repositórios institucionais convencionais tão disseminados no meio acadêmico no nosso país demandam saberes, princípios e

aplicações da Ciência da Informação, os repositórios digitais de objetos de aprendizagem contém especificidades cujo gerenciamento também deve ser estudados sob a égide da Ciência da Informação, sob pena de não serem armazenados e recuperados eficientemente e seu reuso, dessa forma, comprometido.

Ainda de acordo com Cervone (2012), os repositórios digitais institucionais não são adequados para armazenamento de objetos digitais de aprendizagem, uma vez que esses objetos têm características diferentes. Primeiramente, o autor menciona a equipe multidisciplinar que deve estar envolvida na ODA, e não apenas bibliotecários. Estão envolvidos no processo de criação de ODA: técnicos em informática, especialistas em ensino, designers instrucionais, além dos docentes da disciplina em questão.

Nesse contexto, a proposta deste trabalho é sugerir que o ciclo de vida para objetos digitais proposto pelo *Digital Curation Center* seja utilizado para objetos digitais de aprendizagem em repositórios específicos para esses objetos, e que a metodologia do *Design Thinking* seja aplicada em suas fases de curadoria.

2 Iniciativas Brasileiras de Repositórios de Objetos Digitais de Aprendizagem

No Brasil, algumas iniciativas de sucesso encorajam a ideia de repositórios abertos de objetos digitais de aprendizagem que possam ser utilizados por docentes em todo o país. Repositórios educacionais podem ser definidos como os ambientes digitais que armazenam e organizam conteúdos digitais de aprendizagem para que possam ser utilizados em sala de aula por meio de articulação ao plano de aula do professor (RODRIGUES et.al., 2012).

O desenvolvimento de recursos educacionais digitais, que permitam a interoperabilidade entre sistemas e a reusabilidade desses recursos é de fundamental importância para a educação (AFONSO, et.al, 2011). Além disso, muitos dos objetos educacionais são custosos para produzir e sua reutilização justificaria também a produção desses materiais.

Em 2008 o MEC, em parceria com universidades brasileiras - entre elas a UNESP e a UnB - e instituições internacionais, criou o Banco Internacional de Objetos Educacionais, um repositório de acesso aberto de objetos digitais de aprendizagem. Afonso et.al. relatam a experiência na construção e definição de um conjunto de ações para o tratamento informacional dos recursos educacionais

digitais no Banco Internacional de Objetos Digitais (BIOE), de acordo com os procedimentos internacionais para catalogação e indexação.

De acordo com Afonso, et.al., “Cabe ao BIOE localizar, catalogar, avaliar, disponibilizar, manter e compartilhar objetos educacionais digitais de acesso livre, disponíveis em diferentes formatos, considerados relevantes e adequados à comunidade educacional brasileira e internacional” (p. 152-153)

O BIOE foi lançado nacionalmente em 2008 e tem como objetivo a disponibilização de objetos digitais de acesso livre, em diferentes idiomas e formatos a fim de estimular o seu uso na educação, fomentando a produção de informação de forma colaborativa. O BIOE utiliza o *DSpace*, que por sua vez, usa o protocolo para coleta de metadados da *Open Access Initiative* (OAI-PMH v2.0). Cabe ressaltar que o *DSpace* permite a utilização de diferentes formatos de materiais digitais, como texto, imagem, vídeo, áudio, *software*, simulações visuais, etc.

De acordo com Afonso et.al., os objetos submetidos são avaliados por um comitê formado por professores e pesquisadores das áreas de ensino abrangidas pelo BIOE. “O Comitê tem a finalidade de garantir a qualidade e a pertinência do recurso, adequando-o ao nível de ensino correto. Após a avaliação positiva do recurso pelo comitê, é autorizada sua publicação. O BIOE não permite, ainda, o auto arquivamento dos recursos, estando a submissão atrelada às universidades participantes.” (p. 154) Na experiência dos autores, a submissão de um objeto digital de aprendizagem no BIOE consistiu em três fases, o preenchimento dos metadados, carregamento do arquivo e conclusão.

Para a catalogação, o BIOE adota um subconjunto de elementos do DC Qualificado, proposta do DCMI6, versão 1.1, adaptada pelas normas: ISO 15836-2003 (fevereiro de 2003) e ISO Standard Z39.85-2007 (maio de 2007)” (p. 155). Para a descrição temática, a linguagem natural foi escolhida, e “para a descrição física dos recursos educacionais, foram utilizados padrões internacionais, como o AACR2 adaptado ao seu contexto e necessidades de descrição.” (p. 156)

Outra iniciativa considerada bem sucedida é a da Universidade Federal de São Carlos. Em 2011, a SEaD/UFSCar implantou o Repositório Digital de Recursos Educacionais Livre Saber (LISa), repositório de acesso livre, com acervo de materiais produzidos para os cursos da modalidade EaD da UFSCar, organizado

para acesso público sob licença *Creative Commons* (SANTOS, MASSETO, OTSUKA, FURNIVAL, 2014).

De acordo com as autoras, é importante que os materiais armazenados em repositórios digitais de aprendizado sejam autossuficientes uma vez que professor e aluno estão separados espacial e temporalmente e o aluno interage solo com o material de EaD. Assim, o objeto digital de aprendizagem deve possibilitar um entendimento imediato do conteúdo e permitir a construção do conhecimento de forma independente.

A SEaD-UFSCar possui uma equipe multidisciplinar de apoio aos docentes no planejamento e na elaboração de suas disciplinas. Os materiais produzidos com o apoio das equipes da SEaD-UFSCar incluem: livros, ilustrações, gráficos, tabelas, animações, videoaulas em diversos formatos, conteúdos interativos, jogos, *podcasts*, entre outros recursos textuais e audiovisuais que ajudam o professor a compor um conjunto de materiais didáticos de boa qualidade para a apresentação do conteúdo de sua disciplina (OTSUKA et.al., 2011 apud SANTOS, et.al., 2014).

Após um ano de cursos oferecidos na modalidade EaD, o conjunto de recursos educacionais era diversificado e em diversos formatos, assim a necessidade de armazená-los e disseminá-los para a comunidade acadêmica ensejou a criação do repositório Livre Saber (LISa) pela SEaD/USCar.

De acordo com Santos, et.al. (2014), o repositório conta com mais de 1000 objetos compostos por vídeos, animações, textos, podcasts, ilustrações, e outros, que podem ser utilizados tanto como apoio quanto independentemente dos processos educacionais, agregando valor a esses processos.

O LISa da UFSCar é parceira da Federação de Repositórios Educa Brasil¹, uma federação de repositórios brasileiros de objetos de aprendizagem. De acordo com as autoras,

um dos processos mais importantes relacionados à disponibilização desses materiais no repositório é a realização de uma catalogação condizente, responsável por prover metadados específicos para os materiais, garantido a correta e eficiente recuperação dos conteúdos, seja na busca direta no repositório, na federação ou através de sites buscadores, como o Google, capazes de encontrar os resultados graças a utilização do protocolo OAI-PMH, que através da coleta dos metadados providos promove a interoperabilidade e melhora a visibilidade dos conteúdos. (p. 4)

¹ Disponível em: <<http://feb.ufrgs.br/>>. Acesso em 16/05/2017.

Assim como o BIOE, o LISa utiliza o *DSpace*, software livre, desenvolvido em 2002 por uma parceria entre o *Massachusetts Institute of Technology* (MIT) e a *Hewlett Packard* (HP).

3 Curadoria Digital para Repositórios de Objetos Digitais de Aprendizagem

Os Objetos Digitais de Aprendizagem (ODA) devem ser desenhados para reuso, ou seja, precisam ser criados de forma que permitam seu uso em contextos múltiplos para que a duplicação e recriação do material seja evitado. Por outro lado, para permitir que esse objeto seja utilizado em múltiplas instâncias, a customização desse material deve ser possível. No entanto, o ODA precisa ser mantido de forma que possa ser claramente relacionado ao seu objeto original, de maneira perpétua.

Outra permissão importante envolve a funcionalidade social do ODA, tal como permitir a comunicação com o internauta/cursante/tutor. Partindo do pressuposto que um repositório é considerado de qualidade dependendo dos objetos que ele armazena, essa característica é essencial para garantir que o repositório seja utilizado. Cervone (2012) exemplifica que o repositório MERLOT² disponibiliza essa funcionalidade possibilitando que um processo de revisão informal seja realizado, permitindo que membros da comunidade MERLOT submetam comentários sobre a sua experiência utilizando os objetos.

