

MOREIRA, F. M.; BOTEGA, L. C.; SANTAREM
SEGUNDO, J. E.; SANT'ANA, R. C. G.



Acesso a Dados e a Ciência da Informação

Aplicações,
tendências e
reflexões

Fábio Mosso Moreira
(*coord.*)

Leonardo Castro Botega
José Eduardo Santarem Segundo
Ricardo César Gonçalves Sant'Ana
(*orgs.*)

Acesso a Dados e a Ciência da Informação

Aplicações, tendências e reflexões

Tupã-SP
Faculdade de Ciências e Engenharia UNESP – Campus de Tupã
2022

Copyright © 2022 GPTAD – Grupo de Pesquisa Tecnologia de Acesso a Dados

Coordenador | Coordinador | Coordinador

Fábio Mosso Moreira

Organizadores | Organizers | Organizadores

Leonardo Castro Botega

José Eduardo Santarem Segundo

Ricardo César Gonçalves Sant’Ana

Normalização bibliográfica | Bibliographic normalization | Normalización bibliográfica

Elizabete Cristina de Souza de Aguiar Monteiro

Apoio de registro editorial | Editorial Registration Support | Soporte de registro editorial

Eliana Katia Pupim

Capa | Cover | Tapa

Fábio Mosso Moreira

Dados Internacionais de Catalogação na Publicação (CIP)

A174 Acesso a Dados e a Ciência da Informação: Aplicações, tendências e reflexões / Fábio Mosso Moreira (coord.), Leonardo Castro Botega, José Eduardo Santarem Segundo, Ricardo César Gonçalves Sant’Ana (orgs.). – Tupã : Faculdade de Ciências e Engenharia UNESP – Campus de Tupã
350 p. : il. gráfs., tabs.

Inclui bibliografia.

ISBN 978-65-88374-03-0

1. Acesso a dados. 2. Tecnologia da informação e comunicação. 3. Ciência da Informação. I. Moreira, Fábio M. II Botega, Leonardo C. III Santarem Segundo, José E. IV. Sant’Ana, Ricardo C. G. V. Título.

CDD 020

Acesso: dadosabertos.info/events/dti/4dti.php

Apresentação

O conteúdo desta obra resulta das exposições realizadas no IV Encontro Dados, Tecnologia e Informação (IV DTI), ocorrido no dia 16 de Julho de 2021. Promovido no âmbito do Programa de Pós-Graduação em Ciência da Informação – PPGCI (UNESP), com apoio do Grupo de Pesquisa Tecnologia de Acesso a Dados – GPTAD (UNESP), o evento proporcionou uma reunião online de pesquisadores, alunos e comunidade científica em prol do debate acerca das aplicações, tendências e reflexões resultantes de investigações que abordam a inter-relação entre a tríade ‘Dados’, ‘Tecnologia’ e ‘Informação’.

As palestras e os trabalhos apresentados no IV DTI materializam este livro eletrônico, intitulado ‘*Acesso a Dados e a Ciência da Informação: Aplicações, tendência e reflexões*’, cuja publicação encontra-se motivada pela necessidade de dar visibilidade aos temas da Ciência da Informação que possuem intersecções com os aspectos relacionados às tecnologias de acesso a dados e informação, e, ao mesmo tempo, de estabelecer relações interdisciplinares deste arcabouço teórico com segmentos da sociedade que ganharam destaque no cenário pandêmico atual (COVID-19), como, por exemplo, a saúde, a educação, a cultura e a economia.

Para estruturar o fio condutor desta obra, o livro foi organizado em três eixos temáticos. O primeiro congrega os capítulos que tratam sobre questões que dão ênfase aos aspectos ligados ao acesso e uso dos **dados** (Capítulos 1 ao 7). Algumas abordagens ganham destaque, como, por exemplo: o estudo das convergências com o contexto bibliotecário; as implicações para o cenário científico atual; e aspectos relacionados a qualidade dos conjuntos de dados e às ferramentas de visualização.

No segundo eixo temático foram agrupados os Capítulos que possuem vertentes direcionadas ao uso das **tecnologias informacionais** (Capítulos 8 ao 11). Neste bloco, encontram-se capítulos que tratam sobre: questões de interoperabilidade e integração; aspectos ligados ao desenvolvimento de interfaces considerando a experiência do usuário; e aplicação de tecnologias informacionais no contexto governamental.

O terceiro e último eixo temático apresenta os capítulos (Capítulos 12 ao 21) que se dedicam ao estudo de aspectos ligados a **informação**, ou seja, de unidades informacionais que possuem maior carga de elementos semânticos, apresentadas de forma menos estruturada quando comparada aos conjuntos de dados. Destacam-se, neste eixo, temas relacionados com: a representação da informação (ex: uso de vocabulários controlados; padrões de metadados; estruturação conceitual de teses e dissertações); com elementos contextuais da Sociedade da Informação (ex: desinformação, capitalismo de vigilância, assimetria informacional); e com os novos formatos de disseminação de informações (ex: Redes Sociais Online).

A convergência dos três eixos delimitados permite ao leitor refletir sobre as implicações dos imperativos tecnológicos considerando as distintas manifestações informacionais, tanto o conteúdo informacional quando os conjuntos de dados. Isso se dá pela demonstração de *cases* que ilustram aplicações teóricas e práticas, acompanhadas de proposições que vislumbram as principais tendências investigativas da Ciência da Informação, o que o torna um prato cheio para os entusiastas e especialistas da área.

Uma boa leitura a todos.

Fábio Mosso Moreira

Sumário

Visualização de dados e de informação em bibliotecas: estudo exploratório preliminar para pesquisa sobre Inteligência Artificial

Antonio Victor Wolf Tadini, Ricardo César Gonçalves Sant'Ana e José Eduardo Santarem Segundo.....9

Bibliotecário de dados: enfoques e perspectivas nacionais de estudos na Ciência da Informação

Raimunda Fernanda dos Santos e João Victor Moraes Sena.....28

Os dados científicos e os dados de pesquisa na Ciência da Informação Brasileira: uma análise a partir da BRAPCI

Ananda Fernanda de Jesus, William Pires de Castro e José Eduardo Santarém Segundo.....45

What we have to do for Responsible Data Science?

Morgana Carneiro de Andrade, Paula Regina Ventura Amorim Gonçalves, Decio Wey Berti Junior, Ana Alice Rodrigues Pereira Baptista and Caio Saraiva Coneglian.....60

Os princípios FAIR no Ciclo de Vida dos Dados para a Ciência da Informação: maximização do acesso, uso e reúso de dados científicos

Anderson Rafael Castro Simões, Renata Lemos dos Anjos e Guilberme Ataíde Dias.....79

Visualização de dados: a autonomia do usuário na construção de gráficos

Paulo George Miranda Martins, Danila Fernandes Alencar e Ricardo César Gonçalves Sant'Ana.....95

Qualidade de dados no processo decisório de educação a distância

Lafaiete Henrique Rosa Leme, Felipe Stefani Correia dos Santos, Késsia Rita da Costa Marchi e Leonardo Castro Botega..... 110

Interoperabilidade e integração entre DSpace e Archivemática

Laura Marques de Magalhães e Felipe Augusto Arakaki..... 125

Capítulo retirado por solicitação dos autores [pgs. 141-155]

O encapsulamento informacional enquanto fator de aprimoramento na Experiência do Usuário (UX)

Guilherme Oliveira Pironi..... 156

Governança pública e as Tecnologias da Informação e Comunicação: um breve olhar sobre a Lei de Acesso à Informação e a Política de Dados Abertos no Brasil

Kelren Cecília dos Santos Lima da Mota, Diego Leonardo de Souza Fonseca e Roberto Lopes dos Santos Junior..... 167

Análise e gestão de riscos dos documentos digitais existentes na curadoria digital

Fábio Eder Cardoso e Mibno Dgil Pinto de Brito..... 182

Avaliação de controle de vocabulários em repositórios institucionais: uma proposta de aplicação metodológica intrínseca e extrínseca

Lizandra de Souza Santos Alves, Roberta Cristina Dal'Evedove Tartarotti e Mariângela Spotti Lopes Fujita..... 207

Análise focada no padrão MTD3-BR sob a luz dos Princípios FAIR

Keli Rodrigues do Amaral Benin..... 223

Uma discussão sobre o Capitalismo de Vigilância Informacional no contexto da Ciência da Informação

Wilson Roberto Veronez Júnior e Bianca Savegnago de Mira..... 240

Os memes como um gênero discursivo em Serviços de Rede Social Online

*Débora Matni Fonteles, João Batista Ernesto de Moraes, Gilberto Gomes
Cândido e Fernando de Assis Rodrigues.....255*

A desinformação sob a ótica da qualidade da informação

João Pedro Silva de Albuquerque e Valdirene Aparecida Pascoal.....278

**O trabalho colaborativo entre Ciência da Informação e
Assistência Técnica como fator mitigante da assimetria
informacional na agricultura familiar**

Luciana Sekito de Freitas Zambelli e Cristiane Pantoja de Moraes.....295

**Impacto da qualidade da informação no prontuário do paciente:
Reflexões sobre a fase de coleta**

*Fabrcio Amadeu Gualdani, Nathália Adriele de Lima e Leonardo Castro
Botega.....314*

**Diagrama estrutural para teses e dissertações: uma proposta
didática**

Ricardo César Gonçalves Sant'Ana.....329

Visualização de dados e de informação em bibliotecas: estudo exploratório preliminar para pesquisa sobre Inteligência Artificial

Antonio Victor Wolf Tadini^a, Ricardo César Gonçalves Sant'Ana^b e José Eduardo Santarem Segundó^c

1. Introdução

A contemporaneidade tem se caracterizado pela valorização da informação e dos dados, contexto em que têm sido tomados como ativos informacionais, insumos estratégicos para as mais variadas finalidades: inteligência competitiva; desenvolvimento de ciência e tecnologia; gestão pública e transparência governamental; diagnósticos em saúde; monitoramento na agricultura e na indústria; apenas para citar alguns exemplos.

As bibliotecas, como espaços privilegiados para a circulação de conhecimento, são em essência ambientes de informação, profícuos para a utilização de dados. A visualização de dados, por exemplo, tem muito a contribuir na dinamização e potencialização desses ambientes, uma vez que “ajuda o cérebro

-
- a Mestrando em Ciência da Informação pela Universidade Estadual Paulista – UNESP. E-mail: antonio.vw.tadini@unesp.br. ORCID: <https://orcid.org/0000-0002-7234-4617>.
 - b Doutor em Ciência da Informação pela Universidade Estadual Paulista – UNESP. Professor na Universidade Estadual Paulista – UNESP. E-mail: ricardo.santana@unesp.br. ORCID: <https://orcid.org/0000-0003-1387-4519>.
 - c Doutor em Ciência da Informação pela Universidade Estadual Paulista – UNESP. Professor na Universidade de São Paulo – USP. E-mail: santarem@usp.br. ORCID: <https://orcid.org/0000-0003-3360-7872>.

humano a processar dados de modo mais veloz e mais eficaz do que informações baseadas em texto” (CHEN, 2017c, p. 24).

Os dados são inúteis se não formos capazes de visualizar o que estamos procurando [...] e, para isso, é necessário ter as melhores ferramentas de visualização. Com o tempo é verdade que os modelos de inteligência artificial e *machine learning* serão aqueles que processam essas enormes quantidades de dados (já é feito, de fato, em alguns campos), mas sempre haverá um fator humano decisivo (COUTO, 2019, p. 220).

Sobre a diferença entre a visualização de dados e a visualização da informação, pode-se afirmar que, apesar de serem conceitos distintos, tendo campo próprio cada um, com interesses próprios, têm muitos pontos em comum. Kim, Zhu e Chen (2016, p. 123), em estudo que compara as tendências desses domínios entre 2000 e 2014, concluem que

Na visualização de dados, tópicos emergentes e novos progressos abarcam a redução da dimensionalidade e aplicações de técnicas visuais [...]. A pesquisa de visualização de informação está examinando aspectos cognitivos e teóricos. Em conclusão, a visualização das informações e a visualização dos dados se envolveram uma à outra. Ao mesmo tempo, ambos os campos estão se desenvolvendo distintamente com seus próprios interesses científicos.

Esses aspectos cognitivos são referidos por Saracevic (1996) para descrever as relações interdisciplinares da Ciência da Informação com a Biblioteconomia e também com as Ciências

Cognitivas, dentro das quais se encontra o campo da Inteligência Artificial.

Segundo Santaella (2019, p. 392; p. 402), a inteligência artificial, a despeito de remontar a meados do século XX, encontra-se em uma fase de “acelerado aumento de complexidade” nos resultados de suas pesquisas, no que considera uma “explosão até agora bem-sucedida”.

No contexto das bibliotecas não é diferente. A *American Library Association* (ALA), assim como publicou um número inteiro do seu periódico *Library Technology Reports* dedicado à visualização da informação, em 2017, publicou mais recentemente, em 2019, um sobre inteligência artificial em bibliotecas. Evidentemente, ambos os temas constituem tendências na área.

Sendo assim, este trabalho tem como objetivo geral explorar, em nível teórico, possibilidades de aplicação de tecnologias de visualização de dados e de informação no âmbito das bibliotecas, como um levantamento preliminar que forneça subsídios para pesquisa futura sobre inteligência artificial em bibliotecas.

2. Procedimentos metodológicos

A pesquisa se caracteriza como qualitativa (STRAUSS; CORBIN, 2008) e, quanto ao tipo, como exploratória conforme seu objetivo geral, bem como bibliográfica quanto aos procedimentos técnicos empregados (GIL, 2002). Para coleta de dados, procedeu-se a um levantamento bibliográfico nas bases de dados orientado pela questão de pesquisa: quais são as

possibilidades de aplicação de tecnologias de visualização de dados e de informação em bibliotecas, considerando eventuais contribuições de tecnologias de inteligência artificial para essas atividades? O Quadro 1 apresenta a estratégia de busca adotada para recuperação dos textos.

Foram selecionados estudos considerando a ordem de relevância oferecida pelas bases de dados, e de acordo com o exame do título, das palavras-chave e do resumo. A leitura do texto completo foi utilizada subsidiariamente para efeito de seleção. Em seguida, os textos selecionados foram analisados para realização da revisão de literatura.

Quadro 1 – Expressões de busca nas bases de dados

Base de dados	Expressão de busca
Base de Dados em Ciência da Informação (BRAPCI)	Todos os campos: “visualização da informação”
	Todos os campos: “visualização de dados”
	Campo Título: “inteligência artificial”
Web of Science (WoS)	<i>Todos os campos: (“information visualization” OR “data visualization”) AND (“artificial intelligence”) AND (“information science” OR “library science” OR librarianship)</i>
Library and Information Science Abstracts (LISA)	Todos os campos, exceto texto completo: (“ <i>information visualization</i> ” OR “ <i>data visualization</i> ”) AND (“ <i>artificial intelligence</i> ”)

Fonte: Autores.

São três os principais tópicos abordados, conforme os quais se estrutura o trabalho: (i) visualização de dados e de informação em bibliotecas; (ii) inteligência artificial em bibliotecas; e (iii)

resultados preliminares sobre possibilidades com inteligência artificial e visualização de dados e de informação em bibliotecas.

3. Visualização de dados e informação em bibliotecas

Sobre o escopo da visualização de dados, Yang (2017, p. 852) explica que se trata de “uma disciplina abrangente que atravessa a computação gráfica, a interação homem-computador, a estatística, a psicologia”.

Técnicas para visualização de dados e de informação consistem na sua transformação em representações imagéticas (tipos de gráficos, linhas do tempo, infográficos, etc.) com o objetivo de beneficiar a compreensão humana, uma vez que, como pontua Couto (2019), “a capacidade humana de reconhecimento de padrões visuais é inata”.

A visualização ajuda a cognição devido às formas como ela aumenta a memória e o processamento dos recursos disponíveis através de propriedades pré-atentivas, Leis de Gestalt e muitas outras propriedades perceptivas (CHEN, 2017b, p. 9).

Dentro disso, existem dois sentidos para o conceito, que podem ser relacionados à função da visualização: (i) a visualização de dados “estuda principalmente como aproximar a percepção humana sobre os dados da percepção natural da visualização humana” (YANG, 2017, p. 852); e (ii) a visualização de dados “também estuda a interação visual da expressão de dados para aumentar a consciência humana, exibindo as informações

implícitas dos dados. Visa explorar a lei dos dados” (YANG, 2017, p. 852).

Nesse segundo sentido, a visualização se serve à tarefa de apoiar a exploração de dados. É uma demanda crescente tendo em vista a consolidação do paradigma da ciência orientada por dados, em que pesquisas exploratórias partem dos dados dispensando a colocação de uma hipótese de pesquisa.

Uma experiência bastante interessante nesse sentido é a relatada por Hayes, Yi e Villaveces (2009), e referida também por Chen (2017c). A Biblioteca de Ciências da Saúde da Universidade da Carolina do Norte, em parceria com o Instituto Renaissance de Computação, implementou um centro de visualização com um painel para interação em tempo real, com função exploratória, isto é, de apontar para novas questões de pesquisa e direcionamentos. Os pesquisadores passaram a poder consultar e colaborar com especialistas em visualização para desenvolver aplicações customizadas para análise de dados.

A visualização pode então ensejar raciocínios, conferindo mais valor aos dados e informações, uma vez que, “ao simplificar a busca de informações e melhorar o reconhecimento de padrões, a visualização permite operações de inferência perceptiva” (CHEN, 2017b, p. 9).

Essa possibilidade se torna especialmente interessante quando se está lidando com grandes conjuntos de dados ou volumes de informação textual, por exemplo.

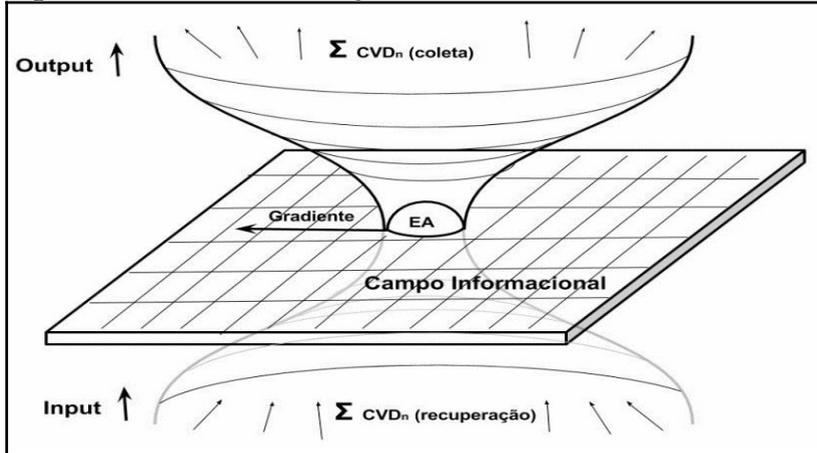
[...] a capacidade para processar e interpretar mais facilmente os dados através do uso da visualização da informação é significativa na retransmissão de mensagens e temas e proporciona uma forma mais fácil de se compreender para grandes quantidades de informação (CHEN, 2017c, p. 25).

É como se as técnicas para beneficiamento da visualização encurtassem um percurso de “camadas de abstração que emergem dos processos de interpretação quando da transdução informacional dos conteúdos registrados nos suportes digitais altamente estruturados até o conteúdo próprio para resposta aos usuários” (SANT’ANA, 2019, p. 117).

Conforme Sant’Ana (2019), a dinâmica dos dados pode ser representada na forma de um ciclo, ao qual se refere como “ciclo de vida dos dados”. O ciclo é composto por quatro fases: (i) coleta; (ii) armazenamento; (iii) recuperação e (iv) descarte. Dentro desse dispositivo de análise, a visualização de dados se insere como um elemento integrante da fase de recuperação, e a fase de recuperação do ciclo de vida dos dados se caracteriza como uma dimensão *input*, como se pode visualizar na Figura 1.

As intervenções que beneficiam a visualização de dados podem funcionar como catalisadores do processamento da informação realizado tanto pelos usuários das bibliotecas, quanto pelos bibliotecários. Ambos podem estar na posição de alvo da dimensão *input* do campo informacional, como mentes interpretadoras, que realizam transdução informacional.

Figura 1 – Dimensões do campo informacional



Fonte: Sant’Ana (2019).

Com os usuários na posição de alvo da dimensão *input* pode-se citar o uso de diversos tipos de dados – referentes à coleção, à circulação de itens, a serviços prestados, acesso de recursos digitais, etc. – para comunicação com os usuários, por meio de ferramentas de visualização em seu website (TERRIL, 2018).

A comunicação pode se dar também de modo interativo, estimulando-se o *storytelling* e o engajamento dos usuários, como descrito por Chen (2017c), apontando para um caminho que as bibliotecas poderiam seguir tomando como exemplo uma ferramenta de visualização da ‘CNN.com’ destinada à interação de e com familiares de combatentes falecidos no Iraque e no Afeganistão, conforme Figura 2. Couto (2019, p. 219) ressalta que

é “recomendável adicionar interação em nossas visualizações de dados”.

Figura 2 – Ferramenta de visualização do projeto “*Home and Away*” da CNN.com



Fonte: Chen (2017c).

Ainda com o usuário na posição alvo da dimensão *input*, outro caso muito interessante é a experiência de Legrady e Hoetzlein (2019) na Biblioteca Pública de Seattle – referido também por Chen (2017c). O projeto chama “*Making Visible the Invisible*”. Nesse projeto, seis grandes telas de LCD foram instaladas atrás do balcão principal de informações dos bibliotecários.

Neste caso, as representações cumprem a função de expressar os cenários de circulação dos itens da biblioteca, sendo geradas em tempo real a partir da coleta de dados de saída dos itens. São considerados sua classificação na CDD, seu código na coleção e seus termos de indexação. No início, pode até parecer algo trivial, mas com um trabalho estético primoroso, a prática

demonstrou que, para a dinamização e potencialização da biblioteca, é muito interessante que indicadores relacionados àquilo que a comunidade de clientes considera informação útil em um determinado momento esteja acessível visualmente.

Figura 3 – Projeto “*Making Visible the Invisible*” na Biblioteca Pública de Seattle



Fonte: Legrady e Hoetzlein (2019). © George Legrady.

Como mencionado, o bibliotecário também pode estar na posição de alvo da posição *input*, interpretando a partir das representações de visualização de dados e informação. Podem ser utilizadas, por exemplo, para avaliação de dados da biblioteca, para melhoria dos processos de gestão e tomada de decisão. Murphy (2013) relata a experiência com o programa de avaliação das bibliotecas da Universidade de Ohio, que utiliza dados sobre a contagem diária dos portões e sobre os serviços de pesquisa oferecidos, além de uma métrica da *Association for Research Libraries* denominada ARL Ranking.

4. Inteligência Artificial em bibliotecas

Griffey (2019a, p. 9), que é bibliotecário e consultor em tecnologia, defende que “a IA e os sistemas de aprendizado de máquina [*machine learning*] têm o potencial de alterar funções básicas dentro das bibliotecas” e que, em se tratando de tecnologias emergentes, os bibliotecários devem compreendê-las, experimentá-las, e construir algumas de suas próprias ferramentas.

Esforços nesse sentido são relatados na literatura, e tais tecnologias têm sido utilizadas para diversas finalidades. São alguns exemplos: recomendação na recuperação da informação (ASEMI; ASEMI, 2018; YELTON, 2019); serviço de referência (ASEMI; ASEMI, 2018; KIM, 2019; LANCASTER, 1994); representação descritiva (ASEMI; ASEMI, 2018; KIM, 2019; LANCASTER, 1994); indexação automática de documentos (GRIFFEY, 2019b; KIM, 2019; SCHREUR, 2020); classificação automática de documentos (SCHREUR, 2020); entre outras tarefas.

Desse modo, o próximo tópico apresenta resultados preliminares para subsidiar pesquisas futuras que apontem em que medida podem ser relacionadas tecnologias de inteligência artificial com tarefas de visualização de dados e de informação em bibliotecas.

5. Possibilidades com Inteligência Artificial e visualização de dados e de informação em bibliotecas

Uma confluência possível entre a visualização e o *machine learning* – que é um subconjunto da inteligência artificial – é a descrita em Harvey e Porter (2016), em estudo sobre um aspecto da interação humano-computador chamado de estratégias de amostragem orientadas para o usuário. A pesquisa investiga, em vez de situações nas quais o computador seleciona quais dados devem ser rotulados a cada iteração, situações em que quem faz isso é o usuário.

Tanto a análise visual quanto o *machine learning* tentam aproveitar os pontos fortes complementares dos seres humanos e das máquinas para resolver tarefas complexas de exploração de dados. Estes campos se sobrepõem mais significativamente quando o treinamento está envolvido: a visualização ou a ferramenta de *machine learning* melhoram com o tempo, explorando as observações da interação humano-computador (HARVEY; PORTER, 2016, p. 64).

Em artigo sobre redes neurais – que é um tipo de *machine learning*, e, portanto, de inteligência artificial –, Yelton (2019) relata que tem como objetivo a visualização de dados em interfaces que funcionem com base nesse tipo de tecnologia.

A rede neural codifica informação sobre ligações entre textos, ligações que, no entanto, não são facilmente exploráveis apenas em uma interface de texto. Imagine, em vez disso, um mapa onde círculos maiores, mais ou menos espaçados, mostrem os *clusters* de significado

presentes no corpus. “Dando um zoom” nos *clusters*, seria possível ver as ligações individuais entre textos que os compõem (YELTON, 2019, p. 14).

Sobre as redes neurais em sua relação com a tradição dos metadados em bibliotecas, a autora mostra algumas disparidades: (i) redes neurais não resultam em categorias; (ii) redes neurais podem resultar em *clusters*, mas esses *clusters* não têm (e às vezes não podem ter) cabeçalhos de assunto (*subject headings*); e (iii) redes neurais podem operar em espaços onde metadados tradicionais são indisponíveis ou inadequados (YELTON, 2019).

Jiang e Carter (2018, p. 16) sugerem que “a visualização de dados em bibliotecas no futuro pode dispor de inteligência artificial para fins de navegação”, de modo a ampliar suas funções, na medida em que,

com o auxílio de inteligência e a busca de bancos de dados, visualizações de dados devem ser capazes de conduzir os usuários para os resultados adaptados por máquinas que correspondam com precisão ao que ele/ela propõe (JIANG; CARTER, 2018, p. 17).

Os autores também apresentam o seguinte cenário: “a análise de dados levantados para avaliação da biblioteca geralmente encontra respostas ausentes ou incompletas” e utilizar esses dados desbalanceados ou apagar os registros incompletos pode gerar imprecisões e, assim, viciar a análise. Como solução eles mostram que, com uma técnica de inteligência artificial chamada imputação,

é possível preencher as células ausentes no conjunto de dados e depois alimentar a visualização com um conjunto de dados “completados” artificialmente. O pré-processamento deste tipo não está disponível na maioria das aplicações de visualização de dados (JIANG; CARTER, 2018, p. 16-17).

Note-se que, por essa perspectiva, os algoritmos de inteligência artificial atuam no pré-processamento dos dados, de modo que, em um primeiro momento se encontram na posição de alvo da dimensão *input* do campo informacional, e em seguida realizam uma tarefa de tratamento sobre os dados. O objetivo final é gerar uma visualização, como facilitadora da compreensão por um humano como alvo da dimensão *input*.

A visualização aqui contribui na medida em que, para tratar os dados, o algoritmo opera estruturando e codificando os conteúdos, sobrepondo camadas que o colocam em uma relação de antagonismo com a compreensão humana.

As diversas interfaces envolvidas neste processo transformativo, ao mesmo tempo em que viabilizam os seus usos e interpretação, necessariamente erguem barreiras à percepção dos usuários sobre as transduções exigidas para que os conteúdos possam ser tratados pelos dispositivos digitais (SANT’ANA, 2019, p. 122).

Sendo assim, um caminho que preliminarmente começa a se delinear é uma composição em que as tecnologias de inteligência artificial e de visualização promovam processos transformativos complementares, com o objetivo de estabelecer um processo

maior que beneficia a interpretação e a exploração dos recursos por humanos.

Essa complementaridade pode ser observada na descrição que Gonzalez Aguilar et al. (2017) apresentam para o processo de visualização analítica, como um método analítico que combina técnicas de análise automatizada com visualizações interativas para melhor compreensão sobre grandes conjuntos de dados complexos, de modo a favorecer o raciocínio e a tomada de decisões.

6. Considerações finais

Este trabalho se propôs a explorar, em nível teórico, possibilidades para desenvolvimento e implementação de atividades de visualização de dados e informação no âmbito das bibliotecas. A pesquisa bibliográfica realizada permite afirmar que as tecnologias e práticas de visualização de dados e de informação constituem oportunidades e desafios contemporâneos para as bibliotecas, no que se refere à melhoria tanto de seus processos, como dos produtos e serviços oferecidos.

O trabalho lançou também alguns indícios para pesquisas futuras sobre como tecnologias de inteligência artificial podem eventualmente contribuir e se combinar a atividades de visualização em bibliotecas – seja na “clusterização” de documentos, no beneficiamento da navegação do usuário, ou no pré-processamento de dados, por exemplo, conforme apresentado.

Referências

ASEMI, A.; ASEMI, A. Artificial intelligence (AI) application in library systems in Iran: a taxonomy study. **Library Philosophy and Practice**, [s.l.], e1840, 2018.

CHEN, H. M. An overview of information visualization. **Library Technology Reports**, Chicago, v. 53, n. 3, p. 5-7, 2017.

CHEN, H. M. Information visualization: principles, techniques, and software. **Library Technology Reports**, Chicago, v. 53, n. 3, p. 8-16, 2017.

CHEN, H. M. Real-world uses for information visualization in libraries. **Library Technology Reports**, Chicago, v. 53, n. 3, p. 21-25, 2017.

COUTO, F. C. S. Visualização de dados: passado, presente e futuro. **Liinc em Revista**, Rio de Janeiro, v. 15, n. 2, p. 205-223, 2019.

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

GONZALEZ AGUILAR, A. *et al.* **Visualização de dados, informação e conhecimento**. Florianópolis: Ed. UFSC, 2017.

GRIFFEY, J. Introduction. **Library Technology Reports**, Chicago, v. 55, n. 1, p. 5-9, 2019.

GRIFFEY, J. Conclusion. **Library Technology Reports**, Chicago, v. 55, n. 1, p. 26-28, 2019.

HARVEY, N.; PORTER, R. User-driven sampling strategies in images exploitation. **Information Visualization**, [s.l.], v. 15, n. 1, p. 64-74, 2016.

HAYES, B.; YI, H.; VILLAVECES, A. Information visualization services in a library?: a public health case study. **Bulletin of the American Society for Information Science and Technology**, [s.l.], v. 35, n. 5, p. 13-18, 2009.

JIANG, Z.; CARTER, R. Visualizing library data interactively: two demonstrations using R language. **Library Hi Tech News**, [s.l.], v. 35, n. 5, p. 14-17, 2018.

KIM, B. AI and creating the first multidisciplinary AI lab. **Library Technology Reports**, Chicago, v. 55, n. 1, p. 16-20, 2019.

KIM, M. C.; ZHU, Y.; CHEN, C. How are they different?: a quantitative domain comparison of information visualization and data visualization (2000-2014). **Scientometrics**, [s.l.], v. 107, n. 1, p. 123-165, 2016.

LANCASTER, F. W. Ameaça ou oportunidade?: o futuro dos serviços de biblioteca à luz das inovações tecnológicas. **Revista da Escola de Biblioteconomia da UFMG**, Belo Horizonte, v. 23, n. 1, p. 7-27, 1994.

LEGRADY, G.; HOETZLEIN, R. Making visible the invisible: a data-driven media artwork, in continuous operation for 15 years. *In: SIGGRAPH*, 2019, Los Angeles. **Short Art Papers**. [S. l.]: ACM, 2019. e14. p. 1-6.

MURPHY, S. A. How data visualization supports academic library assessment: three examples from the Ohio State University libraries using Tableau. **College and Research Libraries**, [s. l.], v. 76, n. 9, p. 482-486, 2015.

SANT'ANA, R. C. G. Transdução Informacional: impactos do controle sobre os dados. *In: MARTÍNEZ-ÁVILA, D.; SOUZA, E. A.; GONZALEZ, M. E. Q. (Orgs.). Informação, conhecimento, ação autônoma e big data: continuidade ou revolução?* São Paulo: Cultura Acadêmica: FiloCzar, 2019. p. 117-128.

SANTAELLA, L. A concepção ampliada da mente segundo C. S. Peirce. **Cognitio**, São Paulo, v. 20, n. 2, p. 392-403, 2019.

SARACEVIC, T. Ciência da Informação: origem, evolução e relações. **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 1, n. 1, p. 41-62, 1996.

SCHREUR, P. E. Artificial intelligence and complicit bias. **Technicalities**, [s. l.], v. 40, n. 5, p. 1, 4-6, 2020.

STRAUSS, A. L.; CORBIN, J. M. **Pesquisa qualitativa: técnicas e procedimentos para o desenvolvimento de teoria fundamentada**. 2. ed. Porto Alegre: Artmed, 2008.

TERRIL, L. J. Telling their story with data: what academic research libraries share on their websites. **Journal of Web Librarianship**, [s. l.], v. 12, n. 4, p. 232-245, 2018.

YANG, M. Research progress on data analysis in big data technology. **Advances in Computer Science Research**, [s. l.], v. 74, p. 851-854, 2017.

YELTON, A. HAMLET: neural-net-powered prototypes for library discovery. **Library Technology Reports**, Chicago, v. 55, n. 1, p. 10-14, 2019.

Bibliotecário de dados: enfoques e perspectivas nacionais de estudos na Ciência da Informação

Raimunda Fernanda dos Santos^a e João Victor Moraes Sena^b

1. Introdução

A partir do século XXI, com o desenvolvimento da Web, o barateamento e acesso aos computadores pessoais e dispositivos móveis, o mundo conectou-se de forma jamais imaginada. Foi observada uma explosão de dados gerados por usuários em todos os lugares, a cada minuto por câmeras de monitoramento, GPS, fotos, vídeos, *posts*, e-mails, mensagens entre outros. Segundo um relatório produzido pela multinacional americana IDC, especializada em armazenamento de dados, (tradução livre) de 2005 até 2020, o universo digital crescerá por um fator de 300 (trezentos), de 130 (cento e trinta) exabytes para 40.000 (quarenta mil) exabytes, ou 40 (quarenta) trilhões de gigabytes (mais de 5.200 gigabytes para cada homem, mulher e criança em 2020). De agora até 2020, o universo digital dobrará a cada dois anos.

Tais informações evidenciam que a produção de dados é exorbitante pelo acesso massivo da Web. De acordo com Manyika et al. (2011, p.1), esse fenômeno é conceituado como Big Data, que se configura como “conjuntos de dados que estão além da

a Doutora em Ciência da Informação pela Universidade Federal da Paraíba – UFPB. Professora na Universidade Federal do Rio de Janeiro – UFRJ. E-mail: fernanda@facc.ufrj.br. ORCID: <https://orcid.org/0000-0002-7750-3269>.

b Graduando em Biblioteconomia e Gestão de Unidades de Informação pela Universidade Federal do Rio de Janeiro – UFRJ. E-mail: joaosena7@gmail.com.

capacidade de captura, armazenamento e análise das ferramentas de software típicas”.

Com o aumento exponencial de sua produção, os dados tornaram-se um ativo de suma importância para as organizações, que passaram a dar atenção ao tratamento dos dados, fundamentando a base para as tomadas de decisões. Portanto, surge a necessidade de profissionais qualificados para lidar com a consequência dos dados e das informações geradas, tornando cada vez mais necessário os profissionais capacitados(as) que tenham habilidades para realizar as atividades que esse novo paradigma exige. Esses profissionais são os(as) cientistas de dados, os(as) analistas de dados, *business intelligence* (inteligência de negócios), engenheiro(a) de dados, bibliotecário(a) de dados, cada um com sua especialidade, mas com sua devida e importante contribuição para as organizações.

No senso comum, a sociedade tem uma visão do(a) bibliotecário(a) de que somente atua com livros ou documentos físicos, isolado dentro de uma biblioteca organizando estantes de livros. Porém a profissão passou por grandes evoluções durante as décadas, rompendo com barreiras conservadoras e atualizando sua atuação em novos ambientes e com as novas tecnologias. A partir dessas atualizações, o(a) bibliotecário(a) adquire novas habilidades e competências para atuar além da forma tradicional e uma das novas formas de atuação é a biblioteconomia de dados.

De acordo com Semeler (2017), a Biblioteconomia de Dados é uma disciplina que aplica a Ciência da Informação, e-Science¹ e Data Science² em Unidades de Informação. Seguindo essa linha de raciocínio, o(a) bibliotecário(a) de dados, contribui diretamente na área de dados, tanto na Ciência de Dados, quanto na Gestão de Dados. Conforme Corrêa (2016, pág. 391) “a coleta, organização e preservação da informação tem sido confiada a bibliotecários(as) de diferentes setores que trabalham com os fluxos de informação; a articulação dos planos de gestão de dados é simplesmente uma manifestação moderna destas funções”. Diante disso, faz-se o seguinte questionamento: Como a Biblioteconomia e a Ciência da Informação estão discutindo aspectos relativos a (ao) bibliotecário(a) de dados?

A partir do questionamento feito anteriormente esta pesquisa tem como objetivo analisar a produção científica nacional sobre os aspectos concernentes à atuação do(a) bibliotecário(a) no contexto dos dados.

A relevância desta pesquisa decorre, em linhas gerais, da necessidade de compreender como os pesquisadores da área de Biblioteconomia e Ciência da Informação têm discutido a temática, considerando a sua relevância para a formação de futuros profissionais bibliotecários(as). Além disso, esta pesquisa também evidencia o status das pesquisas sobre o tema,

1 Compreendida como “a base para o pensamento das ciências orientadas ao uso intensivo dos dados” (SEMELER, 2017, p. 7).

2 “Conjunto de habilidades, métodos, técnicas e tecnologias da Estatística e da Ciência da Computação, utilizadas para extrair conhecimento de dados e para a criação de produtos e serviços a partir de dados” (SEMELER, 2017, p. 7).

elucidando os principais enfoques dos estudos sobre Biblioteconomia de dados.

2. Ciência da Informação e Ciência de Dados

Durante a Segunda Guerra Mundial um grande número de pessoas passou a trabalhar em processos de coleta, seleção, processamento e disseminação de informações que fossem relevantes para o esforço de ganhar a guerra (FREIRE, 2006, p.10). A partir desse fato, o conceito de Ciência da Informação (CI) surgiu após os problemas gerados pelo volume e pelo valor das informações produzidas durante e após o conflito. A Ciência da Informação está preocupada com o corpo de conhecimentos relacionados à origem, à coleção, à organização, ao armazenamento, à recuperação, à interpretação, à transmissão, à transformação e à utilização da informação (BORKO, 1968, p. 3).

De acordo com Saracevic (1996, p.42) a Ciência da Informação consiste em uma área interdisciplinar, estando inextricavelmente ligada à tecnologia da informação. Ainda de acordo com o autor, a Ciência da Informação é um campo dedicado às questões científicas e à prática profissional, voltadas para os problemas da efetiva comunicação do conhecimento e de seus registros entre os seres humanos, no contexto social, institucional ou individual do uso e das necessidades de informação.