Outra diferença entre repositórios institucionais tradicionais e repositórios de ODA é a maneira como são organizados. Além do uso de palavras-chave, tópico/assunto, nível educacional, tipo de formato (pdf, PowerPoint, jpg, etc), para possibilitar o reuso por docentes, outros campos, além dos bibliográficos devem ser incluídos: objetivos de aprendizagem, público-alvo, requisitos de *software* (por exemplo, se o material necessita da instalação do *Flash Player*), data do depósito, porcentagem do conteúdo que é específico para um contexto particular.

Outra questão envolvendo os materiais, diz respeito aos direitos autorais. No Brasil, devido ao rigor da lei de direito autoral, os ODA devem carregar uma licença *Creative Commons* para minimizar questões de reuso e modificações do conteúdo do material.

Com relação ao formato dos materiais, os ODA devem ser desenvolvidos em formatos universalmente aceitos, como HTML5 e *Open Document Format*, o que permite o acesso por diferentes plataformas. De acordo com Cervone (2012),

² Disponível em: <www.merlot.org/merlot/index.htm>. Acesso em 12/05/2017

repositórios tradicionais, tais como *DSpace*, podem ser utilizados para ODA simples. No entanto, repositórios tradicionais não fornecem os níveis de suporte para padrões de objetos de aprendizagem, tais como os estabelecidos pelo IMS, padrões de metadados IMS e ADL que são tipicamente requeridos.

Porque os ODA têm características únicas, e demandam saberes específicos para o gerenciamento como um todo, impossível não pensar na Curadoria Digital para abordar o tema.

A Curadoria Digital para objetos de aprendizagem consistiria em transferir as abordagens curatoriais de dados proposta pelo *Digital Curation Center* para os objetos de aprendizagem.

Considera-se que Curadoria Digital compreende não apenas a preservação de recursos digitais ou a manutenção de uma coleção acessível, mas que tenha certo valor e conhecimento agregado, evitando e prevenindo obsolência e duplicação na criação de objetos, e assim motivando reuso dos recursos digitais.

McGreal (2004) definiu objetos de aprendizagem como qualquer recurso digital reutilizável que esteja encapsulado em uma lição ou conjunto de lições agrupados em unidades, módulos, cursos e até mesmo programas. O IEEE (Institute of Electrical and Electronic Engineers) definiu Objeto de Aprendizagem como “qualquer entidade, digital ou não, que pode ser usada, reutilizada ou referenciada durante o aprendizado apoiado pela tecnologia” (Braga, Menezes, 2015, p. 13). Nota-se que as definições enfatizam os usos educacionais evidenciando as características de reuso.

O modelo de ciclo de vida proposto pelo DCC prevê ações de curadoria digital genéricas que podem ser aplicadas em diferentes contextos, inclusive o de repositórios de objetos digitais de aprendizagem.

De acordo com Siebra et. al.(2013, p.2), a Curadoria Digital emerge "como uma área de pesquisa e prática interdisciplinar que reflete uma abordagem holística para o gerenciamento do objeto digital e inclui atividades que abrangem todo o ciclo de vida desse objeto".

Para o Digital Curation Center (DCC)³, a curadoria digital é um processo complexo que envolve a manutenção, preservação, avaliação, re-avaliação, uso e re-uso bem como agregação de valores aos dados digitais de pesquisas científicas

³ Disponível em: <<http://www.dcc.ac.uk/>>. Acesso em 21/05/2017.

por todo seu ciclo de vida, uma vez que o gerenciamento ativo de dados de pesquisa reduz as ameaças a longo prazo ao valor da pesquisa e diminui o risco de obsolescência digital. Enquanto isso, o dado curado em repositórios digitais confiáveis pode ser compartilhada entre a ampla comunidade de pesquisa. "Assim como reduzir os esforços de duplicação na criação de dados de pesquisa, a curadoria aumenta o valor a longo prazo de dados existentes, tornando-os disponíveis para posterior pesquisa de alta qualidade."

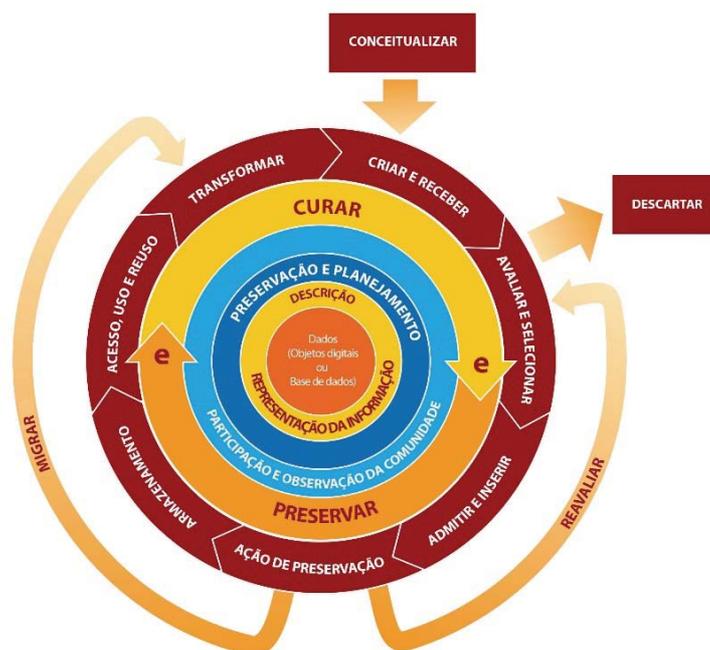
Trazendo a definição para o contexto deste projeto, pode-se aplicar os mesmos princípios da Curadoria Digital de objetos digitais para os objetos digitais de aprendizagem produzidos pela equipe interdisciplinar que inclui designers da informação e produtores de conteúdo de aprendizagem, por exemplo.

As ações são previstas, mas não se esgotam, na curadoria de objetos digitais de aprendizagem e incluem: 1) Conceituar: conceber e planejar a criação de objetos digitais, incluindo métodos de capturar dados e opções de armazenamento, 2) Criar: produzir objetos digitais e designar metadados arquivísticos administrativos, descritivos, estruturais e técnicos, 3) Acessar e usar: assegurar que usuários possam facilmente acessar objetos digitais continuamente. (Alguns objetos digitais de aprendizagem podem estar disponíveis publicamente, enquanto outros podem estar protegidos por login); 4) Avaliar e selecionar: avaliar objetos digitais e selecionar aqueles que requerem preservação e curadoria a longo prazo; 5) Descartar: livrar os sistemas de objetos digitais não selecionados para curadoria e preservação a longo prazo. Implica nas orientações documentadas, políticas e requerimentos legais, que podem requerer a destruição segura desses objetos; 6) Ingerir: transferir objetos digitais para um arquivo, repositório digital confiável, centro de dados ou similar, novamente aderindo a orientação documentada, políticas ou requerimentos legais; 7) Preservar: realizar ações para assegurar a preservação e retenção a longo prazo da natureza de autoridade de objetos digitais; 8) Re-avaliar: retornar objetos digitais que falhem aos procedimentos de validação para nova avaliação e re-seleção; 9) Armazenar: manter os dados de uma maneira segura como delineados por padrões relevantes; 10) Acessar e re-utilizar: assegurar que dados sejam acessíveis para cursantes os usem uma primeira vez e possam re-utilizar (Alguns materiais podem estar publicamente disponíveis, enquanto outros podem ser protegidos por login); 11) Transformar: criar novos objetos digitais de

aprendizagem a partir do original, por exemplo, pela migração em um formato diferente.

O ciclo de vida da Curadoria Digital está ilustrado por Higgins (2008)⁴ e adotado pelo DCC. A Figura 1 apresenta os processos e agentes do ciclo de vida da Curadoria Digital.

Figura 1 - Ciclo de vida da curadoria digital proposto por Higgins (2008)



Fonte: Higgins, 2008. Traduzido pelas autoras.

O modelo apresentado sistematiza as ações de curadoria que podem ou não ser aplicadas em sua totalidade na curadoria dos objetos digitais de aprendizagem, no entanto, depende-se do modelo que uma equipe multidisciplinar deve se envolver no processo de curadoria.

A partir do esquema de Higgins, é possível identificar duas equipes multidisciplinares trabalhando em uma mesma curadoria digital nesse processo de ciclo de vida: uma equipe centrada no núcleo da curadoria, qual seja, para descrição e representação da informação, definição dos dados e metadados nos repositórios digitais, preocupando-se com a preservação e o planejamento da informação. A outra equipe, focada nas etapas de preservação (admitir, preservar, armazenar) e

⁴ Disponível em: <<http://www.dcc.ac.uk/sites/default/files/documents/publications/DCCLifecycle.pdf>>. Acesso em 20/04/2016.

curadoria (acessar, usar, re-usar, transformar, criar ou receber, avaliar e selecionar). Nota-se que a comunidade e o planejamento de preservação estão incluídos no núcleo, devem ter papel essencial no ciclo de vida da curadoria digital.

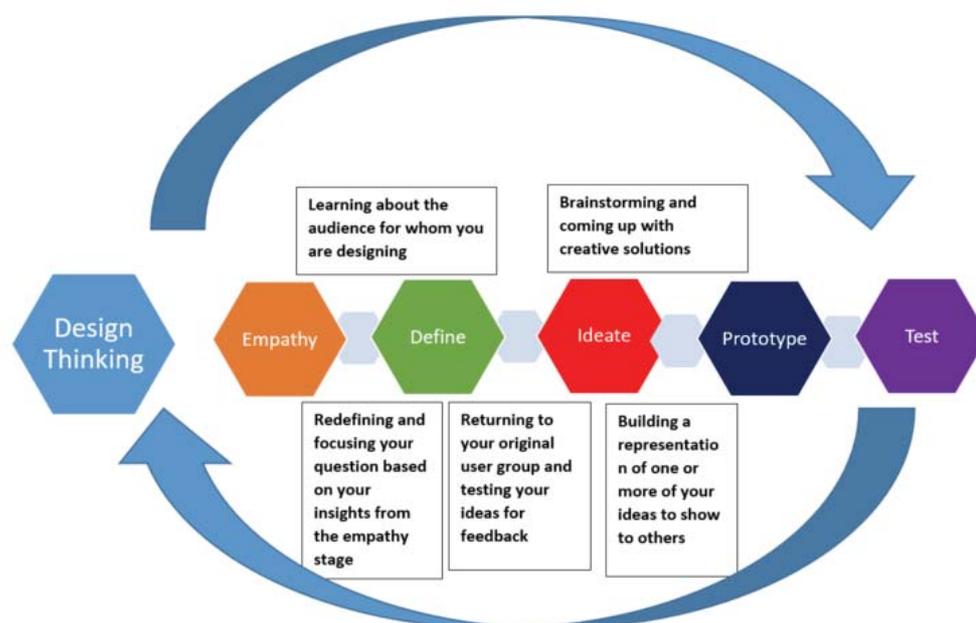
4 O *Design Thinking* para a Curadoria Digital de Repositórios de Objetos Digitais de Aprendizagem

De acordo com Tim Brown (2009), o *Design Thinking* (DT) oferece uma abordagem efetiva e amplamente acessível para alcançar soluções para negócios e para a sociedade; trata-se uma metodologia que permite ser integrada a diferentes questões e que motiva as pessoas a criarem e gerarem ideias possíveis de serem realmente implementadas. Portanto, a metodologia do DT deve colocar as ferramentas nas mãos das pessoas que nunca antes se imaginaram como designers e poder aplicar essas ferramentas.

A ciência do *Design* não é apenas centrado no humano, ela é o humano, conta com nossa habilidade de sermos intuitivos, nossa habilidade natural de reconhecer padrões, construir ideias que tenham significado emocional, de nos expressarmos. Obviamente contar apenas com o lado emocional não é o ideal, mas considerar apenas o lado racional e analítico dos problemas pode ser ineficiente. A abordagem integrada da metodologia do DT propõe uma terceira maneira de abordar a solução dos problemas de curadoria digital de repositórios de objetos digitais de aprendizagem: balancear as emoções e o racional.

A metodologia do DT pode ser utilizada em todas e em cada fase do ciclo de vida da curadoria digital, quando a equipe de curadoria do repositório se deparar com um problema a ser resolvido. Erroneamente, o Design está associado apenas com as questões estéticas e de comunicação visual de uma interface digital. No entanto, quando se compreende que o Design é uma mentalidade (*thinking*), a metodologia sistematizada proposta nos projetos de Design contribuem além da comunicação eficiente de conteúdo informacional. A Figura 1 descreve as fases do DT.

Fig 1 - Fases do *Design Thinking*



Fonte: https://commons.wikimedia.org/wiki/File:Design_thinking.png#filelinks

A primeira fase do DT é a fase de empatia. A equipe multidisciplinar trabalhando no projeto de curadoria digital deve imergir no problema para poder criar empatia com os problemas a serem resolvidos na curadoria dos objetos digitais de aprendizagem. Assim, por exemplo, a equipe deve se engajar para conhecer o conteúdo informacional do objeto, o público alvo para quem esse objeto está sendo disponibilizado, como essas pessoas vão entrar em contato com o objetos, em que situações, etc.

Como exemplo, podemos imaginar a metodologia do DT na fase de ingestão de determinado objeto. Decidir se ele será preservado a longo prazo ou não, considerando o volume crescente de objetos produzidos consiste no problema a ser resolvido. A metodologia poderia ser sistematizada para a solução do problema utilizando as fases de empatia, definição, ideação, protótipo e teste, exemplificadas nas seguintes ações: considerar, estudar, analisar as pessoas que utilizariam o objeto, como elas o utilizariam e a finalidade educacional do objeto (empatia, definição e ideação). Essas ações são essenciais para decidir se o objeto será preservado ou não, pois dependendo dos resultados obtidos ao final do processo de DT, a equipe terá conhecimento de como a interação está ocorrendo, tanto sob a perspectiva do professor quanto do aluno. A equipe de curadoria do repositório

então deve se reunir e pensar em como resolver o problema de ingestão. A última ação a ser tomada poderia consistir na realização de testes com a observação do comportamento e emoções das pessoas no momento da interação com o objeto, compreendendo as fases de protótipo e testes da metodologia do DT.

O uso do DT para aprimoramento de serviços e programas em centros de informação já é uma realidade. O uso do DT para bibliotecas é relatado tanto na literatura internacional quanto na nacional.

A IDEO, renomada empresa de consultoria em design e inovação criou, em parceria com a Fundação Bill & Melinda Gates e bibliotecas nos EUA e Dinamarca, um *toolkit* intitulado *Design Thinking for Libraries* com o intuito de orientar os bibliotecários e motivar o uso da metodologia do *Design Thinking* em bibliotecas para a resolução de problemas e criação de inovação. Observa-se a ênfase do *toolkit* em explicitar que a metodologia pode ser aplicada em diversas situações, incluindo os programas, serviços, espaços e sistemas das instituições, pois quando a metodologia do DT é entendida e incorporado pelas pessoas, e seu valor integrado nas ações, ela se torna mais do que uma abordagem, ela se transforma em uma mentalidade.

De acordo com Oliveira (p. 39), sobre o *toolkit*:

A ideia do livro é servir como uma caixa de ferramentas envolvendo as etapas do processo de *design thinking*, um conjunto de técnicas do *Design* estendidas para diferentes áreas do conhecimento, centrada nas necessidades das pessoas. Parte-se de uma inspiração para um problema (*inspiration*), criação e aprendizado ao gerar ideias e torná-las tangíveis (*ideation*) e finalmente prototipá-las e testá-las com os usuários (*iteration*). [...] O objetivo não é criar um produto final rapidamente, mas sim utilizar o protótipo para fomentar respostas que contribuem para resolver um problema.

Além disso, Oliveira ainda menciona a experiência da comunidade *Blended Librarian*, uma criação inovadora de Bell e Shank (2007), resultado da aplicação do *Design Thinking*. Os autores relataram que a comunidade *online* cresceu de cerca de 50 pessoas em 2004 para cerca de 5.000 em 2011.

No Brasil, Ramírez e Zaninelli (2017) apresentaram casos de sucesso da abordagem do DT e sua importância para a Ciência da Informação, bem como as vantagens que a abordagem representa não apenas para o contexto de negócios, mas também para o acadêmico. O *Design Thinking* como ferramenta para inovação em uma biblioteca também foi relatado por Juliani, Cavaglieri e Machado em 2015.