A Ciência de Dados, por sua vez, consiste em um campo interdisciplinar que envolve três domínios de conhecimento: Programação de Computadores; Estatística e Matemática; e

Domínio do Conhecimento (RAUTENBERG; CARMO, 2019). Com esses domínios, os cientistas de dados devem saber extrair informação e conhecimento úteis a partir de grandes volumes de dados estruturados ou não, resolvendo problemas de negócio e melhorando a tomada de decisão. Esse profissional deve desenvolver diversas habilidades relacionadas aos domínios do conhecimento supracitados como, por exemplo, análises estatísticas, mineração de dados, uso de aprendizado de máquina, ferramentas de visualização de dados (criação de gráficos e painéis com *Tableau*, *Power BI*), habilidades em programação (*Python*, *R*, *SQL*) e uso de *frameworks* de *Big Data*, como *Hadoop* e *Apache Spark*. Suas principais responsabilidades são desenvolver, testar e manter algoritmos de aprendizado de máquina; realizar análise exploratória de dados detectando e tratando anomalias nos dados; fazer inferência e interpretação das análises e apresentar os resultados. Objetivando dar prosseguimento a essas discussões, a seguir serão apresentados aspectos concernentes à Biblioteconomia de Dados.

2.1 Biblioteconomia de dados

A biblioteconomia de dados é um termo que vem sendo discutido desde a década de 1960 na América do Norte e Europa, ou seja, não é um termo novo (SEMELER; PINTO, 2020). Segundo os autores, a biblioteconomia de dados tem ênfase na criação de serviços relacionados ao uso de dados digitais, repositórios de dados, curadoria de dados e gestão de dados nas bibliotecas, além de atividades ligadas ao gerenciamento e à

curadoria de todos os tipos de dados, sendo seu foco o tratamento, a gestão e a curadoria de dados de pesquisas em qualquer disciplina científica (SEMELER, 2017, p.65).

Uma segunda definição encontra-se no site da *Australian National Data Service* (ANDS), porém mais restritiva quanto à atuação do(a) bibliotecário(a) de dados, sendo “[...] profissionais bibliotecários engajados na gestão de dados de pesquisa, usando dados de pesquisa como recurso e dando suporte a pesquisadores nas suas atividades” (ANDS, 2018, tradução nossa). Também se ressalta que a biblioteconomia de dados é um campo interdisciplinar, sendo uma interseção entre Ciência da Informação, *e-Science* e Ciência de dados (SEMELER; PINTO, 2020).

Cabe ressaltar que a biblioteconomia de dados não deve ser somente restrita aos dados de pesquisas (provenientes de pesquisas científicas), diante do fato de que as organizações utilizam dados em seus modelos de negócio e atuação para geração de informação e tomadas de decisão. Portanto, o papel do(a) Bibliotecário(a) de Dados é essencial para que haja um uso efetivo da informação (neste caso, codificada em conjuntos de dados) pelas organizações (como as governamentais), no processo de busca por soluções de problemas (CALLONI, 2020).

Considerando a incipiência e atualidade do tema, bem como o fato de que a Biblioteconomia de Dados não deve ser restrita apenas aos dados de pesquisa, faz-se necessário compreender de que maneira os pesquisadores da área de Biblioteconomia e

Ciência da Informação têm discutido o tema na literatura nacional. Para tanto, serão utilizados os procedimentos metodológicos descritos a seguir.

3. Procedimentos metodológicos

Esta pesquisa é caracterizada como bibliográfica, exploratória, descritiva com abordagem qualitativa. Para atender o objetivo geral deste estudo, foi realizada a recuperação, análise e mapeamento da produção científica nacional sobre a temática a partir das pesquisas bibliográfica, exploratória, descritiva com abordagem qualitativa. Com isso, foram realizadas buscas na Base de Dados em Ciência da Informação (BRAPCI) e na Biblioteca Digital Brasileira de Teses e Dissertações (BDTD), no mês de maio de 2021, visando recuperar as produções científicas nacionais através do emprego dos seguintes descritores em ambas as fontes citadas anteriormente: “Biblioteconomia de Dados” e “Bibliotecário de Dados”.

A seleção das bases de dados BRAPCI e BDTD se deu pelo fato de que a primeira agrega as produções científicas publicadas em artigos de periódicos da área de Biblioteconomia e Ciência da Informação, atualmente disponibilizando 19.255 (dezenove mil duzentos e cinquenta e cinco) textos publicados em 57 (cinquenta e sete) periódicos nacionais impressos e eletrônicos. A segunda fonte, por sua vez, integra e dissemina, em um só portal de busca, os textos completos de Teses e Dissertações defendidas nas instituições brasileiras de ensino e pesquisa em diversas áreas do

conhecimento, dentre elas a Biblioteconomia e a Ciência da Informação.

Seguindo essa linha de raciocínio, o total de produções científicas recuperadas na BRAPCI mediante o emprego dos termos citados anteriormente foi no total de 105 (cento e cinco) artigos de periódicos. A maioria dessas produções científicas recuperadas (oitenta), apesar de apresentar pelo menos um dos termos pesquisados, não se articulavam com este estudo, tendo seus conteúdos dedicados à avaliação de bases de dados; estudos bibliométricos realizados em bases de dados; bem como mapeamento do uso de metadados. Desse total, apenas 25 (vinte e cinco) artigos de periódicos possuíam articulação direta com a Biblioteconomia de Dados e o(a) Bibliotecário(a) de Dados, essas produções foram consideradas nas análises desta pesquisa.

Em nítido contraste, na BDTD foram recuperadas apenas duas produções científicas, sendo uma Dissertação e uma Tese sobre a temática citada anteriormente. Nessa perspectiva, o corpus final para a realização da análise da produção científica nacional foi constituído de 27 (vinte e sete) trabalhos, sendo 25 (vinte e cinco) artigos de periódicos, uma Dissertação e uma Tese.

Como procedimento para a realização da análise dessas pesquisas utilizou-se o método da análise de conteúdo, possibilitando tornar compreensíveis os principais aspectos enfatizados nos estudos sobre a temática. Essa análise foi realizada manualmente através de uma leitura técnica baseada na norma ABNT NBR 12676, que trata especificamente sobre

métodos para análise de documentos e determinação dos seus assuntos. Para tanto, foram considerados, especialmente: título e subtítulo, resumo, palavras-chave, introdução, ilustrações, diagramas, tabelas e explicações, resultados, considerações finais e referências bibliográficas das produções científicas analisadas.

Com isso, foi possível identificar o status da pesquisa sobre o tema, bem como evidenciar os principais enfoques, tendências e perspectivas de estudos sobre o assunto, aspectos que são elucidados na seção seguinte.

4. Análise da produção científica nacional sobre Biblioteconomia de Dados

Diante das análises realizadas, verificou-se que as produções científicas nacionais sobre Bibliotecário de Dados estão distribuídas em um intervalo de tempo entre 2004 a 2020, havendo uma incidência de publicações do ano de 2016 em diante. Os anos de 2019 e 2020 ganham destaque como os períodos que apresentaram maior quantidade de publicações científicas, conforme é possível observar no Quadro 1.

Seguindo a linha de raciocínio discutida no referencial teórico desta pesquisa, acredita-se que a incidência de publicações nos últimos anos acerca da temática pode ser decorrente da necessidade emergente da atuação do(a) Bibliotecário(a) nesse cenário, haja vista que o seu perfil contempla diversas competências e habilidades que podem contribuir para o ciclo de vida dos dados, mediante aplicação de estratégias em diferentes contextos.

Quadro 1 – Quantidade de publicações por ano

Ano e publicação	Quantidade
2004	1
2012	1
2014	1
2015	1
2016	3
2017	2
2018	3
2019	8
2020	7

Fonte: Autores.

Os periódicos científicos eletrônicos e anais de eventos técnico-científicos se configuraram como fontes importantes para a comunicação científica, tornando disponível para a comunidade os resultados das pesquisas em diferentes áreas do conhecimento. Com base nos resultados da pesquisa realizada na BRAPCI, também foi possível perceber os periódicos da área de Ciência da Informação cujos trabalhos analisados encontram-se publicados, são eles: Múltiplos olhares em Ciência da Informação; Informação e Informação; Ciência da Informação; Em Questão; Biblos; Revista ACB: Biblioteconomia em Santa Catarina; Recis: Revista Eletrônica de Comunicação, Informação e Inovação em Saúde; Revista Digital de Biblioteconomia e Ciência da Informação (RDBCI); Perspectivas em Ciência da Informação; Encontros Bibli; Informação em Pauta; AtoZ: novas práticas em Informação e Conhecimento; Anais do Encontro Regional de

Estudantes de Biblioteconomia, Documentação, Gestão e Ciência da Informação (2018). O Quadro 2 apresenta essas informações de maneira sintetizada e organizada com base na incidência das produções.

Quadro 2 – Quantidade de produções por canais de publicação

Fontes das publicações	Quantidade de produções
Informação e Informação	6
Ciência da Informação	4
Em Questão	3
Reciis- Rev. Eletr. Comun. Inf. Inov. Saúde	2
RDBCI: Rev. Dig. Bibliotecon. Cienc. Inf.	2
Múltiplos olhares em Ciência da Informação	1
Biblos	1
Revista ACB: Biblioteconomia em Santa Catarina	1
EREBD 2018	1
Perspectivas em Ciência da Informação	1
Encontros Bibli	1
Inf. Pauta	1
AtoZ	1

Fonte: Autores.

Observa-se que o periódico que possui maior número de publicações sobre Bibliotecário de Dados é Informação e Informação, periódico científico eletrônico do Programa de Pós-Graduação em Ciência da Informação da Universidade Estadual de Londrina (PPGCI/UEL) com um total de seis publicações acerca da temática. Em seguida, a revista Ciência da Informação,

periódico científico do Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT) aparece no quadro com um destaque de três publicações sobre a temática. O periódico “Em Questão” (PPGCI/UFRGS), por sua vez, apresenta o total de três publicações sobre Bibliotecário de Dados e, em seguida, aparecem as demais revistas citadas anteriormente.

Verificou-se ainda que a maioria das produções científicas recuperadas e analisadas concentram-se, em linhas gerais, em investigar questões concernentes à gestão de dados de pesquisa. Embora tenha sido mencionado anteriormente que a Biblioteconomia de dados não deve ser restrita apenas aos dados de pesquisas (provenientes de pesquisas científicas), constatou-se que os pesquisadores da área de Ciência da Informação têm dedicado os seus estudos a esse enfoque temático.

Além das produções científicas disponíveis em periódicos científicos eletrônicos e anais de eventos técnico-científicos, durante a busca e recuperação das produções científicas na BDTD foi possível identificar dois trabalhos provenientes de Programas de Pós-Graduação em Ciência da Informação de duas Universidades do Brasil, são elas: Universidade Federal de Santa Catarina (UFSC) e Universidade Federal de Sergipe (UFS).

O Quadro 3 apresenta os enfoques temáticos das publicações e os respectivos autores cujos trabalhos foram analisados mediante a realização da análise de conteúdo. Em linhas gerais, é possível visualizar a diversidade de enfoques temáticos nas produções científicas analisadas, uma vez que

foram identificadas 12 (doze) temas nas pesquisas analisadas. Percebe-se que a maioria das publicações nacionais (22%) concentram-se em investigar aspectos concernentes à gestão de dados de pesquisa, sobretudo no que diz respeito às atividades do(a) Bibliotecário(a) e as competências necessárias para a sua atuação nesse campo. Essas pesquisas ressaltam a importância de esse(a) profissional atuar desde as fases que antecedem a pesquisa até a pesquisa finalizada, preocupando-se com a organização, tratamento, gestão e preservação dos dados e com os direitos do acesso, uso e local de armazenamento.

Quadro 3 – Enfoques temáticos das produções científicas analisadas

Enfoques temáticos	Autores
Gestão de Dados de Pesquisa	Santa Anna; Dias; Maculan (2019) Carvalho e Queiroz (2019) Sales et al. (2019) Lima; Pinto e Farias (2019) Sales e Sayão (2012) Corrêa (2016) Resende e Bax (2020)
Repositórios Digitais de Dados de Pesquisa	Sayão; Sales (2016) Tartarotti; Dal'Evedove; Fujita (2019) Paganine; Amaro (2020)
Práticas de gerenciamento e publicação de dados no contexto da Web Semântica com dados abertos interligados	Silveira (2020) Guandaline e Santos (2018)
Proteção e privacidade de dados em Unidades de Informação	Vignoli et al. (2020) Serra (2015)

Enfoques temáticos	Autores
Governança de Dados	Espíndola et al. (2018)
Dados de pesquisa no contexto da Biblioteconomia de Dados	Semeler e Pinto (2019)
Tratamento de dados terminológicos	Xavier et al. (2004)
Mineração de dados como estratégia para a composição dos catálogos de periódicos <i>open access</i>	Silva; Queiroz e Simões (2019)
Mineração de dados científicos	Ferreira e Corrêa (2020)
Modelos de busca, acesso, recuperação e apropriação da informação na Web de Dados	Palleta e Gonzales (2019)
A atuação do bibliotecário no tratamento dos dados provenientes da <i>e-Science</i>	Costa e Cunha (2014)
<i>e-Science e Data Science</i> no contexto da Ciência orientada a dados	Bufrem et al. (2016)

Fonte: Autores.

Em consonância de evidências, 13% das produções científicas recuperadas tratam sobre repositórios de dados de pesquisa, os quais são considerados como sistemas de recuperação da informação contemporâneos voltados para apoiar o armazenamento, tratamento, recuperação, acesso e uso dos dados de pesquisa.

De outro modo disposto, 9% das produções científicas recuperadas tratam sobre as práticas de gerenciamento e publicação de dados na Web semântica com dados abertos interligados, perspectivas que devem ser consideradas nos

sistemas contemporâneos. Por outro lado, 9% dos trabalhos recuperados preocupam-se em investigar questões relacionadas à proteção e privacidade de dados em unidades de informação.

Os demais trabalhos recuperados e analisados possuem os seguintes enfoques: dados de pesquisa no contexto da Biblioteconomia de dados; tratamento de dados terminológicos; mineração de dados como estratégia para a composição de catálogos de periódicos *Open Access*; mineração de dados científicos; modelos de busca, recuperação; acesso e apropriação da informação na web de dados; a atuação do Bibliotecário(a) no tratamento dos dados provenientes da *e-Science*; *e-Science* e *Data Science* no contexto da Ciência orientada a dados.

Diante do exposto, é possível observar que as produções científicas evidenciam a importância da atuação do(a) Bibliotecário(a), além do contexto dos dados de pesquisa, nas práticas de governança, gerenciamento, mineração, tratamento, organização, busca, acesso, uso, reúso e ligação dos dados, sobretudo no contexto da Web Semântica.

5. Considerações finais

Ao analisarmos as produções recuperadas, foi possível perceber o status das pesquisas nacionais sobre Bibliotecário de Dados. Constatou-se um grande foco nos últimos anos com relação às pesquisas voltadas para a gestão de dados de pesquisa e repositórios digitais de dados de pesquisa; bem como para as práticas de gerenciamento e publicação de dados no contexto da Web Semântica com dados abertos interligados.

Nesse sentido, constata-se a necessidade de investigação e apresentação das competências e habilidades do(a) Bibliotecário para o desenvolvimento de práticas concernentes à alfabetização de dados; bem como a organização, tratamento e recuperação de dados, sobretudo no contexto da Web semântica e dos Sistemas híbridos de Organização do Conhecimento. Esses enfoques podem ser trabalhados de maneira verticalizada por profissionais e pesquisadores da área em pesquisas futuras.

Referências

BORKO, H. Information Science: What is it? **American Documentation**, v.19, n.1, p.3-5, Jan. 1968.

CALLONI, R. Bibliotecário de Dados. In: Silva, F.C.C. **O perfil das novas competências na atuação bibliotecária**. Florianópolis, SC: Rocha Gráfica e Editora, 2020. p. 339-362.

CORRÊA, F. C. O papel do bibliotecário na gestão de dados científicos. **Revista Digital de Biblioteconomia e Ciência da Informação**, Campinas, v.14, n.3, p. 387-406, 2016.

FREIRE, G. Ciência da informação: temática, história e fundamentos. **Perspectivas em Ciência da Informação**, Belo Horizonte, v.11, n.11, p. 6-19, 2006.

MANYICA, J.; CHUI, M.; BROWN, B.; BUGHIN, J.; DOBBS, R.; ROXBURGH, C.; BYERS, A. H. **Big data**: The next frontier for innovation, competition, and productivity, 2011.

RAUTENBERG, S.; CARMO, P. R. Big Data e Ciência de Dados: complementariedade conceitual no processo de tomada de decisão. **BJIS: Research Trends**. v.1, n.13, p.56-67, 2019.

SARACEVIC, T. Ciência da Informação: origem, evolução e relações. **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 1, n. 1, p. 41-62, jan./jul. 1996.

SEMELER, A. R. Ciência da Informação em contextos de e-Science: Bibliotecários de Dados em tempos de Data Science. 2017. **Tese** (Doutorado em Ciência da Informação) - Universidade Federal de Santa Catarina, 2017.

SEMELER, A. R.; PINTO, A. L. Data librarianship as a field study. **Transinformação**, Campinas, v. 32, out. 2020.

Os dados científicos e os dados de pesquisa na Ciência da Informação Brasileira: uma análise a partir da BRAPCI

Ananda Fernanda de Jesus^a, William Pires de Castro^b e José Eduardo Santarém Segundó^c

1. Introdução

Como resultado dos avanços científicos e tecnológicos no início do século XX, ocorreu uma revolução que influenciou o desenvolvimento de novos suportes informacionais e o aumento da produção tecnocientífica com o estabelecimento das publicações seriadas. (SARACEVIC, 2009).

Esse aumento alterou a relação entre pesquisadores e informações pertinentes para o desenvolvimento de suas pesquisas, tornando inviável para muitos campos do conhecimento a obtenção e tratamento de todo material publicado a respeito das temáticas de interesse.

Nesse cenário, agravado com o estabelecimento do ambiente Web, os dados adquirem um papel de destaque no desenvolvimento das pesquisas, seja influenciando nos processos

-
- a Doutoranda em Ciência da Informação pela Universidade Estadual Paulista – UNESP. E-mail: af.jesus@unesp.br.
 - b Graduando em Ciência da Computação pelo Centro Universitário Eurípedes de Marília – UNIVEM. E-mail: wpdcastro@gmail.com.
 - c Doutor em Ciência da Informação pela Universidade Estadual Paulista – UNESP. Professor na Universidade de São Paulo – USP. E-mail: santarem@usp.br. ORCID: <https://orcid.org/0000-0003-3360-7872>.

de recuperação e seleção de referências, seja atuando como insumo no desenvolvimento das pesquisas. Com base no cenário apresentado, questiona-se: Como os dados científicos e os dados de pesquisa têm sido abordados nos artigos brasileiros da ciência da informação?

Neste sentido, objetiva-se discutir como a comunidade brasileira da ciência da informação tem abordado os dados científicos e os dados de pesquisa, buscando compreender a relação entre os dois conceitos, definições conceituais, identificando as principais temáticas abordadas e as lacunas de pesquisa.

Como método optou-se por um mapeamento sistemático dos artigos nacionais publicados a respeito da temática. Na próxima seção são apresentados os procedimentos metodológicos adotados para o desenvolvimento da pesquisa.

2. Procedimentos metodológicos

Caracterizada como exploratória, descritiva e qualitativa, essa pesquisa adotou como método de coleta o Mapeamento Sistemático da Literatura (MSL). Segundo Kitchenham et al. (2010), os MSL são realizados com uma abordagem ampla, sendo o principal objetivo classificar os estudos de uma temática específica em categorias bem definidas, que permitam um olhar estruturado para a temática em busca de assuntos comuns e assuntos pouco explorados.

Para orientar as buscas, a seleção dos artigos e a análise dos resultados foi elaborado o seguinte protocolo de busca ilustrado no Quadro 1.

Quadro 1 – Protocolo de busca do mapeamento sistemático

Elemento	Protocolo de busca
Bases de dados	BRAPCI
Idiomas	Português
Crítérios de Inclusão	(I) Discute dados científicos
Crítérios de exclusão	(E) Não está em português; (E) Não está no formato estabelecido (artigo); “(E) Apenas mencionam a temática de interesse; (E) Apenas mencionam a temática de interesse (E) Não é possível ter acesso ao artigo completo (E) Aborda o uso de dados para validação de resultados
Categorias de análise	(1) Enfoque do artigo (2) Definições de dados científicos (3) Definição de dados de pesquisa (4) Lacunas de pesquisa
Data da coleta	08/05/2021
Forma de análise dos resultados	Os documentos foram selecionados de acordo com os critérios estabelecidos, sendo divididos os resultados entre os autores responsáveis pela coleta dos dados, em caso de dúvida quanto a seleção do artigo esse foi discutido em conjunto entre os autores. A coleta nos documentos aceitos foi realizada seguindo as categorias de análise apresentadas. Foi elaborado um quadro resumo contendo autor, título, ano da publicação, e os resultados de cada categoria. Os resultados foram categorizados, sendo identificados padrões de temáticas de nos documentos, levantadas as lacunas e agrupadas as definições.

Fonte: Autores.

Para a realização do mapeamento sistemático seguiu-se as seguintes etapas: (i) Planejamento: Preenchimento do protocolo de busca que orientará a pesquisa; (ii) Execução: Busca nas bases de dados e seleção dos documentos; e (iii) Sumarização: agrupamento dos documentos por semelhança, criação de

categorias para classificação dos resultados, sistematização das informações de interesse em imagens e quadros.

Após a seleção os documentos foram classificados em categorias temáticas. Essas categorias foram estabelecidas a posteriori, levando em consideração principalmente os objetivos e os problemas de pesquisa apresentados pelos autores dos documentos, também foram checados os resultados obtidos como forma de verificar se os objetivos eram condizentes com os resultados apresentados, visando assim a inclusão dos documentos em categorias representativas. A nomenclatura aplicada na criação das categorias segue terminologia adotada pelos autores. Após estabelecimento das categorias, buscou-se realizar o agrupamento de categorias semelhantes, sempre levando em consideração a nomenclatura utilizada pelos próprios autores.

3. Resultados e discussões

As buscas na BRAPCI retornaram 392 documentos, sendo aceitos 51 para a análise. Ao observar os documentos rejeitados é possível identificar os fatores que levam ao alto número de rejeição. Os documentos rejeitados com base no critério “(E) aborda o uso de dados para validação de resultados” mencionam o uso de bases de dados ou conjuntos de dados científicos que atuaram como insumo para o desenvolvimento da pesquisa, como a identificação de bases de dados úteis a determinadas áreas, ou o levantamento de um conjunto de dados para solucionar uma questão de pesquisa. Esses documentos foram

recusados pois não abordaram ou discutiram o conceito ou a importância dos dados científicos ou dos dados de pesquisa, a sua necessidade de registro e possibilidade de reutilização, sendo esses dados utilizados de maneira mais operacional.

Já os documentos classificados no critério “(E) apenas mencionam a temática de interesse” reúnem artigos que citaram os dados científicos como um entre outros temas de interesse da CI, ou de alguma forma citaram os dados científicos para fins de exemplo ou contextualização, mas não se aprofundam na discussão da temática. A partir da análise dos documentos aceitos foram estabelecidas 15 categorias que refletem como o tema de dados pesquisa tem sido abordado nos artigos científicos da ciência da informação no Brasil. O quadro 2 apresenta essas categorias e o número de documentos que discutiram a temática, podendo um documento ser enquadrado em mais de uma categoria.

Quadro 2 – Categorias temáticas

Categoria	Quantidade de artigos	Definição
Gestão de dados de pesquisa	19	Documentos que buscam levantar as etapas, selecionar e aplicar ciclos de vida de dado de pesquisa, discutir aspectos de produção, compartilhamento e reutilização, de dados de pesquisa. Inclui ainda artigos que discutem os impactos no avanço da ciência, governança de dados, relação entre gestão de dados e fundamentação teórica da ciência da informação.
Curadoria de dados	9	Discutem a concepção, elaboração,

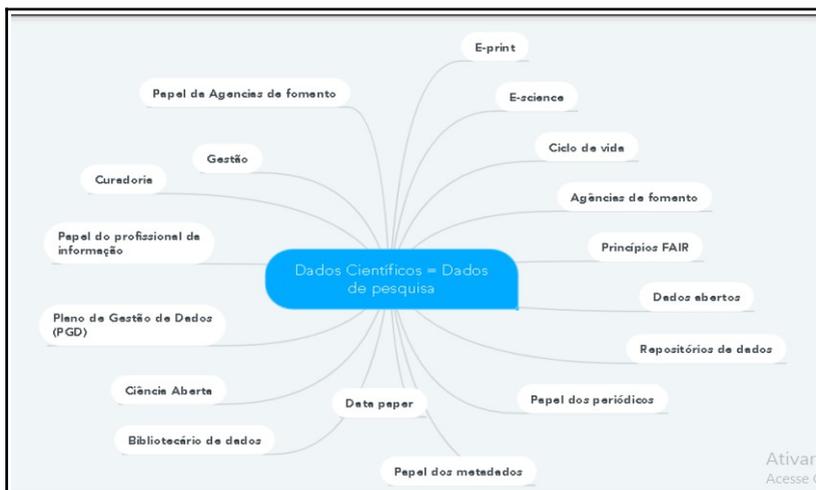
Categoria	Quantidade de artigos	Definição
de pesquisa		preservação e supervisão de dados científicos, sendo em ambientes físicos ou digitais.
Boas práticas, políticas e princípios para a gestão de dados de pesquisa	4	Pesquisas que abordam a necessidade de estabelecimento de princípios, boas práticas e políticas para a gestão de dados, que fazem o levantamento de documentos como esses disponíveis e os benefícios da aplicação de determinado instrumento, como por exemplo os princípios FAIR, Linked Data, o movimento de dados abertos.
Percepção e comportamento dos pesquisadores acerca de dados de pesquisa	4	Documentos que buscam analisar o comportamento dos pesquisadores na gestão e disponibilização dos dados derivados de suas pesquisas, bem como a utilização de dados elaborados por outros pesquisadores.
Elaboração do Plano de Gestão de Dados de Pesquisa (PGD)	1	Pesquisas teóricas acerca da importância do PGD, bem como pesquisas mais aplicadas, discutindo a elaboração e aplicação do plano.
Papel dos profissionais da informação na gestão de dados de pesquisa	3	Documentos que buscam identificar como bibliotecários, arquivistas e cientistas da informação podem auxiliar no processo de gestão de dados de pesquisa e quais as competências que precisam ser aprendidas para que esses profissionais possam atuar de maneira eficiente nesse mercado.
Armazenamento dos dados de pesquisa /repositório de dados	2	Documentos que discutem a importância e os desafios do estabelecimento de repositórios, que buscam mapear repositórios úteis para determinado domínio, que analisa a viabilidade e

Categoria	Quantidade de artigos	Definição
		necessidade de criação de um repositório institucional de dados de pesquisa.
Influências externas na disponibilização de dados	2	Documentos que buscam relacionar o impacto que influências externas, tais como buscadores, agências de fomento, periódicos científicos e instituições de ensino e pesquisa, têm sobre as práticas de gestão e
Relação de dados de pesquisa com e-Science	2	Documentos que abordam o papel dos dados de pesquisa no contexto da e-Science.
Proteção autoral no reúso de dados de pesquisa	2	Documento que propõe soluções para garantir que a autoria dos conjuntos de dados de pesquisa seja respeitada.
Elaboração de dados de pesquisa	2	Documento que aborda a criação de dados de pesquisa
Papel dos metadados na recuperação dos dados científicos	1	Documento que discute a importância da adoção de metadados consistentes e estruturados para a recuperação dos dados de pesquisa.

Fonte: Autores.

Além dos temas centrais tratados nos documentos, buscou-se mapear assuntos que aparecem com frequência nas discussões, embora não sejam o objetivo central dos documentos. Os temas mapeados são apresentados na Figura 1.

Figura 1 – Temas correlatos



Fonte: Autores.

A primeira observação traçada em relação à figura diz respeito aos termos “dados científicos” e “dados de pesquisa”. Essa pesquisa partiu da hipótese de que “dados científicos” constituiria um termo mais amplo, incluindo os dados de pesquisa, mas também os metadados de publicações científicas, dados dos pesquisadores (como os disponíveis no currículo Lattes), e indicadores de produção. Entretanto, a análise dos artigos nacionais não confirmou essa hipótese. Na literatura levantada o termo “dados científicos” é abordado como sinônimo de “dados de pesquisa”.

Em relação aos temas que circundam a discussão de dados de pesquisa, um grande destaque é o conceito de e-Science, caracterizando um novo paradigma do desenvolvimento de

pesquisas, baseado no uso de tecnologias da informação e comunicação e na preservação, curadoria e compartilhamento de dados de pesquisa (ANJOS, DIAS, 2019; GARU, 2005 *apud* SALES et al., 2013).

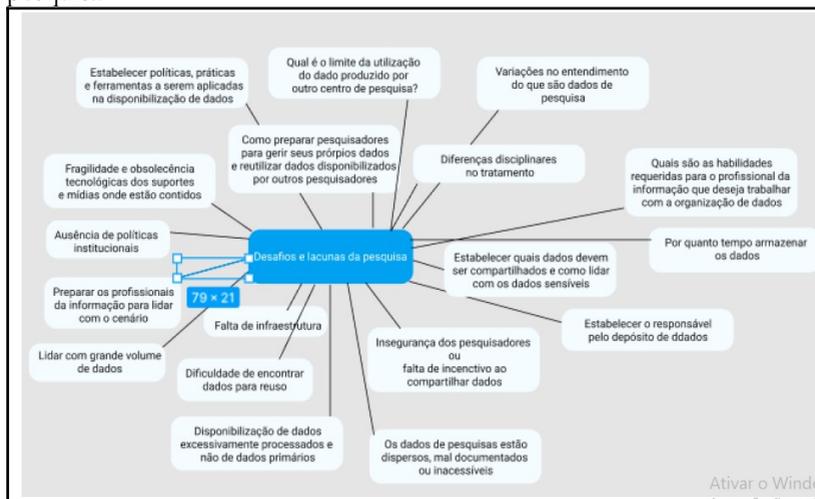
Muitos documentos trazem a preocupação com a curadoria, gestão, estabelecimento dos metadados, escolha da base de dados e disponibilização, aparece de maneira secundária se preocupando piamente em estabelecer o ciclo de vida dos dados como norte.

Sant’Ana (2016, p.119) define o ciclo de vida dos dados como “uma delimitação de fases (momentos em que distintas necessidades e competências são necessárias) envolvidas no acesso e uso dos dados”, o autor descreve as fases como: processo de coleta, recuperação, armazenamento e descarte.

Muito se discute nesses documentos acerca das influências externas, o papel das agências de fomento, das instituições de ensino e pesquisa, dos periódicos científicos no incentivo formal ao compartilhamento de dados de pesquisa, no estabelecimento de políticas e de diretrizes que apoiem o compartilhamento dos dados. Também se destaca nos documentos a preocupação com a ciência aberta e com os movimentos de disponibilização aberta dos dados. Existe um destaque para o entendimento de que quando produzidos com financiamento público, esses dados constituem um bem público que deve ser organizado e disponibilizado de maneira estruturada, seguindo licenças de acesso aberto, preferencialmente a partir de formatos não proprietários.

Além dos temas correlatos foram levantados as dificuldades e lacunas apontadas pelos documentos a respeito da temática, apresentadas na Figura 2.

Figura 2 – Lacunas e desafios relacionados à temática de dados de pesquisa



Fonte: Autores.

Nota-se que a atuação dos pesquisadores na gestão dos dados é uma grande preocupação quando se discute dados de pesquisa. É possível observar, nos artigos incluídos na categoria “percepção e comportamento dos pesquisadores acerca de dados de pesquisa” que os pesquisadores entendem a importância de disponibilizar esses dados, mas, na prática, seguem não compartilhando-os. Alguns dos motivos para o não compartilhamento desses dados aparecem como lacunas de

pesquisa, tais como as preocupações como perder a vantagem competitiva e os direitos autorais desses dados e ainda a falta de conhecimento acerca do procedimento de disponibilização.

Essa dificuldade tem início logo na definição de dados de pesquisa, que varia entre as áreas do conhecimento e entre os próprios pesquisadores, existindo uma tendência de entender como dados de pesquisa apenas os dados que são excessivamente processados e tratados. Também existe uma preocupação sobre o processo de disponibilização desses dados, levando a questões como: Quais dados devem ser disponibilizados? Como lidar com dados sensíveis? Por quanto tempo armazenar esses dados? Onde disponibilizar? Como descrever esses dados?

Ainda em relação ao papel dos pesquisadores, também se identifica uma preocupação em como preparar esses pesquisadores para reutilizar esses dados, com dúvidas como: onde encontrar dados de interesse? Como selecionar esses dados? Como lidar com grandes volumes de dados? Qual o limite para a reutilização de dados de outros pesquisadores no estabelecimento de uma nova pesquisa?

Também se destaca entre os desafios a falta de infraestrutura e de diretrizes para auxiliá-los nesse processo, a gestão dos dados de pesquisa demanda muita atenção e tempo dos pesquisadores, que esperam que as instituições de pesquisa providenciem a infraestrutura pessoal e tecnológica necessária para apoiar esse processo. Também se espera que essas instituições disponibilizem diretrizes para embasar a gestão dos dados de pesquisa. Nota-se

que algumas instituições têm se movimentado para discutir e estabelecer essas diretrizes, mas que elas em sua maioria ainda não estão estabelecidas e divulgadas.

4. Considerações finais

Essa pesquisa teve por objetivo discutir como a comunidade brasileira da ciência da informação tem abordado os dados científicos e os dados de pesquisa. Nesse sentido conclui-se que o termo dados científico tem sido abordado como um sinônimo de dados de pesquisa, e que a definição de dados de pesquisa é ampla, exigindo mais discussões.

Em relação a como a temática tem sido discutida, conclui-se que abordagens voltadas para gestão, curadoria e recuperação dos dados são assuntos de destaque, sendo abordados em grande parte dos documentos. É possível observar que entre as duas categorias de maior destaque, sendo elas “Gestão de dados de pesquisa” e “Curadoria de dados de pesquisa” existem muitas semelhanças, não sendo clara a diferenciação entre as duas abordagens. Optou-se por respeitar a nomenclatura adotada pelos autores para definir seus trabalhos, entretanto estudos futuros são necessários para discorrer sobre a relação entre os dois conceitos.

Outrora se sofria da escassez de dados para pesquisas, hoje o problema é o contrário, quantidades exorbitantes de dados espalhados de formas diversas e sem uma política de gestão de dados predefinida, fato esse que pode também ser responsável pela falta de compartilhamento de dados gerados pelos próprios

pesquisadores, criando mais um gargalo que é a recuperação e seleção de dados já compartilhados.

Todos os documentos que trazem os desafios de se lidar com dados científicos abordam em essência a necessidade de se preparar o pesquisador para disponibilizar seus dados e seguir com a utilização de dados de terceiros. Essa falta de compartilhamento gera gargalos nas provas de trabalhos e em continuidades dos mesmos, podendo gerar assim inconsistências veladas, trazendo lentidão ao avanço científico em uma área.

Para pesquisas futuras, propõe-se um aprofundamento na discussão acerca do conceito de dados de científicos, analisando como incluir os demais dados provenientes do processo de comunicação científica, tais como os metadados descritivos, os dados dos pesquisadores, os indicadores de produção, levanto o papel e a importância exercido por esses dados como insumos para a compreensão do cenário científico. Propõe-se ainda: o aprofundamento acerca do conceito de dados de pesquisa, a ampliação do levantamento realizado, incluindo trabalho para eventos e monografia, para um panorama mais completo do cenário nacional e a ampliação para o cenário internacional, realizando uma comparação entre o desenvolvimento da temática na ciência da informação no Brasil e no exterior.

Nesse sentido conclui-se que ainda existem muitos aspectos a serem explorados a respeito da relação entre dados científicos e dados de pesquisa, e que esse trabalho pode atuar como um ponto de partida para o desenvolvimento de novas pesquisas.

Referências

KITCHENHAM, B. et al. Systematic literature reviews in software engineering: a tertiary study. **Information and Software Technology**, [S.l.], v. 52, p. 792–805, 2010.

SARACEVIC, T. Information science. **Encyclopedia of Library and Information Sciences**, [S.l.: s.v.: s.n.], p. 2570-2585, 17 dez. 2009. Disponível em: <https://tefkos.comminfo.rutgers.edu/SaracevicInformationScienceELIS2009.pdf>. Acesso em: 11 maio. 2021

ANJOS, R. L. D.; DIAS, G. A. Atuação dos profissionais da informação no ciclo de vida dos dados – dataone: um estudo comparado. **Informação & Informação**, v. 24, n. 1, p. 80-101, 2019. Disponível em: <http://www.uel.br/revistas/uel/index.php/informacao/article/view/31085>. Acesso em: 17 maio. 2021.

SANT’ANA, R. C. B. Ciclo de vida dos dados: uma perspectiva a partir da ciência da informação. **Informação & Informação**, [S.L.], v. 21, n. 2, p. 116, 20 dez. 2016. Disponível em: <http://www.uel.br/revistas/uel/index.php/informacao/article/view/27940/20124>. Acesso em: 18 maio 2021.

SALES, L. F.; SAYÃO, L. F.; SOUZA, R. F. Publicações ampliadas: um novo modelo de publicação acadêmica para o ambiente de e-Science. In: Encontro Nacional de Pesquisa

em Ciência da Informação (ENANCIB), 14., 2013,
Florianópolis. **Anais...** Florianópolis: ENANCIB, 2013.
Disponível em:
[http://eprints.rclis.org/20666/1/APRESENTA
%C3%87%C3%83O%20-%20ENANCIB%202013.pdf](http://eprints.rclis.org/20666/1/APRESENTA%C3%87%C3%83O%20-%20ENANCIB%202013.pdf). Acesso
em: 23 jun. 2017.

What we have to do for Responsible Data Science?

*Morgana Carneiro de Andrade^a, Paula Regina Ventura Amorim
Gonzalez^b, Decio Wey Berti Junior^c, Ana Alice Rodrigues Pereira Baptista^d
and Caio Saraina Coneglian^e*

1. Introduction

Increasing data volume has changed the way research, governance, socialization, and business are being conducted (HILBERT; LÓPEZ, 2011, KEMPER; KOLKMAN, 2018). This data is generated by different segments of society, government, industry, research institutes, and universities. How they are collected, stored, and disseminated is also carried out in different ways and by different bodies (EUROPEAN DATA SCIENCE ACADEMY, 2019; STOYANOVICH; HOWE, 2018; TAYLOR, 2017b). Investments, whether financial or human resources have

-
- a Doutora em Tecnologias e Sistemas de Informação pela Universidade do Minho – UMINHO. Bibliotecária e Documentalista na Universidade Federal do Espírito Santo – UFES. E-mail: morganaandrade@gmail.com.
 - b Doutora em Ciência da Informação pela Universidade Estadual Paulista – UNESP. Professora na Universidade Federal do Espírito Santo – UFES. E-mail: paula.gonzalez@ufes.br. ORCID: <https://orcid.org/0000-0002-5480-4106>.
 - c Doutor em Gestão e Organização do Conhecimento pela Universidade Federal de Minas Gerais – UFMG. Analista de Sistemas na Universidade Estadual de Londrina – UEL. E-mail: decio@uel.br. ORCID: <https://orcid.org/0000-0003-4313-2727>.
 - d Doutora em Engenharia de Sistemas de Informação pela Universidade do Minho – UMINHO. Professora na Universidade do Minho – UMINHO. E-mail: analice@dsi.uminho.pt. ORCID: <https://orcid.org/0000-0003-3525-0619>.
 - e Doutor em Ciência da Informação pela Universidade Estadual Paulista – UNESP. Professor no Centro Universitário Eurípides de Marília – UNIVEM. E-mail: caio.coneglian@gmail.com. ORCID: <https://orcid.org/0000-0002-6126-9113>.

been made in data science, artificial intelligence, and data technologies. In parallel to this, there is a concern shared by society, governments, and researchers regarding the responsible use of data.