O objetivo da utilização do DT na biblioteca da Universidade do Estado de Santa Catarina foi gerar serviços inovadores na biblioteca. O estudo concluiu que a abordagem possibilitou a disponibilização de informações de qualidade e promoveu acessibilidade informacional da biblioteca universitária.

5 Considerações Parciais

É consenso na literatura da área que os recursos educacionais digitais são mais eficientemente aproveitados quando organizados, catalogados e armazenados em um repositório (TAROUCO, FABRE e TAMUSIUNAS, 2007; AFONSO, et.al, 2011). Os objetos digitais de aprendizagem são criados por docentes com o intuito de utilização imediata e posterior reutilização, assim, um repositório para coletar, cadastrar, indexar, armazenar para posterior recuperação é o ideal para assegurar posterior recuperação e utilização.

Embora os ODA estejam sendo estudados e pesquisados em contexto tecnológico-educacional, é essencial que a Ciência da Informação, como área interdisciplinar, se preocupe com os Repositórios de Objetos de Aprendizagem, uma vez que as preocupações e objetivos desses repositórios, como a armazenagem, indexação, recuperação e interoperabilidade, por exemplo, são comumente pesquisadas na CI e o conhecimento produzido por essa ciência, eficientemente aproveitado.

Uma vez que os repositórios não precisam armazenar os arquivos dos objetos, uma função essencial desses repositórios é que eles identifiquem a localização dos objetos e formação um sistema de indexação que permita uma busca eficiente e recuperação dos objetos. Com o crescente número de repositórios e iniciativas no Brasil, ações específicas para manter esses objetos ao longo do tempo e torná-los disponível para assegurar seu uso e reuso por gerações futuras devem ser tomadas, tornando, assim, importante que os estudos em Curadoria Digital e a aplicação do *Design Thinking* colaborem com essas ações.

Referências

AFONSO, M.C.L., EIRÃO, T.G., MELO, J.H.M., ASSUNÇÃO, J.S., LEITE, S.V. Banco Internacional de Objetos Educacionais (BIOE): tratamento da informação em um repositório educacional digital. *Perspectivas em Ciência da Informação*, v.16, n.3, p. 148-158, 2011.

BRAGA, Juliana (Org.). *Objetos de Aprendizagem Volume 1: introdução e fundamentos*. Santo André: UFABC, 2015. 157 p. Disponível em: <pesquisa.ufabc.edu.br/intera/?page_id=370>.

BRAGA, J., MENEZES, L. *Introdução aos Objetos de Aprendizagem. Objetos de Aprendizagem Volume 1: introdução e fundamentos*. Santo André: UFABC, 2015. 157 p. Disponível em: <pesquisa.ufabc.edu.br/intera/?page_id=370>.

CERVONE, H. F., (2012) "Digital learning object repositories", *OCLC Systems & Services: International digital library perspectives*, Vol. 28 Issue: 1, pp.14-16, doi: 10.1108/10650751211197031 Permanent link to this document: <http://dx.doi.org/10.1108/10650751211197031>

DESIGN THINKING FOR LIBRARIES: a toolkit for patron-centered design. **IDEO**, Palo Alto, 2015. Disponível em: <<http://www.designthinkingforlibraries.com/>>. Acesso em: 21 mai. 2017.

HIGGINS, S. The DCC Curation Lifecycle Model. *The International Journal of Digital Curation*, Edinburgh, v.3,. n.1, 2008. Disponível em: <doi10.2218/ijdc.v3i1.48>.

JULIANI, J. P.; CAVAGLIERI, M.; MACHADO, R. B. Design thinking como ferramenta para geração de inovação: um estudo de caso da Biblioteca Universitária da UDESC. **INCID - Revista de Ciência da Informação e Documentação**, São Paulo, v.6, n. 2, 2015. Disponível em: <<http://www.revistas.usp.br/incid/article/view/100887>>. Acesso em: 21 mai. 2017

KURAMOTO, H. Acesso Aberto. Para entender a ciência da informação / Lúcia Maria Batista Brandão Toutain: org. Salvador: EDUFBA, 2007. 242 p.

MCGREAL, R.:. Learning objects: A practical definition. Disponível em: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.118.9253&rep=rep1&type=pdf#page=25> (2004)

OLIVEIRA, J. A. D. B. **A Ciência da Informação e o Design da Informação: perspectivas interdisciplinares**. 2015. Dissertação (Mestrado em Ciência da Informação) - Faculdade de Filosofia e Ciências de Marília, Universidade Estadual Paulista, Marília, 2015. Disponível em: <https://www.marilia.unesp.br/Home/Pos-Graduacao/CienciadaInformacao/Dissertacoes/oliveira_jadbe_me_mar.pdf>. Acesso em: 21 mai. 2017

PALAVITSINIS, N., MANOUSELIS, N., SANCHES-ALONSO, S. Preliminary Discussion on a Digital Curation Framework for Learning Repositories. Disponível em: <<http://ceur-ws.org/Vol-681/paper06.pdf>>. Acesso em: 21 maio 2017.

RAMÍREZ, D. M. B.; ZANINELLI, T. B. O uso do design thinking como ferramenta no processo de inovação em bibliotecas. **Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação**, v. 22, n. 49, p. 59-74, maio/ago., 2017. Disponível em: <<http://dx.doi.org/10.5007/1518-2924.2017v22n49p59>>. Acesso em: 21 mai. 2017.

SANTOS, E.G.C dos, MASSETO, D.C., OTSUKA, J.L., FURNIVAL, A.C.M., Acesso Aberto na Educação a Distância: a experiência da SEaD-UFSCar. Anais... Simpósio Internacional de Educação a Distância. Universidade Federal de São Carlos, 2014.

TAROUCO, L. M. R.; FABRE, M.-C. J. M.; TAMUSIUNAS, F. R. Reusabilidade de objetos educacionais. 2003. Disponível em: <http://www.cinted.ufrgs.br/renote/fev2003/artigos/marie_reusabilidade.pdf>



PRINCÍPIOS E NORMAS QUE FUNDAMENTAM OS DADOS ABERTOS: um estudo de caso do software ICA-AtOM

Nandia Letícia Freitas Rodrigues

Lucineia da Silva Batista

Maria José Vicentini Jorente

RESUMO: Na contemporaneidade, com o aumento exponencial dos fluxos informacionais e a complexidade que envolve os processos de produção, organização, preservação e disseminação da informação em ambientes digitais, muitos *softwares* têm sido desenvolvidos e aprimorados no intuito de promover melhor assimilação, uso e reuso das informações pelos indivíduos em sociedade. O presente artigo indaga-se se o AtoM atende aos princípios de dados abertos e busca por meio da Ciência da Informação (CI) analisar o *software* para apresentar as possibilidades da sua utilização através da representação da informação no ambiente *Web*. Neste sentido, a metodologia aplicada concentra-se em estudo teórico exploratório da literatura da área e converge a disciplina *Design da Informação (DI)* para analisar o design do AtoM em seus vários níveis de descrição, considerando que o software atende a grande parte das necessidades para a abertura de dados nas instituições de custódia de informação. Recomenda-se, na finalização estudos mais aprofundados sobre o tema.

Palavras-chave: Dados Abertos; ICA-AtOM, Design da Informação; Informação e Tecnologia.

ABSTRACT:

In contemporary times, with the exponential increase of information flows and the complexity that surrounds the processes of production, organization, preservation and dissemination of information in digital environments, many softwares have been developed and improved in order to promote better assimilation, use and reuse of information Individuals in society. The present paper investigates whether the AtoM complies with the principles of open data and searches through the Information Science (CI) to analyze the software to present the possibilities of its use through the representation of information in the Web environment. In this sense, the methodology Applied focuses on an exploratory theoretical study of the literature of the area and converges the discipline of Information Design (DI) to analyze the design of the AtoM in its various levels of description, considering that the software meets the great part of the needs for the opening of data Institutions of custody of information. Further study on the subject is recommended at the finalization.

Keywords: *Open Data; ICA-AtOM, Information Design; Information Technology.*

1 INTRODUÇÃO

Na contemporaneidade o acesso aos dados impacta diretamente na vida das pessoas e conseqüentemente o convívio social na cultura contemporânea. Disponibilizar dados abertos à sociedade, entre muitas outras possibilidades, auxilia o desenvolvimento social, a resolução de problemas cotidianos e principalmente na produção de novas informações e conhecimentos.