For Kemper & Kolkman (2018, p. 5),

Data science can only be effective if people trust the results and can correctly infer and interpret the outcomes. Data Science should thus not be viewed as a black box that magically transforms data into value. Many design choices need to be made in a typical 'data science pipeline' [...].

Recent examples of data collection, treatment, and use have been widely reported: the use of digital and bibliometric record data by India and China to track and monitor citizens (TAYLOR, 2017b); the collection of Facebook personal data used for political purposes without users' consent by Cambridge Analytica (FACEBOOK, 2019). Actions such as these warn against the use and publication of data without respect for privacy or leading to unfair and/or biased conclusions from different citizens (AIMS, 2017; OHM, 2014; TAYLOR, 2017b).

The current landscape shows that companies make profits, governments influence decision making concerning their citizens and people are induced into concepts based on data that are not imbued with the principles of ethics and justice (STOYANOVICH; HOWE, 2018; TAYLOR, 2017b).

Amid this problem arises, as strong demand, the application of individual and corporate rights and rules that safeguard the

privacy of the individual and the transparency of information (MOEREL; PRINS, 2016). In this sense, the Responsible Data Science (RDS) project arises, presented as a possibility to address issues regarding impartiality, accuracy, reliability, and transparency, regarding Big Data and Data Science. These four principles are represented by the acronym FACT (Fairness, Accuracy, Confidentiality, Transparency) (VAN DER AALST, 2016).

RDS seeks to develop methods and techniques to support the publication and use of data following the FACT, to provide technology to ensure these principles at the development stage, and to create multidisciplinary platforms. The adoption of FACT, still in the conception phase of the technologies and methods applied to Big Data and Data Science, is an alternative to guarantee the quality and security of data use (STOYANOVICH; HOWE, 2018; VAN DER AALST, 2016).

In view of the challenges of the fair and sustainable incorporation of ethical values into individual rights, the technological innovations involving Data Science and the potential benefits of the effort to meet new global challenges that the whole of humanity is engaged in, this study analyses the FACT guidelines.

This work is descriptive research in which the bibliographic search was performed from the Scopus and Web of Science databases and with the Google Scholar search engine. The documentary research was developed on websites of organizations that address the present theme: Responsible Data

Science, Data Science Center Eindhoven, and GE 301-Group 7. The searches were conducted using the terms “Responsible Data Science”, “Fairness, Accuracy, Confidentiality, Transparency + Data Science”, FACT + Data Science; FAIR + FACT.

This article is structured in four sections: Introduction, Section 1, where the theme is presented; Section 2, which contains a literature review around Responsible Data Science and the FACT; The discussions in Section 4; and, in the end, the Final Considerations and References.

2. Literature review

For Mayo (1996) and Chalmers (2013), cited by Van der Aalst (2016), in the context of the Web, there is a change in the way knowledge is generated. When knowledge is based on data, [...] follows the logic of the new experimentalism, in which knowledge is derived from experimental observations, not theory. In this new paradigm, concerns arise about data, especially how data may be used irresponsibly.

A group of researchers from different institutions in the Netherlands has started a new precept: RDS which, according to van der Aalst (2016), aims to produce positive solutions rather than avoiding the use of data. The RDS consortium was launched in 2016 with the aim of “[...] tackling ethical and legal challenges, and developing data science techniques, infrastructures, and approaches that are responsible for fair, accurate processes and confidential, design-transparent data” (AIMS, 2017; MOEREL; PRINS, 2016).

This initiative involves disciplines such as data mining, digital humanities, ethics, information retrieval, knowledge representation, law, machine learning, natural language processing, security, statistics, and visualization (VAN DER AALST, 2016; VAN DER AALST; BICHLER; HEINZL, 2017) and can be applied to the media, safety and health domains.

For the European Academy of Data Science³, “The ‘Data Age’ produces exponentially diverse data whose challenges relate to the collection, storage, and analysis of these data”. In this environment, powerful techniques such as data mining, machine learning, database, and visualization have been developed that aim to improve people's lives by offering new services and products with higher quality and efficiency (GE 301 GROUP 7, 2017; STOYANOVICH et al., 2017; STOYANOVICH; HOWE; JAGADISH, 2018).

In Piersman's (2018) view, RDS “[...] is one of the foundations of digital transformation” and can contribute to the enforcement of the United Nations Sustainable Development Goals (PAPPALARDO; GROSSI; PEDRESSHI, 2021). Consequently, there is a need to adopt guidelines towards Responsible Data Science. According to Pappalardo, Grossi e Pedreschi (2021, p. 261), “Social Mining and Big Data Ecosystems for Open, Responsible Data Science are necessary, which not only strengthen the existing clusters of excellence in social data mining research but also strive at creating global,

3 EDSA Project. See: <http://edsa-project.eu/overview/aboutedsa/>. Access: 20 oct. 2021

interdisciplinary communities of social data scientists, fostered by extensive training, networking, and innovation activities⁴.

Although initiatives such as the mentioned above are encouraging, however, reality shows that automatic data decisions may be unfair and not transparent; confidential data may be unintentionally shared or misused by a third party; inaccuracies can be generated from the incorporation of biases by algorithms, among other problems (OHM, 2014; STOYANOVICH & HOWE, 2018). These facts cause great concern regarding the production and use of data, namely regarding false information, biased conclusions, disclosure of private and non-transparent information, which may hinder the application of scientific data (AIMS, 2017), for example, or compromise the safety of people and the personal and professional life of the individual (GE 301 GROUP 7, 2017; OHM, 2014; TAYLOR, 2017; TAYLOR; PURTOVA, 2019).

Richard and King (2014), referenced by Lodder (2016), professor at Vrije Universiteit, Amsterdam, considers that the construction of this new digital society, which he calls Big Data Society, and the balance of human values such as privacy, confidentiality, transparency, identity, free will, associated with new digital structures, will impact on the construction of individuals. In this sense, it warns about the importance of preserving these values for the benefit of innovation and convenience.

4 Social Big Data. See: <http://www.sobigdata.eu/index>. Access: 20 oct. 2021.

These values are reiterated by the RDS initiative which, as outlined, aims to develop a sustainable model for data science based on ethical values and the application of techniques under the requirements of impartiality, accountability, transparency, and confidentiality by conception (DATA SCIENCE CENTER EINDHOVEN, 2018).

For the Data Science Center Eindhoven Research Program,

[...] societal values are an intrinsic part of the value of big data. Building values into data science ‘by design’ is a challenging and fascinating interdisciplinary field of research with many practical applications [...]. Creating societal value through data science requires an understanding of context and effective governance of data! (DATA SCIENCE CENTER EINDHOVEN, 2018).

Thus, Research Data Science emerges based on four pillars – impartiality, accuracy, confidentiality, and transparency, represented by the FACT acronym – to address issues that affect society in different ethical, economic, and/or social aspects.

Tayla and Purtova (2019) bring light to the lack of clarity regarding the concept of Responsible Data Science when applied to the various contexts, which according to the authors may cause difficulties establishing goals and limits for the scientific practice of the data. Therefore the motivation to explore FACT principles.

These principles can be defined as follows:

- a) **Fairness:** The term is defined as being the attitude of those who consider, objectively, without passions or prejudices, a certain fact in all its aspects. Those who are impartial do not sacrifice truth and justice for the convenience or convenience of others for personal gain, nor do they weigh in the judgment previous facts or information that may harm or favor them individually (ÁVILA, 1967). By way of impartiality, fairness of data, Taylor (2017a, p. 1) states that the availability of digital data, especially from the use of technological devices and services by people, “[...] has political and practical implications for the way people are treated by the community, the state, and the private sector”. However, the level of awareness and implementation of mechanisms that combat potential discrimination is lower compared to the speed with which data processing technologies are developed.
- b) **Accuracy:** Accuracy of information or measurement is its quality of being true or correct, even in small details (ACCURACY..., 2019). The production of inaccurate results is independent of the amount of data analyzed. Variables are used to predict a result. For this, there must be a positive or negative correlation between variables and results to obtain better accuracy (GE 301 GROUP 7, 2017).

- c) **Confidentiality:** It is considered a duty to safeguard all information that concerns a person, that is, their privacy. Confidentiality includes the preservation of private and intimate information (GOLDIM, 1997).
- d) **Transparency:** It means “[...] that all codes and all data must be public” (STOYANOVICH et al., 2018, p. 2165) and is seen as a requirement for algorithmic accountability.

Annany e Crawford (2016) suggest that transparency cannot be a characteristic of an algorithmic model. Rather the opacity of algorithms should be considered with a sensitivity for the contexts of their use; transparency is performed by socio-technical assemblages of algorithms and people. However, transparency is not always complete, either because of guarantees associated with commercial rights or because of the privacy of citizens who, if not respected, violate the laws (STOYANOVICH et al., 2018).

The principles cited here are upheld by the RDS and are associated with the legality, the normative, and the ethical that, in addition to the FAIR principles (Findability, Accessibility, Interoperability, and Reusability), appear as the main research subjects in Big Data. It should be noted that the FACT guidelines, although not covered, are complementary to FAIR, mainly because they focus on essential aspects of fairness and data re-use.

The discussion on FAIR has involved professionals and institutions from different domains and is in a more advanced stage, therefore with more publications. Regarding the FACT guidelines, because they have been presented recently or because the actors working in the area do not understand the dimension of these elements, they have not reached the same level of discussion, although their implications involve society in a more present and impacting way. Advocates of the FACT guidelines emphasize that they should be applied not only in the use of algorithmic models but also in the design and development phases.

FACT handles methodological issues for the management of data, which point out the risks in the Big Data analysis and the concern with preparing and ensuring the data so that responsibility and ethics can be verified.

3. Discussion

The RDS initiative represents a group that looks for alternatives to the implementation of FACT and the actions in search of results are contemplated by Green Data Science, a term coined by one of the founders of the RDS group.

Although of unparalleled importance for answering questions related to ethics, economics, and social aspects, the signaling of FACT guidelines in the literature is still timid. In the articles consulted, transparency is the principle most commented on by most articles consulted, which denotes the concern of the authors with the society's demand on the subject, closely related

to the public availability of data. The principle of confidentiality was the second most addressed, followed by impartiality and precision. In analyzing these results, we note that there are difficulties in adopting the principles equally and effectively. Examples are actions that, while preserving confidentiality and impartiality, affect levels of transparency.

According to Taylor e Purtova (2019), these principles are far from informing a coherent system of norms and standards concerning data governance. The position of Kemper e Kolkman (2018) is that there should be more empirical studies on algorithmic models to test algorithmic responsibility, because only then will it be possible to develop sound guidelines appropriate to existing practices. This should be a major challenge, as the existence of legislation does not guarantee the effectiveness of the use of these principles (MOEREL, 2016; OHM, 2014).

For Ohm (2014), most privacy laws focus on collection and disclosure, not use. On the other hand, there are still differences in laws from one country to another. United States privacy law is different from the European Union Data Protection Directive, even when it comes to aspects whose damage may be global. The rule that data must be processed for a specific purpose and subsequently used or communicated only to the extent that it is not incompatible with the usefulness of the transfer also seems intangible. Working with big data is working with the unpredictable. So, how to ensure the final result of data analysis?

Questioning motivates discussion and investigation in the data universe.

On the other hand, professionals such as analysts, programmers, data scientists, information scientists and the government play a vital role in safeguarding these principles to promote an egalitarian society. As stated in this study, actions to ensure the four principles are designed more slowly than technologies developed for data analysis and dissemination.

In addition to the proposals identified by the study, the theme still requires a great joint and interdisciplinary effort from social and technological areas so that they can happen actually.

In the age of liquid society, where everything dissolves - following Bauman's concept – and where each individual is a producer, diffuser, and advocate of his idiosyncrasies, which takes for truth and proclaims them, the reliability of information has acquired a kind of own quotation, a valuation that varies according to the capacity of the individual who needs it, which, for higher levels of intellectual education, even opens the way for a new professional: the one who investigates, checks and makes available, considered reliable, attending to the consumer's needs of this reliability.

Observing the historical period that precedes this phase, called Bauman (2001) “solid modernity”, the author presents ideas that point out constancy and security, because the transformation of values occurred slowly and noticeably, giving

the sensation of certainty and certainty. control over what was happening in the world.

Thus, when we consider, through this approach of liquid modernity, the directions that most denote the social side, we find that they may be more complex, especially in matters of data discrimination and sensitivity. According to the initiatives that show how RDS technology is being worked, it is inferred that data processing systems demand greater flexibility to adapt to the vertiginous changes in direction caused by the flow resulting from the pressure exerted. conditions common to “net life” (BAUMAN, 2001).

4. Conclusion

The concept of RDS presents a significant advance in the context of Big Data and Data Science, bringing important elements that favor reflection by data and information scientists. In particular, FACT identifies the aspects and facets linked to this concept, as well as demonstrating what these actors should consider during the processes that make up the phases of the Data Lifecycle.

We note that the actions promoted by the RDS group concerning FACT guidelines can contribute to safeguarding individual rights without the need for measures that summarily prevent access and reuse of data.

Thus, we emphasize that data analysis techniques, together with semantic technologies, have been widely used from a technical perspective, aiming to generate value for organizations.

In this sense, the study and approach of the social sciences to these themes are relevant for social issues to be considered during technological processes. While there are relevant moves to pass laws that guarantee individual rights in the context of Big Data, the use of technology is imperative for these rights to be secured.

Given the above, there is a need for empirical studies on algorithm models, in which FACT principles are analyzed in practice, so that those involved in data generation, storage, and processing adopt mechanisms that assure them confidence in the services provided, as well as the data accessed.

The development of such data analysis technology according to the FACT model of Responsible Data Science approach is of crucial importance because failure to implement this approach in data technologies can result in undesirable consequences for society. impacting on three elements: privacy, security, and responsibility. It should be noted that the four FACT principles will hardly be contemplated simultaneously, however, a balance should be sought in which compensations are considered to prioritize one over the other(s). If confidentiality is ensured, accuracy and transparency are likely to be compromised. Therefore, the search for a balance between these four principles should be pursued by both developers and data managers.

Referências

ACCURACY. **Collins dictionary**. See: <https://www.collinsdictionary.com/pt/dictionary/english/accuracy>. Access: 29 may. 2021.

AIMS. **Responsible data science. Ensuring, fairness, accuracy, confidentiality, transparency.** 2017. See: <http://aims.fao.org/news/responsible-data-science-ensuring-fairness-accuracy-confidentially-transparency-fact>. Access: 29 may. 2021.

ANNANY, M.; CRAWFORD, K. Seeing without knowing: limitations of the transparency ideal and its application to algorithmic accountability. **New Media & Soc.**, v. 20, n.3, p. 973–989. 2016.

ÁVILA, F. B. S. J. **Pequena enciclopédia de moral e civismo.** Rio de Janeiro: MEC, 1967.

BAURMAN, Z. **Modernidade líquida.** Rio de Janeiro: Zahar, 2001.

DATA SCIENCE CENTER EINDHOVEN. **Responsible Data Science:** Ensuring fairness, accuracy, confidentiality & transparency by design. DSCE, 2019. See: <https://www.tue.nl/en/research/research-areas/data-science/responsible-data-science/>. Access: 10 jan. 2019.

EUROPEAN DATA SCIENCE ACADEMY. **About EDSA.** 2019. See: <http://edsa-project.eu/overview/about-edsa/>. Access: 28 jan. 2019.

FACEBOOK: Cambridge analytica data scandal. **Wikipedia:** The Free Encyclopedia. See: https://en.wikipedia.org/wiki/Facebook%E2%80%93Cambridge_Analytica_data_scandal#:~:text=The%20Facebook%E2%80%93Cambridge%20Analytica%20data,be%20used%20for%20political%20advertising.&text=The%20app%20harvested%20the%20data%20of%20up%20to%2087%20million%20Facebook%20profiles. Access: 26 apr. 2019.

GE 301 GROUP 7. **Responsible Data Science**. 2017. See: <http://ge301.bilkent.edu.tr/fall2017group7/>. Access: 25 may. 2021

GOLDIM, J. R. **Confidencialidade** (1997-2003). See: <https://www.ufrgs.br/bioetica/confiden.htm>. Access: 15 mar. 2019.

HILBERT, M., LÓPEZ, P. The world's technological capacity to store, communicate, and compute information. **Sciencexpress**, p. 1-7, 2011. See: <https://www.ris.org/uploadi/editor/13049382751297697294Science-2011-Hilbert-science.1200970.pdf>. Access: 18 mar. 2019.

KEMPER, J., KOLKMAN, D. Transparent to whom? No algorithmic accountability without a critical audience. **Inf. Commun. Soc.**, v.14, 1–16, 2018.

LODDER, G. M. A., SCHOLTER, R. H. J., GOOSSENS, L., ENGELS, R. C. M. E., VERHAGEN, M. Loneliness and the

social monitoring system: emotion recognition and eye gaze in a real-life conversation. **Br. J. Psychol**, v. 107, n. 1, p. 135-153, 2016.

MOEREL, L. **GDPR conundrums: the data protection officer requirement**. 2016. See: <https://fddocuments.in/document/gdpr-conundrums-the-data-protection-officer-requirement-mediafilesarticles201607160719gdprgdpr>. Access: 10 jan. 2019.

MOEREL, E. M. L., PRINS, C. **Privacy for the homo digitalis: proposal for a new regulatory framework for data protection in the light of Big Data and the Internet of Things**. 2016. See: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2784123. Access: 10 jan. 2019.

OHM, P. Changing the rules: general principles for data use and analysis. In: Lane, J., Stodden, V., Bender, S., & Nissenbaum, H. (Ed.). **Privacy, big data, and the public good: frameworks for engagement**. Cambridge: Cambridge University Press. v.1, p. 96-111, 2014.

PAPPALARDO, L.; GROSSI, V.; PEDRESCHI, D. Introduction to the special issue on social mining and big data ecosystem for open, responsible data science. **International Journal of Data Science and Analytics**. v. 11, p.261-262, 2021.

PIERSMA, N. Data in urban environments. In: PIERSMA, N. (Ed.). **Through the clouds: urban analytics for smart cities**. Amsterdam: Hogeschool van Amsterdam. p. 11-21, 2018.

STOYANOVICH, J., HOWE, B. **Follow the data!** Algorithmic transparency starts with data transparency, 2018. See: <https://ai.shorensteincenter.org/ideas/2018/11/26/follow-the-data-algorithmic-transparency-starts-with-data-transparency>. Access: 21 mar. 2019.

STOYANOVICH, J., HOWE, B., JAGADISH, H. V. Special Session: A technical research agenda in data ethics and responsible data management. In *Proceedings... International Conference on Management of Data*, p. 1635–1636, 2018. See: <https://dl.acm.org/doi/10.1145/3183713.3205185>. Access: 23 mar. 2021.

STOYANOVICH, J., HOWE, B., ABITEBOUL, S., MIKLAU, G., SAHUGUET, A., WEIKUM, G. (2017). Fides: Towards a platform for responsible data science. In *SSDBM'17. 29th International Conference on Scientific and Statistical Database Management*, 2017, Chicago, United States. See: <https://hal.inria.fr/hal-01522418/document>. Access: 10 jan. 2019.

STOYANOVICH, J., HOWE, B., JAGADISH, H. V., MIKLAU, G. Panel: a debate on data and algorithmic ethics. *Proceedings of the VLDB Endowment*, v.11, n.12, p. 2165–7, 2018. See: <https://dl.acm.org/doi/10.14778/3229863.3240494> Access: 14 abr. 2019.

TAYLOR, L. **Data, visibility and justice**, 2017a. See: <https://redasci.org/wp-content/uploads/2016/10/Linnet-Taylor-RDS-16.3.17.pdf>. Access: 22 feb. 2021.

TAYLOR, L. What is data justice? The case for connecting digital rights and freedoms globally. **Big Data & Society**. p. 1-14, 2017b.

TAYLOR, L., & PURTOVA, N. What is responsible and sustainable data science? **Big Data & Society**, v.6, n. 2, 2019.

VAN DER AALST, W. M. Green data science: using big data in an “environmentally friendly” manner. **In 18th International Conference on Enterprise Information Systems (ICEIS 2016)**. p. 25-28, 2016. Rome: SciTePress.

VAN DER AALST, W. M., BICHLER, M., & HEINZL, A. Responsible data science. **Bus. Inf. Syst. Eng.** v. 59 n. 5: p. 311–313, 2017.

Os princípios FAIR no Ciclo de Vida dos Dados para a Ciência da Informação: maximização do acesso, uso e reúso de dados científicos

Anderson Rafael Castro Simões^a, Renata Lemos dos Anjos^b e Guilherme Ataíde Dias^c

1. Introdução

O acesso, uso e reúso dos dados científicos por meio do seu compartilhamento, vêm se destacando na comunidade acadêmico-científica. Pode-se mencionar como exemplo deste fato, as políticas editoriais de vários periódicos científicos eletrônicos, que passaram a solicitar que os dados científicos gerados e/ou coletados no processo de investigação científica sejam devidamente referenciados e fornecidos no momento da submissão do artigo.

De modo semelhante, algumas agências de fomento à pesquisa passaram a solicitar de forma mandatória, que as solicitações de financiamento para projetos de pesquisa incluam um documento conhecido como Plano de Gestão de Dados

-
- a Doutorando em Ciência da Informação pela Universidade Federal da Paraíba – UFPB. E-mail: anderson.simoes@estudantes.ufpb.br. ORCID: <https://orcid.org/0000-0002-8268-2783>.
- b Mestre em Ciência da Informação pela Universidade Federal da Paraíba – UFPB. E-mail: renata.anjos@academico.ufpb.br. ORCID: <https://orcid.org/0000-0002-4599-3494>.
- c Doutor em Ciências da Comunicação/Ciência da Informação pela Universidade de São Paulo – USP. Professor na Universidade Federal da Paraíba – UFPB. E-mail: guilhermeataide@ccsa.ufpb.br. ORCID: <https://orcid.org/0000-0001-6576-0017>.

(PGD). Este documento planeja e detalha previamente todas as etapas do processo de gestão dos dados ao longo do seu ciclo de vida, essas etapas contemplam os cuidados a longo prazo, contribuindo desta forma para facilitar e simplificar os processos de acesso, uso e reúso a posteriori (SAYÃO; SALES, 2012; WILKINSON et al., 2016; FAPESP, 2017; PLOS MEDICINE EDITORS, 2016; ANJOS; AUTRAN, 2018).

Atualmente, existem inúmeros modelos de ciclo de vida dos dados, elaborados para áreas específicas, com propósitos específicos e, portanto, cada um com suas particularidades. Esses modelos permitem a percepção de etapas cíclicas, características, requisitos e fatores que transcendem essas etapas (SANT'ANA, 2016).

Considerando que a Ciência da Informação pode e deve atuar neste cenário de acesso, uso e reúso de dados, Sant'Ana (2016) propõe o Ciclo de Vida dos Dados para a Ciência da Informação – CVD-CI, composto pelas fases de coleta, armazenamento, recuperação e descarte permeadas por fatores que devem ser considerados em todas elas, como a privacidade, integração, qualidade, direito autoral, disseminação e preservação (SANT'ANA, 2016).

Do mesmo modo, de forma a orientar pesquisadores em todo o processo de gestão, agindo como alicerce para o acesso, uso e reúso desses dados, e de modo a maximizar o valor das publicações acadêmicas que compartilham seus dados, os Princípios FAIR foram criados por partes representantes da

academia, indústria e agências de fomento, em face da necessidade de aprimorar a infraestrutura de suporte para o compartilhamento de dados científicos (WILKINSON et al., 2016).

Estes princípios, cujo acrônimo refere-se a *Findable, Interoperable, Accessible e Reusable*, têm como objetivo fazer com que os dados científicos sejam mais facilmente encontráveis, acessíveis, interoperáveis e reutilizáveis. Ressalta-se que os princípios consideram em suas diretrizes não apenas humanos, mas também máquinas visto que estes dependem cada vez mais do suporte computacional, sobretudo, para lidar com os dados. E referem-se à três entidades, a saber: dados, metadados e infraestrutura (WILKINSON et al., 2016; GO FAIR, online).

De acordo com Jacob (2019), as diretrizes dos princípios FAIR, podem melhorar o ciclo de vida dos dados, sobretudo, usufruindo do uso de tecnologias como, por exemplo, por meio do uso de identificadores persistentes. Nessa perspectiva, o *Consortium of European Social Science Data Archives – CESSDA*, criou um guia denominado “*Data Management Expert Guide*” elaborado por especialistas europeus para auxiliar os pesquisadores das ciências sociais a tornar os seus dados mais “justos”, ou seja, FAIR, durante todo o seu ciclo de vida, passando pelo seu planejamento, organização, documentação, processamento, armazenamento e proteção, até o compartilhamento e publicação (CESSDA, 2020). Ainda nessa perspectiva, Miranda (2021) afirma que 90% das diretrizes dos

princípios FAIR envolvem metadados, da mesma forma que, praticamente, os metadados estão presentes em todas as instâncias do ciclo de vida dos dados.

A partir destas breves considerações, este estudo buscou identificar como as diretrizes que compõem os Princípios FAIR podem ser aplicadas nas etapas do Ciclo de Vida dos Dados para a Ciência da Informação (CVD-CI) de maneira a maximizar o acesso, uso e reúso dos dados científicos.

2. Procedimentos metodológicos

Configura-se como uma pesquisa exploratória, bibliográfica, de cunho qualitativo, fazendo uso do método comparativo apontando convergências entre os fatores estudados, no caso entre os Princípios FAIR e o CVD-CI, por meio da inferência da análise de conteúdo (MARKONI; LAKATOS, 2003; BARDIN, 2016).

Inicialmente abordaremos as principais diretrizes propostas pelos princípios FAIR e as fases componentes do CVD-CI. Posteriormente, relacionaremos os Princípios FAIR às fases do CVD-CI, por meio do método comparativo e da análise de conteúdo. Ressalta-se que, neste momento, utilizamos as questões fundamentais que devem ser pensadas no decorrer do CVD-CI propostas por Sant'Ana (2016).

Em seguida, observou-se que os fatores a serem considerados em todas as fases do CVD-CI, também possuem relação com os princípios FAIR, as quais serão elucidadas. Por fim, elaboramos uma figura evidenciando e detalhando as

diretrizes dos Princípios FAIR que podem ser aplicadas no CVD-CI (SANT'ANA, 2016).

3. Resultados e discussões

O primeiro passo para o acesso, uso e reúso dos dados é tornar os dados encontráveis (*Findable*), sendo uma condição para a efetivação dos próximos princípios. Para serem encontráveis, os (meta)dados devem ser atribuídos a um identificador persistente, devem ser descritos com metadados ricos, indexados em recursos pesquisáveis e estar disponíveis online (WILKINSON et al., 2016; MIRANDA, 2021; GO FAIR, online).

Encontrados os dados, faz-se necessário saber como podem ser acessados (*Accessible*), inclui-se aqui o acesso por meio do identificador persistente, acessibilidade dos dados mediante termos de acesso, autenticação e autorização, além da acessibilidade dos metadados quando o acesso livre aos dados não é permitido (WILKINSON et al., 2016; MIRANDA, 2021; GO FAIR, online).

O terceiro princípio (*Interoperable*) ressalta a importância da ligação entre (meta)dados de forma a estabelecer relações semânticas entre os diversos *datasets*, também orienta o uso de formatos abertos e preferenciais, e (meta)dados relevantes por meio da utilização de recursos/instrumentos de controle terminológico como vocabulários controlados que permitem a interoperabilidade entre máquinas (WILKINSON et al., 2016; MIRANDA, 2021; GO FAIR, online).

Por fim, através do quarto princípio (*Reusable*) viabiliza-se o reúso dos dados, permitindo que sejam reutilizados por novos usuários, demandas e aplicações, licenças de uso devem ser explicitamente informadas, assim como questões éticas e legais que regem e protegem os dados. Além do registro, por meio de (meta)dados (relevantes ao domínio para o qual se destinam), sobre o contexto de como os dados foram gerados, incluindo-se questões de proveniência e fluxo de trabalho (WILKINSON et al., 2016; MIRANDA, 2021; GO FAIR, online).

O CVD-CI tem início na fase de Coleta, na qual os dados são obtidos a partir de uma necessidade informacional que determina quais dados serão necessários, como esses dados serão localizados, quais serão os mecanismos para sua obtenção, e as metodologias e ferramentas necessárias para obtenção dos mesmos (SANT'ANA, 2016).

Observa-se que a fase de coleta do CVD-CI associa-se com os quatro princípios FAIR. Com o primeiro princípio *Findable*, quando esse aborda aspectos acerca da encontrabilidade dos dados que facilitam o processo de coleta, por meio da atribuição de um identificador persistente, descrição com metadados ricos que possibilitam a localização dos dados, e a indexação dos mesmos em recursos pesquisáveis e online. Associa-se ao princípio *Accessible* quando a fase de coleta depende do saber de como os dados podem ser acessados, de acordo com os termos de acesso, quais são as questões de autenticação e autorização que regem os dados. O princípio *Interoperable*, associa-se a essa fase,

quando facilita esse momento com o uso de formatos abertos, o uso de recursos/instrumentos de controle terminológico que vão permitir a interoperabilidade entre máquinas, caso a coleta seja realizada por agente(s) computacionais, além de possibilitar a descoberta de outras fontes de coleta visto a vinculação a outros (meta)dados. Por fim, o princípio *Reusable*, rege como os dados coletados poderão ser reutilizados em termos legais e éticos, além de elucidar como esses dados coletados foram gerados por meio de proveniência detalhada.

Quanto aos fatores a serem considerados na coleta, observou-se que os fatores da privacidade e do direito autoral devem ser considerados diante da necessidade em identificar, nas fontes utilizadas para a obtenção dos dados, possíveis aspectos que possam quebrar a privacidade (dos dados pessoais e institucionais obtidos) podendo comprometer as próximas fases do ciclo. Os quais relacionam-se ao princípio *Reusable* que orienta a atribuição de questões éticas e legais na regência de dados sensíveis para que os mesmos sejam reutilizados de forma a respeitar o arcabouço jurídico sustentando este acesso.

Outro fator a ser considerado na coleta, é a integração, na qual os atributos responsáveis pela identificação unívoca de cada registro sejam válidos, sendo relacionada ao *Interoperable*, que recomenda a utilização de recursos/instrumentos de controle terminológico, que vão auxiliar nesse processo de identificação unívoca da integração. O fator da disseminação e qualidade devem ser considerados na fase de coleta quando proporciona

uma maior encontrabilidade e acesso aos dados, relacionando-se ao *Findable*, quando orienta o uso de metadados ricos que devem incluir informações contextuais, sobre a qualidade, condição ou características dos dados. A preservação pode ser considerada quando os dados devem ser preservados e utilizados mesmo com possíveis alterações, que devem ser detalhadas (proveniência detalhada), proposto pelo princípio *Reusable*.

Após a fase de coleta dos dados, percebe-se que os mesmos podem ser posteriormente utilizados em novas análises ou ainda na interação com outros dados. Manifestando-se a necessidade do armazenamento, a próxima fase do CVD-CI. Dessa forma, a fase de armazenamento possui um enfoque tecnológico, que define os aspectos que vão garantir a reutilização desses dados (SANT'ANA, 2016). Ressalta-se que para o CESSDA (2020), a prática do armazenamento e publicação dos dados de maneira adequada (seguindo os princípios FAIR), é o ponto central para tornar os seus dados mais justos, tornar os seus dados mais FAIR.

As diretrizes recomendadas do princípio *Findable* podem ser aplicadas no momento da definição do local de armazenamento (fontes pesquisáveis), e na atribuição dos metadados ricos no momento do armazenamento de forma a facilitar o processo de recuperação (próxima fase). O princípio *Accessible* pode ser aplicado no momento da permissão do acesso apenas aos metadados, em casos específicos (ex. embargo aos dados), e na definição de quais desses dados que foram armazenados podem ser acessados, e em caso positivo, sob quais termos de acesso e

autenticação. O princípio *Interoperable* pode reger a fase de armazenamento no momento da garantia da sua estruturação lógica (uso de recursos/instrumentos de controle terminológico), de como esses dados serão interligados a outros e como essas interligações serão mantidas (referências qualificadas). Por fim, o *Reusable* pode ser aplicado, no que se refere as definições legais e éticas acerca do direito de armazenar esses dados, se os mesmos apresentam riscos a privacidade dos indivíduos ou instituições neles mencionados, além de outros aspectos que garantam sua utilização a posteriori como na garantia de qualidade desses dados com detalhes da proveniência, fluxo de trabalho, e controles de versões.

De forma semelhante a fase de coleta, os fatores da privacidade, direito autoral e qualidade na fase de armazenamento, também se relacionam ao princípio *Reusable*. Na fase de coleta esses fatores devem ser observados no momento em que são coletados, já no momento do armazenamento, estes devem ser observados e realizados de maneira prévia ao armazenamento, sendo definidos para que proporcionem a reutilização dos dados que estão sendo armazenados respeitando as questões éticas e legais referentes a privacidade e direito autoral.

Os fatores da qualidade e preservação também podem ser otimizados por meio do uso de proveniência detalhada, fluxo de trabalho e controle de versões, diretrizes do *Reusable*. Quanto ao fator da integração, este pode ser otimizado considerando as

diretrizes do *Accessible* de como esses dados podem ser acessados (termos de acesso, autenticação e autorização), e do *Interoperable* diante do uso de formatos abertos, na necessidade do uso de recursos/instrumentos de controle terminológicos para identificação e interpretação das entidades e atributos desses dados, no uso de estruturas, padrões de metadados, e na identificação de possíveis relacionamentos, ligações entre dados (referências qualificadas).

A disseminação na fase do armazenamento é fundamental para garantia do acesso futuro, dessa forma, pode ser otimizada com as diretrizes do princípio *Accessible* e *Interoperable* quando deve-se prever meios para que os dados sejam acessíveis, como o uso de identificadores persistentes, descrição com metadados acessíveis, e os termos de acesso, questões de autenticação e autorização, além do uso de recursos/instrumentos de controle terminológicos de forma a garantir a interpretação por máquinas.

Seguindo o CVD-CI, e uma vez que os dados coletados foram armazenados, têm-se a fase da recuperação, que se direciona para a disponibilização desses dados para acesso, uso e reúso, sob o enfoque do responsável pela manutenção desses dados para recuperação (SANT'ANA, 2016).

As diretrizes recomendadas pelo *Findable*, podem ser aplicadas no momento da definição de que, se os dados armazenados serão disponibilizados, aonde e de que forma serão disponibilizados (fontes pesquisáveis), como viabilizar a encontrabilidade desses dados se por meio do uso de

identificadores persistentes e metadados ricos. Quanto ao *Accessible* pode ser aplicado na possível definição de um público-alvo e/ou quem poderá acessá-los mediante autorização e concordância aos termos de acesso. O *Interoperable* pode ter suas diretrizes aplicadas no momento da definição de como dar-se-á a integração entre as estruturas dos dados (padrões de metadados), se haverá a ligação para com outros dados de forma a complementar esse conjunto (referências qualificadas). E o *Reusable* no momento da verificação das questões éticas e legais, sobre o direito de disponibilizar esses dados, considerando riscos à privacidade dos indivíduos e/ou entidades neles mencionados. Bem como, na orientação de proveniência detalhada, informando a frequência de atualização dos dados (controle de versões).

De forma semelhante as fases de coleta e armazenamento, na fase de recuperação, os fatores da privacidade e direito autoral, também se relacionam ao princípio *Reusable*, referentes a sensibilidade dos dados e questões éticas e legais acerca do acesso, uso e reúso, facilitando e assegurando a sua utilização e replicação. As diretrizes do *Interoperable* podem ser aplicadas no fator da integração por meio do uso de padrões metadados, recursos/instrumentos de controle terminológicos, e ligação entre os dados, possibilitando uma análise geral desses dados. A qualidade pode ser otimizada com a aplicação das diretrizes do *Reusable* sobretudo no controle da qualidade, de versões e da proveniência detalhada, bem como do *Findable* e *Accessible*, adotando elementos que ampliem a usabilidade e acessibilidade

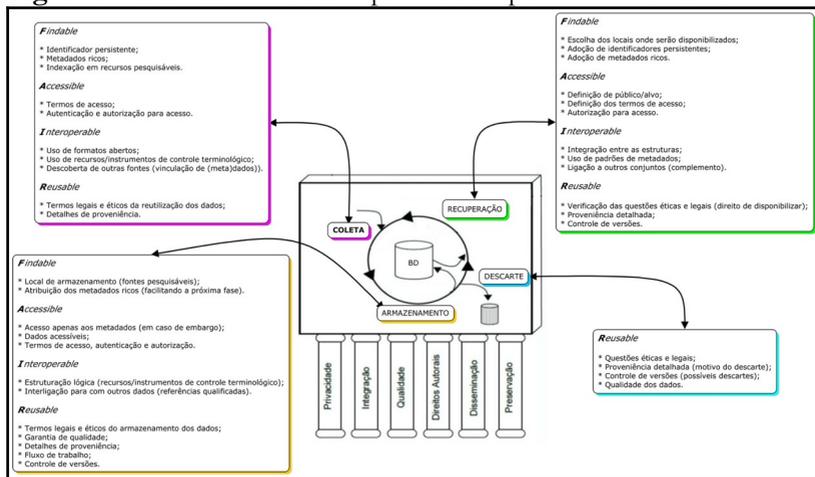
dos dados. De modo semelhante, a disseminação, objetivando a encontrabilidade e (re)uso deve adotar as diretrizes do *Findable* e *Accessible* nas estratégias de localização, de acesso, e sobretudo, no uso de mecanismos para o acesso e a visualização (como formatos e metadados). Por fim, a preservação, atrelada nessa fase ao conceito de interpretação, pode aplicar as diretrizes do *Reusable*, de forma a controlar a qualidade, versões e proveniência, evitando a geração de resultados distintos ao longo do tempo no momento da recuperação.

Finalizando o CVD-CI temos a fase do descarte, consistindo no descarte dos dados julgados desnecessários à determinada pesquisa, ou ainda, aqueles dados que estão acima da capacidade eficiente de tratamento. Nesta última fase, observou-se que apenas o princípio *Reusable* pode ter as suas diretrizes aplicadas, no momento do registro nos controles de versão desse (s) descarte (s), na proveniência detalhada, explicitando o porquê desses dados terem sido descartados, e se, esse descarte afeta de algum modo a interligação a outros dados, ou ainda, a sua encontrabilidade e acesso. Além de dispor do direito, baseando-se em questões éticas e legais, de descarte desses dados (SANT'ANA, 2016).

Quanto aos fatores que devem ser considerados nas fases, a privacidade alega que o indivíduo tem o direito de solicitar que os dados sejam retirados de determinado conjunto, o que pode basear-se nas diretrizes do *Reusable*. O fator da integração, qualidade, do direito autoral, também podem basear-se no

Reusable relacionando-se as questões éticas e legais, mas sobretudo, no registro da proveniência de forma a registrar tal descarte, não gerando insegurança legal em trabalhos que já reutilizaram e/ou referenciaram esses dados (dados interligados). O fator da disseminação e preservação podem ser otimizados com o uso do *Accessible* quando os dados descartados ainda podem ser recuperados por meio dos metadados acessíveis, os quais, disporão sobre o descarte dos mesmos.

Figura 1 – Diretrizes dos Princípios FAIR aplicadas no CVD-CI



Fonte: Autor adaptado de Sant’Ana (2016) e GO FAIR (online).

Para melhor visualização das relações e/ou aplicações desenvolveu-se a Figura 1, na qual os princípios foram ligados nas fases do CVD-CI em que podem ter suas diretrizes aplicadas como forma de otimização do ciclo durante todas as suas fases.

4. Considerações finais

Os princípios FAIR, por meio de diretrizes, objetivam maximizar o compartilhamento e o reuso de dados mais justos, possibilitando, assim, que os mesmos sejam mais encontráveis, acessíveis, interoperáveis e reutilizáveis. De modo semelhante, o ciclo de vida dos dados descreve todos os momentos pelos quais os dados científicos atravessam. No caso do CVD-CI, desde a sua coleta, armazenamento, recuperação, podendo ser uma nova coleta, até o seu descarte.

Como forma de otimizar e maximizar o acesso, uso e reuso dos dados científicos, esta pesquisa em concordância com outras investigações (JACOB, 2019; CESSDA, 2020; MIRANDA, 2021) propôs a adoção das diretrizes recomendadas pelos princípios FAIR no decorrer do ciclo de vida dos dados.

Observou-se que, em sua maioria, os princípios podem ser amplamente adotados no decorrer do Ciclo de Vida dos Dados para a Ciência da Informação (SANT'ANA, 2016), e que essa implementação possibilita a maximização do acesso, uso e reuso dos dados científicos, visto que, enriquece todo o ciclo de vida dos dados, viabilizando todas as suas etapas.

Referências

ANJOS, R. L. dos; AUTRAN, M. M. M. Política de Compartilhamento de Dados Científicos: a adoção nos periódicos da Ciência da Informação. *In: WORKSHOP DE INFORMAÇÃO, DADOS E TECNOLOGIA*, 2., 2018, João Pessoa. **Anais** [...]. João Pessoa: II WiDat, 2018, p. 269-276.

Disponível em:

<http://www.ufpb.br/widat/contents/documentos/anaiswidat2018.pdf>. Acesso em: 19 maio 2021.

CONSORTIUM OF EUROPEAN SOCIAL SCIENCE DATA ARCHIVES – CESSDA. **Data Management Expert Guide**.

2020. Disponível em:

https://zenodo.org/record/3820473#.YK_ypZNKhN1. Acesso em: 18 maio 2021.

FUNDAÇÃO DE AMPARO À PESQUISA – FAPESP. Plano de gestão de dados se incorporam a projetos de pesquisa no

Brasil. Revista Pesquisa FAPESP, [S. l.], 2017. Disponível:

<https://revistapesquisa.fapesp.br/planos-de-gestao-de-dados-se-incorporam-a-projetos-de-pesquisa-no-brasil/>. Acesso em 21 maio 2021.

GO FAIR. **FAIR Principles**. *Online*. Disponível em:

<https://www.go-fair.org/fair-principles/>. Acesso em: 20 maio 2021.

JACOB, D. **FAIR principles, an new opportunity to improve the data lifecycle**. 2019. Disponível em: <https://hal.archives-ouvertes.fr/hal-02070883/>. Acesso em: 20 maio 2021.

MIRANDA, P. **Como aplicar os princípios FAIR ao ciclo de vida dos dados?** Publicado pelo canal Projeto RCAAP. [S. l.: s. n.], 2021. 1 vídeo (42 minutos e 27 segundos). Disponível em:

<https://www.youtube.com/watch?v=HZM-JDEaXBc&t=1283s>.
Acesso em: 21 maio 2021.

PLOS MEDICINE EDITORS. Can data sharing become the path of least resistance?. 2016. **PLoS Medicine**. Disponível em: <https://doi.org/10.1371/journal.pmed.1001949>. Acesso em: 19 maio 2021.

SANT'ANA, R. C. G. Ciclo de vida dos dados: uma perspectiva a partir da ciência da informação. **Informação & Informação**, [S.l.], v. 21, n. 2, p. 116–142, dez. 2016. ISSN 1981-8920. Disponível em: <http://dx.doi.org/10.5433/1981-8920.2016v21n2p116>. Acesso em: 19 maio 2021.

SAYÃO, L. F.; SALES, L. F. Curadoria digital: um novo patamar para preservação de dados digitais de pesquisa. **Informação & Sociedade**, v. 22, n. 3, 2012. Disponível em: <http://www.periodicos.ufpb.br/index.php/ies/article/view/12224/8586>. Acesso em 19 maio 2021.

WILKINSON, M. D. et al. The FAIR Guiding Principles for scientific data management and stewardship. **Scientific Data**, v. 3, n. 1, p. 1 – 9, 2016.

Visualização de dados: a autonomia do usuário na construção de gráficos

Paulo George Miranda Martins^a, Danila Fernandes Alencar^b e Ricardo César Gonçalves Sant'Ana^c

1. Introdução

O volume de dados gerados nas últimas décadas têm proporcionado mudanças significativas na forma de como os dados são recuperados e utilizados por usuários de sistemas informatizados, o que tem se tornado um desafio cada vez maior no mundo contemporâneo frente a abundância de dados e informações que, em sua maioria, não estão estruturados adequadamente na internet, necessitando de tratamento especializado.

Considerando tal cenário, a informação desempenha um papel fundamental no desenvolvimento humano, seja no compartilhamento de informações entre os diversos grupos sociais, seja no uso estratégico pelas organizações e/ou unidades de informação. O acesso a dados públicos constitui-se um fator

-
- a Doutorando em Ciência da Informação pela Universidade Estadual Paulista – UNESP. Inspetor Sênior do Controle da Qualidade na LATAM Airlines Brasil. E-mail: paulo.george@unesp.br. ORCID: <https://orcid.org/0000-0002-3922-5069>.
 - b Mestranda em Ciência da Informação pela Universidade Estadual Paulista – UNESP. Arquivista no Centro de Documentação da Universidade de Marília – UNIMAR. E-mail: danila.fernandes@unesp.br.
 - c Doutor em Ciência da Informação pela Universidade Estadual Paulista – UNESP. Professor na Universidade Estadual Paulista – UNESP. E-mail: ricardo.santana@unesp.br. ORCID: <https://orcid.org/0000-0003-1387-4519>.

fundamental para a geração de conhecimento e, em um contexto democrático de ações e informações transparentes, o “[...] acesso a dados e as muitas possibilidades de construção de visualização para estes dados”, constitui-se um elemento fundamental para a “[...] construção de uma participação cidadã e para a concretização de uma relação mais transparente”, entre os mais diversos órgãos governamentais e a sociedade civil (SANT’ANA; RODRIGUES, 2013, p. 58).

Para o *Open Government Working Group* (OGD), grupo de estudo multidisciplinar formado em 2007 na cidade americana de Sebastopol, o acesso aos dados públicos de forma transparente e livre parte do pressuposto de que tais informações devem estar em formato aberto. Para nortear o uso dos dados abertos, o grupo estabeleceu um conjunto de 8 Princípios na qual os dados devem ser: completos, primários, atuais, acessíveis, processáveis por máquinas, de acesso não-discriminatório, de formato não-proprietário e com licenças livres (ODG, 2007).

Nesse contexto, a Ciência da Informação desempenha papel fundamental na perspectiva de acessibilidade informacional ao proporcionar melhores práticas nos processos de recuperação de recursos e na instrumentalização através da visualização de informações que podem ser utilizadas para a prática decisória nas instituições informacionais, empresariais, indivíduos e/ou para o próprio processo de retroalimentação de coleta informacional.

O processo de visualização dos dados deve ser calcado em métodos estatísticos como o modelo proposto por Card,

Mackinlay e Shneiderman (1999), que ilustra a sequência de transformação de dados brutos até sua forma visual, passando por fases de tratamento e transformações de dados e o mapeamento visual necessário para apresentação gráfica.

Nessa perspectiva e considerando o cenário de múltiplas formas em se obter dados, a presente pesquisa apresenta a seguinte questão: quais são os pré-requisitos necessários para ampliação da autonomia dos usuários na construção de visualizações gráficas de dados?

O objetivo central da pesquisa é identificar barreiras na construção de visualização gráfica de dados a partir da utilização do software R, e para tanto, os objetivos específicos são: a) selecionar um conjunto de dados disponibilizados em formato que permita reutilização, b) construção de uma nova visualização gráfica dos dados e c) identificar pré-requisitos necessários para a construção do gráfico.

Esse estudo pretende trazer elementos para uma reflexão sobre a autonomia do usuário na construção e visualização gráfica de dados, além de subsídios para o entendimento sobre a ferramenta.

2. Procedimentos metodológicos

Por se tratar de uma análise a partir da recuperação de dados, a metodologia adotada para o desenvolvimento da pesquisa se caracteriza como exploratória, através da coleta de conjunto de dados disponíveis no sítio Terrabrasilis a partir dos resultados obtidos nos termos de busca “Desmatamento” e “Cerrado”, no

mecanismo de busca, tendo como ponto específico as premissas de disseminação da informação, destacadas no Ciclo de Vida dos Dados no âmbito da CI (SANT'ANA, 2016).

Caracteriza-se, também, como pesquisa de natureza aplicada uma vez que objetiva gerar “conhecimentos para aplicações práticas” (SILVEIRA & CÓRDOVA, 2009, p. 35) e busca “[...] compreender e analisar a realidade dos enfoques da pesquisa” (TREVINHO, 1987, p. 117).

Sant'Ana e Rodrigues (2013) pontuam que, o acesso aos dados primários favorece a construção de uma visualização mais próxima possível dos valores originais representados, evitando falhas de interpretação e permitindo caminhos variados no processo de análise, dada a liberdade de combinações que um dado primário oferece.

Para este estudo, realizou-se a coleta de dados sobre o desmatamento no Brasil a partir da fonte gráfica denominada “Incremento de desmatamento – Cerrado – Estados”, disponível no portal TerraBrasilis⁵, uma plataforma de disseminação de dados de desmatamento do INPE.

Para construção do arcabouço teórico foram pesquisadas fontes de informação primárias e secundárias, como artigos, livros e de estudos apresentados, além de pesquisas em bases de dados especializadas na temática abordada.

5 TerraBrasilis. Disponível em: <http://terrabrasilis.dpi.inpe.br/>. Acesso em: 10 out. 2021.

3. Coleta de dados e o processo de visualização

A coleta dos dados foi realizada no sítio TerraBrasilis, cuja disposição atende os princípios de acessibilidade a dados públicos de forma transparente e livre, estabelecidos pela OGD (2007) de que os dados devem ser acessíveis e processáveis por máquinas.

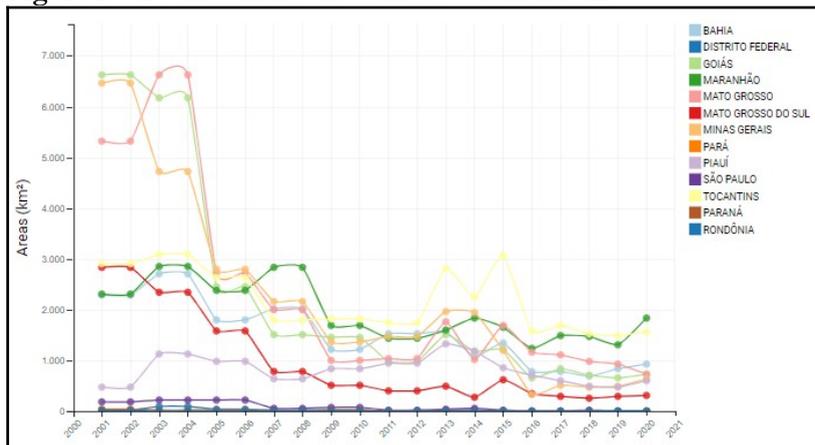
Os dados coletados foram armazenados em uma planilha de cálculo, para gerar um arquivo no formato “.xlsx” estruturado para posterior importação para o software R.

Sant’Ana (2016) destaca que algumas etapas predefinidas são essenciais para o tratamento pleno dos dados, que devem obedecer a um ciclo de vida estruturado para que gerem valor e atendam as expectativas de seus usuários, possibilitando desenvolver melhores formas de identificar, planejar e implementar serviços e produtos alinhados com as novas perspectivas tecnológicas contemporâneas.

Para a pesquisa foi selecionado o gráfico referente a linha do tempo de desmatamento e os dados correspondem aos nomes dos estados, áreas desmatadas em km² e ao ano referente aos dados levantados, como apresentado na Figura 1.

No intuito de propiciar condições para que o usuário possa ter liberdade de solucionar eventuais questões sobre os dados, percebe-se a necessidade de empoderá-los com ferramentas para que ele possa usufruir dos potenciais dados disponibilizados em expressiva quantidade na web.

Figura 1 – Desmatamento no Cerrado – TerraBrasilis



Fonte: TerraBrasilis PRODES (2021).

Para Wickham e Grolemun (2016), um processo de visualização eficiente mostrará coisas que o leitor não esperava ou fomentar a elaboração de novas questões sobre os dados, podendo, inclusive, indicar que o direcionamento a ser adotado está correto ou não, ou se é preciso coletar dados diferentes para consolidar as informações já processadas.

Os dados por si só não são o suficiente para representar algum contexto, sendo necessário, portanto, intervenções tecnológicas e humanas para dar sentido aos mesmos, transformando-os em informações relevantes que façam sentido e gerem valor para aqueles que os detêm.

Para o escopo dessa pesquisa, será utilizado o software R. Considerado um sistema para computação estatística e geração de gráficos, criada em 1993 por Ross Ihaka e Robert Gentleman, o

software R fornece, entre outras coisas, uma linguagem de programação, gráficos de alto nível, interfaces para outras linguagens e recursos de depuração (FARIA; PARGA, 2020).

4. Resultados e discussões

A partir do gráfico apresentado na Figura 1, buscou-se construir um gráfico similar, em linhas, e outro, com pequenas alterações no conjunto de dados, em barras. Dado o escopo deste trabalho, busca-se demonstrar qual nível de autonomia o usuário, mesmo com conhecimento mínimo sobre a ferramenta, possa ter para criar visualizações gráficas.

Os dados coletados foram armazenados em uma planilha de cálculo, gerando um arquivo “.xlsx” de valores dispostos em linhas e colunas de forma estruturada para que fosse possível a leitura no software R. O arquivo “.xlsx” foi usado para gerar os gráficos que correspondem, em síntese, ao gráfico da Figura 1 disponível no site TerraBrasilis. Neste exemplo, foram selecionados nove Estados que apresentam índices que compreendem os anos de 2001 até 2020.

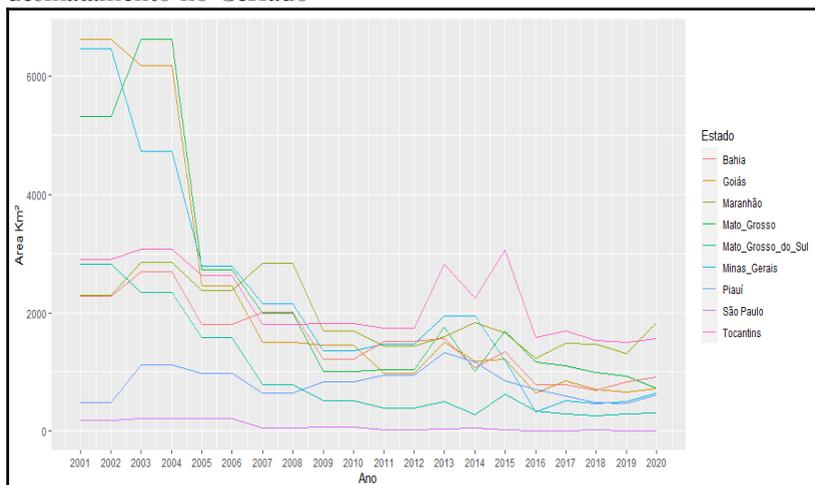
Foram utilizadas duas bibliotecas para importar arquivos e criar a visualização de dados, sendo: “*library(readxl)*”, biblioteca necessária para leitura de arquivos no formato Excel; e “*library(ggplot2)*”, usados para criação de visualizações.

O pacote “*ggplot2*” oferece uma ampla variedade de construção gráfica a partir de uma sintaxe compacta, possibilitando criar gráficos com uma melhor visualização, sendo

considerado especialmente útil para iniciantes na ferramenta R (WICKHAM, 2016).

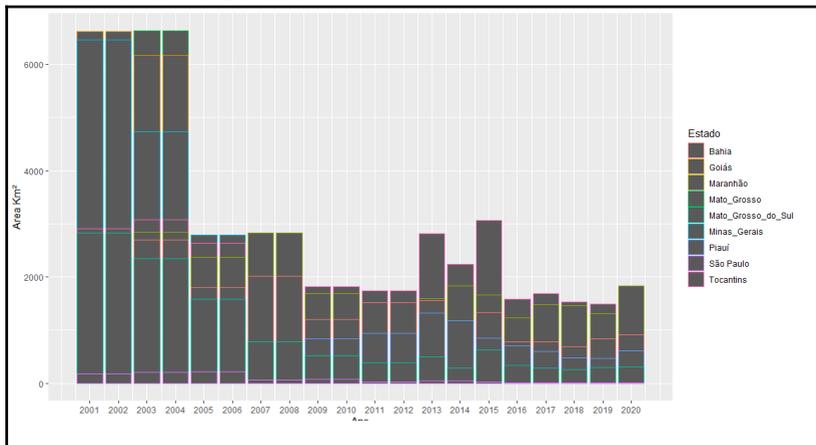
Após importar o arquivo “.xlsx” para o R, foram elaborados dois gráficos distintos, sendo um de linhas (Figura 2), e outro de barras verticais (Figura 3), com a finalidade de demonstrar que com as mesmas funções, fazendo-se pequenas alterações nas linhas de comando é possível gerar outras formas de visualizações.

Figura 2 – Gráfico desenvolvido no software R para ilustrar o desmatamento no Cerrado



Fonte: Autores.

Figura 3 – Gráfico desenvolvido no software R para ilustrar o desmatamento no Cerrado



Fonte: Autores.

O código utilizado para a elaboração dos gráficos foi construído em linhas de comando repetidas 2 vezes com parâmetros diferentes para carregar as bibliotecas de funções (*library*).

Para execução do gráfico existem apenas duas linhas de comando: uma para carga do conteúdo da planilha para a variável de memória e outra para execução do gráfico. A primeira com 1 parâmetro e a segunda com 9 parâmetros, sendo que um deles (*geom_line*) foi necessário mais de uma vez (Figura 4).

Figura 4 – Código utilizado para construir o gráfico da Figura 2.

```
1. library(readxl)
2. library(ggplot2)
3. dados <- read_excel("arquivo.xlsx")
4. ggplot(dados)+aes(x=2001:2020,y=Goiias)+
5.   geom_line(aes(y = Mato_Grosso,col="Mato_Grosso")) +
6.   geom_line(aes(y = Goias,col="Goiás")) +
7.   geom_line(aes(y = Minas_Gerais,col="Minas_Gerais")) +
8.   geom_line(aes(y = Tocantins,col="Tocantins")) +
9.   geom_line(aes(y = Maranhao,col="Maranhão")) +
10.  geom_line(aes(y = Bahia,col="Bahia")) +
11.  geom_line(aes(y = Piaui,col="Piauí")) +
12.  geom_line(aes(y = Mato_Grosso_do_Sul,col="Mato_Grosso_do_Sul")) +
13.  geom_line(aes(y = Sao_Paulo,col="São Paulo")) +
14.  scale_x_continuous("Ano", labels = as.character(2001:2020), breaks = 2001:2020)+
15.  labs(x = "Ano",y = "Area Km²",color="Estado")
```

Fonte: Autores.

Para a construção de um gráfico alternativo (Figura 3), foram alterados os parâmetros da segunda linha de comando, modificando o parâmetro “*geom_line*”, no gráfico em linhas, para “*geom_col*”, no gráfico de colunas, demonstrando-se, nesse caso, a flexibilidade em se obter diferentes visualizações, substituindo apenas um parâmetro, conforme Figura 5.

Figura 5 – Código utilizado para construir o gráfico da Figura 3

```
1. library(readxl)
2. library(ggplot2)
3. dados <- read_excel("arquivo.xlsx")
4. ggplot(dados)+aes(x=2001:2020,y=Goiias)+
5.   geom_col(aes(y = Mato_Grosso,col="Mato_Grosso")) +
6.   geom_col(aes(y = Goias,col="Goiás")) +
7.   geom_col(aes(y = Minas_Gerais,col="Minas_Gerais")) +
8.   geom_col(aes(y = Tocantins,col="Tocantins")) +
9.   geom_col(aes(y = Maranhao,col="Maranhão")) +
10.  geom_col(aes(y = Bahia,col="Bahia")) +
11.  geom_col(aes(y = Piaui,col="Piauí")) +
12.  geom_col(aes(y = Mato_Grosso_do_Sul,col="Mato_Grosso_do_Sul")) +
13.  geom_col(aes(y = Sao_Paulo,col="São Paulo")) +
14.  scale_x_continuous("Ano", labels = as.character(2001:2020), breaks = 2001:2020)+
15.  labs(x = "Ano",y = "Area Km²",color="Estado")
```

Fonte: Autores.

Apresenta-se no Quadro 1, uma síntese das linhas de comando, funções, parâmetros e pré-requisitos utilizados para a criação da visualização gráfica da Figura 2 e Figura 3. Além dos pré-requisitos necessários, apresentados no quadro, e com base nas barreiras encontradas, verificou-se que os usuários devem possuir competências básicas no idioma predominante do R e raciocínio lógico para correlacionar os parâmetros do software.

Quadro 1 – Pré-requisitos para construção dos gráficos

Linhas de comando	Função	Parâmetros	Descrição
<i>library(readxl)</i> <i>library(ggplot2)</i>	<i>library()</i>	Pacote de bibliotecas <i>readxl</i> <i>ggplot2</i>	Permite carregar bibliotecas de funções Conjunto de funções para acesso e interação com planilhas no Excel Conjunto de funções para construção de gráficos
<i>dados <- read_excel</i> <i>("arquivo.xlsx")</i>	<i>read_excel()</i>	"nome_do_arquivo"	Retorna o conteúdo do arquivo que neste caso é atribuído a uma variável denominada 'dados'
<i>ggplot(dados [...])</i>	<i>ggplot()</i>	<i>dados</i> , <i>aes()</i> , <i>geom_line()</i> , <i>geom_col()</i> , <i>scale_x_continuous()</i> , <i>labs()</i>	Executa o gráfico com as definições estabelecidas nos respectivos parâmetros

Fonte: Autores.

Considerado uma linguagem de programação bem desenvolvida, o R é um software livre que oferece um ambiente para análise interativa de dados, podendo ser utilizado em uma ampla variedade de plataformas, sendo considerado simples e eficaz que inclui condicionais com funções recursivas definidas pelo usuário (R FOUNDATION, 2021).

Contudo, ressalta-se que do ponto de vista da práxis, faz-se necessário um conhecimento mínimo de ferramentas tecnológicas que propiciam o desenvolvimento de uma matriz estruturada dos dados em planilhas de cálculos, como Excel, que seja interpretável por uma ferramenta como o R. Para seu uso se faz necessário conhecer as bibliotecas que podem ser utilizadas para atender demandas específicas, e qual a sintaxe para operar as funções disponíveis no pacote. E, a partir das funções que se pretende usar, quais os parâmetros e os argumentos necessários para executar a função.

Assim, para que o usuário tenha autonomia na construção de visualizações gráficas, é necessário um esforço considerável, uma vez que requer competências que demandam conhecimento específico da ferramenta que será utilizada para que se possa interpretar os conjuntos de elementos intrínsecos ao software.

5. Considerações finais

Compreender as implicações que envolvem as tecnologias utilizadas para o tratamento de grandes volumes de dados em suas aplicações, não se restringe apenas ao uso sistêmico da

informação, mas também para as relações que os usuários têm com as ferramentas que realizam tais operações.

Neste trabalho, buscou-se avaliar barreiras para construção de visualizações gráficas a partir do software R e na busca pela autonomia do usuário, que pretende usar a ferramenta para gerar visualização de dados.

Constatou-se que, o uso do R em suas funções básicas requer, ainda, um considerável volume de conhecimento sobre suas principais funções e parâmetros, além, é claro, sobre seus pacotes de bibliotecas, bem como, conhecimento relacionado à essência dos processos que envolvem as etapas do Ciclo de Vida dos Dados como a coleta, armazenamento e recuperação dos dados. Assim, mesmo com todas as vantagens de simplificações oferecidas por recursos como o R, ainda existe um longo caminho na construção do encapsulamento da complexidade a ser percorrido na busca por maior autonomia do usuário na construção de visualizações de dados.

Referências

BATINI, C.; SCANNAPIECO, M. **Data and information quality: Dimensions, Principles and Techniques**. Cham, Switzerland: Springer International Publishing, 2016.

CARD, S. K.; MACKINLAY, J. D.; SHNEIDERMAN, B. **Readings in information visualization: using vision to think**. San Francisco, CA: Morgan Kaufmann Publishers Inc., 1999.

FARIA, P. D.; PARGA, J. P. F. A. **Introdução à Linguagem R: seus fundamentos e sua prática**. 1. ed. Belo Horizonte: [s.n.], 2020. Disponível em: <https://pedro-faria.netlify.app/pt/publication/book/introducao-linguagem-r/>. Acesso em: 21 maio 2021.

OGD – Open Government Data. **The Annotated 8 Principles of Open Government Data**. 2007. Disponível em: <https://opengovdata.org/>. Acesso em: 21 maio 2021.

R FOUNDATION. **What is R?**. 2020. Disponível em: <https://www.r-project.org/about.html>. Acesso em: 24 maio, 2021.

SANT'ANA, R. C. G. Ciclo de vida dos dados: uma perspectiva a partir da ciência da informação. **Informação & Informação**, [S.l.], v. 21, n. 2, p. 116–142, dez. 2016. Disponível em: <http://www.uel.br/revistas/uel/index.php/informacao/article/view/27940/20124>. Acesso em: 5 jul. 2020.

SANT'ANA, R. C. G.; RODRIGUES, F. A. Acessando dados para visualização de afinidades nas votações entre parlamentares do senado. **Informação & Sociedade: Estudos**, v. 23, n. 1, 2013. Disponível em: <http://hdl.handle.net/20.500.11959/brapci/92985>. Acesso em: 21 maio 2021.

SILVEIRA, D. T., CÓRDOVA, F. P. **A pesquisa científica**. In: Gerhardt, T. E., & Silveira, D. T. (Orgs.). Métodos de pesquisa. Porto Alegre: Editora da UFRGS, 2009.

TREVIÑOS, A. N. S. **Introdução à pesquisa em ciências sociais**: a pesquisa qualitativa em educação. São Paulo: Atlas, 1987, p.117.

TERRABRASILIS. **PRODES Desmatamento**. 2021.

Disponível em:

<http://terrabrasilis.dpi.inpe.br/app/dashboard/deforestation/biomes/cerrado/increments>. Acesso em: 20 maio 2021.

WICKHAM, H.; GROLEMUN, G. **R for Data Science**. 2016.

Disponível em: <https://r4ds.had.co.nz>. Acesso em: 24 maio, 2021.

WICKHAM, H. **ggplot2**: Elegant Graphics for Data Analysis. 2^a Edition. Springer, 2016.

Qualidade de dados no processo decisório de educação a distância

Lafaiete Henrique Rosa Leme^a, Felipe Stefani Correia dos Santos^b, Késsia Rita da Costa Marchi^c e Leonardo Castro Botega^d

1. Introdução

O campo educacional tem incorporado Tecnologias de Informação e Comunicação (TIC) em suas atividades referentes aos processos de ensino e aprendizagem. Por consequência disto, são produzidos cada vez mais dados que demandam de análises detalhadas e voltadas para um melhor planejamento e execução das ações educacionais.

Neste contexto, os sistemas de gestão de aprendizagem (em tradução livre de *Learning Management Systems* – LMS) geram muitos dados que podem ser analisados e utilizados para auxílio da tomada de decisão pedagógica, gerando informações úteis a partir dos dados coletados neste ambiente. Contudo, estes dados gerados nem sempre são dados de qualidade, podendo então

-
- a Mestre em Ciência da Computação pela Universidade Estadual de Maringá – UEM. E-mail: lafaiete.leme@unesp.br.
 - b Mestrando em Ciência da Informação pela Universidade Estadual Paulista – UNESP. E-mail: felipe.stefani@unesp.br. ORCID: <https://orcid.org/0000-0002-9359-9272>.
 - c Doutoranda em Ciência da Informação pela Universidade Estadual Paulista – UNESP. E-mail: kessia.marchi@unesp.br. ORCID: <https://orcid.org/0000-0001-5504-5685>.
 - d Doutor em Ciência da Computação pela Universidade Federal de São Carlos – UFSCar. Professor na Universidade Estadual Paulista – UNESP. E-mail: leonardo.botega@unesp.br. ORCID: <https://orcid.org/0000-0003-1495-5935>.

gerar informações as quais não tem aplicação ou não tem confiabilidade podendo culminar em uma má interpretação da realidade e gerar uma falsa impressão, o que atrapalha a tomada de decisão, inclusive no contexto pedagógico.

Diante do exposto, este trabalho apresenta uma revisão de literatura que tem como objetivo identificar dimensões de qualidade de dados abordadas, direta ou indiretamente, em artigos que tratam de tomadas de decisão pedagógica baseada em dados provenientes de LMS. Ao desenvolver este trabalho foram apresentados os conceitos de qualidade de dados, informações e as dimensões de qualidade, identificadas no contexto da educação, visando contribuir com uma análise crítica sobre estas abordagens. Ao considerar os trabalhos é possível observar uma escassez de artefatos que remetem à qualidade dos dados, o que levou a identificar lacunas que requerem atenção.

2. A qualidade de dados e informação

O conceito de qualidade de dados apresenta variações e pode ser interpretado de acordo com o contexto em que é aplicado. Floridi e Illari (2014) e Samitsch (2015) definem que uma informação é de qualidade quando atende ao seu propósito. Batini e Scannapieco (p. 5, 2016) apresentam algumas definições para qualidade que, de forma geral, compreendem a adequação do uso, a conformidade com os requisitos e a satisfação do usuário.

Desta forma, a qualidade é reconhecida como um problema multidimensional e impacta diretamente na performance de

processos operacionais, atividades de tomadas de decisão e de requisitos de cooperação inter-organizacionais. Samitsch (2015) afirma que tomadas de decisões baseadas em dados de alta qualidade proporcionam melhorias na produtividade organizacional.

Diante deste contexto, Wang e Strong (1996) e Wang (1998) abordaram que a qualidade dos dados não precisa ser absoluta, mas deve atender a critérios que possibilitam tomadas de decisões significativas e eficientes, considerando que uma vez que os stakeholders perdem confiança na qualidade dos dados, o sistema está relegado como não confiável, talvez para sempre (REDGATE, 2016).

Para alcançar a qualidade dos dados é necessária maior compreensão do ambiente organizacional e entender as necessidades dos stakeholders. Para isto, são propostas dimensões de qualidade que permitem avaliar os dados e caso atendam tais dimensões, podem ser classificados, então, como de qualidade.

Batini e Scannapieco (2016) afirmam que para avaliar a qualidade dos dados é necessário verificar as dimensões de qualidade, quando descreve seu aspecto de acordo com o contexto que este dado está inserido. São encontradas diversas dimensões de qualidade de dados na literatura. Diante desta variedade, Wang e Strong (1996), Wang (1998) e Batini e Scannapieco (2016) evidenciam que não existe um consenso relativo ao conjunto de dimensões que devem ser considerados para realizar uma avaliação de qualidade.

Neste contexto e ao considerar a complexidade, o aumento da produção e a diversidade de dados em LMS, se faz necessário compreender o progresso e a performance dos alunos a partir destes dados, com a finalidade de descobrir relações não óbvias que apoiam tomadas de decisões pedagógicas. Para tanto, é necessária uma preocupação com a qualidade desses dados adquiridos, considerando que se percebe carência na atenção com a qualidade desses dados.

Visando esta contribuição justifica-se avaliar como e quais dimensões de qualidade de dados vêm sendo abordadas em trabalhos que discutem tomada de decisão baseada em dados educacionais, o que possibilita reconhecer o panorama sobre estas dimensões e identificar lacunas que requerem maiores esforços.

3. Procedimentos metodológicos

Os procedimentos metodológicos constituem, do ponto de vista da sua natureza em uma pesquisa básica e pela natureza de seus objetivos, uma pesquisa exploratória. No que diz respeito aos procedimentos técnicos, este trabalho se compõe de uma revisão de literatura que incluiu artigos publicados em periódicos, em que as principais áreas que dialogam com a abordagem proposta foram as áreas de ciência da informação, ciência da computação e educação. A Figura 1 apresenta as etapas de desenvolvimento deste artigo.

Figura 1 – Etapas de desenvolvimento



Fonte: Autores.

O levantamento de artigos ocorreu no mês de maio de 2021 e foram inseridos os termos [“quality of data” in education* * “quality of data” in education management* *data quality lms management*], em bases de dados como a SCIELO, ACM, ScienceDirect, IBICT e Google Scholar. A partir dos resultados foram considerados 10 artigos internacionais, publicados nos anos de 2010 a 2020.

Para elencar os artigos candidatos deste estudo considerou-se pesquisas que trabalham com dados provenientes de ambientes educacionais à distância. Finalizada a coleta, procedeu-se a leitura classificatória em duas partes, a partir da perspectiva da interação autor-texto-leitor na qual, conforme afirmam Di Raimo e Petermann (2016), se faz necessário definir, inicialmente, os objetivos e estratégias de leitura, assim como ativar conhecimentos prévios.

Diante disso, no primeiro critério foram considerados trabalhos que versam sobre tomada de decisão baseada em dados, posteriormente a abordagem direta ou indireta das dimensões da qualidade de dados apresentadas por Wand e Wang (1996) e Wang e Strong (1996). Subsequentemente, as dimensões consideradas foram mapeadas e apontadas na próxima seção, bem como uma discussão dos resultados. Procedeu-se a

identificação quantitativa do uso de cada dimensão nos artigos, com o subsídio de conceitos científicos presentes na literatura.

4. Resultados e discussões

Nos trabalhos apreciados foram identificadas seis dimensões de qualidade. O Quadro 1 apresenta a categorização dos artigos de acordo com estas dimensões seguindo a ordem de colunas: 1) Granularidade, 2) Completude, 3) Acurácia, 4) Precisão, 5) Temporalidade e 6) Consistência. Foi observado que as dimensões granularidade e completude estão presentes em 70%, seguida pela dimensão acurácia, encontrada em 60% destes trabalhos.

Quadro 1 – Categorização dos artigos de acordo com os critérios de qualidade de dados

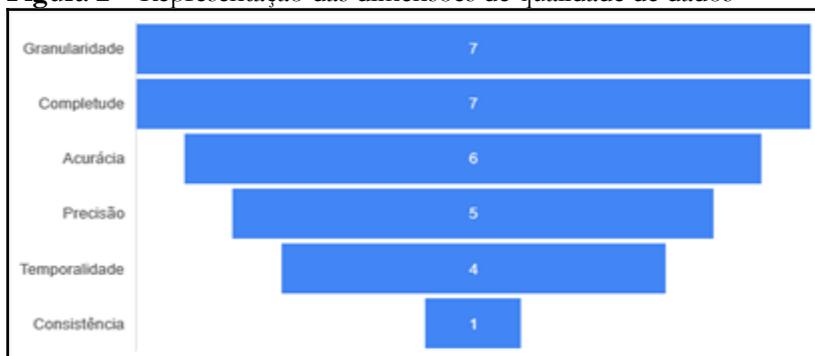
Artigos	Gran.	Comp.	Acur.	Prec.	Temp.	Con.
Ranjeeth (2020)		x	x			x
Villegas-ch (2020)	x	x	x			
Schildkamp (2019)	x	x	x		x	
Henrie (2018)	x	x	x	x	x	
Kadoic (2018)			x		x	
Cerezo (2016)	x	x		x		
You (2016)		x		x	x	
Dias (2015)	x			x		
Kazanidis (2012)	x			x		
Macfadyen (2010)	x	x	x			

Fonte: Autores.

A dimensão de precisão foi identificada em 50% dos trabalhos e as dimensões menos consideradas foram a temporalidade, observada em 40% dos trabalhos e a consistência, contemplada em 10% dos trabalhos. A Figura 1 exibe um gráfico contendo a representação das dimensões de qualidade de dados constatadas nos trabalhos.

As dimensões de qualidade de dados identificadas foram mapeadas e categorizadas (Quadro 1) conforme suas características intrínsecas. Inicialmente, foram detectadas seis dimensões (Figura 2), e se observa granularidade e completude abordadas em sete trabalhos.

Figura 2 – Representação das dimensões de qualidade de dados



Fonte: Autores.

A dimensão granularidade se refere a sua capacidade de distinguir objetos (FURBER, 2016). Para Santos e Santana (2013) a granularidade está relacionada ao conteúdo disponível no conjunto de dados e se refere ao número de atributos que

compõem este conjunto e a sua diversidade de conteúdos. Os autores complementam que a granularidade impacta diretamente nos processos de acesso e de tratamento dos dados.

A completude é definida por Wand e Wang (1996) como o grau no qual uma coleção de dados inclui referências diante da descrição de um conjunto de objetos que se relaciona com o mundo real. Batini e Scannapieco (p. 70, 2016) comentam que são propostas diferentes definições para completude, todas coerentes e complementares. Entretanto, os autores ressaltam que muitas vezes estas definições se referem a diferentes níveis de granularidade ou de modelos de dados.

A acurácia é citada em seis trabalhos e, segundo Wang e Strong (1996), é obtida quando um dado é considerado correto, confiável e certificado. Esta dimensão permite observar a sua relação com a confiabilidade, o que ressalta a sua importância, uma vez que avaliar a confiabilidade, bem como determinar se um dado é correto e certificado é de extrema relevância para o contexto da educação.

A precisão, considerada em cinco trabalhos, pode variar de acordo com suas medições sobre um determinado valor. Wang e Strong (1996) afirmam que a precisão é obtida a partir da medida em que os conteúdos estejam corretos, confiáveis e livres de erros.

A temporalidade, identificada em quatro trabalhos, é definida por Wand e Wang (1996) como o limite para o uso apropriado do dado. Bovee et al (2003) abordam dois componentes para a

dimensão de temporalidade. Primeiramente a idade, quão velho um determinado dado é, e depois, a volatilidade, que aborda sobre a instabilidade e frequência de mudança de um atributo pertencente a uma entidade. Para Liu e Chi (2002), esta dimensão é utilizada de forma a analisar se os dados estão atualizados suficientemente para realizar uma tarefa.

Por fim, a dimensão consistência, identificada em apenas um trabalho, tem sua relevância no sentido de determinar conflitos entre os dados, verificando se sua função e a posição são válidas a partir de dados anteriores (WANG; STRONG, 1996). Para alcançá-la é necessário considerar todo o contexto e avaliar a sua utilização diante do cenário. No contexto de qualidade é uma dimensão que pode impactar diretamente o resultado, uma vez que sua abordagem se baseia mediante regras semânticas sobre um determinado conjunto de dados.