Segundo a BSA (2015), os dados abertos e interoperáveis tem ocasionado mudanças e progressos significativos em diversos âmbitos na sociedade, viabilizando, por exemplo: previsões de crises climáticas, ocorrência de tsunamis e outros catástrofes naturais com maior antecedência; redução de tempos de trajeto nas rodovias contribuindo para a diminuição dos congestionamentos; maior produtividade agrícola, uma vez que acessando dados sobre sementes, equipamentos e condições climáticas o agricultor pode melhorar as formas de cultivo; planejamento de edifícios eficientes no uso da energia, utilizando ferramentas de dados é possível otimizar a arquitetura e assim projetar ambientes que utilizem o máximo de luz solar, por exemplo; melhorias no setor de aviação com a análise e planejamento dos vôos; construção de cidades inteligentes; e uma infinidade de outras possibilidades de resoluções de problemas a partir da disponibilização e uso do acesso dos dados abertos nos ambientes da *Web*.

Contudo, se o compartilhamento de dados e informações produzidos diariamente pela sociedade e disponibilizado na *Web*, possibilita acesso à informação para um grupo mais amplo de indivíduos, também apresenta problemas de desinformação, gerados por um excesso de informações e dados com problemas de organização ou de gestão em sua apresentação em ambientes digitais da *Web*.

Neste cenário paradigmático de abundância informacional, como salienta Davenport et al. (2012), trabalhar com grandes quantidades de dados, tratando, filtrando, e disponibilizando o acesso ao público é um dos maiores anseios dos últimos anos, embora a produção extraordinária de dados seja apenas um dos grandes desafios atuais da ciência moderna.

Em resposta, avanços técnicos, científicos e tecnológicos no contexto das Tecnologias de Informação e Comunicação (TIC) têm contribuído significativamente no processo de compreensão dos dados e sua transformação em respostas reais e práticas para as questões do dia a dia. Surgem, para responder aos problemas

gerados, estratégias de Curadoria em ambientes digitais que buscam responder também à uma legislação para implementação das Políticas Públicas de Informação no mundo globalizado.

Destaca-se que, no Brasil, a Lei nº 12.527 de 2011, conhecida como Lei de Acesso à Informação (LAI), fez crescer as iniciativas para o acesso aberto da informação pública aos cidadãos, bem como a criação e investimento em políticas de acesso à informação. Para que a LAI, que rege a abertura dos dados de caráter público e a transparência da gestão pública e suas equivalências internacionais sejam cumpridas, contudo, fica evidente que os dados produzidos e disponibilizados no ambiente digital necessitam ser organizados, classificados, categorizados, estruturados e descritos sintática e semanticamente, para que sejam facilmente acessíveis pelas pessoas. O Brasil, além da criação e aprovação da LAI, que no artigo 8º reconhece a necessidade de dados governamentais disponíveis em formato aberto, também destacou-se por fundar, em 2011, a *Open Government Partnership*, do qual participam vários países na construção de um portal para disponibilizar dados governamentais, de acordo os princípios de dados abertos, sendo esses parte da política de acesso à informação do governo federal (INDA), de 2012.

O acesso à informação tornou-se, dessa maneira, pela primeira vez no Brasil, um direito constitucional, fundamental e assegurado pela LAI, que regulamenta “[...] o direito, previsto na Constituição, de qualquer pessoa solicitar e receber dos órgãos e entidades públicos, de todos os entes e Poderes, informações públicas por eles produzidas ou custodiadas” (ASPECTO GERAIS DA LEI, [s. d]). No Estado democrático de direitos, após sancionada a LAI, é legitimado como direito fundamental de que todo e qualquer cidadão o contato com os documentos e as informações públicas. Por conseguinte, o acesso não mais se restringe a somente uma parcela minoritária de indivíduos, isto é, toda gestão e informação de caráter público passa a ser acessível e transparente à sociedade em geral.

Assim, como reflexo desse cenário, nas ciências em geral e na CI, em particular, tem surgido inúmeras pesquisas com enfoques direcionados a criar mecanismos de Curadoria a partir da atribuição de metadados estruturais e semânticos que visem descrever, organizar, salvaguardar, preservar e garantir o acesso e uso dessa crescente massa de dados. A partir delas visualizaram-se

nesse sentido, sistemas digitais abertos, *softwares* apropriados para a curadoria, que vêm sendo internacionalmente desenvolvidos e aprimorados para facilitar todo o trabalho de organização e representação de dados na *Web*. São sistemas que convergem e interoperam, como o AtoM e o Archivemática, que se destacam na área dos Arquivos em suas três idades, por possibilitar a Curadoria de dados e o acesso à dados e informação via *Web*.

Objeto desse artigo, o AtoM (convergido ao Archivemática) possibilita o acesso à informação aos internautas, a partir de descrição em níveis, representação e apresentação da informação via *Web* - o Archivemática atua sobre os dados digitais para a sua preservação, com autenticidade, confiabilidade que garantem o acesso por longo tempo. Por conseguinte, a problemática do artigo se constrói a partir da indagação: o AtoM atende os princípios de dados abertos? Em resposta, exploramos as possibilidades do *software* no cumprimento dos princípios de abertura de dados. Esse sistema auxilia na organização da informação e interopera com o Archivemática no armazenamento de dados para a disponibilização de informação no ambiente *Web*, tendo em vistas as convergências de normas, interoperabilidade, descrições multiníveis, multilíngue, *Design* simples e a promoção de acesso e difusão de dados.

2 O COMPARTILHAMENTO DA INFORMAÇÃO NA WEB E O ACESSO ABERTO

Segundo Hendges e Sliwka (2013) o compartilhamento da informação em ambiente *Web* e, conseqüentemente, a possibilidade do acesso aberto e do uso dos dados públicos pela sociedade sem limitações legais, contribui para a consolidação da democracia e da transparência pública. No entanto, os dados necessitam ser organizados, representados e apresentados de forma eficiente e eficaz na *Web*.

No contexto arquivístico, alguns desses procedimentos técnicos são característicos do paradigma tecnicista, patrimonialista, custodial e moderno. Esse modelo envolve questões práticas e técnicas da Ciência da Informação (CI) voltadas à organização da informação, pois "a prioridade estava na guarda do patrimônio cultural incorporado e acumulado e não no acesso ou na difusão plena" (RIBEIRO, 2010, p. 65), o que significa uma predominância da guarda e preservação da informação, enquanto o acesso estava relegado ao segundo plano. Segundo Ribeiro (2002, p. 2):

O arquivista passa a ser, sobretudo, um técnico a quem compete fazer estudos de avaliação para legitimar a selecção e eliminação de documentos ou um especialista em descrever documentação para a tornar acessível nos arquivos históricos, através dos denominada "instrumentos de pesquisa", cada vez mais sujeitos a normas internacionalmente aceites.

Na visão de Cook (1993), Silva (2002) e Ribeiro (2005), entretanto, esse modelo entra em colapso com denominada revolução tecnológica, em que se questiona o objeto de estudo da CI, o documento. Contudo, Ribeiro (2002) salienta que o paradigma custodial - também entendido como um "modelo francês", por ser oriundo da Revolução Francesa que propiciou a abertura de bibliotecas e arquivos públicos - ainda é predominante na maioria das situações.

Esses autores defendem, contudo, que o modelo custodial deve ser, e já foi ampliado, abordando o acesso à informação que ficava em segundo plano no modelo anterior. Nesse sentido, não houve um rompimento de paradigmas, e sim a continuidade do pós-custodial em relação ao precedente. Se para possibilitar o acesso contínuo à informação, é necessária a sua preservação a longo prazo, a preservação fundamenta-se pelas necessidades e direitos ao acesso, que dela depende para garantir a sua disponibilidade ao longo do tempo na *Web*.

Por outro lado, o paradigma denominado pós-custodial, pós-moderno ou de acesso, conflui para o consenso, de que o objeto de estudo da CI seja a informação e não somente o documento. As unidades de informação - termo usado para definir instituições que lidam com informação, como Bibliotecas, Arquivos, Museus e Centros de Documentação - ao se depararem com o paradigma pós-custodial, imposto pela realidade digital, encontram grandes desafios, pois segundo Canclini (1998), a modernidade não ocorreu como um todo e não alcançou todos os níveis da sociedade e cultura global¹.

Em relação aos Arquivos Públicos, no Brasil, por exemplo, ao considerar esse período no contexto da CI, percebe-se que, na prática, a grande maioria desses Arquivos se mantiveram em desarranjo organizacional, anterior à digitalidade e que, essa realidade acentua-se quando se fala de ambiente *Web 2.0*.

¹ Em muitos casos isso significa entrar na pós-modernidade, ou até mesmo na denominada Sociedade da Informação, sem antes ter passado pela modernidade, o que se constitui em problemática não tratada neste artigo e deve ser tratado em outra ocasião, complementarmente.

Um diagnóstico realizado por pesquisadores do LADRI² (Laboratório de *Design* e Recuperação da Informação) a partir de um projeto PIBIC (CNPq) que tinha por objetivo pesquisar a realidade dos Arquivos Permanentes de municípios brasileiros com mais de 100.00 habitantes na *Web* - portanto sujeitos à determinação de abertura pela LAI - demonstrou-se que, de 180 municípios, apenas 23 contavam com algum tipo de *Website*. Em relação aos conteúdos disponibilizados, por outro lado, nesses ambientes disponíveis, existiam inúmeras carências - quanto à representação da informação e aos instrumentos de busca e pesquisa - que limitavam o acesso mesmo ao inventário do acervo da instituição.

Em um contexto mais amplo, segundo o estudo global recentemente produzido pela *Business Software Alliance* (BSA, 2015) intitulado "*What's the Big Deal With Data?*" constatou-se que aproximadamente 2,5 quintilhões de *bytes* de dados são criados todos os dias no mundo. Porém, grande parte desses dados não está disponível, e tampouco estruturada para o acesso ao público. Os autores Isotani e Bittencourt (2015, p. 17) salientam que a disponibilização "precária" e a falta de tratamento dos dados fazem com que "a extração de informações e a produção de conhecimentos que poderiam ser úteis para a sociedade não acontecem com a agilidade e a eficácia necessárias".

Apresenta-se de maneira notável e geral uma carência no atendimento aos requisitos de complexidades operacionais, de recuperação, integração, disseminação, intercâmbio e interoperabilidade entre sistemas, princípios estes fundamentais para a construção de um sistema informacional que contemple com excelência às demandas informacionais dos indivíduos que buscam informações nestes ambientes.

Em resposta à essas necessidades, seguindo a perspectiva de ampliação do acesso à informação, ambientes e *softwares* têm sido desenvolvidos e aprimorados no intuito de otimizar a disponibilização de dados a partir da sua estruturação para melhor representação da informação no ambiente digital. O estudo destes *softwares* e ambientes torna-se, portanto, uma importante tarefa dos profissionais da informação e da CI em particular, no sentido de contribuir para uma fruição da sociedade e cultura contemporâneas.

² Trata-se da pesquisa realizada, com fomento do CNPq, pelas bolsistas: Anahí, Solene e Talita. Pesquisa intitulada: "A disponibilização da Informação em *Websites* de Arquivos Públicos Permanentes no Brasil.

3 DEFINIÇÕES, NORMAS E PRINCÍPIOS PARA A CONSOLIDAÇÃO DO ACESSO ABERTO POR MEIO DO ATOM

O movimento de abertura de dados vem ganhando força nos últimos anos. Segundo Machado (2015, p. 201) "Há 23 anos, surgia o *arXiv.org*, que hoje reúne quase 1 milhão de trabalhos, principalmente das áreas de física, matemática, ciência da computação, estatística. [...]", com acesso gratuito e livre à informação.

A expressão **acesso aberto** surgiu em documento de 2002, denominado Iniciativa do Acesso Aberto na *Declaração de Budapeste*. Contudo, a definição de dados abertos ocorreu em 2007, a partir de oito princípios que vieram a "[...] dar mais força ao processo de abertura de dados e à ampliação dos usos e reusos da informação [...]" (MACHADO, 2015, p. 201).

De acordo com a Open Definition (2014), Dados Abertos são

[...] dados que podem ser livremente utilizados, reutilizados e redistribuídos por qualquer pessoa – sujeitos, no máximo, à exigência de atribuição à fonte original e ao compartilhamento pelas mesmas licenças em que as informações foram apresentadas.

Neste contexto, as instituições públicas necessitam de adequação para que possam preservar a informação e também disponibiliza-la aos indivíduos de forma aberta. Essa abertura significa "evitar um mecanismo de controle e restrições sobre os dados que forem publicados, permitindo que tanto pessoas físicas quanto jurídicas possam explorar estes dados de forma livre." (ISOTANI; BITTENCOURT 2015, p. 20).

A abertura de dados, na visão dos autores Isotani e Bittencout (2015) possibilita que diferentes sistemas e organizações trabalhem colaborativamente, a partir da interoperabilidade de dados abertos, desde que respeitem as três normas que fundamentam dados abertos, as quais são: **disponibilidade e acesso; reuso e redistribuição e participação universal**. A disponibilidade e o acesso refere-se à disponibilidade completa aos dados, com possibilidade de *download*; o reuso e redistribuição são termos de permissões favoráveis reutilização e redistribuição; e a participação universal refere-se a possibilidade de qualquer pessoa fazerem uso, reutilização e redistribuição dos dados (OPEN KNOWLEDGMENT, 2010 *apud* ISOTANI; BITTENCOURT, 2015).

A partir das normas que fundamentam os dados abertos surgem princípios a serem seguidos para que os dados possam ser considerados abertos, os quais são:

Completo, Primários, Oportunidade, Acessibilidade, Processável por máquinas, Não-discriminatório, Não-proprietário e Licença livre.

Segundo a *OPEN Data Government Working Group* (2007) o primeiro princípio, que diz respeito à sua completude – **Completo** - refere-se à necessidade de todos os dados públicos, os não restritos de privacidade e segurança, estarem disponíveis para acesso a comunidade.

De acordo com o segundo – **Primários** - os dados devem ser proveniente da fonte geradora e ser detalhado ao máximo;

No terceiro – **Oportunidade** - os dados devem estar disponíveis o quanto antes para que seja preservado o seu valor.

Seguindo o quarto, **Acessibilidade**, a disponibilidade dos dados deve ser para um público variado e amplo e para diversos fins.

O quinto, **Processável por máquinas**, entende-se que os dados necessitam estar estruturados, para que sejam processados de forma automatizada;

O sexto, **Não-discriminatório**, garante acessibilidade à informação sem restrições, como por exemplo, exigência de registro para o acesso ao dado;

O sétimo, **Não-proprietário**, refere-se ao uso de *softwares* que não tenham controle de entidade, isto é, que sejam *softwares* livre;

O oitavo, **Licença**, os dados devem ser livres de direitos autorais, regulamentos ou qualquer tipo de restrição de autoria, salvo os dados que necessitam de privilégios, restrições e segurança.

Algumas barreiras técnicas devem ser superadas para atender a esses princípios. No contexto arquivístico, o *AtoM* (*Access to Memory* ou Acesso a Memória) como sistema de descrição arquivística, de representação e apresentação da informação no ambiente digital, foi criado com base nas normas arquivísticas internacionais, para permitir a disponibilização das informações na *Web*, destacando-se por ser uma alternativa as instituições públicas que desejam estar em conformidade com a LAI e outras leis, normas e políticas de acesso à informação, transparência pública e abertura dos dados.

Esse sistema tem por características a utilização de metadados arquivísticos para representar e apresentar a informação:

- a) É um *software* de fonte aberta (*open source*), livre e gratuito que torna possível ser utilizado por qualquer instituição pública ou privada, arquivísticas ou não;

- b) Está em beta perpétuo, o que permite que o seu *design* seja atualizado e aperfeiçoado com a colaboração e o *feedback* da comunidade que o utiliza;
- c) É multilíngüe, com a possibilidade de alterar sua interface para outras línguas, como inglês, espanhol dentre outras que o administrador do sistema definiu para ser traduzido;
- d) É multi-repositórios, com a aceitação e agregação de descrições de outras instituições e seu sistema, como o Arquivo Nacional do Brasil tem feito.

A construção do AtoM se deu por meio da utilização de várias linguagens e ferramentas de código aberto e software livre, tendo sido finalmente disponibilizado com a licença e ideologia. A-GPL - General *Public License* (Licença Pública Geral), o que implica ser editável gratuitamente e utilizado para qualquer finalidade: estudado, aperfeiçoado, modificado e redistribuído (modificado ou não).