As dimensões abordadas nos trabalhos demonstram a existência de lacunas no que diz respeito a pesquisas relacionadas à qualidade de dados em ambientes educacionais. O contexto educacional, em especial a modalidade à distância, apresenta especificidades próprias, especialmente por requerer uma comunicação bidirecional. Logo, o apoio tecnológico referente a obtenção de dados para subsidiar tomadas de decisões pedagógicas é extremamente importante e conseqüentemente a validação da qualidade dos dados também é fator decisivo para a eficiência destas decisões. Redgate (2016) afirma que, em educação, não pode haver tolerância para dados defeituosos;

quanto maior a acurácia, traz-se mais conforto a estudantes, responsáveis, professores e gestores.

Diante disso, questões relacionadas à atualidade, acessibilidade, corretude, rastreabilidade e recuperabilidade, dimensões estas sugeridas pela ISO/IEC 25012:2008, devem ser consideradas no contexto de sistemas informacionais educacionais na contribuição para uma decisão mais assertiva e que poderá acarretar melhores resultados na performance dos alunos.

5. Considerações finais

Há imensidão de dados educacionais disponíveis e gerados continuamente, traz consigo uma dificuldade de processamento e geração de informações úteis, devido à alta granularidade dos dados e por se tratar de um ambiente que comporta inúmeros contextos. Perante este desafio, é necessária a preocupação com a qualidade desses dados para gerar informações que possibilitem um melhor suporte às tomadas de decisões pedagógicas.

No atual cenário, pesquisas sobre a qualidade dos dados são um importante instrumento para a adequação dos dados e a melhoria dos ambientes educacionais. Por meio delas é possível ampliar os benefícios e o potencial da utilização dos dados. Neste contexto, as evidências encontradas neste estudo permitem concluir que existem poucos trabalhos que abordam a qualidade dos dados, além de poucas dimensões encontradas nestes trabalhos, o que possibilita a conclusão de que ainda há muitas

relações inexploradas que podem trazer maior qualidade informacional no suporte à tomada de decisão pedagógica.

Para realização deste, se faz necessário explorar as maneiras pelas quais a qualidade dos dados nos ambientes educacionais são abordadas. Para tal, a ciência da informação traz uma relevante contribuição, por ser uma área que se caracteriza pela interdisciplinaridade e que é dedicada a estudos relacionados à informação, com foco em atender as necessidades dos usuários da melhor forma possível.

Ao se levar em conta o objetivo deste trabalho, a natureza interdisciplinar e colaborativa da Ciência da Informação e os interesses relacionados aos professores e gestores, a exploração de estudos que se remetem à qualidade devem contemplar diversas óticas em múltiplas dimensões, essas ainda carentes e que podem ser consideradas em trabalhos futuros.

Com isto, acredita-se que a presença da Ciência da Informação em estudos referentes à qualidade de dados pode possibilitar a identificação, o reconhecimento de fatores e características que propiciem uma ampliação da harmonia entre os profissionais envolvidos no processo e na performance da utilização desses dados, o que traz maior desenvolvimento à área.

Referências

BATINI, C.; SCANNAPIECO, M. **Data and information quality: Dimensions, Principles and Techniques**. Cham, Switzerland: Springer International Publishing, 2016.

BOVEE, M.; SRIVASTAVA, R. P.; MAK, B. A conceptual framework and belief-function approach to assessing overall information quality. **International journal of intelligent systems**, v. 18, n. 1, p. 51-74, 2003.

CEREZO, R. et al. Students' LMS interaction patterns and their relationship with achievement: A case study in higher education. **Computers & Education**, v. 96, p. 42-54, 2016.

DI RAIMO, L. C. F.; PETERMANN, R. Perguntas de leitura: uma abordagem na perspectiva da interação. **Entretextos**, v. 16, n. 1, p. 145-169, 2016.

DIAS, S. B. et al. Fuzzy cognitive mapping of LMS users' quality of interaction within higher education blended-learning environment. **Expert systems with Applications**, v. 42, n. 21, p. 7399-7423, 2015.

FLORIDI, L.; ILLARI, P. (Ed.). **The philosophy of information quality**. Cham, CH: Springer, 2014.

FÜRBER, C. Data Quality in the Semantic Web. In: **Data Quality Management with Semantic Technologies**. Springer Gabler, Wiesbaden, 2016.

HENRIE, C. R. et al. Exploring the potential of LMS log data as a proxy measure of student engagement. **Journal of Computing in Higher Education**, v. 30, n. 2, p. 344-362, 2018.

KADOIĆ, N.; OREŠKI, D. Analysis of student behavior and success based on logs in Moodle. In: 2018 41st **International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)**. IEEE, 2018. p. 0654-0659.

KAZANIDIS, I. et al. Measuring and mining LMS data. In: 2012 16th Panhellenic Conference on Informatics. **IEEE**, 2012. p. 296-301.

LIU, L.; CHI, L. Evolutional Data Quality: A Theory-Specific View. In: **ICIQ**. 2002. p. 292-304.

MACFADYEN, L. P.; DAWSON, S. Mining LMS data to develop an “early warning system” for educators: A proof of concept. **Computers & education**, v. 54, n. 2, p. 588-599, 2010.

RANJEETH, S.; LATCHOUMI, T. P.; PAUL, P. V. A survey on predictive models of learning analytics. **Procedia Computer Science**, v. 167, p. 37-46, 2020. REDGATE, Russ. **What Matters the Most with Education Data**. eScholar, 2016. Disponível em: <https://escholar.com/2016/03/why-data-quality-matters/>. Acesso em 27 de maio de 2021.

SAMITSCH, C. **Data Quality and its Impacts on Decision-Making: How Managers can benefit from Good Data**. Germany: Gabler Verlag. DOI 10.1007/978-3-658-08200-0, 2015.

SANTOS, P. L. V. A. da. C.; SANTANA, R. C. G. Dado e Granularidade na perspectiva da Informação e Tecnologia: uma interpretação pela Ciência da Informação. **Ciência da Informação**, v. 42, n. 2, 2013.

SCHILDKAMP, K. Data-based decision-making for school improvement: Research insights and gaps. **Educational research**, v. 61, n. 3, p. 257-273, 2019.

VILLEGAS-CH, W.; ROMÁN-CAÑIZARES, M.; PALACIOS-PACHECO, X. Improvement of an online education model with the integration of machine learning and data analysis in an LMS. **Applied Sciences**, v. 10, n. 15, p. 5371, 2020.

WAND, Y.; WANG, R. Y. Anchoring data quality dimensions in ontological foundations. **Communications of the ACM**, v. 39, n. 11, p. 86-95, 1996.

WANG, R. Y. A product perspective on total data quality management. **Communications of the ACM**, v. 41, n. 2, p. 58-65, 1998.

WANG, R. E.; STRONG, D. Beyond accuracy: What Data quality means to data consumers. **Journal of Management Information System**, v.12, n.4, p.5-33. 1996.

YOU, J. W. Identifying significant indicators using LMS data to predict course achievement in online learning. **The Internet and Higher Education**, v. 29, p. 23-30, 2016.

Interoperabilidade e integração entre DSpace e Archivematica

Laura Marques de Magalhães^a e Felipe Augusto Arakaki^b

1. Introdução

A Internet modificou a forma como a informação é gerada e transmitida entre os meios de comunicação. Os usuários encontram-se em um novo contexto de busca pela informação e o seu mediador, profissional da informação, deve avaliar suas necessidades de forma que o objetivo do fluxo informacional seja alcançado.

Araújo Junior (2007, p.65) afirma que no processo de busca e recuperação da informação, a motivação do usuário está relacionada a sua necessidade informacional, e a recuperação da informação deve estar próxima à expectativa ou demanda informacional do que foi demandado por esse usuário.

Tendo em vista a interação da informação com o meio digital, a *World Wide Web* – WWW (ROBREDO; BRÄSCHER, 2010), criada por Berners-Lee, inicia em 1989 uma inovação no compartilhamento automatizado de informações. Juntamente com a popularização da Web na década de 90, observou-se a preocupação em relação à localização e recuperação da

a Graduada em Arquivologia pela Universidade de Brasília – UnB. E-mail: lauramarques.m@gmail.com.

b Doutor em Ciência da Informação pela Universidade Estadual Paulista – UNESP. Professor na Universidade de Brasília – UnB. E-mail: fe.arakaki@gmail.com. ORCID: <https://orcid.org/0000-0002-3983-2563>.

informação e como alternativa para esse problema, foram adotados metadados para descrever e localizar recursos informacionais.

Os metadados são importantes para organizar os recursos eletrônicos, assim como para a necessidade de constante descobertas e informações a serem disponibilizadas na internet e nas intranets. A utilização de padrões de metadados proporciona aos sistemas de informação e de gestão do conhecimento, que recursos e aplicações sejam integrados e compartilhados (ALVES; SOUZA, 2007).

Os metadados são caracterizados como: estrutura padronizada de descrição (ALVES, 2010, p. 46-47); transformação dos dados em conhecimento (IKEMATU, 2001) e a representação, estruturação, gerenciamento, preservação, uso e reúso de informações (ARAKAKI, 2019).

Um dos padrões que é utilizado mundialmente é o padrão de metadados Dublin Core (DC), que se destina a descrever os objetos digitais, sejam eles textos, sons, imagens, vídeos ou sites da web. O DC, conhecido por ser um padrão de metadados para Web, focou em elementos que pudessem representar a informação de uma forma referencial entre diferentes instituições com vistas a recuperação da informação na internet (GRÁCIO, 2002, p. 31). O DC pode ser utilizado com diversas linguagens como *HyperText Markup Language* (HTML), *Extensible Markup Language* (XML), dentre outras.

Nesse âmbito do contexto dos metadados, apresentam-se sistemas de informação, como o DSpace e o Archivematica, que foram construídos, o primeiro software, com o objetivo de captura, armazenamento e disseminação do objeto digital e o segundo software, foi construído com o objetivo de gestão e acessibilidade do objeto digital.

Para Smith et al., (2003) o DSpace é um sistema de código aberto, disponível gratuitamente para instituições, o qual pode ser adaptado ou expandido conforme necessidade. O gerenciamento dos materiais de pesquisa e de publicações, a partir do momento que se utiliza do sistema, promove uma maior visibilidade e acessibilidade à comunidade e usuários.

O Archivematica, é um software livre assim como o DSpace, e tem o objetivo de auxiliar no processo de armazenamento de documentos em formato digital. Além disso, o Archivematica segue normas e padrões para preservação digital, como o padrão ISO-OAIS.

O Padrão ISO-OAIS é definido como um modelo de referência que a partir de um esquema lógico, orienta um sistema de arquivo dedicado à preservação e manutenção do acesso a informações digitais a longo prazo (CCSDS, 2012).

Os dois softwares livres, são grandes aliados para a área da Ciência da Informação, com funções ligadas ao compartilhamento de dados, preservação e garantia de acesso aos usuários.

O objetivo principal desta proposta é de apresentar a possível interoperabilidade entre os softwares DSpace e Archivematica, de forma que, as informações sejam de fácil acessibilidade para seus usuários e preservadas a longo prazo.

2. Procedimentos metodológicos

Os procedimentos metodológicos desta proposta abarcam buscas e leituras de artigos e textos científicos, por meio de pesquisa exploratória e qualitativa da literatura, a fim de compreender e apresentar conceitos e ligações entre os softwares DSpace e Archivematica. Os termos da estratégia de busca foram por meio de palavras-chave, como “Biblioteca Digital” “Preservação Digital” “Dublin Core” “DSpace” “Archivematica”. Não obstante, foram utilizadas como base de dados: Scielo, Google Acadêmico, página Web das revistas científicas e de publicações e livros, escolhidas pela confiabilidade e acessibilidade, além de citações vindas de cientistas da informação. O recorte temporal compreende os anos entre 2002 e 2020.

3. Resultados e discussões

Nesta seção será apresentado os conceitos do padrão de metadados Dublin Core, dos sistemas de informação e a proposta de integração entre os dois softwares: Archivematica e DSpace.

3.1 Padrão de Metadados Dublin Core

A origem do padrão DC data da 2ª Conferência Internacional *World Wide Web*, de 1994, onde ocorreram discussões sobre semântica e na descoberta de recursos de informação relacionados à Web. Em 1995, com a realização de um evento da OCLC (*Online Computer Library Center*) e o NCSA (*National Center for Supercomputing Applications*), as discussões foram pautadas num conjunto semântico da busca e recuperação de recursos da Web. O resultado dessas discussões culminou nos elementos de metadados intitulado DC cujo nome faz jus à localização do evento em Dublin, Ohio. O “*Core*” corresponde aos princípios do padrão, núcleo de informações, ou seja, núcleo base de informações para localização de documentos na Web. (DUBLIN CORE METADATA INICIATIVE, c2015; HARPER, 2010).

A instituição que administra o DC é a *Dublin Core Metadata Initiative* (DCMI), uma iniciativa que foi concebida em 1995, em Ohio, nos Estados Unidos, que visa o desenvolvimento de padrão de metadados para organização da informação. A Iniciativa de Metadados Dublin Core (*Dublin Core Metadata Initiative* – DCMI) é categorizada como uma organização que promove a adoção e interoperabilidade dos padrões de metadados.

O DC possui dois níveis de descrição. O primeiro, DC Simple, que possui 15 elementos descritivos; e, o outro, DC Qualificado, que possui elementos que qualificam o metadado, caracterizando e atribuindo mais semântica ao metadado. Seus

elementos e qualificadores auxiliam na descrição dos objetos digitais.

Os metadados são fundamentais para promover a interoperabilidade entre sistemas, como o DSpace e o Archivematica. Neste contexto, o DC é uma alternativa recomendada pela *Open Archives Initiative* (OAI) para interoperabilidade entre sistemas de informação e utilizado muito utilizado em Repositórios e Bibliotecas digitais. O DC possui ainda duas (2) ISO que regulamenta e estabelece sua estrutura e metadados como a ISO 15836-2017, “*Information and documentation – The Dublin Core™ metadata element set*”, e a ISO 15836-2, “*Information and documentation – The Dublin Core™ metadata element set – Part 2: DCMI Properties and classes*”.

3.2 Sistemas de informação e preservação digital

Os sistemas de informação buscam de alguma forma, atender os requisitos de um ambiente digital ligado aos centros de informação como o arquivo, biblioteca e/ou museu. A crescente inserção da informação no meio digital levanta algumas questões pontuais, especialmente voltadas para a preservação digital.

Torino (2017) lembra que “[...] apenas a preservação digital poderá assegurar que o documento possa ser utilizado em longo prazo [...]” e que além dos arquivos, os metadados, sejam eles descritivos ou de direito autoral ou técnicos, e a estrutura do documento devem ser considerados no âmbito da preservação.

Para Sayão (2010), a revolução digital transforma constantemente o modo de criação, comunicação e preservação

do conhecimento científico por meio dos acadêmicos, de forma que, a preocupação maior é com a proteção dos conteúdos a longo prazo.

Na área da Biblioteconomia, por exemplo, em muitas instituições o DSpace é utilizado como o software que possibilita a um repositório digital o armazenamento, gerenciamento, preservação e o acesso à produção intelectual. Já a área da Arquivologia, possui normas e padrões internacionais que influenciam diretamente na preocupação com a preservação da informação digital, a obsolescência tecnológica e o armazenamento dos documentos.

As duas áreas de forma síncrona, preocupam-se em preservar o objeto digital, dar acesso aos usuários e aliar sempre que possível aos softwares do mercado para atingir seus objetivos relacionados à informação. Dessa forma, foram abordados dois sistemas de informação: o DSpace e o Archivematica.

O DSpace possui uma arquitetura simples e eficiente, utiliza-se de uma tecnologia voltada ao acesso aberto, com o propósito de atender demandas de Repositórios Institucionais para armazenamento, preservação e disseminação de registros e traduz a estrutura organizacional da institucional mediante seu sistema de comunidades/coleções (ROSA; MEIRELLES; PALACIOS, 2011).

As características mais evidenciadas do DSpace são (VIANA; MÁRDERO ARELLANO; SHINTAKU, 2005): é um software livre; a arquitetura de software é simples e eficiente; uso da

tecnologia de ponta; direcionado para o acesso aberto; e implementado com o intuito de servir intencionalmente de repositório institucional.

A partir disso, pode-se inferir que o DSpace é um dos softwares livres mais utilizado em Repositórios e Bibliotecas digitais, levando em conta suas características de acesso aberto e fácil utilização.

Já o Archivematica configura-se como um software essencial para a preservação e o acesso a longo prazo dos objetos digitais. Em conformidade com o modelo OAIS, este modelo apresenta três tipos de pacotes de informação, com seus devidos metadados específicos para cada pacote. Os pacotes se dividem em:

- SIP: Pacote de informação para submissão – está relacionado à admissão de documentos digitais e seus metadados associados;
- AIP: Pacote de informação para arquivamento – está relacionado ao acondicionamento e armazenamento de documentos digitais e seus metadados associados; e
- DIP: Pacote de informação para disseminação – está relacionado ao acesso aos documentos digitais e seus metadados associados (CONSELHO NACIONAL DE ARQUIVOS, 2015, p.3).

3.3 Proposta de integração entre o DSpace e o Archivematica

Após a análise da literatura sobre a temática de repositórios e bibliotecas digitais, preservação digital, metadados e sistemas de

informação de código aberto, pensou-se em uma proposta de integração entre o DSpace e o Archivematica conforme apresentada na Figura 1.

Figura 1 – Fluxo para integração entre DSpace e o Archivematica



Fonte: Autores.

Na primeira seta, apresenta-se o processo de arquivamento. Nela, um item, uma coleção ou até mesmo uma comunidade do DSpace pode ser escolhida para a integração.

Na segunda seta, há a exportação desse item, coleção ou comunidade para o Archivematica. Um arquivo ZIP é gerado para ser feita a leitura pelo software. Em *transfer*, o Archivematica recebe o pacote do DSpace e o transforma em SIP. Lembrando que, um arquivo “mets.xml” é gerado e nele conterà os metadados Dublin Core.

Na terceira seta, após a leitura do arquivo pelo Archivematica, há uma etapa de *ingest*, na qual o pacote SIP é selecionado e após o pacote ser aceito, o SIP se transformará em um AIP para armazenar o pacote de informação. Para que o diretório “item/coleção/comunidade” seja preservado e acessado futuramente, é obrigatório escolher a opção “*Normalize for preservation and access*” no Archivematica.

Na quarta seta, no processo de recolhimento pelo Archivematica, em *Archival Storage*, o pacote será processado (surgimento do DIP) e disponibilizado para download. Feito o download no computador, têm-se os dois pacotes DIP e AIP. O primeiro pacote de disseminação e o segundo, de arquivamento. No pacote de arquivamento têm-se os arquivos *bagit* que garantem que o arquivo está completo, contendo todos os metadados e demais informações do pacote. No pacote de disseminação, o intuito é o de acesso.

Com a integração entre os dois softwares, DSpace e Archivematica, a preservação digital e o acesso a longo prazo dos objetos digitais estão assegurados, conforme exemplificado na última seta com o SIP, AIP e DIP.

4. Considerações finais

A evolução da Web modificou o processo informacional em que, anteriormente, os dados eram apresentados sem uma relação intrínseca e semântica. Com a inserção dos metadados, utilizados para representar a informação, o processo de busca e recuperação da informação tem sido cada vez mais intuitivo e completo.

Para a área científica, a utilização dos metadados e dos sistemas de informação possibilitam que os princípios da preservação digital a longo prazo sejam cumpridos, e o usuário, principal consumidor e beneficiário, tenha acesso independentemente de fatores como lugar ou tempo.

A Iniciativa Dublin Core dá origem ao padrão DC, que se estabelece como um conjunto de elementos de metadados

(simples e qualificado), utilizado para facilitar a descrição dos recursos da Web. A iniciativa promove a adoção e interoperabilidade do padrão DC pela sua facilidade de uso, simplicidade e aplicabilidade.

A integração demonstra que, por meio da utilização dos dois softwares, DSpace e Archivematica, um item, uma coleção ou até mesmo uma comunidade podem ser preservados e garantido seu acesso futuro pelos usuários.

Do ponto de vista técnico, a integração efetua-se pela exportação dos dados em arquivo ZIP do software DSpace para o software Archivematica, sem a necessidade de outro sistema ou programa para integrar, no qual o próprio arquivo ZIP cumpre o plano de preservação do Archivematica pelo modelo OAIS, e este mesmo arquivo é processado em cada pacote SIP, AIP e DIP para assegurar a confiabilidade e acesso desse arquivo a longo prazo.

Observou-se que para uma interoperabilidade bem-sucedida entre o DSpace e o Archivematica é necessário a adoção de metadados e que a integração pode ser aplicada em qualquer instituição que utilize os dois softwares e queira preservar e dar acesso aos seus objetos digitais.

Alguns dos casos que implementaram a integração é de uma tese do autor Francisco Lopes de Aguiar, da Universidade de São Paulo, no domínio da Sociedade Brasileira para o Progresso da Ciência – SBPC. Há a possibilidade de integração com um terceiro software chamado ArchivesSpace. Esta integração tripla,

melhora a criação e reutilização de metadados e agiliza a ingestão e o depósito de arquivos digitais (ARCHIVESSPACE, 2021).

Agradecimentos

Agradeço o apoio técnico e computacional do Laboratório de Tecnologias da Tomada de Decisão – LATITUDE, da Universidade de Brasília, que conta com apoio do CNPq – Conselho Nacional de Pesquisa (Outorgas 312180/2019-5 PQ-2, BRICS2017-591 LargEWiN e 465741/2014-2 INCT em Cibersegurança), da CAPES – Coordenação de Aperfeiçoamento do Pessoal de Nível Superior (Outorgas 23038.007604/2014-69 FORTE e 88887.144009/2017-00 PROBRAL), da FAP-DF – Fundação de Amparo à Pesquisa do Distrito Federal (Outorgas 0193.001366/2016 UIoT e 0193.001365/2016 SSDDC), do Ministério da Economia (Outorgas 005/2016 DIPLA e 083/2016 ENAP), da Secretaria de Segurança Institucional da Presidência da República do Brasil (Outorga ABIN 002/2017), do Conselho Administrativo de Defesa Econômica (Outorga CADE 08700.000047/2019-14), da Advocacia-Geral da União (Outorga AGU 697.935/2019), do Ministério das Cidades, (Outorga MC 01/2019), do Ministério da Justiça e Segurança Pública, (Outorga MJSP 01/2019) e dos Decanatos de Pesquisa e Inovação e de Pós-Graduação da Universidade de Brasília (DPI/DPG/UnB).

Referências

AGUIAR, F. L. de. Dspace e Archivematica: concepção e criação de um protótipo de repositório digital aplicado no

domínio da SBPC: sob uma perspectiva interdisciplinar entre Arquivística e Organização e Representação do Conhecimento. Tese (Doutorado em Ciência da Informação – Escola de Comunicação e Artes. Universidade de São Paulo), São Paulo, 2018. 667f.

ALVES, R. C. V. **Metadados como elementos do processo de catalogação.** 2010. 132 f. Tese (doutorado) - Universidade Estadual Paulista, Faculdade de Filosofia e Ciências, 2010. Disponível em: <http://hdl.handle.net/11449/103361>. Acesso em 24 mai. 2021.

ALVES, M. das D. R.; SOUZA, M. I. F. Estudo de correspondência de elementos metadados: DUBLIN CORE e MARC 21. RDBCI: **Revista Digital de Biblioteconomia e Ciência da Informação**, Campinas, SP, v. 5, n. 1, p. 20-38, 2007.

ARAKAKI, F. A. Metadados e modelo PROV: perspectivas dos dados de proveniência em contextos digitais. **Informação & Informação**, Londrina, v. 25, n. 3, p. 187-211, jul./set. 2020. Disponível em: <http://www.uel.br/revistas/uel/index.php/informacao/article/view/41869>. Acesso em: 24 mai. 2021.

ARAÚJO JUNIOR, R. H. de. **Precisão no processo de busca e recuperação da informação.** Brasília: Thesaurus, 2007.

ARCHIVESSPACE. **Integrations with Archivesspace.** Disponível em:

<https://archivesspace.org/developers/integrations-with-archivesspace>. Acesso em: 24 jun. 2021.

CCSDS. CONSULTATIVE COMMITTEE FOR SPACE DATA SYSTEM. **Reference Model for an Open Archival Information System (OAIS)**. Blue book (CCSDS 650.0- M-2). Washington, DC, 2012. Disponível em: <http://public.ccsds.org/publications/archive/650x0m2.pdf>. Acesso em: 24 mai. 2021.

CONSELHO NACIONAL DE ARQUIVOS (BRASIL). CÂMARA TÉCNICA DE DOCUMENTOS ELETRÔNICOS. **Diretrizes para a implementação de repositórios arquivísticos digitais – RDC-Arq**. Rio de Janeiro: Arquivo Nacional, 2015. Disponível em: http://conarq.gov.br/images/publicacoes/textos/diretrizes_rdc_arq.pdf. Acesso em: 25 mai. 2021.

DUBLIN CORE METADATA INITIATIVE. **Dublin Core Metadata Element Set, Version1.1**: Reference Description. July 1999. Disponível em: <https://www.dublincore.org/>. Acesso em: 25 mai. 2021.

GRÁCIO, J. C. A. **Metadados para a descrição de recursos da Internet: o padrão Dublin Core, aplicações e a questão da interoperabilidade**. Marília: Unesp, 2002. 104f. Dissertação (Mestrado) - Programa de Pós-Graduação em Ciência da Informação, Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília, 2002.

IKEMATU, R. S. Gestão de metadados: sua evolução na tecnologia da informação. DataGramZero, **Revista de Ciência da Informação**, v. 2 n. 6, dez, 2001. Disponível em: http://www.dgz.org.br/dez01/Art_02.htm. Acesso em: 24 mai. 2021.

ROBREDO, J.; BRÄSCHER, M. (Orgs.). **Passeios pelo bosque da informação**: estudos sobre a representação e organização da informação e do conhecimento – eroic. Brasília. DF: IBICT, 2010. 335 p. ISBN: 978-85-7013-072-3. Disponível em: <http://www.ibict.br/publicacoes/eroic.pdf>. Acesso em: 10 jul. 2021.

ROSA, F.; MEIRELLES, R. F.; PALACIOS, M. Repositório institucional da universidade Federal da Bahia. **Inf. & Soc.:Est.**, João Pessoa, v.21, n.1, p. 129-141, 2011.

SAYÃO, J. Repositórios digitais confiáveis para a preservação de periódicos eletrônicos científicos. **Ponto de Acesso**, Salvador, v. 4, n. 3, p. 68-94, dez. 2010. Disponível em: <http://www.portalseer.ufba.br/index.php/revistaici/article/view/4709/3565>. Acesso em: 25 mai. 2021.

SMITH, M. et al. Dspace: an open source dynamic digital repository. **D-Lib Magazine**, 9, 1 (2003). Disponível em: <http://www.dlib.org/dlib/january03/smith/01smith.html>. Acesso em: 24 mai. 2021

TORINO, E. Políticas em repositórios digitais: das diretrizes à implementação. In: VECHIATO, F. et al. (org.). **Repositórios digitais: teoria e prática**. Curitiba: EDUTFPR, 2017. p. 91-114.

VIANA, C. L. de M.; MÁRDERO ARELLANO, M. A.; SHINTAKU, M. **Repositórios institucionais em ciência e tecnologia uma experiência de customização do DSpace**. 2005. Disponível em:
<http://eprints.rclis.org/archive/00005563/01/viana358.pdf>.
Acesso em: 25 mai. 2021.

**Capítulo retirado por solicitação dos autores [pgs.
141-155]**

Retirado por solicitação dos autores

Retirado por solicitação dos autores

Retirado por solicitação dos autores

Retirado por solicitação dos autores

Retirado por solicitação dos autores

Retirado por solicitação dos autores

Retirado por solicitação dos autores

O encapsulamento informacional enquanto fator de aprimoramento na Experiência do Usuário (UX)

Guilherme Oliveira Pironi^a

1. Introdução

Dentro de um universo informacional cada vez mais mediado e dominado por dispositivos digitais operados através de Interfaces Humano Computador (IHC) fica evidente de forma crescente a necessidade de facilitar o acesso a tal, corroborando com os apontamentos de INTERPARES (2017) que afirma não haver mais forma de refrear o levante digital vivido atualmente. Sendo assim áreas do conhecimento como a Ciência da Informação e o Design de Interface e Experiência do Usuário (UI/UX Design) se veem compelidas a trabalhar de forma conjunta para avançar em direção a tal objetivo por meio das mais diversas técnicas e ferramentas.

Uma estratégia a ser explorada dentro deste recorte é o de encapsulamento da informação, que segundo Flores e Santos (2015) pode ser sintetizado como o agrupamento dos arquivos digitais em sua forma originária, por exemplo bits e linhas de código, podendo englobar inclusive o software requerido para sua visualização e interpretação, ou seja, o agrupamento de uma série de funções e informações dentro de um mesmo objeto e

a Especialista em Mídias Digitais Interativas pela Universidade do Oeste Paulista – UNOESTE. E-mail: guilherme.pironi@unesp.br.

disponibilizá-lo pronto para executar determinada função informacional sem a necessidade daquele que o manipula, o usuário, conhecer os pormenores de seu funcionamento interno.

Adotando a definição de Busse e Uflacker (2007) que defendem que uma boa experiência do usuário está ligada a fatores como um design de produto que permita que a interação com este seja o mais simples possível, este artigo tem por objetivo relacionar ambos os conceitos no intuito de explorar o seguinte questionamento: seria o encapsulamento um fator positivo quanto à melhora da experiência do usuário ao manipular uma interface?

2. Desenvolvimento

Todos temos documentos guardados em nossas casas, escritórios comerciais ou mesmo repartições públicas como cartórios ou bibliotecas; basta procurá-los no local específico onde foram deixados pela última vez e lá estão eles em sua inegável materialidade, prontos para nos fornecer todas suas qualidades informacionais sejam quais forem: cores, formato de cada caractere, logotipos, marcas d'água e obviamente as informações geradas a partir da interpretação destes sinais supracitados. Se tratando de tais documentos físicos é fácil perceber a totalidade de suas qualidades tanto materiais quanto de conteúdo, porém ao adentrar o universo digital o panorama se transforma, Arellano (2008) destaca que neste caso o processo ganha camadas extras de complexidade onde a preservação digital passa a depender também de repositórios de dados que garantam

a longevidade e acesso a tais conteúdos. Podemos elencar alguns níveis para abarcar e administrar estes objetos digitais seja enquanto objeto físico, codificação lógica, objetos conceituais ou significativos para os sujeitos e enquanto agrupamento de partes essenciais a serem resguardadas com a finalidade de oferecer a essência do objeto (WEBB, 2003).

Seguindo esta lógica começam a surgir estratégias para que não seja necessário contemplar e manipular diretamente todas as minúcias contidas na cadeia de procedimentos e informações que ocorrem para que o usuário possa interagir com a informação digital desejada. Entre elas o presente artigo destaca o encapsulamento, um conceito que pode ser sintetizado como a “preservação da informação necessária para o desenvolvimento de conversores, emuladores e visualizadores de objetos digitais.” (ARELLANO, 2008, p. 353), quer dizer, agrupar informações, dados, funções de software e quaisquer outras qualidades que necessárias ao acesso a tal conteúdo de maneira a eliminar uma série de intermediários e assim diminuir a complexidade da tarefa.

Em se tratando do funcionamento do aparato digital no qual busca-se acessar a informação, podemos adaptar a definição acima para relacionar-se melhor com a programação orientada a objeto, assim Barros et al. (2003) aponta o encapsulamento como uso de uma interface que oculta o detalhes de acesso e modificação dos dados, permitindo diversidade em seus usos sem afetar o código original encapsulado pois estes serão definidos pela própria interface porém sem a necessidade de revelar os

detalhes de sua implementação. A estratégia do encapsulamento apresenta pontos positivos significativos quanto à preservação e facilitação de acesso a dados e informações, destacando entre as mais significativas a sua aplicabilidade por permitir a aplicação desta em informações das mais diversas naturezas e formatos Flores e Santos (2015). Vale lembrar que dificilmente a estratégia de encapsulamento é utilizada sozinha, costumando vir acompanhada de outras como migração, emulação e conversão, porém, para os fins deste artigo, estas não serão exploradas.

A passagem do modelo de documento físico, supracitado, para o modelo digital sugere uma mudança no prisma pelo qual este é encarado, já que o contato com este segundo modelo difere-se principalmente por conta do elemento da interatividade ao lidar com os dados e informações (LÉVY, 1993).

Esta interação sugere um tipo específico de experiência denominada de Experiência do Usuário (*User Experience*, UX, em inglês) que vem a estar relacionada com o conceito de usabilidade do sistema que, segundo Ferreira e Leite (2003), pode ser entendida como a referente à utilização eficiente de um sistema a fim de realizar as tarefas para as quais ele foi desenhado, visando sempre um número mínimo de erros e o menor grau possível de dificuldade. Vista disso podemos adaptar esta definição ao nosso contexto como o acesso aos dados e informações requeridos por intermédio de uma experiência com as características supracitadas. Dentro deste recorte é preciso esclarecer que a Experiência do Usuário tem seu foco na nas emoções

experimentadas pelo utilizador, enquanto a usabilidade se atém às facetas intrínsecas dos produtos e da tecnologia utilizada (NORMAN, 2008).

São diversos os fatores que influenciam na boa UX e estes podem variar dependendo da ação a ser realizada, porém Cerejo (2001) defende alguns princípios básicos que devem permear qualquer uma delas como por exemplo a redução do número de etapas no processo o tanto quanto possível sem que haja perda na sua efetividade de sua execução, a simplificação destes processos e, por consequência, a facilidade em sua completude.

Desta forma fica sugerido que a Experiência do Usuário tem grande valia na interação entre humano e computador ao explorar aspectos não só voltados para o patamar objetivo, como é o caso da disciplina de usabilidade, como também subjetivos ao lançar luz sobre as emoções e sensações mais profundas do interador. Nesta prerrogativa Ferreira (2018) conclui que o campo ajuda a suprir as necessidades informacionais do usuário no momento em que ocorre a interação deste com o ambiente informacional.

Computadores e outros aparatos digitais funcionam a partir de uma lógica bem definida onde é necessária uma ação específica para obter um resultado específico correspondente, ou como explica Tanenbaum (1992), fornecemos um conjunto de instruções familiares à máquina que determinarão como esta deve executar a tarefa.

O cerne das instruções em questão não será fornecido de forma natural como seria feito em uma conversa entre duas

peças e sim utilizando a realidade lógica da máquina, gerando o problema apontado por Domingues (2001) que se caracteriza no fato de tais instruções fazerem sentido para o computador ou, com sorte, para um profissional da informática, mas dificilmente terão algum respaldo para o usuário comum. É importante adicionar que Domingues (2001) esclarece que a tarefa realizada através de tais instruções pode ser caracterizada dentro de um amplo leque que vai desde a manipulação de um arquivo de texto, ao tratamento de uma imagem ou mesmo um código de programação em uma linguagem como Pascal ou Cobol.

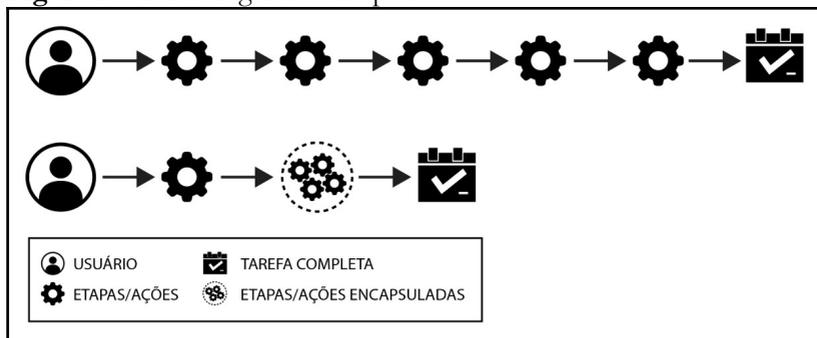
A relação destas ideias com o encapsulamento e a experiência do usuário pode ser feita por meio da noção proposta por Shneiderman et al. (2016) de que para que seja realizada uma tarefa o usuário deve se concentrar diretamente nela e não na manipulação da ferramenta sendo utilizada para sua realização, de forma a se sentir em contato direto e real com a informação ou dado manipulado, portanto ao se valer do encapsulamento é possível diminuir essa sensação de manipulação da ferramenta já que muitos destes processos estão encapsulados e ocorrerão de forma automática quando necessário.

Ao compreender o encapsulamento como um conjunto de dados, funções e informações a serem requisitados e executados quando necessário, junto da ideia também apresentada anteriormente de que uma boa experiência do usuário é aquela que permite que ele consiga realizar a ação requerida de maneira simples e sem mais esforço do que o necessário, logo podemos

sugerir que ao não precisar lidar com toda a complexidade informacional do conteúdo encapsulado, obtendo apenas os desenlaces práticos e diretos do resultado das interações deste conjunto encapsulado, o usuário estaria sendo poupado de uma complexidade desnecessária e portanto, apoiado na relação de simplicidade da ação e redução do número de etapas para sua completude proposto por Cerejo (2001), este se beneficiaria com uma experiência mais agradável na sua interação com a máquina.

Para ilustrar melhor este conceito a Figura 1 apresenta uma representação de uma tarefa que, se executada normalmente do início ao fim pelo usuário, exigiria um total de 5 etapas em comparação com a mesma tarefa executada exigindo apenas 2 etapas aparentes ao usuário pelo fato de que 4 delas foram encapsuladas, ou seja, o usuário passaria por menos etapas e conseqüentemente a ação se torna mais simples de ser executada e promove uma experiência menos custosa.

Figura 1 – Tarefa regular e encapsuladas



Fonte: Autor.

Trazendo esta ideia para um plano mais factual podemos imaginar uma situação onde o usuário deseja enviar uma mensagem de texto que acaba de redigir, ao apertar o botão “enviar” ela será direcionada ao destinatário de forma aparentemente automática porém estão encapsuladas neste botão uma série de etapas como, por exemplo, estabelecer uma conexão com o servidor que mediará a entrega da mensagem, codificar o conteúdo da mensagem de linguagem escrita para a linguagem binária da máquina e enviar este código ao servidor. É plausível assumir que cumprir todas estas etapas de forma não automatizada seria muito mais trabalhoso e por consequência menos prazeroso ou talvez até impossível para o usuário leigo, portanto sua experiência seria bastante prejudicada no processo.

3. Considerações finais

Este artigo apresentou uma breve análise da possível relação do encapsulamento de informações e dados com uma boa experiência do usuário executar tarefas por intermédio de uma Interface Humano Computador (IHC).

Foi observado que o encapsulamento tem como uma de suas características a junção de uma porção de informações e dados sob um mesmo conjunto a fim de torná-lo acessível quando necessário para que determinadas tarefas possam ser cumpridas.

Concomitantemente o conceito de Experiência do Usuário foi definido como as sensações e sentimentos que são despertados em um usuário ao interagir com determinado objeto a fim de realizar uma ação, sendo caracterizados como fatores

que podem levar a uma boa experiência a simplicidade e facilidade para realizar a tarefa desejada.

Com relação ao objetivo proposto é visto que se a simplicidade e facilidade podem ser catalisadores para uma boa Experiência do Usuário, portanto estratégias de gerenciamento de dados que promovam tais aspectos, como, por exemplo, um dos objetos de estudo aqui apresentado, o encapsulamento, nos direcionam a esta boa experiência.

Referências

BARROS, E. A. R.; PAMBOUKIAN, S. V. D.; ZAMBONI, L. C. **C++ Builder para Universitários**. São Paulo: Páginas & Letras, 2^a ed., 2003.

CEREJO, L. **Creating a hierarchy of user-experience needs**. ClickZ, 2001. Disponível em: <http://www.clickz.com/clickz/column/1696253/creatinghierarcy-user-experienceneeds>. Acesso em: 25 mai. 2021.

DOS SANTOS, H. M.; FLORES, D. Preservação de documentos arquivísticos digitais: reflexões sobre as estratégias de emulação. **Encontros Bibli**: revista eletrônica de biblioteconomia e ciência da informação, v. 20, n. 43, p. 3-19, 2015.

DOMINGUES, D. G. **O uso de metáforas na computação**. 2001. Tese. Universidade de São Paulo.

FERREIRA, A. M. J. F. da. C. Contribuições da experiência do usuário para a arquitetura da informação. 2018.

FERREIRA, S. B. L.; LEITE, J. C. S. P. Avaliação da usabilidade em sistemas de informação: o caso do sistema submarino. **Revista de Administração Contemporânea**, v. 7, n. 2, p. 115-136, 2003.

INTERPARES. INTERPARES 2 PROJECT. **Diretrizes do Produtor: a elaboração e a manutenção de materiais digitais: diretrizes para indivíduos. 2007.**

LÉVY, P. **As tecnologias da inteligência: o futuro do pensamento na era da informática.** Rio de Janeiro, Ed. 34, 1993. Coleção Trans., 2004.

MÁRDERO ARELLANO, M. Á. **Critérios para a preservação digital da informação científica. 2008.**

NORMAN, D. A. **Design emocional: por que adoramos (ou detestamos) os objetos do dia-a-dia.** Rio de Janeiro: Rocco, 2008.

UFLACKER, M.; BUSSE, D. **Complexity in enterprise applications vs. simplicity in user experience.** In: International Conference on Human-Computer Interaction. Springer, Berlin, Heidelberg, 2007. p. 778-787.

SHNEIDERMAN, B., PLAISANT, C., COHEN, M., JACOBS, S., ELMQVIST, N., DIAKOPOULOS, N. **Designing the user interface: strategies for effective human-computer interaction.** Pearson, 2016.

TANEMBAUM, A. S. **Organização Estruturada de Computadores.** Rio de Janeiro, PHC, 1992

WEBB, C. et al. **Guidelines for the preservation of digital heritage.** 2003.

Governança pública e as Tecnologias da Informação e Comunicação: um breve olhar sobre a Lei de Acesso à Informação e a Política de Dados Abertos no Brasil

Kelren Cecília dos Santos Lima da Mota^a, Diego Leonardo de Souza Fonseca^b e Roberto Lopes dos Santos Junior^c

1. Introdução

Desde a Constituição de 1988, o Brasil vem passando por um processo de mudança quanto a forma de gestão nas organizações públicas, visto que, durante o período da ditadura civil-militar, os direitos dos cidadãos foram coibidos, em especial o relacionado ao acesso às informações do Estado.

A implantação da redemocratização, a partir de 1985, possibilitou a participação da sociedade brasileira nas tomadas de decisões, dando origem a nova gestão pública baseada por resultados, melhores desempenhos e a participação mais ativa da sociedade, em prol da eficiência no emprego da verba pública e

-
- a Mestranda em Ciência da Informação pela Universidade Federal do Pará – UFPA. E-mail: kelrenlima27@gmail.com. ORCID: <https://orcid.org/0000-0002-5490-6460>.
 - b Doutorando em Ciência da Informação pela Universidade Estadual de Londrina – UEL. Bibliotecário e Documentalista na Biblioteca do Instituto Federal do Amazonas – IFAM. E-mail: diego.leonardo@uel.br. ORCID: <https://orcid.org/0000-0002-0105-1577>.
 - c Doutor em Ciência da Informação pela Universidade Federal do Rio de Janeiro – UFRJ. Professor na Universidade Federal do Pará – UFPA. E-mail: robertolopes@ufpa.br. ORCID: <https://orcid.org/0000-0001-6063-920X>.

incrementando a *accountability*⁶ governamental (responsabilidade com ética). Campos (1990) observa que a responsabilização das instituições e dos indivíduos pelos seus atos na gestão pública também envolve a necessidade de manter uma organização mais confiável e transparente.

Como forma de alcançar os objetivos propostos por essa nova gestão, adotou-se os princípios da governança pública, que visam gerenciar os recursos econômicos e sociais na perspectiva do desenvolvimento. Para que essa governança seja implementada nas instituições públicas, a mesma precisa dispor de instrumentos normativos que a potencializam, como por exemplo, a Lei de Acesso à Informação (LAI) e a Política de Dados Abertos.

Considerando o avanço dos recursos tecnológicos, as normativas mencionadas orientam quanto ao uso das Tecnologias de Informação e Comunicação (TIC) como ferramentas que potencializam o processo de democratização da informação, viabilizando o acesso à informação aos dados abertos por meio de canais no meio eletrônico como, por exemplo, e-SIC, websites, plataformas, bases de dados, dentre outros.

O percurso metodológico da pesquisa foi delineado em dois aspectos tipológicos: pesquisa documental e pesquisa bibliográfica. Foram analisadas a LAI e a Política de Dados Abertos, de forma preliminar e introdutória. Inicialmente, foi feita a contextualização de aspectos relacionados ao acesso à

6 Termo designado para diferentes campos do conhecimento que se trata sobre a necessidade de responsabilidade de indivíduos e organizações sobre determinado desempenho em relação a práticas de exercício de poder (SCHEDLER, 2004).

informação e aos dados abertos presentes nas normativas por meio da pesquisa documental e levantamento bibliográfico. Quanto a abordagem, trata-se de uma pesquisa qualitativa, cujos objetivos estão embasados em uma pesquisa descritiva e exploratória.

O estudo justifica-se pela importância de se discutir a evolução do processo de democratização da informação no âmbito brasileiro, a partir do processo de implementação de normativas vigentes como a LAI e a Política de Dados Abertos, visto que estas normativas visam garantir direitos fundamentais como acesso à informação, privacidade e liberdade, previstos na Constituição Federal de 1988.

2. Acesso a informação pública

O direito de acesso à informação possui, como essência, o binômio direito à verdade e direito à memória, ambas sendo premissas que caracterizam a reconstrução democrática após períodos ditatoriais e de relativização dos conceitos de dignidade humana (MOURA, 2014).

No contexto brasileiro, a Constituição Federal de 1988 prevê o acesso à informação pública como direito fundamental, presente em seu art. 5, inciso XXXIII, onde estabelece que:

[...] todos têm direito a receber dos órgãos públicos informações de seu interesse particular, ou de interesse coletivo ou geral, que serão prestadas no prazo da lei, sob pena de responsabilidade, ressalvadas aquelas cujo

sigilo seja imprescindível à segurança da sociedade e do estado [...] (BRASIL, 1988).

A Constituição de 1988 também consolidou a gestão de documentos como princípio constitucional que pode viabilizar e otimizar o acesso à informação pública. De acordo com o parágrafo 2º, do art. 216, “Cabem à administração pública, na forma da lei, a gestão da documentação governamental e as providências para franquear sua consulta a quantos dela necessitem” (BRASIL, 1988).

Jardim (2013) afirma que a ausência de políticas públicas arquivísticas e a forte tendência do Estado brasileiro inviabilizaram o processo de implantação da Lei de Arquivos nas esferas federal, estadual e municipal. Contudo, o autor também afirma que, mesmo diante das dificuldades, a Lei possibilitou avanços significativos na visibilidade dos acervos arquivísticos e da importância da gestão de documentos.

De acordo com Moura (2014) a denominada “Era de Transparência” deu poder e voz à sociedade civil, tornando-a mais ativa e exigente, garantindo então a efetivação do cumprimento legal e constitucional de acesso à informação. Portanto, o processo de institucionalização da transparência amplia e dá força ao cidadão, fortalecendo a trajetória democrática da sociedade, por meio do reconhecimento de que a desinformação não apazigua, mas reforça práticas como a injustiça e a impunidade.

Nesse cenário construído a parti da viabilização do acesso à informação em caráter legal pautado na política pública, cita-se o

Portal da Transparência do Governo Federal, lançado em 2014 pela Controladoria-Geral da União (CGU), sendo então considerado uma representação emblemática de iniciativa do poder público em disponibilizar, de forma acessível e viável, os dados do poder público para o cidadão⁷.

2.1 Tecnologias da Informação e Comunicação (TIC)

A Lei Federal nº 12.527 foi promulgada em 18 de novembro de 2011, também conhecida como a Lei da Transparência ou Lei de Acesso à Informação Pública, que regulamenta o direito de acesso à informação, previsto desde a Constituição Federal de 1988, e no qual dispõe de procedimentos e diretrizes relacionados ao acesso à informação pública, que devem ser seguidos por todas as esferas de governo: União, Estados, Distrito Federal e Municípios (BRASIL, 2011)

Conforme Jardim e Miranda (2015), o direito de acesso à informação pública previsto pela LAI, garante o direito do cidadão de ser informado e se informar quanto aos atos governamentais, além de enfatizar a dimensão epicêntrica da informação nas relações entre Estado e Sociedade Civil. A Lei nº 12.527/ 2011 tem como propósito fortalecer e cultivar a transparência e o controle social na gestão pública, a partir de diretrizes que obrigam as instituições divulgarem informações públicas de forma proativa, independente de solicitação, e em local de fácil acesso, denominada como transparência ativa. Ela

⁷ Portal da Transparência. Disponível em: <http://www.portaltransparencia.gov.br>. Acesso em: 07 jul. 2021.

também orienta que a informação deve ser divulgada de forma atualizada, eficiente e em linguagem de fácil compreensão (MOURA, 2014).

Ressalta-se que, o princípio da transparência, também é entendido como sendo o desdobramento do princípio constitucional da publicidade. No entanto, a LAI trouxe um novo conceito de “transparência”, ou seja, a “transparência ativa”, denominada como o princípio que exige de órgãos e entidades públicas a divulgação de informações de interesse geral, independentemente de terem sido solicitadas ou não. No entanto, é necessário que a disponibilização das informações e dos dados seja realizada de maneira clara, objetiva, estruturada e padronizada. (BRASIL, 2011)

Apesar do avanço ocorrido a partir da promulgação desta Lei relacionada à transparência, cita-se também o direito à privacidade, direito fundamental consubstanciado na Constituição Federal de 1988, em seu artigo 5º. Ligada a essa questão, em 2018, foi criada a Lei nº 13.709, de 14 de agosto de 2018, denominada de Lei Geral de Proteção de Dados Pessoais (LGPD), que dispõe sobre o tratamento de dados pessoais, inclusive nos meios digitais, com o objetivo de proteger os direitos fundamentais de liberdade, privacidade, e o livre desenvolvimento da personalidade da pessoa natural (BRASIL, 2018).

Com relação aos dados públicos, Sant’Ana (2009) afirma que garantir o acesso a esses dados possibilita ampliar a participação

da população no controle vertical da dinâmica entre o Estado e a Sociedade, contribuindo para o controle social. Os preceitos da transparência pública e do controle social estão contemplados na LAI. Para auxiliar no cumprimento da LAI no âmbito do Poder Executivo Federal, foram elaborados dois Decretos: nº 7.724/2012 e nº 8.777/2016, que ampliam e especificam o acesso à informação nas instituições públicas.

Mais recentemente, em 2021, de acordo com os dados do Painel de Acesso à Informação (21/06/2021)⁸, existem mais de 1.038.770 pedidos recebidos embasados na LAI, com um total de 99,48% respondidos pelos órgãos públicos, 0,50% em tramitação e 0,03% omissos. Esses dados reúnem todas as autarquias, fundações e órgãos diretamente ligados ao Governo Federal.

2.2 Política de Dados Abertos

Em 2016, foi sancionado o Decreto nº 8.777, que instituiu a Política de Dados Abertos do Poder Executivo Federal, estabelecendo normas para a livre utilização das bases de dados, e assegurando a possibilidade de uso e reúso dos dados públicos por parte da sociedade. Em 2019, foi criado o Decreto nº 9.903/2019 que altera a Política de Dados Abertos do Poder Executivo Federal, transferindo a responsabilidade pela gestão da política para a CGU que anteriormente estava a cargo do Ministério da Economia (BRASIL, 2016).

8 Dados coletados no Painel de Acesso à Informação. Disponível em: <http://paineis.cgu.gov.br/lai/index.htm>. Acesso em: 30 out. 2021.

No Decreto nº 9.903/2019, a CGU, além da competência de acompanhar e monitorar, de forma sistemática, o cumprimento dos Planos de Dados Abertos (PDAs), também adquiriu a responsabilidade pela gestão da Política de Dados Abertos do Poder Executivo Federal. O decreto também prevê que os dados publicados pelas entidades governamentais são de livre utilização pelo Estado e pela sociedade, disponibilizando gratuitamente as bases de dados das informações disponibilizadas (VIVARTA; MARTINS, 2009).

A Política de Dados Abertos obriga todas as instituições da administração pública federal direta, autárquica e fundacional a disponibilizarem as informações contidas em suas bases de forma aberta, ou seja, permitindo o livre acesso, utilização, modificação e compartilhamento (BRASIL, 2016). Cita-se aqui, também, o Portal Brasileiro de Dados Abertos, que é uma ferramenta criada pela CGU para encontrar e utilizar os dados públicos e as informações públicas, sendo de fundamental importância promover a mediação entre o poder público e a sociedade.

No entanto, para atender a demanda de dados primários é importante que o acesso aos dados governamentais siga alguns princípios que garantam a disponibilidade dos dados, sem restrições. Para que seja feito de forma eficiente cita-se, no tópico a seguir, o papel das tecnologias da informação e comunicação e a transparência pública.

3. Tecnologias da Informação e Comunicação e a transparência pública

As Tecnologias da Informação e Comunicação (TIC) estão presentes em várias áreas do conhecimento. Dando destaque à Ciência da Informação, entende-se que ela pode contribuir nas fases do projeto de construção das TIC com relação à informação e aos estudos em “camadas de abstração”, no qual essas camadas tratam do uso, armazenamento, recuperação e acesso à informação (RODRIGUES, 2012).

As TIC também estão presentes na gestão pública, sendo foco tanto dos gestores públicos quanto dos de empresas privadas. Com relação ao Estado, o uso dessas tecnologias classifica-se como interno (auxiliam nos processos internos de administração) e externo (contribui na relação entre a organização pública com entidades externas, podendo ser empresas, cidadãos ou grupo de cidadãos e outras organizações públicas) (SANT’ANA, 2009).

Na perspectiva de Rodrigues (2012), alguns órgãos governamentais ainda não utilizam as TIC de uso externo para essa finalidade, e sim recursos tecnológicos disponíveis no ambiente informacional digital, no intuito de disponibilizar informações tais como quadros de avisos, páginas institucionais, listas com contatos telefônicos e eletrônicos e localizações físicas (com logradouros e códigos postais).

Apesar dessa utilização, sabe-se que as TIC permitem uma interação mais direta e eficiente entre a Sociedade e o Estado, e

consequentemente, aumentam o nível de transparência entre ambos os lados, fortalecendo a democracia. frisa-se que, para o progresso e sucesso desse processo faz-se necessário a participação de lideranças políticas, investimentos em TIC, o conhecimento técnico dos artefatos tecnológicos, e a colaboração entre Estado e sociedade civil (RODRIGUES, 2012).

De acordo com Vivarta e Martins (2009), normativas como a LAI e a Política dos Dados Abertos, foram elaboradas para auxiliarem no processo de implementação da cultura da transparência pública nas organizações públicas, com o auxílio das TIC, para viabilizar o processo de democratização dos dados abertos e das informações públicas de interesse para a sociedade.

4. Resultados e discussões

A contextualização do acesso à informação e aos dados abertos ocorreu a partir das normativas Lei Nº12.527/2011 e a Política de Dados Abertos – Decreto Nº 8.777/2016, citadas anteriormente e, a partir de diferentes artigos e dissertações, cita-se que a LAI, de diferentes formas, contribuiu para a consolidação do regime democrático brasileiro e no fortalecimento das políticas de transparência pública, sendo apontada como uma ação governamental disruptiva na garantia do acesso a dados públicos.

Conforme discutido em tópico anterior, a LAI também regulamenta o direito constitucional de acesso à informação pública por parte dos cidadãos, aplicável a todas as esferas de governo: União, Estados, Distrito Federal e Municípios; a

implementação de termos como transparência ativa e passiva; definiu mecanismos, prazos e procedimentos para a entrega das informações solicitadas pelos cidadãos; determinou um rol de informações que as instituições públicas são obrigadas a divulgar de forma proativa por meio da internet; orienta sobre a criação do Serviço de Informação ao Cidadão (SIC), do Sistema Eletrônico de Informação ao Cidadão (e-SIC); e confirma o website institucional como canal de acesso à informação.

Compreende-se então, a partir do atual contexto da Política Nacional de Dados Abertos, e a implementação de um compêndio de iniciativas para a ampliação das políticas de transparência na gestão pública, que o uso das TIC surge como um caminho viável para o fortalecimento desse processo. O Decreto Nº 8.777/2016, ou Política de Dados Abertos, apresenta, como composição, a necessidade de fomentar o acesso público aos dados governamentais como medida de Estado.

É oportuno, a partir deste estudo, observar o papel da Ciência da Informação no processo de fortalecimento das políticas de acesso aberto no âmbito governamental. Conforme observa Jardim (2013), as boas práticas da gestão da informação na estrutura organizacional pública são importantes para tornar a informação mais acessível, de modo que ela seja disseminada e disponibilizada para todos. Posto isso, a relação dos recursos e ferramentas na gestão das TIC, juntamente com a gestão da informação no contexto público, tende a possibilitar a garantia do acesso informacional público confiável, transparente e eficiente.

5. Considerações finais

O presente estudo, a partir dessa análise preliminar, identificou que as normativas estudadas – a LAI e a Política de Dados Abertos – contribuem para a implementação dos preceitos da transparência pública, controle social e privacidade no âmbito público. Estas normativas consolidam e caracterizam o Estado Democrático de Direito, previsto na Constituição Federal de 1988.

No que se refere a Política de Dados Abertos, tem como propósito contribuir para o aumento da transparência pública, a partir da divulgação de dados que possibilitem o controle social das ações governamentais, além de contribuir para o melhor uso da verba pública e propiciar o monitoramento e avaliação das políticas públicas. Identificou-se que, para além do seu objetivo de instrumentalizar a promover o acesso à informação, o Decreto N° 8.777/2016 delinea um caminho estratégico, a longo prazo, para consolidar no Estado brasileiro (Poder Executivo) uma cultura sistêmica de transparência pública e acessibilidade informacional.

Observou-se, quanto ao uso das TIC para implementação da transparência e do controle social nas instituições públicas, uma perspectiva de adoção a nível nacional sobre o contexto dos dados abertos, considerando o avanço dos recursos tecnológicos e a forte adesão do poder público e sociedade. Posto isto, pode-se mencionar o Portal Brasileiro de Dados Abertos e o Painel da Lei de Acesso à Informação, que são instrumentos práticos de

consolidação do processo de transparência e disseminação da informação pública.

O estudo é uma pesquisa preliminar, que pretende identificar as normativas mencionadas (LAI e Políticas de Dados Abertos), orientadas quanto ao uso das TIC para a implementação da transparência e do controle social em universidades públicas federais da Região Norte: Universidade Federal do Pará (UFPA), Universidade Federal Rural da Amazônia (UFRA), Universidade Federal do Oeste do Pará (UFOPA) e a Universidade Federal do Sul e Sudeste do Pará (UNIFESSPA).

Referências

BRASIL. Constituição da República Federativa do Brasil.

1988. Disponível em:

http://www.planalto.gov.br/ccivil_03/constituicao/constituicao.htm. Acesso em: 05 fev. 2021.

BRASIL. Decreto N° 8.777, de 11 de maio de 2016. Institui a Política de Dados Abertos do Poder Executivo federal.

Disponível em: http://www.planalto.gov.br/ccivil_03/ato2015-2018/2016/decreto/d8777.htm. Acesso em: 22 jan. 2021.

BRASIL. Lei nº 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 no 8.112, de 11 de dezembro de 1990; revoga a Lei no 11.111, de 5 de maio de 2005, e dispositivos da Lei no 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/112527.htm. Acesso em: 28 jan. 2021.

BRASIL. **Lei nº 13.709, de 18 de agosto de 2018**. Lei Geral de Proteção de Dados Pessoais (LGPD). Disponível em: http://www.planalto.gov.br/ccivil_03/ato2015-2018/2018/lei/13709.htm. Acesso em: 30 mar. 2021.

CAMPOS, A. M. Accountability: quando poderemos traduzi-la para o português? **Revista de Administração Pública**, Rio de Janeiro, v. 2, n. 24, p. 30-50, fev./abr. 1990. Disponível em: <http://bibliotecadigital.fgv.br/ojs/index.php/rap/article/view/9049/8182>. Acesso em: 20 maio 2021.

JARDIM, J. M. A implantação da lei de acesso à informação pública e a gestão da informação arquivística governamental. **Liinc em Revista**, Rio de Janeiro, v. 9, n.2, p. 383-405, nov. 2013. Disponível em: <http://revista.ibict.br/liinc/index.php/liinc/article/view/639>. Acesso em: 11 abr. 2021.

JARDIM, J. M.; MIRANDA, V. L. A implantação da Lei de Acesso à Informação nas Universidades Federais do Estado do Rio de Janeiro. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 16., 2015, João Pessoa. **Anais...** João Pessoa: UFPB, 2015.

MOURA, M. A. Introdução. In: _____ (Org.). **A construção social do acesso público à informação no Brasil**: contexto,

historicidade e repercussões. Belo Horizonte: UFMG, 2014. p.19-23.

RODRIGUES, F. de. A. **Mapeamento de tecnologias informacionais sobre os dados abertos em saúde pública: destino de repasses financeiros federais.** 2012. 208 f. Orientador: Ricardo César Gonçalves Sant’Ana. Tese (Doutorado em Ciência da Informação) - Faculdade de Filosofia e Ciências. Universidade Estadual Paulista “Júlio de Mesquita Filho”. Marília, SP., 2012.

SANT’ANA, R. C. G. **Tecnologia e Gestão Pública Municipal: mensuração da interação com a sociedade.** São Paulo: Cultura Acadêmica Editora – UNESP, 2009. v. 1. 181p.

SCHEDLER, A. Que és lá rendición de cuentas? **Cuadernos de Transparencia del Instituto Federal de Acceso a la Información Pública**, v.3, p. 1-44, 2004.

VIVARTA, V.; MARTINS, P. Acesso à informação como direito humano fundamental. In: CANELLA, Guilherme; NASCIMENTO, Solano (Coord.). **Acesso à informação e controle social das políticas públicas.** Brasília: ANDI/Artigo 19, 2009. p. 8-31.

Análise e gestão de riscos dos documentos digitais existentes na curadoria digital

Fábio Eder Cardoso^a e Mihno Dgil Pinto de Brito^b

1. Introdução

Com as facilidades que os avanços tecnológicos proporcionam, surgem inúmeros desafios, devido às questões como: capacidade de armazenamento, segurança, deterioração, obsolescência de softwares e hardwares que dão suporte aos objetos digitais.

Neste sentido, destacamos o modelo de ciclo de vida “*Digital Curation Centre*”, na perspectiva da preservação digital e tem-se como objetivo, utilizar normas e métodos de análise e gestão de riscos para contribuir com o gerenciamento dos documentos arquivísticos digitais, garantido condições necessárias para a sua preservação.

É abordado o uso da norma ABNT NBR ISO 31000:2018 – Análise e Gestão de Risco e o método APP – Análise Preliminar de Perigos (AGUIAR, s.d.), aplicados a preservação digital, alinhados ao processo DCC – *Digital Curation Centre*.

O problema de pesquisa deste trabalho se norteia pela seguinte indagação: Como identificar os riscos, causas e efeitos

a Doutorando em Ciência da Informação pela Universidade Estadual Paulista – UNESP. E-mail: fabio.eder@unesp.br.

b Mestrando em Ciência da Informação pela Universidade Estadual Paulista – UNESP. E-mail: mihno.dgil@unesp.br.

dos documentos digitais no modelo de ciclo de vida da DCC – *Digital Curation Centre*? O profissional que utiliza a curadoria digital para a preservação dos documentos, tem certa dificuldade em identificar, nos modelos de ciclo de vida de documentos digitais, os riscos nos quais a informação está passível.

Para tal, foi realizada uma revisão de literatura, no sentido interdisciplinar com o foco na Ciência da Informação, adotando para isso uma pesquisa qualitativa e exploratória.

Pretende-se, com essa pesquisa, suscitar questões ou pontos de vista acerca dos riscos em que a curadoria digital pode ter, visto que, ao combinar análise de riscos e preservação de acervos digitais, obtém-se uma gestão proativa, prevenindo e identificando possíveis ameaças que podem se concretizar junto aos serviços de informação, contribuindo, assim, com o desenvolvimento da Ciência da Informação.

2. Gestão e análise de risco

Gerenciar riscos é iterativo e auxilia as organizações no estabelecimento de estratégias, no alcance de objetivos e na tomada de decisões fundamentadas. (ABNT ISO 31000:2018, 2018, p. 6).

De acordo com Ganoulis (2009) o risco possui dois elementos que o definem, sendo-os: a gravidade e a frequência do perigo. Vesley (1984) define o risco como o perigo decorrente da probabilidade ou possibilidade da ocorrência de um infortúnio.

A gestão de risco prevê situações que podem expor os dados de qualquer organização a situações de perigo. Baseando-se em

estimativas científicas, este tipo de gestão oferece subsídios para o controle dos riscos através do estabelecimento de prioridades e orientação na tomada de decisão de uma organização. Segundo Vaughan (1997), este tipo de gestão é uma abordagem científica para lidar com riscos, capaz de antecipar eventuais perdas acidentais, e também projetar e implementar procedimentos que minimizam a ocorrência da perda e de seus impactos.

A gestão de risco pode ser aplicada em qualquer organização nas suas mais variadas áreas e níveis, uma vez que o risco pode ser sinônimo de desafios e oportunidades (BERNSTEIN, 1996, p. 7). Sua aplicação pode ser implementada, também, em todas as áreas do conhecimento, entretanto, para cada tipo de análise têm-se distintos significados e percepções, podendo gerar certa confusão, pois suas terminologias são transmitidas de uma área de conhecimento para outra sem as devidas modificações. (GANOULIS, 2009).

Neste sentido, Vaughan (1997, p. 8) define a noção de risco caracteriza-se como a “condição na qual existe uma possibilidade de variações desfavoráveis de um resultado desejado e esperado”. Dessa maneira, conceitualmente, entende-se que o risco se relaciona à possibilidade de efeitos adversos indesejáveis ou imprevistos, resultando na perda ou prejuízos causados por exposição a um perigo.

Pode-se considerar que a gestão de risco se baseia em atividades de planejamento de risco, ou seja, identificação, análise e priorização de ameaças. Essas atividades compreendem uma

sequência de fases que podem aprimorar a qualidade do serviço, sendo elas: caracterização da instalação; identificação dos perigos; identificação dos cenários acidentais; quantificação das consequências de cada cenário; estimativa das frequências de cada cenário; estimativa dos riscos; e planejamento de emergência, se necessário (VAUGHAN, 1997).

Considerar a probabilidade de que o evento venha a ocorrer e estimar a extensão dos danos que o mesmo pode causar, são parâmetros para se avaliar o risco. O processo de gestão de riscos envolve a aplicação sistemática de políticas, procedimentos e práticas para as atividades de comunicação e consulta, estabelecimento do contexto e avaliação, tratamento, monitoramento, análise crítica, registro e relato de riscos, a Figura 1 ilustra todas as fases do processo de gestão de risco (ABNT ISO 31000:2018, 2018, p. 9).

Cabe ressaltar que o processo de gestão de riscos seja parte integrante da gestão e da tomada de decisão, e seja integrado na estrutura, operações e processos da organização. Pode ser aplicado nos níveis estratégico, operacional, de programas ou de projetos (ABNT ISO 31000:2018, 2018, p. 9).

Como se observa na Figura 1, o cerne do processo de gestão de riscos contém as fases de Identificação de riscos, Análise de riscos e Avaliação de riscos. O propósito da identificação de riscos é encontrar, reconhecer e descrever riscos que possam ajudar ou impedir que uma organização alcance seus objetivos. Informações pertinentes, apropriadas e atualizadas são

importantes na identificação de riscos. (ABNT ISO 31000:2018, 2018, p. 12).

Figura 1 – Processo de gestão de riscos



Fonte: ABNT ISO 31000:2018, 2018, p. 15

O propósito da análise de riscos é compreender a natureza do risco e suas características, incluindo o nível de risco, onde apropriado. A análise de riscos envolve a consideração detalhada de incertezas, fontes de risco, consequências, probabilidade, eventos, cenários, controles e sua eficácia. Um evento pode ter múltiplas causas e consequências e pode afetar múltiplos objetivos (ABNT ISO 31000:2018, 2018, p. 13).

O propósito da avaliação de riscos é apoiar decisões. A avaliação de riscos envolve a comparação dos resultados da

análise de riscos com os critérios de risco estabelecidos para determinar onde é necessária ação adicional (ABNT ISO 31000:2018, 2018, p. 13).

Ao se implementar análise e gestão de riscos em ambientes de repositórios digitais, faz uso de políticas de segurança que também devem ser consideradas. Nesse contexto, deve-se considerar procedimentos de segurança da informação, visto que é de vital importância para a efetividade do processo de curadoria digital.

3. Preservação digital

A preservação digital é um dos temas que vem sendo estudado na Ciência da Informação, presente em publicações nacionais e internacionais, buscando soluções de maneira interdisciplinar para as questões que envolve esse tema.

Por meio de estudos que abordam as especificidades desse tipo de informação que circula em meio digital em seus processos relativos ao armazenamento, recuperação e acesso dessas informações para a geração atual e principalmente preservando-a para as gerações futuras.

Para a preservação digital é importante entender os objetos digitais nos seus diferentes níveis de abstração: o físico, isto é, a inscrição de dígitos binários em um suporte; o lógico, ou seja, as regras que configuram o seu formato e; o conceitual, ou o produto da interpretação dessas regras pelo software (Ferreira, 2006).

Nesse sentido Palleta e Gonçalves (2016), também ressaltam sobre a fragilidade da preservação digital frente a evolução tecnológica e como garantir seu acesso às gerações futuras, destacando aí a necessidade de desenvolver uma conscientização dos profissionais da área para a importância da curadoria digital. Os documentos arquivísticos digitais, requerem uma gestão voltada à preservação, com processos padronizados respeitando as especificidades de cada etapa da organização, buscando alcançar soluções e criando condições de um ambiente digital seguro, a partir de diretrizes em modelos de gestão que tem como ênfase a preservação digital com base na curadoria digital.

4. Curadoria digital

O termo “curadoria”, tem sua origem relacionado ao termo latino *curare*, que significa “cuidado para” que expressa a custódia e preservação e está atrelado a uma ampla gama de atividades, tais como: cuidar, preservar e salvaguardar. Atualmente esse termo vem ganhando ampla repercussão nas atividades relacionadas a objetos digitais, por meio de suas diretrizes de gestão voltado para organizar o crescente aumento do fluxo de informação, buscando tornar acessível a informação para os usuários e que no futuro essas informações sejam acessadas de forma que a integridade do conteúdo informacional esteja assegurada. (LONGAIR, 2015).

Nessa perspectiva alguns países têm buscado soluções que garantam a preservação dos objetos digitais. Uma das soluções encontradas foi a utilização de metadados de preservação. Grácio (2012, p. 125) esclarece que:

Os metadados de preservação são essenciais não somente para a descrição do conteúdo de um objeto digital, mas também para as atividades de gestão relacionadas à preservação digital e devem apoiar todo o processo ao longo do tempo, registrando todas as estratégias aplicadas e as mudanças ocorridas.

A utilização de metadados de preservação auxilia na descrição detalhada dos objetos digitais, buscando simplificar a investigação e a identificação de suas respectivas fontes de informação, representando suas estruturas de acesso e possíveis modificações dentro do fluxo documental. Entre os modelos existentes, Sant’Ana (2013, p. 11) destaca o *Digital Curation Centre*:

Considerando a necessidade de descrever os processos envolvidos no acesso a dados, vale destacar o conceito de curadoria de dados. O *Digital Curation Centre – DCC*, que tem como foco a capacitação para a gestão de dados de pesquisa em toda comunidade de pesquisa de ensino superior do Reino Unido.

4.1 Digital Curation Centre – Ciclo de vida da curadoria digital (DCC)

Esta proposta tem base no ciclo de vida da curadoria digital, que busca garantir que o objeto digital esteja acessível, desde o momento da sua criação, até a sua destinação final, levando em consideração fatores como a obsolescência tecnológica, propondo uma visão geral das etapas necessárias ao processo de curadoria e preservação dos objetos digitais. (DIGITAL CURATION CENTRE, 2012).

Descrição e representação da informação	Coleta e atribui informações de representação necessária para entender o dado digital e sua apresentação. Envolve a atribuição de metadados administrativos, descritivos técnicos, estrutural e de preservação, utilizando padrões adequados, para assegurar a descrição e controle adequado a longo prazo.
Planejamento da preservação	É onde é feito o planejamento da preservação ao longo do ciclo de vida de curadoria do objeto digital. Isso inclui planos de gestão e administração de todas as ações do ciclo de vida de curadoria.
Acompanhamento e participação da comunidade	Enfatiza a necessidade de atenção para as atividades que se desenrolam no âmbito das comunidades envolvidas com o problema de curadoria, bem como a necessidade de participação no desenvolvimento de padrões , de ferramentas e de <i>software</i> adequados ao problema e que possam também serem compartilhados.
Curadoria e preservação	Estar continuamente alerta e empreender as ações administrativas e gerenciais planejadas para a curadoria e preservação por todo o ciclo de vida da curadoria.
Ações sequenciais	
Conceitualização	Conceber e planejar a criação do dado , incluindo os métodos de captura e as opções de armazenamento.
Criação e/ou Recebimento	Criar dados, incluindo metadados administrativos, descritivos, estruturais e técnicos. Metadados de preservação podem, também, serem adicionados no momento da criação. Receber dados, de acordo com políticas bem documentadas, de criadores de dados, de outros arquivos, de repositórios ou de centros de dados e, se necessário atribuir metadados apropriados.
Avaliação e seleção	Avaliar o dado e selecionar o que será objeto dos processos de curadoria e de preservação por longo prazo. Manter-se aderente tanto às boas práticas quanto às políticas pertinentes e também às exigências legais.
Arquivamento	Transferir os dados para um arquivo, ou repositório , ou centro de dados ou outro custodiante apropriado.
Ações de	Empreender ações para garantir a preservação a longo

preservação	prazo e a retenção do dado de natureza oficial. Ações de preservação devem garantir que o dado permaneça autêntico, confiável e utilizável, mantendo a sua integridade.
Armazenamento	Armazenar os dados de forma segura mantendo a aderência as normas relevantes.
Acesso, uso e reúso	Garantir que o dado possa ser acessado tanto pela sua comunidade alvo, quanto pelos demais usuários interessados na reutilização do dado.
Transformação	Criação de novos dados a partir do original.
Ações ocasionais	
Eliminação/ Descarte	Eliminar os dados que não foram selecionados para curadoria de acordo com políticas documentadas, diretrizes e/ou exigências legais.
Reavaliação	Retornar ao dado cujos procedimentos de avaliação foram falhos para nova avaliação e possível seleção para curadoria.
Migração	Migrar os dados para um formato diferente.

Fonte: Autores adaptado de Higgins (2008).

O Modelo de Ciclo de Vida de Curadoria do DCC, foi projetado visando apoiar atividades realizadas por gestores de acervos de dados digitais, todas as fases têm como foco a preservação. O modelo permite ainda que os curadores ou gestores identifiquem pontos fracos em políticas ou lacunas na cadeia documental. Podendo suas diretrizes e orientações serem aplicadas em intuições que gerenciam grandes quantidades de documentos digitais, onde uma adequada aplicação dessas orientações podem melhorar as atividades relativas à circulação, avaliação, seleção, descarte, armazenamento e acesso, apoiadas por processos e ações de preservação dos dados garantido sua integralidade para as gerações atuais e futuras.

5. Procedimentos metodológicos

Com o objetivo de mapear e ressaltar as etapas que contemplam a análise e gestão de risco com foco na segurança dos documentos digitais dentro do ciclo de vida da Curadoria Digital, foi realizada uma revisão de literatura.

O primeiro passo foi a escolha do termo de busca, onde foi definido o termo “Análise e Gestão de Risco”, pois o foco desta pesquisa é mapear as etapas referentes a esse termo dentro do ciclo de vida da curadoria digital.

O passo seguinte, foi a escolha das bases de dados de referência internacional vinculada a Ciência da Informação, onde foram selecionados os artigos na base de dados SSCI (*Social Sciences Citation Index*) do Web of Science. O acesso foi feito através do portal de periódicos da Capes.

Após definida as fases da pesquisa e a base de dados Web of Science, buscou mapear o processo referente a análise e gestão de risco, nos artigos recuperados, para subsidiar as fases posteriores, conforme apresentado no Quadro 2.

Como parâmetro de atuação e análise comparativa foi utilizado parte do processo DCC, ou seja, os pontos relacionados às ações sequenciais: conceitualização, criação e/ou recebimento, avaliação e seleção, arquivamento, ações de preservação, armazenamento, acesso, uso e reúso, e, transformação.

Quadro 2 – Fases da pesquisa

Fases	Procedimento	Resultados obtidos
1º	Definição do termo de busca Análise e	BRAPCI: 24

Fases	Procedimento	Resultados obtidos
	Gestão de Riscos e das bases de dados BRAPCI e na base de dados SSCI (Social Sciences Citation Index) do Web of Science.	Web of Science: 17.949 IBICT(BD'TD): 49
2º	Definição dos filtros para refinar a busca: a) Espaço temporal entre 2000 e 2021; b) Termos de busca presente nos títulos. c) No Web of Science o termo de busca foi traduzido para " <i>Risk Analysis</i> " e " <i>Risk Management</i> " e utilizando o operador <i>booleando</i> "or"; nas categorias foram escolhidos os artigos que estão no ítem: " <i>Information Science Library Science</i> ";	BRAPCI: 0 Web of Science: 56
3º	Leitura e tradução dos títulos e resumo para análise de relevância para a pesquisa.	BRAPCI: 16 Web of Science: 26
4º	Análise dos dados Principais termos encontrados nos artigos correlatos a essa pesquisa	Gerenciamento de risco em: projeto de software, sistemas de informação (SI), tecnologia em <i>cloud</i> , processamento de informações, gestão do conhecimento, proteção dos dados, análise de risco em (SI), apoio na tomada de decisão, Redução de custos, segurança da comunicação, responsabilidade corporativa, estratégias organizacionais, inovação.

Fonte: Autores.