O AtoM tem, assim, características muito próximas ao que é proposto pelos princípios e política de dados abertos, apresentando vários instrumentos tecnológicos convergentes a serem utilizados na organização e armazenamento de dados. Realizou-se uma exploração do sistema em relação aos princípios de dados abertos, o qual são apresentados os oito princípios e como podem ser atendidos por esse *software*.

Primeiro princípio: Completo. Todos os dados públicos devem ser disponibilizados. Dados públicos são aqueles que não estão sujeitos a restrições de privacidade, segurança ou privilégios de acesso.

O AtoM possibilita que a informação seja disponibilizada de forma multinível, o que pode dar o acesso completo a informação no ambiente digital: a descrição pode ser somente dos fundos arquivísticos preservados pela instituição ou descer até os itens documentais, o que garante o acesso efetivo de toda a informação para a sociedade. Esse processo de descrição pode ser no arquivo permanente, ou até mesmo a partir do corrente, desde que se tenha uma boa gestão documental. Contudo, para visualização do documento, que ainda são produzidos nas instituições públicas em formato papel, é necessário que esse seja digitalizado e preservado em um sistema que tenha todos os requisitos necessários de preservação, como por exemplo, o Archivematica.

Esse é um sistema de preservação da informação arquivística, que baseou-

se, tanto na estrutura quanto no funcionamento, o modelo OAIS (*Open Archival Information System*) e converge vários padrões para a preservação, como o *Mets*, o *Premis*, o *Dublin Core*, o da Biblioteca do Congresso especificação *BagIt* e outros padrões. Normas estas que garantem a integridade, a autenticidade, a confiabilidade e a preservação digital dos dados. O Archivemática, também utiliza-se de *software*, ferramentas e linguagens livres, de código aberto. Destaca-se por haver transparência informacional em todas as atividades que o documento é submetido dentro do sistema. É possível, assim, visualizar todo o procedimento de preservação digital. Após o tratamento, gera-se um pacote DIP (Pacote de Informação de Disseminação) *linkado* ao AtoM, para que este reproduza a imagem do documento digitalizado ou digital.

A convergência desses dois sistemas faz-se necessária para a garantia de acesso a longo prazo, o que depende de haver no sistema recursos de preservação e pelo AtoM ser voltado ao acesso (sem o tratamento de preservação). Por outro lado, ao fazer o *upload* de grande quantidade de documentos para o sistema por meio do AtoM, esse torna-se lento no processamento da informação.

Segundo Princípio: **Primários**. Os dados devem ser coletados na fonte, com o maior nível de detalhamento possível, e não de forma agregada ou modificada. Os documentos arquivísticos, por si só são considerados fonte primária, oriundos das atividades de uma instituição. Ao se utilizar o AtoM possibilita-se que os documentos sejam descritos e representados com máximo de informação possível referente ao item (o que dependerá, por outro lado, das competências do profissional da informação que irá descrever os documentos no sistema).

Terceiro Princípio: Oportunidade. Sua disponibilidade deve ser feita tão rapidamente quanto necessário para preservar o valor dos dados. Neste sentido, a descrição e disponibilização dos dados no AtoM pode ser aplicável no Arquivo Permanente, e também nas fases anteriores - corrente e intermediária - o que necessitará de um controle e gerenciamento. O sistema informa o indivíduo em que fase se encontra, mesmo não disponibilizando o documento gerado e descrito no AtoM, e garante portanto, a transparência pública, sem contar com a possibilidade de controle administrativo.

Quarto Princípio: Acessibilidade. Os dados devem estar disponíveis para a mais ampla gama de usuários e as mais diversas finalidades. Uma vez publicado no AtoM, qualquer indivíduo poderá ter acesso à informação sem que necessite de se

identificar. O sistema possibilita as pessoas visualizar e baixar (*download*) o documento.

Quinto Princípio: Processável por máquinas. Os dados devem ser razoavelmente estruturados, de modo a permitir o processamento automatizado. O sistema, por utilizar o formato PDF, Machado (2015, p. 218) entende que:

Os dados precisam ser corretamente codificados para que possam ser amplamente utilizados. [...]. Deve haver documentação sobre o formato e a codificação dos dados, assim como os significados de cada um dos itens para que os usuários possam conhecer o sentido e o contexto dos dados.

Contudo, Isotani e Bittencout (2015) apresenta um esquema de cinco estrelas em que é possível ocorrer a abertura de dados; segundo essas estrelas

A única exigência para que o dado seja considerado aberto é que os dados sejam disponibilizados com licenças abertas. Mesmo um documento sendo publicado em PDF ou PNG, se ele utiliza uma licença aberta, então pode ser considerado um dado aberto. (ISOTANI, BITTENCOUT, 2015, p. 47)

Neste sentido, além do uso do PDF³ – formato proprietário - o AtoM utiliza-se de metadados para que o documento possa ser bem representado no ambiente digital, bem como, de ontologias por meio da construção de vocabulário controlado para possibilitar o acesso a informação, mesmo que esse ainda seja trabalhado de forma apenas sintática.

Também se utiliza do *Dublin Core*, EAD, EAF, EAC e SKOS para que as páginas possam ser representadas, de acordo com suas especificidades. O *Dublin Core* auxilia na interoperabilidade dos dados, uma vez que é composto por um conjunto de metadados gerais que descrevem a página *Web*; a EAD (*Encoded Archival Description*), é baseada no esquema *Dublin Core*, representa, codifica, exporta e importa as descrições arquivísticas; A EAF é um padrão que realiza as mesmas atividades da EAD, porém é voltado para registros de autoridades; a EAC, semelhante às anteriores, diferencia-se por ser utilizada para funções arquivísticas; e o SKOS (*Simple Knowledge Organization System*) é usado para a construção de Thesaurus, esquemas de classificação, listas de títulos de assuntos e Taxonomias,

³ Segundo informações disponíveis no site da *Adobe Acrobat DC*, criadora do formato PDF, este agora é um padrão aberto mantido pela *International Organization Standardization* (ISO). No entanto, a norma ISO 32000-1 alega que o cumprimento desse formato pode haver uso de patentes relativas à criação, modificação, exibição e processamento de arquivos PDF. Disponível em: <<https://www.loc.gov/preservation/digital/formats/fdd/fdd000277.shtml>> Acesso em 20 maio 2017. Disponível em: <<https://acrobat.adobe.com/br/pt/why-adobe/about-adobe-pdf.html>>. Acesso em 20 maio 2017.

e no AtoM auxilia na importação e exportação de XML do vocabulário construído. Mesmo que não sejam estruturados semanticamente, há indícios de possibilidade de construção de semântica no sistema, ainda que superficial. O que acredita-se ser processáveis por máquinas.

Sexto Princípio: **Não discriminatório** - os dados devem estar disponíveis para qualquer pessoa, sem necessidade de registro. O AtoM permite que a informação seja acessada gratuitamente e não exige a identificação, cadastro ou qualquer tipo de exigência nesse sentido. Assim, qualquer pessoa poderá acessar a informação, pois uma vez publicada, estará disponível na *Web* de maneira aberta.

Sétimo Princípio: **Não-Proprietário**. Os dados devem estar disponíveis em um formato sobre o qual nenhuma entidade tem o controle exclusivo. O AtoM por ser um *software* livre, gratuito e *open source*, garante que a informação será acessada gratuitamente, sem a necessidade de se pagar para manter a informação acessível. De acordo com Machado (2015, p. 219):

Quando uma empresa produz o programa que é necessário para que um arquivo onde dados armazenados podem ser lidos, o acesso do usuário a tais informações passa a ser dependente do programa. O uso de formato proprietário cria a possibilidade de o programa estar disponível ao público somente mediante a cobrança de um determinado valor.

Neste contexto, o AtoM foi desenvolvido por uma empresa, porém, de maneira conjunta e consensuada com órgãos importantes como a UNESCO e o ICA (Conselho Internacional de Arquivos). O grande investimento na construção do AtoM resultou em um sistema de qualidade que pode ser utilizado por qualquer instituição pública, gratuitamente, atendendo a políticas de acesso a informação, de dados abertos e também, no Brasil, a LAI.

Oitavo Princípio: os dados não estão sujeitos a quaisquer direitos de autor, patentes, marcas comerciais ou regulamento secreto. Pode ser permitida uma razoável privacidade e restrições de privilégio e segurança.