No que se refere a análise e gestão de risco foi utilizado o método proposto por Aguiar (s.d.) denominado Análise Preliminar de Perigos – APP, sendo capaz de quantificar e qualificar riscos e ameaças em relação à preservação de arquivos digitais. Nesse contexto, Aguiar (s.d.) define dimensões de análise preliminar de risco para a preservação digital, dentre elas: Risco, Causa e Efeito, essas, utilizadas nesta pesquisa e descritas a seguir:

- Risco: Esta dimensão deve conter os perigos identificados para o sistema em estudo, ou seja, eventos que podem causar danos às instalações, materiais digitais e ao patrimônio institucional. Portanto, os perigos referem-se a eventos tais como: descargas elétricas e inundações; obsolescência tecnológica; falta de preparo dos profissionais; dentre outros.
- Causa: As causas de cada perigo devem ser discriminadas nesta dimensão e podem envolver falhas provocadas por agentes internos e externos.
- Efeito: O resultado de uma ou mais causas é definido como efeito. Os possíveis efeitos danosos de cada risco identificado devem ser listados. Os principais efeitos dos acidentes envolvendo a guarda da memória registrada em mídias digitais incluem, entre outros: Ilegibilidade da informação e Danificação das mídias. O Quadro 3 apresenta a correlação entre os processos das ações sequenciais do DCC – *Digital Curation Centre* e a APP – Análise Preliminar de Perigos.

Quadro 3 – Ações sequenciais do DCC com a análise e gestão de risco

Ações sequenciais do DCC		Identificação do Risco
Conceitualização	Conceber e planejar a criação do dado; Métodos de captura e as opções de armazenamento.	Risco: A escolha não ponderada de estratégias para a preservação dos materiais digitais Causas: Falta de conhecimento técnico daqueles que trabalham com os repositórios digitais Efeitos: Maximização de esforços desnecessários à preservação de objetos digitais
Criação e/ou Recebimento	Criar dados; Metadados (administrativos, descritivos, estruturais, técnicos e de preservação).	Risco: A escolha incorreta de estratégias para a criação e/ou recebimento dos metadados Causas: Falta de conhecimento técnico dos profissionais de informação. Vulnerabilidade dos dados Efeitos: Não garantia da autenticidade, integridade, e disponibilidade dos recursos digitais.
Avaliação e seleção	Preservação por longo prazo; Boas práticas a políticas pertinentes; Adequação as exigências legais.	Risco: Condições inadequadas do ambiente no qual estão depositados e do manuseio dos mesmos. Causas: Condições ambientais, econômicas e políticas no território nacional, que se apresentam de formas mais adversas a prática de preservação digital. (GALINDO, 2005) Efeitos: Levam ao desgaste e até mesmo a perda de leitura das informações registradas em objetos digitais. Considerado um dos principais causadores da perda da

Ações sequenciais do DCC		Identificação do Risco
		informação em hardwares e softwares.
Arquivamento	Transferência dos dados.	Risco: Degradação de mídias digitais. Causas: Inadequado acondicionamento dos materiais digitais nas instalações dessas instituições. Efeitos: Ilegibilidade das informações registradas nessas mídias
Ações de preservação	Preservação a longo prazo; Retenção do dado de natureza oficial; Ações de preservação; Autenticidade, confiabilidade e integridade.	Risco: Obsolescência de hardware e software; Causas: Mercado altamente competitivo da tecnologia da informação Efeitos: Leva a ciclos de renovação de tecnologia a cada 3 a 5 anos (HEDSTRON, 1998)
Armazenamento	Armazenar os dados de forma segura.	Risco: A falta de gestão de riscos que envolvem o armazenamento digital Causas: Falta de capacitação em relação aos objetos digitais e o gerenciamento dos riscos envolvidos na sua preservação Efeitos: Vulnerabilidade dos processos e sistemas de preservação digital, bem como dos dispositivos físicos.

Ações sequenciais do DCC		Identificação do Risco
Acesso, uso e reúso	Garantir que o dado possa ser acessado; Garantir a reutilização do dado.	Risco: Instabilidade de suporte por longo prazo; Causa: Falta de uso de padrões e formatos de arquivos de dados abertos, com amplo acesso e assistência técnica (THOMAZ E SOARES, 2004) Efeito: Complexidade no momento de selecionar e aplicar estratégias de preservação, minimizando sua efetividade.
Transformação	Criação de novos dados a partir do original.	Risco: Redundância e Inconsistência Causas: Desconhecimento prévio das fontes dos dados Efeitos: Dados sem qualidade, imprecisos.

Fonte: Autores.

6. Resultados e discussões

A partir dos dados obtidos, foi feita a correlação entre análise e gestão de risco e a fase referente as ações sequências do DCC, referente a segurança da informação dos dados digitais.

A análise de riscos pode ser realizada com vários graus de detalhamento e complexidade, dependendo do propósito da análise, da disponibilidade e confiabilidade da informação, e dos recursos disponíveis. As técnicas de análise podem ser qualitativas, quantitativas ou uma combinação destas, dependendo das circunstâncias e do uso pretendido. (ABNT ISO 31000:2018, 2018, p. 19)

Os pontos analisados no ciclo DCC e das ações sequenciais foram: conceitualização, criação e/ou recebimento, avaliação e

seleção, arquivamento, ações de preservação, armazenamento, acesso, uso e reúso, e, transformação. As dimensões utilizadas da análise de risco foram: risco, causas e efeitos, assim, será apresentada em cada ponto do ciclo DCC as considerações após análises baseadas nas normas ABNT ISO 31000:2018, 2018.

Na fase de conceitualização, onde ocorre a concepção e o planejamento da criação do dado; deve-se observar os métodos de captura e as opções de armazenamento. Nesse ponto o risco levantado foi a escolha não ponderada de estratégias para a preservação dos materiais digitais, visto que este é um fator que pode influenciar negativamente a criação dos dados. As causas podem ter origem na falta de conhecimento técnico daqueles que trabalham com os repositórios digitais, e os efeitos podem derivar da maximização de esforços desnecessários à preservação de objetos digitais.

Na criação e/ou recebimento dos dados, deve-se considerar a preservação dos metadados, sendo eles: administrativos, descritivos, estruturais, técnicos e de preservação.

De acordo com Saramago (2004, p.1) compreende-se por metadados de preservação “como informação de apoio aos processos associados com a preservação digital de longo prazo”. O risco que pode afetar esse processo de preservação a longo prazo está contido na escolha incorreta de estratégias para a criação e/ou recebimento dos metadados. As possíveis causas estão relacionadas à insciência de conhecimento técnico dos profissionais de informação e na vulnerabilidade dos dados. Os

efeitos resultantes da incapacidade pessoal na gestão dos metadados resultam em não garantir a autenticidade, integridade, e disponibilidade dos recursos digitais.

Na fase de avaliação e seleção, o processo de preservação por longo prazo, boas práticas e políticas pertinentes e adequação as exigências legais, dentre elas a LGPD – Lei Geral de Proteção de Dados, o risco encontrado foi sobre as condições inadequadas do ambiente no qual estão depositados e do manuseio dos mesmos. As causas são condições ambientais, econômicas e políticas no território nacional, que se apresentem de formas mais adversas a prática de preservação digital e os efeitos resultam no desgaste e até mesmo a perda de leitura das informações registradas em objetos digitais. Sendo considerado um dos principais causadores da perda da informação contidas em hardwares e softwares.

A fase de arquivamento, onde ocorre a transferência dos dados, o risco encontrado está relacionado a degradação das mídias digitais, as causas podem estar relacionadas ao acondicionamento inadequado dos dispositivos digitais nas instalações dessas instituições e os efeitos podem ser a ilegibilidade das informações registradas nessas mídias.

Nas ações de preservação, onde deve ser considerada: a preservação a longo prazo, a retenção do dado de natureza oficial, as ações de preservação e a autenticidade, confiabilidade e integridade. O risco proposto se relaciona à obsolescência de hardware e software, as causas estão associadas ao mercado altamente competitivo da tecnologia da informação e hardware e

como efeitos, ou consequências, são os ciclos de atualização tecnológica de três a cinco anos (HEDSTRON, 1998).

Em relação ao armazenamento, o que deve ser considerado, primeiramente, são os processos de armazenamento dos dados de forma segura. O risco principal dessa fase está contido na falta de gestão de riscos que envolvem o armazenamento digital. Uma das principais causas está na falta de capacitação em relação aos objetos digitais e o gerenciamento dos riscos envolvidos na sua preservação. Os efeitos podem resultar em vulnerabilidade dos processos e sistemas de preservação digital, bem como dos dispositivos físicos.

No nível de acesso, uso e reúso, deve-se garantir que o dado possa ser acessado e, também, garantir o reúso do dado. O risco encontrado se relaciona à instabilidade de suporte por longo prazo, a causa pode se dar pela alta de uso de padrões e formatos de arquivos de dados abertos, com amplo acesso e assistência técnica e os efeitos são a complexidade no momento de selecionar e aplicar estratégias de preservação, minimizando sua efetividade.

Durante a fase de transformação há a criação de novos dados a partir do original, dessa maneira um risco muito comum é a redundância e inconsistência. Suas causas se dão pelo desconhecimento prévio das fontes dos dados e seus efeitos resultam em dados sem qualidade e imprecisos.

Um dos elementos mais relevantes da Preservação Digital está relacionado à representação das técnicas de armazenamento,

gerenciamento de arquivos digitais, alocação de recursos e tecnologia. Entretanto, frente a diversidade de técnicas de preservação, não há como identificar e definir uma única técnica totalmente eficaz. Sendo assim, os diversos métodos de análise na preservação digital oferecem resultados complementares.

Existe uma relação muito forte entre preservação digital e a redução e controle dos riscos, assim, nesse estudo, buscou relacionar, um processo já consolidado de curadoria, o DCC – *Digital Curation Centre* e as normas e procedimentos de gestão e análise de riscos, vislumbrando a propositura de um novo olhar para a preservação dos repositórios digitais.

Dessa maneira, na tomada de decisões estratégicas apoiadas pelos recursos tecnológicos existentes, tais como armazenamento em *cloud*, por exemplo. São ferramentas fundamentais no momento, que requerem gerenciamento voltado para preservação, abordando todas as especificidades que os dados digitais possuem, sendo o modelo DCC, uma ferramenta que pode orientar o gestor nesse processo.

Essa pesquisa identificou as etapas referentes a análise de risco dos documentos digitais em conformidade com o *Digital Curation Centre* e a norma ABNT ISO 31000:2018, 2018, nesse sentido, confirma-se que é de suma importância, uma gestão voltada para objetos digitais. Pois tal atitude pode garantir que o objeto digital tenha todas as ações necessárias para esse fim e que as informações sejam identificadas, planejadas e implementadas na sequência correta. Podendo garantir dessa forma princípios

referentes a manutenção da autenticidade, confiabilidade, integridade e usabilidade do objeto digital.

Como limitações do estudo realizado e sugestões para estudos futuros, é necessário um melhor aprofundamento dos elementos comparativos da pesquisa podendo abrir possibilidades de novas pesquisas, ampliando o escopo com novos dados e novas perspectivas referentes a análise e gestão de risco relacionado com a preservação e segurança de dados digitais.

Referências

AGUIAR, L. A. de. **Metodologias de análise e riscos APP e Hazop**. Rio de Janeiro, RJ: [s.n., s.d.]. Disponível em: http://files.visaosegura.webnode.com/200000056-584dc5947a/APP_e_HAZOP.pdf. Acesso em: 08 jun. 2021.

ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS.
NBR ISO 31000: Gestão de Riscos – Diretrizes. Rio de Janeiro, 2018.

ARAÚJO, W. J.; BELARMINO, V. F. **Análise de vulnerabilidade computacionais em repositórios digitais**. Universidade Federal da Paraíba - UFPB. 2014

BERNSTEIN, P. **Desafio aos deuses: a fascinante história do risco**. 3. ed. Rio de Janeiro: Campus, 1996.

DCC. Digital Curation Centre. **DCC Curation Lifecycle Model**. Edinburgh: University of Edinburgh. 2012. Disponível em:

<http://www.dcc.ac.uk/resources/curation-lifecycle-model>.

Acesso em: 20 mai. 2021

FERREIRA, M. **Introdução à preservação digital: conceitos, estratégias e actuais consensos**. Universidade do Minho, Escola de Engenharia, 2006.

GALINDO, M. Tragédia da Memória. **Massangana**, Recife, v. 2, n. 1, p. 29-31, 2005.

HEDSTROM, M. Digital preservation; a time bomb for digital libraries. **Computer and the Humanities**, v.31, n.3, 1998, p.189-202. Disponível em:

http://deepblue.lib.umich.edu/bitstream/2027.42/42573/1/10579_2004_Article_153071.pdf. Acesso em: 09 jun. 2021.

HIGGINS, S. The DCC curation lifecycle model. **International Journal of Digital Curation**, v. 3, n. 1, p. 134-140, 2008.

Disponível em:

<http://www.ijdc.net/index.php/ijdc/article/viewFile/69/48>.

Acesso em: 23 mai 2021.

HIGGINS, S. **Digital Curation**: The Emergence of a New Discipline. 2011. Disponível em:

<http://www.ijdc.net/article/view/184/251>. Acesso em: 23 mai 2021.

LONGAIR, S. Cultures of Curating: The Limits of Authority. **Museum history journal**, v. 8, n. 1, p. 1-7, Jan. 2015. Disponível em:
<http://www.maneyonline.com/doi/pdfplus/10.1179/1936981614Z.00000000043>. Acesso em: 23 mai 2021.

PALLETA, F. C.; GONÇALVES, V. “Curadoria digital: O papel das bibliotecas na sociedade em rede”. **Pesq. Bras. em Ci. da Inf. e Bib.**, vol. 11, no. 2, 2016.

PENNOCK, M. Digital Resources: To Infinity and Beyond? Curating and Preserving Digital Library Resources (slides from presentation given at British & Irish Association of Law Libraries (BIALL) **Annual Conference**, Sheffield). 2007.

SANT’ANA, R. C. G. Ciclo de vida dos dados e o papel da ciência da informação. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 14., 2013, Florianópolis. **Anais...** Florianópolis, 2013. Disponível em:
<http://repositorios.questoesemrede.uff.br/repositorios/bitstream/handle/123456789/2477/CICLO%20DE%20VIDA%20DOS%20DADOS.pdf?sequence=1>. Acesso em: 25 maio 2021.

SARAMAGO, M. L. Metadados para preservação digital e aplicação do modelo OAIS. **Actas dos Congressos Nacionais de Bibliotecários, Arquivistas e Documentalistas**, Lisboa, n. 8, p. 1-6, 2004. Disponível em:
<http://www.bad.pt/publicacoes/index.php/congressosbad/article/view/640/637>. Acesso em: 10 jun. 2021.

SOUSA, R. T. B. de. O arquivista e as políticas públicas de arquivo. In: CONGRESSO NACIONAL DE ARQUIVOLOGIA, 2., 2006, Porto Alegre. **Anais...** Porto Alegre: ABARQ-UnB, 2006.

THOMAZ, K. P.; SOARES, A. J. A preservação digital e o modelo de referência open archival information system (oais). **DataGramZero**, v. 5, n. 1, 2004. Disponível em: <http://hdl.handle.net/20.500.11959/brapci/6608>. Acesso em: 09 jun. 2021.

VAUGHAN, E. **Risk Management**. New Baskerville: John Wiley & Sons, 1997. p. 67.

VESELY, W. E. Engineering risk analysis. In: RICCI, P. F.; SAGAN, L. A.; WHIPPLE, C. G. **Technological risk assessment**. Hingham, (MA): Martinus Nijhoff, 1984.

Avaliação de controle de vocabulários em repositórios institucionais: uma proposta de aplicação metodológica intrínseca e extrínseca

Lizandra de Souza Santos Alves^a, Roberta Cristina Dal'Evedove Tartarotti^b e Mariângela Spotti Lopes Fujita^c

1. Introdução

Com o crescimento dos recursos informacionais em formato digital e com a evolução das Tecnologias de Informação e Comunicação (TIC), os repositórios institucionais se tornaram ambientes importantes para guarda, preservação e disseminação da produção informacional científica. Fujita (2018) evidencia que os repositórios digitais em universidades são, atualmente, recursos de grande relevância na organização e gestão do conhecimento gerado pela produção científica, tecnológica, artística e administrativa.

Nestes sistemas de recuperação da informação, o uso de vocabulários controlados possibilita melhor precisão na representação e recuperação por assuntos. Essas informações são

-
- a Graduanda em Biblioteconomia pela Universidade Estadual Paulista – UNESP. E-mail: lizandra17alves@gmail.com.
 - b Doutora em Ciência da Informação pela Universidade Estadual Paulista – UNESP. Bibliotecária na Universidade Estadual de Campinas – UNICAMP. E-mail: roberta_tartarotti@yahoo.com.br. ORCID: <https://orcid.org/0000-0002-0246-4971>.
 - c Doutora em Ciências da Comunicação pela Universidade de São Paulo – USP. Professora na Universidade Estadual Paulista – UNESP. E-mail: mariangelasl57@gmail.com. ORCID: <https://orcid.org/0000-0002-8239-7114>.

armazenadas e preservadas, porém, para que haja eficácia na representação e disseminação desses recursos informacionais, é preciso verificar se há a disponibilização de vocabulário controlado para o uso do profissional indexador e analisar como se realiza o controle desses vocabulários em repositórios institucionais, pois a falta de controle de vocabulário pode gerar vários tipos de inconsistências determinantes para a imprecisão e irrelevância na recuperação de informações em qualquer sistema de busca (FUJITA; TOLARE, 2019, p. 94).

O uso de vocabulários controlados em repositórios institucionais possibilita melhor precisão na representação e recuperação por assuntos. No entanto, verifica-se ausência na literatura de procedimentos metodológicos voltados para a avaliação do controle de vocabulários nestes sistemas de recuperação da informação. Nesse sentido, o objetivo da pesquisa foi realizar a avaliação do controle de vocabulários controlados em repositórios institucionais, contribuindo com a melhoria da representação e recuperação por assuntos em repositórios institucionais.

2. Metodologias para avaliação de vocabulário controlado em repositórios institucionais

Leite, Amaro, Batista e Costa (2012) destacam que os repositórios institucionais levam em conta reunir, armazenar, organizar, preservar, recuperar e, sobretudo, a ampla disseminação da informação científica produzida na instituição acadêmica. No que tange à relevância de se utilizar um

vocabulário controlado em repositórios institucionais, Terra e Vasconcelos (2015) destacam que o controle de vocabulário permitirá a identificação dos documentos, visando à recuperação da informação por parte dos usuários de forma mais precisa.

Dahlberg (2006) acentua que a Organização do Conhecimento é uma ciência que cuidará de organizar e sistematizar os conceitos, sempre se preocupando com os valores dos conteúdos dos objetos ou assuntos que estão sendo organizados. Partindo disso, criam-se os Sistemas de Organização do Conhecimento, que servirão de auxílio na hora de realizar a interpretação do objeto e sistematizar seus conceitos para a representação. Para realizar essa representação de forma consistente, o uso de vocabulários controlados tem imensurável relevância para que esse processo seja realizado de forma padronizada conforme o domínio.

Para Harpring (2016), vocabulário controlado é um arranjo organizado de palavras e frases usadas para indexar e/ou recuperar conteúdos. Promovem consistência em termos preferidos e a atribuição dos mesmos termos a um conteúdo semelhante, tendo suas funções mais importantes de agrupamento de termos variantes e sinônimos em conceitos e a ligação dos conceitos em uma ordem lógica ou sua classificação em categorias. Kobashi (2008) salienta que o vocabulário controlado é uma linguagem artificial construída de termos que representarão um ou vários campos do conhecimento.

No entanto, para que haja consistência na representação e recuperação pelo uso desses vocabulários, é necessário realizar constante atualização e também aplicar metodologias avaliativas direcionadas tanto para a estrutura do vocabulário, quanto para a análise da precisão em relação à representação e recuperação. O vocabulário controlado é essencial para o momento da representação e recuperação de um documento, facilitando a comunicação entre a linguagem do indexador e dos usuários que realizam a busca. Todavia, para esse fim é importante a aplicação de avaliações dos vocabulários para que se tenha maior controle da padronização dos termos e assim os objetivos sejam alcançados de forma consistente. Sua avaliação é indispensável, pois, como um vocabulário controlado é composto por termos e o conhecimento está sempre em movimento, está sujeito a ficar desatualizado.

Essas avaliações precisam ser aplicadas visando-se o vocabulário controlado propriamente dito, buscando verificar se os elementos de sua estrutura estão cumprindo a função de auxiliar no momento da representação de maneira consistente; e avaliar como está sendo o seu impacto perante a recuperação da informação. Em vista disso, ao longo dos anos, mesmo que em pouca proporção, na literatura são apresentadas metodologias voltadas à avaliação de vocabulários controlados, podendo ser citados os estudos de Lancaster (2002), a Norma ANSI/NISO Z39.19 (2005), Souza e Campos (2007), Gil Leiva (2008) e Alves

(2021). Tomando-se como base estes estudos, foram elencados os procedimentos metodológicos, descritos a seguir.

3. Procedimentos metodológicos

Os procedimentos metodológicos foram realizados em duas etapas: (i) levantamento bibliográfico sobre a temática de pesquisa e; (ii) exploratório-explicativa com estudo de caso, sendo aplicados critérios avaliativos em seis repositórios institucionais que integram em seu sistema de vocabulários controlados estruturados, como os tesouros, em duas categorias de análise: avaliação intrínseca, contendo critérios voltados para o repositório propriamente dito e a avaliação extrínseca, com critérios voltados para a análise da representação e da recuperação da informação nesses repositórios.

Gil (2002, p. 41) salienta que a pesquisa exploratória proporciona maior familiaridade do pesquisador com o problema, propiciando ideias ou descobertas de intuições, sendo bastante flexível no que tange à possibilidade de consideração de diversos aspectos relativos ao fato estudado. Para a coleta de dados na primeira etapa, foi delineada a técnica de levantamento bibliográfico que tem por base os materiais já elaborados (GIL, 2008, p. 50). A revisão de literatura foi realizada a partir do corpus informacional advindo do levantamento bibliográfico em fontes bibliográficas nacionais e internacionais. Esse levantamento se deu a partir de livros, artigos, trabalhos apresentados em eventos de Ciência da Informação, teses e

dissertações em repositórios institucionais, bases de dados, periódicos e Internet.

A segunda etapa, exploratório-explicativa com estudo de caso, continuou a investigar o tema de pesquisa buscando identificar fatores que contribuam para a ocorrência de fenômenos. Na presente pesquisa, o controle de vocabulário em repositórios institucionais é considerado o fenômeno com o qual a pesquisa explicativa aprofunda o conhecimento da realidade (GIL, 2002, p. 42). O estudo de caso é caracterizado pelo estudo de um ou de poucos objetos, de maneira a permitir o seu conhecimento amplo e detalhado (GIL, 2008, p. 57-58).

A seleção dos repositórios institucionais que utilizam vocabulários controlados para a sua representação e recuperação da informação pautou-se em investigação de Fujita e Tolare (2019) voltada para a identificação de repositórios institucionais que empregam vocabulários controlados em seus sistemas e dos tipos e das principais características dos vocabulários controlados utilizados. Para o levantamento dos repositórios institucionais, as autoras realizaram pesquisa eletrônica na lista do Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT) composta com um total de 107 repositórios. Após o levantamento desses 107 repositórios, as autoras realizaram análise documental de cada um deles e a partir disso “foram selecionados 86 em função de haver 13 repositórios com websites indisponíveis, 7 bibliotecas digitais diferentes de repositórios e

um diretório de repositórios não selecionados” (FUJITA; TOLARE, 2019, p. 113).

Posteriormente à localização dos diferentes repositórios, as referidas autoras passaram para o processo de verificar a disponibilização de vocabulário controlado nesses repositórios. Os vocabulários controlados foram empregados à representação e recuperação a partir dos seguintes tipos: tesouros em 6 repositórios; listas de termos em ordem alfabética (utilizados em 70 repositórios); listas de áreas do conhecimento ou listas classificadas utilizadas por 8; e taxonomias utilizadas por 7 repositórios. Para a presente pesquisa, utilizou-se do resultado dos 6 repositórios institucionais que utilizam 5 diferentes tesouros: DeCS, Thesagro, NAL Thesaurus, VocaUSP e Tesouro do Senado Federal. No que compete aos repositórios institucionais utilizados como universo de pesquisa, cada repositório é referenciado pela letra “R” e um número, da seguinte maneira: R1: DeCS; R2: Thesagro e NAL Thesaurus; R3: VocaUSP; R4: Tesouro do Senado Federal; R5: Tesouro do Senado Federal; R6: Thesagro e NAL Thesaurus.

Para que fosse realizada a aplicação da metodologia de avaliação de controle de vocabulário em repositórios institucionais, foram separadas duas categorias de análise contendo critérios avaliativos, cada uma delas, para que fossem aplicados nos repositórios institucionais selecionados.

Gil Leiva (2008, p. 213) aponta duas metodologias de avaliação: uma tomando como objeto de análise o próprio

vocabulário controlado, chamada de avaliação intrínseca, e a outra que estuda o papel do vocabulário controlado na representação e recuperação de informações, denominada de avaliação extrínseca. A partir dessas duas metodologias propostas pelo referido autor, foram criadas as duas categorias de análise e critérios avaliativos que vão compor essas categorias, tendo como embasamento os pontos de metodologias apresentadas na literatura. Cabe ressaltar que os critérios foram adaptados para serem aplicados nos repositórios institucionais.

3.1 Categoria de análise com enfoque na avaliação intrínseca

Considerando-se que a avaliação intrínseca estuda o interior de algo, essa categoria trará critérios de avaliação voltados para o repositório institucional propriamente dito. Nesta categoria, foram delineados os seguintes critérios avaliativos:

- a) Disponibilidade de link para o vocabulário controlado: corresponde ao fornecimento pelo repositório institucional, em sua interface, de um link que direcione o usuário prontamente para o vocabulário controlado (ALVES, 2021).
- b) Vocabulário correspondente ao domínio: se o vocabulário controlado empregado confere ao domínio contemplado pelo repositório institucional (ALVES, 2021).
- c) Autoarquivamento: Disponibilização do autoarquivamento. Em caso positivo, analisar se é utilizada a linguagem natural ou se há o direcionamento

para o vocabulário controlado para realizar a escolha de termos advindos da linguagem padronizada correspondente ao recurso informacional. Explorar no repositório institucional se está disponível algum manual de autoarquivamento (ALVES, 2021).

3.2 Categoria de análise com enfoque na avaliação extrínseca

A avaliação extrínseca volta-se para o impacto do vocabulário controlado nos sistemas de informação que utilizam a linguagem padronizada para realizar o processo de indexação e auxiliar na recuperação. Os critérios avaliativos serão pautados exclusivamente à recuperação da informação nos repositórios institucionais:

- a) Controle terminológico: voltado para a análise de imprecisões advindas de ambiguidade, sinonímia, polissemia e homonímia na recuperação dos recursos informacionais.
- b) Forma dos termos: corresponde à forma com que os termos foram apresentados na recuperação. Se aparecem em plural ou singular, se apresenta entrada direta ou inversa, o idioma adotado, se os termos são simples ou compostos e se são adotadas abreviaturas (SOUZA; CAMPOS, 2007).
- c) Grau de especificidade: análise do nível de especificidade terminológica no momento da recuperação dos recursos informacionais, isto é, verificar se a especificidade da

recuperação está de acordo com os usuários potenciais (GIL LEIVA, 2008).

4. Resultados e discussões

Os resultados e discussão foram pautados nas categorias de análise avaliação intrínseca (que contem critérios voltados para o repositório propriamente dito) e avaliação extrínseca (cujos critérios foram voltados para a análise da representação e da recuperação da informação) nos repositórios institucionais, conforme apresentados nos procedimentos metodológicos, obtendo-se os seguintes resultados:

4.1 Avaliação intrínseca

a) Disponibilidade de link para o vocabulário controlado

Nesse critério a busca pelo resultado foi feita por toda a interface do repositório institucional. Dos seis repositórios analisados, apenas o R1 fornece acesso direto ao vocabulário utilizado. Foi possível destacar a falta do direcionamento ao vocabulário controlado integrado ao sistema para conhecimento dos usuários e também para uma possível pesquisa por algum pesquisador interessado, sem necessidade de contato com os gestores.

b) Vocabulário correspondente ao domínio

Os resultados do segundo critério demonstraram ampla integração de vocabulários controlados correspondentes ao domínio que o repositório institucional é pautado. Esse resultado

é relevante, pois quanto mais especializado o vocabulário controlado for em relação ao domínio de especialidade, maior será a precisão no momento da representação e recuperação por assuntos dos recursos informacionais que compõem a coleção.

c) Autoarquivamento

Em relação ao autoarquivamento da produção acadêmica pelos autores, o resultado foi parecido com o resultado do primeiro critério, pois apenas o R1 possibilita aos autores realizarem o autoarquivamento de seus próprios materiais. Porém, para que seja realizado o autoarquivamento, é necessária autorização. Não foi possível o acesso ao manual que instrui sobre o preenchimento dos dados, não podendo ser constatada se a representação dos recursos informacionais é realizada via linguagem natural ou se os autores são direcionados para o vocabulário controlado integrado ao repositório institucional.

4.2 Avaliação extrínseca

a) Controle terminológico

Referente ao controle terminológico, verificou-se o correto controle de quatro repositórios, sendo eles: R2, R3, R4 e R6. Os outros dois repositórios apresentaram variações terminológicas em relação ao termo de busca. O R1 teve “hipertensão” como o termo escolhido e obteve quatro resultados por “hipertensão” e seis por “hipertensos”. Já no R5, o termo escolhido foi “psicoterapia” e no momento da busca houve a recuperação de 7 recursos informacionais pelo termo escolhido, 2 recursos

informativos não tinham nenhuma relação com o termo e um estava descrito como “psicologia” e “psicólogos”.

b) Forma dos termos

Quanto à forma como os termos são apresentados, foi verificado se o idioma do termo de busca apresentava singularidade ou pluralidade, se os recursos informativos eram recuperados sozinhos ou se apresentavam alguma composição ou abreviações do termo. No que tange à recuperação com o idioma proposto, todos recuperaram recursos informativos em português sem necessidade de busca avançada. Os repositórios R1 e R3 apresentaram singularidade e pluralidade. Os demais apenas recuperaram materiais na singularidade do termo. Em relação à composição do termo escolhido com outros termos, o R1 só apresentou composição ao termo quando o mesmo estava no singular e os demais apresentaram composição independentemente de singular ou plural e nenhum dos repositórios institucionais apresentaram abreviações na recuperação do termo.

c) Grau de especificidade

Quanto à verificação se recursos informativos recuperados abrangiam a especificidade em relação aos usuários potenciais do sistema de recuperação, todos os repositórios institucionais atenderam ao grau de especificidade na recuperação dos recursos informativos.

De forma geral, com a aplicação dos critérios avaliativos nos repositórios institucionais, observou-se que o repositório classificado como R1 apresenta melhor representação e recuperação da informação, mesmo que todos os outros tenham apresentado resultados muito semelhantes.

5. Considerações finais

Os vocabulários controlados são uma linguagem em construção com estrutura relacional, que tem como função recuperar recursos informacionais e controlar a linguagem natural. Além de serem úteis para o momento da recuperação por assuntos, os vocabulários controlados são utilizados para a realização da representação do recurso informacional por meio do processo de indexação para melhorar a comunicação entre sistema e usuário. Controlando essa comunicação, será possível um maior nível de precisão na recuperação, diminuindo ruídos na recuperação dos recursos informacionais.

Para que se tenha consistência na representação e na recuperação por assuntos dos recursos informacionais são indispensáveis a atualização e a avaliação do vocabulário controlado, tendo em vista que é um instrumento dinâmico e que precisa estar em consonância com o conhecimento científico que vem se modificando e se atualizando constantemente. Neste contexto, é necessária a avaliação e atualizações constantes do vocabulário controlado, por meio de uma metodologia de avaliação de controle de vocabulário nos repositórios

institucionais para que a representação e a recuperação dos recursos informacionais sejam consistentes.

A aplicação de critérios avaliativos nos repositórios institucionais analisados corrobora a importância da utilização de um vocabulário controlado na representação dos recursos informacionais para melhorar a precisão na recuperação por assuntos no momento da busca pelos usuários e da integração ao repositório institucional de um vocabulário controlado que integre termos de seu domínio. Quanto mais especializado o vocabulário controlado, maior a precisão na representação e recuperação por assuntos dos recursos informacionais.

Referências

BRASIL. **Constituição da República Federativa do Brasil**. 1988. Disponível em: http://www.planalto.gov.br/ccivil_03/constituicao/constituicao.htm. Acesso em: 05 fev. 2021.

ANSI/NISO. Z39.19 – 2005. **Guidelines for the construction, format, and management of monolingual controlled vocabularies**. Bethesda: NISO Press, 2005. 184 p.

ALVES, L. de S. S. **Avaliação de controle de vocabulários em repositórios institucionais**: uma proposta de aplicação metodológica intrínseca e extrínseca. 2021. Trabalho de Conclusão de Curso (Graduação em Biblioteconomia) - Faculdade de Filosofia e Ciências, Unesp, Marília, 2021.

DAHLBERG, I. Knowledge Organization: A New Science? **Knowledge Organization**, v.33, n.1, p. 11-19, 2006.

FUJITA, M. S. L. **Controle de vocabulário na representação e recuperação da informação em repositórios universitários**. 2017. Projeto de Pesquisa (Bolsa de Produtividade em Pesquisa) - CNPq, Programa de Pós-Graduação em Ciência da Informação, Unesp, Marília, 2018.

FUJITA, M. S. L.; TOLARE, J. B. Vocabulários controlados na representação e recuperação da informação em repositórios brasileiros. **Informação & Informação**, Londrina, v. 24, n. 2, p. 93-125, 2019.

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

GIL, A. C. **Métodos e técnicas de pesquisa social**. 6. ed. São Paulo: Editora Atlas, 2008.

GIL LEIVA, I. **Manual de indización: teoría y práctica**. Gijón: Trea, 2008.

HARPRING, P. **Introdução aos vocabulários controlados: Terminologia para arte, arquitetura e outras obras culturais**. São Paulo: Secretaria da Cultura do Estado de São Paulo, Pinacoteca de São Paulo, ACAM Portinari, 2016.

KOBASHI, N. Y. **Vocabulário controlado**: estrutura e utilização. Brasília, Escola Nacional de Administração Pública, 2008.

LANCASTER, F. W. **El control del vocabulário em la recuperación de información**. 2. ed. Espanha: Universitat de València, 2002.

LEITE, F. C. L.; AMARO, B.; BATISTA, T.; COSTA, M. **Boas práticas para a construção de repositórios institucionais da produção científica**. Brasília: Ibict, 2012.

SOUZA, J. C. C. E.; CAMPOS, M. L. A. Avaliação de linguagem de indexação aplicada à informação jornalística: estudo de caso. *In*: **ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO**, 18, 2007, Salvador. **Anais** [...]. Salvador: 2007. p. 1-17.

TERRA, G. M.; VASCONCELOS, T. R. Vocabulário controlado em repositórios de instituições de ensino superior brasileira: estudo comparativo. **RACIn**, João Pessoa, v. 3, n. 1, p. 32-51, jan./jun. 2015.

Análise focada no padrão MTD3-BR sob a luz dos Princípios FAIR

Keli Rodrigues do Amaral Benin^a

1. Introdução

“O movimento de Acesso Aberto tem como principal objetivo fomentar a leitura e obtenção de conhecimento para todos. Além disso, busca estimular o diálogo entre cientistas de maneira mais rápida e eficaz, uma vez que todos terão acesso àquela publicação” (PERIODICOS DE MINAS, 2021). Para isso o movimento de Acesso Aberto à informação científica criou duas estratégias: a Via Verde (autoarquivamento em repositórios eletrônicos, repositórios institucionais) e a Via Dourada (periódicos científicos de acesso livre).

Nessa mesma linha de movimento de Acesso Aberto, a Biblioteca Digital de Teses e Dissertações (BDTD), do Instituto Brasileiro de Informação em Ciência, Tecnologia e Inovação (IBICT), foi o primeiro sistema a utilizar os conceitos de Arquivos Abertos.

A BDTD/IBICT tem como objetivo de “integra e dissemina, em um só portal de busca, os textos completos das teses e dissertações defendidas nas instituições brasileiras de ensino e pesquisa” (BDTD, 2021).

a Especialista Informação, Conhecimentos e Sociedade pela Universidade Estadual de Londrina – UEL. E-mail: keliamaral@hotmail.com. ORCID: <https://orcid.org/0000-0002-2564-2531>.

No processo de criação da BDTD/IBICT vislumbrou-se a criação de um padrão de metadados que é fundamental para que se garanta a interoperabilidade entre os sistemas de informação, o Padrão Brasileiro de Metadados para Teses e Dissertações (MTD-BR).

Os padrões de metadados apresentam-se como importante ferramenta na organização dos dados em contextos específicos, conforme defendem Nogueira, Vaz e Souza (2010), a utilização de metadados para descrição de objetos contribui para minimizar as dificuldades inerentes às atividades de localização, descoberta, documentação, seleção e recuperação de informação.

Ainda dentro do movimento de Acesso Aberto, vem de encontro os Princípios FAIR, conjuntos preciso e mensurável de qualidades que uma boa publicação de dados deve apresentar. Com a adoção dos Princípios FAIR nos dados, metadados e infraestruturas traz os benefícios como:

- Ajudar os colegas e o próprio futuro a compreender o projeto de pesquisa e os dados;
- Facilitar o compartilhamento de dados e colaborações;
- Aumentar a visibilidade da pesquisa e pode levar a mais citações;
- Melhorar a transparência, confiabilidade e reprodutibilidade da pesquisa;
- Prevenir a perda de dados;
- Maximizar o potencial dos ativos de dados; e
- Maximizar o impacto da pesquisa.

O foco principal da recomendação para os Princípios FAIR é para ambientes de gestão de dados de pesquisa, contudo, considerando que o FAIR se aplica a dados, metadados e infraestruturas, pode-se recomendadas essas boas práticas nas Bibliotecas Digitais de Teses e Dissertações e nos Repositórios Institucionais, que envolvem dados, metadados e infraestruturas.

Moraes e Oliveira (2010) apresentam a Biblioteca Digital Brasileira de Teses e Dissertações como uma das principais fontes de consulta para teses e dissertações brasileiras. Nesse sentido surgiu a seguinte questão, a Biblioteca Digital de Teses e Dissertações (BDTD/IBICT), tem adotado os Princípios FAIR no MTD3-BR? O presente artigo teve como objetivo principal analisar quais metadados do MTD3-BR, já estão adequados nos Princípios FAIR.

2. Metadados e Padrões de Metadados

Os metadados são essenciais para a gestão e a recuperação da informação, por serem elementos usados para identificar, descrever e representar o conteúdo de um recurso informacional, além de descrever a localização do recurso e viabilizar sua busca, acesso e recuperação na Web.

Os autores Souza, Vendrusculo e Melo (2000, p. 93) definem metadados como “[...] dado sobre dado. É a catalogação do dado ou descrição do recurso eletrônico”. Da mesma forma, Takahashi (2000, p. 172) conceitua metadados como: “[...] dados a respeito de outros dados, ou seja, qualquer dado usado para auxiliar na identificação, descrição e localização de informações. Trata-se, em

outras palavras, de dados estruturados que descrevem as características de um recurso de informação”. Neste sentido, metadado refere-se a uma estrutura descritiva da informação sobre outro dado.

Em outras palavras, metadados é um “Conjunto de elementos que descrevem as informações contidas em um recurso, com o objetivo de possibilitar sua busca e recuperação” (GRÁCIO, 2002, p. 23).