O AtoM, por ser um sistema flexível, e que pode ser voltado a outros tipos de Unidades de Informação, como Bibliotecas e Museus, dispõem de uma área voltada para utilização de licenças no objeto digital, por exemplo *Creative Commons*. Geralmente, estes recursos são utilizados em Bibliotecas, para proteção dos direitos dos autores.

No caso da instituição arquivística, dependendo da informação, é mediante

a permissão do autor, doador, ou instituição responsável. Para isso, há um campo de metadados que orienta questões sobre reprodução do objeto informacional.

Machado (2015) defende o uso de licenças livres para que os direitos sejam garantidos, bem como a liberdade e princípios dos dados abertos. Entende-se que o autor critica o tradicional, que ainda utilizam-se do *Copyright*, e esclarece que há uma tradução equivocada desse princípio, pois os "[...] dados precisam estar protegidos com algum tipo de licença para que não sejam apropriados por outra parte que venha restringir a cadeia de inovação com base nessa informação " (2015, p. 220). Neste sentido, o autor trabalha no contexto biblioteconômico, apresentando o que pode acontecer quando está não tem nenhuma proteção legal.

4 CONSIDERAÇÕES FINAIS

O AtoM é um sistema com grande potencial em relação aos dados abertos, ao permitir que as instituições públicas possam armazenar e preservar os dados no Archiavamtica (também um *software* livre), e disponibilizar no AtoM para que todo cidadão possa ter acesso efetivo a informação.

Contudo, o sistema, em si, tem muito mais potencial do que simplesmente dar acesso, permite que a informação seja organizada, o que entendemos que é uma forma de estrutura, classificada, descrita, representada, apresentada de forma multinível, clara e transparente para o indivíduo.

Ao considerar as realidades das instituições públicas na *Web* - que geralmente não descrevem a informação até o item documental, isto é, quando seus acervos não estão desorganizados, sem digitalizar - apresenta-se como forte aliado na implementação das políticas de acesso à informação e política de dados abertos.

Quanto aos princípios, o AtoM atende a todos os requisitos, alguns em sua completude, outros superficialmente, mas cumpre com o básico para se ter dados abertos disponibilizados por esse sistema.

Percebemos também, que o AtoM prepara os dados para uma realidade futura na *Web*, a *Web* semântica, em que esses serão interoperáveis, conectados, estruturados e compreendidos por máquinas. Atualmente, o AtoM possibilita que as instituições públicas se adequem às políticas de acesso à informação no meio digital, bem como, saiam de uma realidade de *Web* 1.0 para a *Web* 2.0, a *Web* Colaborativa em que por meio do acesso se abre um novo universo de colaboração

na produção de informação.

Em tempos pós-custodiais, que configura a entrada da sociedade moderna na utópica "sociedade da informação", o AtoM destaca-se por sua característica de acesso à informação, uma vez que, o desenvolvimento humano, social, econômico, político e intelectual, é dependente dessa disponibilidade de dados.

No Brasil, a LAI, oriunda das políticas de acesso à informação, foi criada pensando nas ações e programas de um país emergente, para a construção de uma sociedade baseada na informação. E em consonância com esses instrumentos políticos, desenvolveram-se políticas de dados abertos para instituições públicas e, juntamente, princípios que devem ser atendidos pelos órgãos que desejam a abertura de dados e a transparência pública.

Por se tratar de um *software* desenhado especificamente para um ambiente *Web* de descrição arquivística, o AtoM possui a grande vantagem de funcionar bem nesse ambiente. Os multiníveis do AtoM permitem que os dados dos documentos descritos neste ambiente atendam significativamente as normas que fundamentam a abertura dos dados e, portanto, aos seus oito princípios básicos.

Observa-se que o AtoM atende parcialmente aos princípios de dados aberto, uma vez que permite a transparência pública e a representação e acesso a informação livre e gratuita ao indivíduos no ambiente digital. Neste sentido, percebe-se que o AtoM, pode ser visto como o ponto de partida para que as instituições comecem a se adequar para a produção de dados abertos no mundo digital mesmo que o *software* ainda não atenda todos os princípios de dados abertos.

Sugere-se, finalmente, como parte das conclusões desse estudo maiores aprofundamentos tendo em vista que as colaborações advindas de novos estudos poderão ser recebidas e incorporadas pelas equipes responsáveis pelo desenvolvimento contínuo do *software* AtoM e suas potencialidades.

Referencias

ASPECTOS GERAIS DA LEI. **Portal de Acesso à Informação: Governo Federal**, Brasília, [s. d]. Disponível em: <<http://www.acessoainformacao.gov.br/perguntas-frequentes/aspectos-gerais-da-lei#1>>. Acesso em: 10 de mar. 2017.

BRASIL. Lei n. 12.527, de 18 de novembro de 2011. Regula o acesso a informações previsto no inciso XXXIII do art. 5º, no inciso II do § 3º do art. 37 e no § 2º do art. 216 da **Constituição Federal**; altera a Lei n. 8.112, de 11 de dezembro de 1990;

revoga a Lei n. 11.111, de 5 de maio de 2005, e dispositivos da Lei n. 8.159, de 8 de janeiro de 1991; e dá outras providências. Disponível em: http://www.planalto.gov.br/ccivil_03/_ato2011-2014/2011/lei/l12527.htm. Acesso em: 10 abr. 2017.

BSA. **What's the Big Deal With Data?**. 2015. Disponível em: <http://data.bsa.org/wp-content/uploads/2015/12/bsadatastudy_en.pdf>. Acesso em: 21 abr. 2017

CANCLINI, N. G. **Culturas híbridas: estratégias para entrar e sair da modernidade**. São Paulo: Editora da Universidade de São Paulo, 1998.

DAVENPORT, T. H.; PATIL, D. J. Data Scientist: The Sexiest Job of the 21st Century. **Harvard Business Review**, v. 1, p. 70-77, 2012. Disponível em: <<https://hbr.org/2012/10/data-scientist-the-sexiest-job-of-the-21st-century>>. acesso em 20 abr. 2017.

HENDGES, C. E. J.; SLIWKA, I. S. O direito constitucional à informação, a Lei 12.527/2011 e a gestão documental no Poder Judiciário. **Autos & Baixas**, Rio Grande do Sul, v. 1, n. 1, 2013.

ISOTANI, S.; BITTENCOURT, I. I. **Dados abertos conectados**. São paulo: Novatec Editora, 2015.

MACHADO, J. Dados abertos e ciência aberta. IN: ALBAGLI, S.; MACIEL, M. L.; ABDO, A. H. (Org.). **Ciência aberta, questões abertas**. Brasília: IBICT; Rio de Janeiro: UNIRIO, 2015, p. 201-229.

OPEN DEFINITION. **Organization Knowledge Foundation**. Disponível em: <http://opendefinition.org/od/>>. Acesso em: 20 mar. 2017.

RIBEIRO, F. Gestão da informação/Preservação da memória na era pós-custodial: um equilíbrio precário? In: MESA REDONDA DE PRIMAVERA: CONSERVAR PARA QUÊ?, 8., 2015, Porto. **Atlas...** Porto: Faculdade de Letras da Universidade do Porto, 2005. Disponível em: <<http://repositorio-aberto.up.pt/handle/10216/39365>>. Acesso em: 29 mar. 2017.

RIBEIRO, F. O desafio da formação profissional: novo paradigma, novo modelo formativo. IN: CONGRESSO INTERNACIONAL DE ARQUIVOS, BIBLIOTECAS, CENTROS DE DOCUMENTAÇÃO E MUSEUS, 2002, Porto. **Atlas...** Porto: Faculdade de Letras da Universidade do Porto, 2002, p.419-440. Disponível em:<<https://repositorio-aberto.up.pt/handle/10216/14022>>. Acesso em: 20 maio 2017.

RIBEIRO, F. **Os Arquivos na era pós-custodial: reflexões sobre a mudança que urge operar**.

RIBEIRO, F. Da mediação passiva à mediação pós-custodial: o papel da ciência da informação na sociedade em rede. **Inf. & Soc.:Est.**, João Pessoa, v.20, n.1, p. 63-70, jan./abr. 2010.

W3C BRASIL. Publicação de dados em formato aberto. Escola de Políticas Públicas, 2013. Disponível em:<<http://cursos.ep.org.br/course/view.php?id=25§ion=1>>. Acesso em: 20 mar. 2017.