“A padronização dos metadados facilita o acesso e a recuperação da informação, assegurando aos usuários a fácil movimentação de dados entre diferentes sistemas e plataformas computacionais” (ALVES e CAFÉ, 2010, p. 185).

Alves (2005, p. 115) esclarece que:

[...] formatos de metadados, também chamados de padrões de metadados, são estruturas padronizadas para a representação do conteúdo informacional que será representado pelo conjunto de dados-atributos (metadados). Em outras palavras, os formatos ou padrões de metadados podem ser considerados como formas de representação de um item documentário (ALVES, 2005, p.115).

“A implantação da Biblioteca Digital Brasileira de Teses e Dissertações (BDTD), que teve as primeiras discussões sobre sua criação a partir de 2001 e seu lançamento oficial em dezembro de 2002, se deu em nível nacional, através do IBICT, com base no *Open Archives*” (MORAES; OLIVEIRA, 2010, p. 74).

No processo de criação da BDTD, o IBICT elaborou e implantou o Padrão Brasileiro de Metadados para Teses e Dissertações (MTD-BR). Desde sua criação, o MTD-BR teve três versões, sendo que a primeira versão lançada em 2002 continha 71 metadados, sendo 38 obrigatórios e 33 opcionais. A segunda versão (MTD2-BR) foi desenvolvida em 2005 e possui oitenta metadados sendo 31 obrigatórios e 49 opcionais. A terceira versão (MTD3-BR) lançada em 2014 tem, portanto, ao todo 38 metadados, sendo 18 metadados obrigatórios e 20 metadados opcionais. A nova proposta do Padrão Brasileiro de Metadados da BDTD foi desenvolvido com base no esquema de metadados do Dublin Core com adaptações para a realidade brasileira.

3. Princípios FAIR

Os princípios FAIR vêm com o movimento atual em direção a e-Science e Ciência Aberta, envolvendo principalmente a necessidade do gerenciamento de dados de pesquisa. O FAIR é um conjunto de princípios orientadores para o bom gerenciamento de dados, com o objetivo de tornar os dados de pesquisa Encontráveis, Acessíveis, Interoperáveis e Reutilizáveis (Quadro 1) (WILKINSON et. al, 2016).

Os princípios se referem a três tipos de entidades: (i) Dados (ou qualquer objeto digital produzido em pesquisa: dados de pesquisa no sentido mais estrito, código, software, apresentações, etc.); (ii) Metadados (informações sobre esse objeto digital); e (iii) Infraestrutura.

Quadro 1 – Diretrizes dos princípios FAIR

Princípio	Especificações
Encontrável (<i>findable</i>)	F.1 (meta) dados devem ter identificadores globais, únicos e persistentes.
	F.2 dados devem ser descritos utilizando metadados ricos (impacta diretamente R1).
	F.3 metadados devem incluir clara e explicitamente os identificadores dos dados que descrevem.
	F.4 (meta) dados devem ser registrados ou indexados em mecanismos de busca
Acessível (<i>accessible</i>)	A.1 (Meta) dados devem ser recuperáveis pelos seus identificadores usando protocolo de comunicação padronizado.
	A.1.1 O protocolo deve ser aberto, gratuito e universalmente implementável.
	A.1.2 O protocolo deve permitir procedimentos de autenticação e autorização, quando necessário.
	A.2 Metadados devem ser acessíveis, mesmo quando os dados não estiverem mais disponíveis.
Interoperáveis (<i>interoperable</i>)	I.1 (Meta) dados devem ser representados por meio de uma linguagem formal, acessível, compartilhada e amplamente aplicável para a representação do conhecimento.
	I.2 (Meta) dados devem usar vocabulários de acordo com os princípios FAIR.
	I.3 (Meta) dados devem incluir referências qualificadas para outros (Meta) dados.
Reusáveis (<i>reusable</i>)	R.1 (Meta) dados são descritos com uma pluralidade de atributos precisos e relevantes.
	R.1.1 (Meta) dados devem ser disponibilizados com licenças de uso claras e acessíveis.
	R.1.2 (Meta) dados devem estar associados à sua proveniência.
	R.1.3 (Meta) dados devem estar alinhados com padrões relevantes do seu domínio.

Fonte: Go FAIR apud Wilkinson et al. (2016, tradução nossa).

Os Princípios não são necessariamente sobre dados abertos. É possível trabalhar de forma FAIR com dados que não se

destinam à disponibilidade pública. Também não são regras ou normas. A adoção dos Princípios FAIR será frequentemente uma adaptação gradual das rotinas de trabalho, mas também pode ser um enorme salto, onde um tipo de infraestrutura é substituído por outro. Caberá às diferentes áreas de pesquisa e comunidades de pesquisa fazer com que os Princípios do FAIR funcionem em seus respectivos contextos.

Além disso, a modularidade dos Princípios e sua distinção entre dados e metadados suportam explicitamente uma ampla gama de circunstâncias especiais. Um exemplo envolve a publicação de objetos de pesquisa sem dados. Fluxos de trabalho analíticos, por exemplo, são um componente crítico do ecossistema acadêmico e sua publicação formal é necessária para alcançar transparência e reprodutibilidade científica. Os Princípios FAIR podem igualmente ser aplicados a esses ativos não relacionados a dados, que precisam ser identificados, descritos, descobertos e reutilizados da mesma maneira que os dados (WILKINSON et al, 2016, tradução nossa)

Quando implementado, será uma gestão e administração mais rigorosas desses valiosos recursos digitais, para o benefício de toda a comunidade acadêmica.

4. Procedimentos metodológicos

A pesquisa se define como exploratória e descritiva, a fim de realizar estudos sobre o Padrão Brasileiro de Metadados para Teses e Dissertações e os Princípios FAIR.

Gil (2008) cita algumas fontes de interesse para a realização de pesquisas científicas, como publicações em obras de referência, teses e dissertações, periódicos científicos e anais de encontros. Para exploração dessas fontes, o presente estudo utilizou-se as seguintes bases de dados eletrônicas: Brapci, Repositório Institucional UNESP, Biblioteca Digital Brasileira de Teses e Dissertações (BDTD), Periódicos Capes e Scielo.

5. Resultados e discussões

Segundo GO FAIR (2021), os Princípios FAIR podem ser aplicados nos metadados, dados e infraestrutura de suporte, a “maioria dos requisitos de encontrável e acessível podem ser alcançados no nível de metadados. Mas a interoperabilidade e a reutilização exigem mais esforços no nível dos dados”. Então, o presente artigo focou analisar apenas os requisitos dos Princípios FAIR ‘Encontrável’ e ‘Acessível’.

O Quadro 2 mostra todos o Padrão MTD3-BR e seus respectivos atributos, foi realizada a análise de cada metadado, destacando os quais estão adequados nos Princípios FAIR dos requisitos Encontráveis e Acessíveis.

Quadro 2 – Padrão Brasileiro de Metadados da Biblioteca Digital Brasileira de Teses e Dissertações (MTD3-BR) e Princípios FAIR

Metadado	Valor	Princípios FAIR
dc.creator	Autor	F.2. Metadado enriquecido
dc.creator.ID	ID do autor	F.2 Metadado enriquecido F.3. ID em metadados

Metadado	Valor	Princípios FAIR
dc.creator.Lattes	Currículo Lattes do autor	F.2. Metadado enriquecido
dc.contributor.advisor1	Orientador1	F.2. Metadado enriquecido
dc.contributor.advisor1ID	ID do orientador1	F.2. Metadado enriquecido F.3. ID em metadados
dc.contributor.advisor1Lattes	ID Lattes do orientador1	F.2. Metadado enriquecido F.3. ID em metadados.
dc.contributor.advisor2	Orientador2	F.2. Metadado enriquecido
dc.contributor.advisor2ID	ID do orientador2	F.2. Metadado enriquecido F.3. ID em metadados
dc.contributor.advisor2Lattes	ID Lattes do orientador2	F.2. Metadado enriquecido F.3. ID em metadados.
dc.contributor.advisor-co1	Co-orientador1	F.2. Metadado enriquecido
dc.contributor.advisor-co1ID	ID do co-orientador1	F.2. Metadado enriquecido F.3. ID em metadados
dc.contributor.advisor-co1Lattes	ID Lattes do co-orientador1	F.2. Metadado enriquecido F.3. ID em metadados.
dc.contributor.advisor-co2	Co-orientador2	F.2. Metadado enriquecido
dc.contributor.advisor-co2ID	ID do co-orientador2	F.2. Metadado enriquecido F.3. ID em metadados
dc.contributor.advisor-co2Lattes	ID Lattes do co-orientador2	F.2. Metadado enriquecido F.3. ID em metadados.

Metadado	Valor	Princípios FAIR
dc.contributor.referee1	Membro da banca1	F2. Metadado enriquecido
dc.contributor.referee1ID	ID do membro da banca1	F2. Metadado enriquecido F.3. ID em metadados
dc.contributor.referee1Lattes	Lattes do membro da banca1	F2 Metadado enriquecido F.3. ID em metadados.
dc.contributor.referee2	Membro da banca2	F2. Metadado enriquecido
dc.contributor.referee2ID	ID do membro da banca2	F2. Metadado enriquecido F.3. ID em metadados
dc.contributor.referee2Lattes	Lattes do membro da banca2	F2 Metadado enriquecido F.3 ID em metadados.
dc.contributor.referee3	Membro da banca3	F2. Metadado enriquecido
dc.contributor.referee3ID	ID do membro da banca3	F2. Metadado enriquecido F.3. ID em metadados
dc.contributor.referee3Lattes	Lattes do membro da banca3	F2 Metadado enriquecido F.3. ID em metadados.
dc.contributor.referee4	Membro da banca4	F2. Metadado enriquecido
dc.contributor.referee4ID	ID do membro da banca4	F2. Metadado enriquecido F.3. ID em metadados
dc.contributor.referee4Lattes	Lattes do membro da banca4	F2 Metadado enriquecido F.3. ID em metadados.
dc.contributor.referee5	Membro da banca5	F2. Metadado enriquecido

Metadado	Valor	Princípios FAIR
dc.contributor.referee5ID	ID do membro da banca5	F.2 Metadado enriquecido F.3. ID em metadados
dc.contributor.referee5Lattes	Lattes do membro da banca5	F.2 Metadado enriquecido F.3. ID em metadados.
dc.identifier.doi	Identificador permanente – DOI	F.1 Identificador permanente F.2 Metadado enriquecido A.1. Protocolo padronizados de comunicação. A.1.1 Protocolo abertos e gratuitos.
dc.description.resumo	Resumo	F.2 Metadado enriquecido
dc.description.abstract	Abstract	F.2 Metadado enriquecido
dc.publisher	Nome da instituição por extenso	N/A
dc.publisher.country	País da instituição	N/A
dc.publisher.department	Nome da faculdade, instituto ou departamento por extenso	N/A
dc.publisher.program	Nome do Programa de Pós-Graduação por extenso	N/A
dc.publisher.initials	Sigla da instituição	N/A
dc.identifier.uri	Identificador persistente	F.1 Identificador permanente F.2 Metadado enriquecido A.1. Protocolo

Metadado	Valor	Princípios FAIR
		padronizados de comunicação. A.1.1 Protocolo abertos e gratuitos.
dc.subject	Palavras-chave que descrevem o assunto do documento	F.2 Metadado enriquecido
dc.subject.cnpq	Área(s) do conhecimento do documento (Tabela CNPq)	N/A
dc.thumbnail.url	URL Arquivo (imagem)	F.2 Metadados enriquecidos A.1.Protocolo padronizados de comunicação. A.1.1 Protocolo abertos e gratuitos.
dc.description.sponsorship	Agência de fomento	N/A
dc.type	Tipo do documento: tese ou dissertação	N/A
dc.title	O título e subtítulo da tese ou dissertação na língua do documento	F.2 Metadado enriquecido
dc.title.alternative	Título alternativo do documento e/ou traduzido em outro idioma	F.2 Metadado enriquecido
dc.rights	Direitos de acesso ao documento	A.1.2 Autentificação A.2. Metadados sempre disponíveis.
dc.date.issued	Data de defesa	N/A
dc.date.available	Data de embargo	N/A
dc.language	Idioma	N/A
dc.relation.references	Referências citadas na tese ou dissertação	F.2 Metadado enriquecido

Metadado	Valor	Princípios FAIR
dc.identifier.citation	Citação do documento	F.2 Metadado enriquecido
dc.format	Formato do documento	
dc.relation	Relação, quando for o caso, com outros materiais produzidos para a elaboração da tese/dissertação ou comentários do próprio autor sobre a tese/dissertação	N/A

Fonte: Autora adaptado de BD'TD.

A partir da análise de todos os elementos do padrão MTD3-BR e seus respectivos atributos apuraram-se os requisitos Encontrável (F.1): Identificador único persistente tanto para o conjunto de dados quanto para os metadados, no padrão MTD3-BR é possível inserir o identificador persistente para o conjunto de dados, como também para alguns metadados, como exemplo: Identificador permanente – DOI, ID Lattes do autor, orientador, co-orientador e dos membros da banca e outros.

Já para os requisitos Encontrável (F.2): Os metadados ricos (metadados completos), tornando metadados de qualidade e simétricos, o padrão MTD3-BR traz vários campos que podem enriquecer o conjunto de dados, mas alguns metadados não alcança a sua completude. Os requisitos Encontrável (F.3): ID em metadados, alguns campos permitem que insira o identificador persistente da descrição dos metadados, como exemplo já citado no F.1.

Os requisitos Encontrável (F.4): Meta (dados) registrados ou indexados em mecanismos de busca, aqui fala sobre o componente da infraestrutura, no caso da MTD3-BR, são as Bibliotecas Digitais de Teses e Dissertações e os Repositórios Instrucionais.

Analisando os requisitos Acessível (A.1): Protocolo padronizados de comunicação, esse requisito se concentra como os dados e metadados poderão ser recuperados por seus identificadores permanentes. Mostrando a sua importância para configurar todo o ambiente e o serviço de dados FAIR, esse requisito tem ligação com o requisito F.1. Os requisitos Acessível (A.1.1) é uma continuidade do (A.1): Protocolo abertos e gratuitos, no caso da Biblioteca Digital de Teses e Dissertações, o acesso sem custo e aberto, qualquer pessoa com um computador e uma conexão com a Internet pode acessar pelo menos os metadados. Os requisitos Acessível (A.1.2) também é uma continuidade do requisito (A.1): Autenticação, quando necessário, muitos casos podem haver restrições de acesso aos dados e/ou metadados um mecanismo de autenticação e autorização para o acesso deve ser liberado pelo protocolo de comunicação.

E finalizando, os requisitos Acessível o (A.2): Metadados sempre disponíveis, mesmo quando os dados não tiverem mais disponíveis ou nunca estiveram disponíveis. Esse requisito está relacionado com o requisito F.4, o componente da infraestrutura e o direito de acesso.

6. Considerações finais

A análise do padrão MTD3-BR, permitiu observar que a existência de vários metadados que já estão adequados aos Princípios FAIR, favorecendo o Acesso Aberto, a recuperação da informação e o compartilhamento da informação.

Também foi observado que a maioria dos metadados descritivos no padrão MTD3-BR estão nos requisitos Encontráveis (F.2) que tem uma importante ligação com os requisitos Reusáveis (R.1), portanto, quanto mais rico e completo forem as informações fornecidas nos metadados, mais fácil ficará do dado ser encontrado e aumentar a reutilização e a citação.

Percebeu-se que alguns casos os requisitos dos Princípios FAIR são bastantes repetitivos, assim reforçando alguns itens importante para a construção do ambiente e serviço de dados FAIR.

O presente trabalho analisou apenas os requisitos Encontrável e Acessível, porém ainda faltam mais estudos aprofundado para aprimorar mais metadados que atendem as boas práticas de gestão de dados nas Bibliotecas Digitais de Teses e Dissertações e Repositórios Institucionais.

Referências

ALVES, R. C. V. **Web semântica:** uma análise focada no uso de metadados. 2005. Dissertação (Mestrado em Ciência da Informação) - Universidade Estadual Paulista, Marília, 2005. Disponível em: <http://hdl.handle.net/11449/93690>. Acesso em: 05 jun. 2021.

ALVES, J. C.; CAFÉ, L. M. A. Análise focada em metadados sob a luz do padrão MTD-BR. **Em Questão**, Porto Alegre, v. 16, n. 2, p. 179-202, jul./dez. 2010. Disponível em: <https://seer.ufrgs.br/EmQuestao/article/view/12930>. Acesso em: 08 jun. 2021.

BDTD – Biblioteca Digital Brasileira de Teses e Dissertações. **O que é a BDTD**. São Paulo, 2021. Disponível em: <https://bdtd.ibict.br/vufind/Content/whatIs>. Acesso em: 07 jun. 2021.

GIL, A. C. **Métodos e técnicas de pesquisa social**. 6. ed. São Paulo: Atlas, 2008.

GO FAIR. **FAIR Principles**. 2021. Disponível em: <https://www.go-fair.org/fair-principles/>. Acesso em: 08 jun. 2021.

GRÁCIO, J. C. A. **Metadados para a descrição de recursos da Internet: o padrão Dublin Core, aplicações e a questão da interoperabilidade**. 2002. Dissertação (Mestrado em Ciência da Informação) - Universidade Estadual Paulista, Marília, 2002. Disponível em: <http://hdl.handle.net/11449/93722>. Acesso em: 05 jun. 2021.

MORAES, A. F.; OLIVEIRA, T. M. de. Experiências relacionadas ao levantamento de teses e dissertações. **Informação & Sociedade: Estudos**, João Pessoa, v.20, n.1, p. 73-81, jan./abr. 2010. Disponível em:

<https://www.arca.fiocruz.br/handle/icict/1990>. Acesso em: 07 jun. 2021.

NOGUEIRA, E. D. A.; VAZ, M. S. M. G.; SOUZA, L. Um Padrão de Metadados para Descrição de Imagens Astronômicas do Tipo FITS. **Revista Ciências Exatas e Naturais**, v.12 n° 1, jan./jun. 2010. Acesso em: http://ri.uepg.br/riuepg/bitstream/handle/123456789/636/AR_TIGO_Padr%C3%A3oMetadados.pdf?sequence=1. Acesso em: 07 jun. 2021.

PERIÓDICOS DE MINAS. **Entenda o que é Acesso Aberto**. 2021. Disponível em: <https://www.periodicosdeminas.ufmg.br/entenda-o-que-e-acesso-aberto/>. Acesso em: 08 jun. 2021.

SOUZA, M. I. F.; VENDRUSCULO, L. G.; MELO, G. C. Metadados para a descrição de recursos de informação eletrônica: utilização do padrão Dublin Core. **Ciência da Informação**, Brasília, v.29, n.1, p.93-102, jan./abr. 2000. Disponível em: <https://doi.org/10.1590/S0100-19652000000100010>. Acesso em: 06 jun. 2021.

TAKAHASHI, T. (Org.). **Sociedade da Informação no Brasil**: o livro verde. Brasília: Ministério da Ciência e Tecnologia, 2000.

WILKINSON, M. D. et al. The FAIR Guiding Principles for scientific data management and stewardship. **Scientific Data**. Nature. 2016.

Uma discussão sobre o Capitalismo de Vigilância Informacional no contexto da Ciência da Informação

Wilson Roberto Veronez Júnior^a e Bianca Savegnago de Mira^b

1. Introdução

A partir do momento em que instituições do setor de tecnologia como a *Cambridge Analytica* passaram a expor e vaziar dados pessoais de seus clientes, e, simultaneamente, personagens notórios da grande mídia como o australiano Julian Assange, da organização transnacional *WikiLeaks*, e o estadunidense Edward Snowden, ex-analista da *National Security Agency* ou Agência de Segurança Nacional (NSA), apresentaram provas documentais, fidedignas e contundentes de que o governo dos Estados Unidos (EUA) e de outros países estavam realizando práticas de espionagem na vida privada de seus cidadãos cresce a urgência de uma discussão aberta e aprofundada sobre o uso e a segurança de dados privados. Meireles (2021), lembra que essa vigilância já estava sendo colocada em prática, e que os eventos ocorridos em 11 de setembro de 2001, em Nova York, representam um dos

a Doutorando em Ciência da Informação pela Universidade Estadual Paulista – UNESP. E-mail: wilson.veronez@unesp.br. ORCID: <https://orcid.org/0000-0003-2939-1917>.

b Doutoranda em Ciência da Informação pela Universidade Estadual Paulista – UNESP. E-mail: bianca.mira@unesp.br. ORCID: <https://orcid.org/0000-0001-7913-4084>.

fatores preponderantes que iniciaram investigações dessa natureza.

No cenário atual, temas como *Fake News*, Pós-Verdade, *Big Data*, Ciência de Dados e Inteligência Artificial têm sido debatidos no âmbito da Ciência da Informação e se tornaram assuntos imprescindíveis para os profissionais da informação que atuam na comunidade científica, em instituições ou em organizações não governamentais, assim como para governos e gestores públicos e privados.

O presente artigo busca analisar a noção de Capitalismo de Vigilância no escopo da Ciência da Informação e, a partir dessa relação dialógica, analisar a manipulação de dados de usuários fornecidos em sistemas manuais ou automatizados, seja por intermédio de formulários, fichas cadastrais, sites de lojas, sites de hospedagens, departamentos, *e-commerce*, redes sociais, entre outros dispositivos essenciais para a captação e, conseqüentemente, a mineração, manipulação, tratamento, organização e disseminação desses dados em multiplataformas digitais.

Koerner (2021), lembra que o Capitalismo de Vigilância está sedimentado sob três fatores: a) a teoria behaviorista de Frederic Skinner; b) a distopia de George Orwell, um dos principais expoentes da ideia de Vigilância de Estado; c) e o desenvolvimento da plataforma de busca Google. Para Zuboff (2019), a última tornou-se a maior e mais bem-sucedida empresa de *Big Data* por ter o site mais visitado e, portanto, possuir a

maior quantidade de *data exhaust*. O Google, de acordo com Arruda (2019), registra quais buscas são realizadas por cada usuário, quais sites são clicados, qual ordem os links clicados estavam na tela, etc. Coletados aos milhões ou bilhões para posterior processamento, esse conjunto de informações poderá revelar, através do reconhecimento de padrões, formas de melhorar a resposta do sistema às demandas dos usuários.

Segundo Arruda (2019) e Zuboff (2019), os sistemas algorítmicos contemporâneos de uso de massa (redes sociais, buscadores, sistemas de recomendação como Amazon e Netflix) capturam dados sobre as atividades e ações realizadas pelos usuários e essa captura serve a dois propósitos – identificação dos modos de uso do sistema, a fim de identificar formas de melhorá-lo; identificação dos modos de uso do sistema, a fim de identificar formas de influenciar o comportamento dos usuários.

Ademais, sabe-se que os dados coletados são utilizados para previsão de comportamento de usuários e organização de *feeds* de notícias e páginas de resultado de busca, dentre outros serviços que efetivamente direcionam as atenções para determinados conteúdos enquanto ocultam outros (ARRUDA, 2019).

Com base nos elementos supracitados, o trabalho procura responder às seguintes questões teóricas: Em que medida o Capitalismo de Vigilância pode ser analisado no âmbito da Ciência da Informação? Podemos falar sobre um conceito de Capitalismo de Vigilância Informacional?

Para respondermos a essas questões, o objetivo geral deste trabalho é analisar o conceito de Capitalismo de Vigilância Informacional e propor uma relação dialógica com a Ciência da Informação.

No que tange aos objetivos específicos, esta pesquisa visa: (i) examinar e analisar de forma aprofundada o conceito de Capitalismo de Vigilância; e (ii) propor uma discussão sobre a ideia de Capitalismo de Vigilância Informacional no contexto da Ciência da Informação.

2. Capitalismo de Vigilância

A fim de contextualizar alguns conceitos que não são comumente utilizados e analisados no campo Ciência da Informação, antes de adentrarmos no conceito de Capitalismo de Vigilância propriamente dito, é salutar apresentarmos a definição de Capitalismo Informacional.

Logo, para tratarmos de um conceito de Capitalismo de Vigilância Informacional e que tem amplo espaço no campo da Ciência da Informação, é preciso retomarmos alguns aspectos conceituais e epistemológicos sobre o objeto de estudo do campo, ou seja, a informação, sendo esse fenômeno compreendido do ponto de vista mentalista, material e organizacional. Nessa perspectiva, Rendón Rojas (2012), lembra que o objeto de estudo da Ciência da Informação é complexo, visto que a informação é o encontro da mensagem com o seu receptor, possibilitando que haja a produção, interação e a socialização do conhecimento.

Outro fator elementar que corrobora com a proposição da ideia de Capitalismo de Vigilância Informacional é a interdisciplinaridade, elemento nuclear no campo da Ciência da Informação, sobretudo em convergência com aspectos relacionados à tecnologia de informação e comunicação. A interdisciplinaridade, conforme descrita por Borko (1968) e Saracevic (1992, 1995, 1996), é o ponto de intersecção e que em um primeiro momento possibilita a relação dialógica em relação ao conceito citado no início deste parágrafo.

É por volta do último quarto do século XX que observa-se o surgimento de uma nova economia mundial, Castells (1997) a define como Capitalismo Informacional e Global e credita esse acontecimento à revolução da tecnologia da informação que propiciou a base material fundamental para que essa nova economia se desenvolvesse. A faceta informacional está relacionada à dependência da capacidade de gerar, processar e aplicar efetivamente informações que se baseiam no conhecimento.

Assim como o capitalismo informacional, o capitalismo de vigilância nasce como um produto de mudanças globais que envolvem cultura, sociedade, política, economia, informação e conhecimento. Zuboff (2019) salienta que ele se formou gradualmente durante a última década e incorporou novas políticas e relações sociais que ainda não haviam sido bem delineadas ou teorizadas. Dessa forma, entender e explorar o fenômeno desse novo desdobramento do sistema capitalista, que

se baseia nas tecnologias de informação e comunicação e depende inerentemente à informação, é do interesse da área da Ciência da Informação.

A definição de Capitalismo Informacional dada por Montoya (2013) aponta que este representa a quarta fase do capitalismo e caracteriza-se como uma sociedade que foi capaz de codificar o saber teórico, anteriormente concentrado na mão dos trabalhadores, de forma que as máquinas passam a ser capazes de não só preservá-lo, mas também aplicá-lo. Essa mudança representa uma virada no processo de apropriação do conhecimento. Zuboff (2019) entende o capitalismo de vigilância como uma nova forma do capitalismo informacional em que se procura prever e modificar o comportamento humano como meio de produzir receitas e controle de mercado.

A ideia de Capitalismo de Vigilância foi apresentada por Shoshana Zuboff em um artigo publicado no ano de 2015. Na referida obra, a autora aponta que essa nova ordem econômica utiliza experiências humanas como matéria prima para fins comerciais, em um processo de desapropriação dos direitos humanos básicos servindo como uma ameaça direta à democracia. A autora ainda ressalta que o Capitalismo de Vigilância não pode ser reduzido às máquinas e sequências algorítmicas, pois o discurso que circula não é somente da ordem da técnica, mas também das relações de poder que se estabelecem nas instituições que produzem os objetos técnicos (ZUBOFF, 2015).

Fornasier e Knebel (2020), ensinam que o capitalismo de vigilância apresenta uma nova forma de exploração da vida e de hiperexploração do trabalho: da mesma forma que a General Motors inventou o capitalismo fordista, a Google é pioneira no capitalismo de vigilância, todavia, seus métodos não estão mais adstritos à competição entre empresas de tecnologia.

Segundo Poter e Souto (2019, p. 33), “o capitalismo de vigilância é uma nova vertente econômica, onde a produção de bens e serviços é dependente do novo modelo global de alterações de comportamento, é a espoliação dos pilares dos direitos humanos”.

Ferreira (2019), comenta que os dados produzidos por usuários, são capturados, armazenados e estão à disposição dos engenheiros da Amazon, entre outras plataformas virtuais, que trabalham para desenvolver o *voice-sniffer algorithm* que pretende interagir automaticamente às palavras gatilho como “comprar”, “amar”, “não gostar”. A partir disso, no entendimento de Evangelista (2017), surge uma forma nova de capitalismo de vigilância informacional que visa prever e modificar o comportamento humano como meio de produzir lucro e controle de mercado”.

Segundo Koerner (2021), o capitalismo de vigilância é uma transformação do capitalismo da informação, que nos coloca diante de um desafio civilizacional. As *Big Techs* – seguidas por outras firmas, laboratórios e governos – usam tecnologias da informação e comunicação para expropriar a experiência humana,

que se torna matéria-prima processada e mercantilizada como dados comportamentais.

Para Koerner (2021), o capitalismo de vigilância viola a privacidade e invade os domínios de existência porque essa é a dinâmica de acumulação do seu setor dominante, o capital de plataforma, que necessita extrair dados para os seus negócios.

De acordo com Evangelista (2017), no capitalismo de vigilância somos monitorados por dispositivos informacionais o tempo todo e isso permite a emergência dessa ideia de inteligência cibernética. O monitoramento inteligente, que é um controle informacional, garante muito mais efetividade, porque opera de outras formas, mais diretas, se estamos ou não seguindo um determinado comportamento, se o contrato está sendo cumprido.

O capitalismo de vigilância e o comércio de dados pessoais aparecem como fatores importantes num tipo de privatização mascarada dos serviços públicos, em que os dados surgem como moeda de troca que serve de base para o financiamento do aluguel da infraestrutura (EVANGELISTA, 2017). Caribé (2018), endossa que ele foi descoberto e consolidado pelo Google, e adotado pelo Facebook, Instagram, e se baseia inclusive no uso de mecanismos ilegítimos de extração, mercantilização e controle de comportamento para produzir novos mercados. Para Arruda (2019), o capitalismo de vigilância, com seu imperativo de extração e suas economias de escala, escopo e ação, teria se

tornado uma nova lógica econômica que se espalha por todos os segmentos da indústria, comércio e serviços.

Segundo Caribé (2019), é possível que o capitalismo de vigilância tenha sido construído a partir de uma sucessão de falhas, eventos e oportunidades, percebidos ou construídos em diversas “cestas” e perspectivas, em contextos distintos, mas de alguma forma coordenados.

3. Relações dialógicas entre Capitalismo de Vigilância e Ciência da Informação

Apesar de ser mencionado em inúmeros trabalhos, nesta proposição, o tópico Ciência da Informação será tratado de maneira breve, ou seja, serão destacados apenas elementos que apresentam vinculação com os estudos sobre manipulação de dados e tecnologia, assim como a temática Capitalismo de Vigilância. Neste caso, a ideia é realizar a intersecção desse conceito com o campo da Ciência da Informação e verificar a possibilidade da discussão sobre a ideia de uma resignificação da noção de Capitalismo de Vigilância Informacional.

Com base em autores que trabalham com o tema Informação, como Borko (1968), Buckland (1991), Saracevic (1996) e Barreto (1991), podemos entender que o Capitalismo de Vigilância, por tratar-se de um conceito ainda não discutido com maior profundidade na Ciência da Informação, se faz necessária a análise, estudo e o desenvolvimento de uma proposta de caráter científico para ser compartilhada junto à comunidade acadêmica, visto que dados e informações são aspectos essenciais para as

necessidades dos usuários, independentemente das plataformas que por eles são utilizadas. Desse modo, é importante destacar o campo da Ciência da Informação como responsável pela produção, análise, organização, tratamento, classificação e disseminação de dados e informação.

No ato de fornecimento dos dados, conforme explica Koerner (2021), o usuário cede gratuitamente as suas informações ao concordar com termos de uso, utilizar serviços gratuitos ou, simplesmente, circular em espaços onde as máquinas estão operando. Em muitos casos, por desatenção ou falta de paciência do usuário, o próprio acaba não se atentando aos termos e mais tarde, se depara com as consequências da extração e manipulação de seus dados por uma determinada empresa. O Google é um grande exemplo de como isso ocorre e as consequências que isso gera para o consumidor, acarretando até em processos jurídicos, visto que o uso de dados por parte da empresa viola a privacidade de seus usuários, ainda mais se o acesso e a manipulação dos dados for realizada sem consentimento.

Nesse sentido, Zuboff (2015, 2019), ao tratar do assunto, explica que esse novo tipo de vigilância informacional incorpora um modelo de capitalismo não menos transcendental do que o capitalismo industrial, e a luta pelo poder e controle na sociedade não está mais associada aos dados ocultos da classe e sua relação com a produção, porém, pelo contrário, aos dados ocultos da modificação do comportamento projetado e automatizado.

Ainda, Zuboff (2015) afirma que o capitalismo de vigilância é a forma dominante de capitalismo, com o Google e o Facebook se tornando a vanguarda de uma dinâmica que podemos verificar em toda a economia. Essas corporações são, sem dúvida, muito poderosas e têm capitalizações de mercado surpreendentes, mas quase todos os seus recursos financeiros provêm da publicidade e prospecção de novos clientes. Assim, esse novo tipo de captação não tem por preocupação apenas a captação e atenção de novos consumidores em potencial, mas também a ideia de criar uma necessidade e fazer com que aquela determinada pessoa seja impulsionada a consumir um produto ou se comportar de forma com que ela possa adquirir um determinado produto.

4. Considerações finais

Com base na literatura, foram apresentados e discutidos os principais aspectos de Capitalismo de Vigilância. Desse modo, a presente investigação, ainda que preliminarmente, buscou estabelecer uma relação dialógica entre Capitalismo de Vigilância e Ciência da Informação, sendo que o primeiro conceito foi desenvolvido no campo da Psicologia Social e propagado com mais veemência no campo da Tecnologia de Informação e Comunicação, com análises em instituições privadas, como o Google, Facebook, Instagram, Twitter, Amazon Prime, Netflix, entre outras multiplataformas digitais. Sobre o segundo conceito, é salutar destacarmos o seu potencial interdisciplinar, epistemológicos e científico e que busca analisar os modos de produção, classificação, tratamento, organização e a disseminação

da informação em todas as suas dimensões, sejam elas culturais, históricas, filosóficas, sociológicas, tecnológicas, entre outras.

Objetivando elucidar a problemática proposta para este trabalho, sustenta-se que a ideia de Capitalismo de Vigilância Informacional é um conceito que precisa ser melhor analisado do ponto de vista teórico e crítico, visto que dados e informações são elementos fundamentais e que são inerentes às práticas humanas e sociais. Além disso, o elemento que possibilita a intersecção dialógica entre o Capitalismo de Vigilância Informacional e a Ciência da Informação é a informação e a interdisciplinaridade.

Na continuidade deste trabalho, a ideia é realizar um aprofundamento na relação entre informação, interdisciplinaridade e tecnologia a fim de estabelecer de maneira mais concreta as relações propostas para esta investigação.

Referências

ARRUDA, R. E. Sistemas algorítmicos e governamentalidade: perspectivas da sociedade de controle e capitalismo de vigilância.

XII Simpósio Nacional da ABCiber. Devires da Cibercultura: políticas e práticas. UFRGS, Porto Alegre, RS. 23 a 25 de julho de 2019.

BARRETO, A. de A. A condição da informação. **São Paulo em Perspectiva**, v. 16, n. 3, p. 67-74, 2002.

BORKO, H. Information science: what is it?. **American Documentation**, v. 19, n. 1, p. 3-5, 1968.

BUCKLAND, M. K. Information as a thing. **Journal of the American Society for information science**, v. 42, n. 5, p. 351-360, 1991.

CARIBÉ, J. C. R. Vigilância cega, o que as pegadas digitais podem revelar sobre o indivíduo. **II Simpósio Internacional Network Science**, 2018.

CARIBÉ, J. C. R. Uma perspectiva histórica e sistêmica do capitalismo de vigilância. **Inteligência Empresarial**, v. 41, p. 5-13, 2019.

CASTELLS, M. La era de la información. Volumen 1: La sociedad red. **Alianza editorial**, 1997.

EVANGELISTA, R. de A. Capitalismo de Vigilância no sul global: por uma perspectiva situada. **5o Simposio Internacional LAVITS | Vigilancia, Democracia y Privacidad en América Latina: Vulnerabilidades y resistencias**. 29 y 30 de noviembre, 01 de diciembre de 2017. Santiago, Chile, p. 243-253.

FERREIRA, A. E. S. C da S. Capitalismo de Vigilância na Sociedade da Transparência: Um estudo discursivo sobre as bibliotecas da Universidade de Berkeley. **Mosaico**, v. 10, n. 16, p. 155-173, 2019.

FORNASIER, M. de O; KNEBEL, N. M. P. “O titular de dados como sujeito de direito no capitalismo de vigilância e mercantilização dos dados na Lei Geral de Proteção de Dados”. **Revista Direito e Práxis, Ahead of print**, Rio de Janeiro, 2020.

KOERNER, A. Capitalismo e vigilância digital na sociedade democrática. **Revista Brasileira de Ciências Sociais**, vol. 36 n° 105 /2021: p. 1-6, 2021.

MEIRELES, A. V. Algoritmos e autonomia: relações de poder e resistência no capitalismo de vigilância. **Opinião Pública**, v. 27, n. 1, p. 28-50, 2021.

MONTOYA, A. N. **Educación y comunicación: del capitalismo informacional al capitalismo cultural**. Universidad Pedagógica Nacional, 2013. Disponível em: <http://biblioteca.clacso.edu.ar/Colombia/dcs-upn/20151022043743/edu.pdf>. Acesso em: 20 out. 2021.

POTER, G. B; SOUTO, C. Capitalismo de vigilância. **Seminário de Tecnologia, gestão e educação**, v. 1, n. 2, p. 31-34, 2019.

RENDÓN–ROJAS, M. Á. Epistemologia da Ciência da Informação: objeto de estudo e principais categorias. **InCID: Revista de Ciência da Informação e Documentação**, v. 3, n. 1, p. 3-14, 2012.

SARACEVIC, T. **Information science: origin, evolution and relations**. In: VAKKARI, Pertti; CRONIN, Blaise; YLIOPISTO, Tampereen. *Conceptions of Library and Information Science Historical, Empirical and Theoretical Perspectives*. 1992. p. 6-27

SARACEVIC, T. Ciência da informação: origem, evolução e relações. **Perspectivas em ciência da informação**, v. 1, n. 1, 1996.

ZUBOFF, S. Big other: surveillance capitalism and the prospects of an information civilization. **Journal of Information Technology**, v. 30, n. 1, p. 75-89, 2015.

ZUBOFF, S. The age of surveillance capitalism: the fight for a human future at the new frontier of power. **New York: Public Affairs**, 2019.

Os memes como um gênero discursivo em Serviços de Rede Social Online

Débora Matni Fonteles^a, João Batista Ernesto de Moraes^b, Gilberto Gomes Cândido^c e Fernando de Assis Rodrigues^d

1. Introdução

Os Serviços de Redes Sociais Online (SRSO) permitem atos comunicativos frequentes, diversos e possibilitados por meio de relações interpessoais tornando-se um campo promissor de identificação e análise de gêneros discursivos. Para Marcuschi (2020, p. 1) os gêneros textuais são “fenômenos históricos” caracterizados pelas relações que mantêm com os meios culturais e sociais. Marcuschi (2020, p. 1) entende que os gêneros textuais são definidos como “[...] entidades sócio-discursivas e formas de ação social, incontornáveis em qualquer situação comunicativa”.

Se compreende que os memes produzidos, postados e compartilhados em SRSO, como: Facebook, Instagram e TikTok, podem ser considerados como gêneros que representam as “entidades sócio-discursivas” (MARCUSCHI, 2020),

- a Mestranda em Ciência da Informação pela Universidade Federal do Pará – UFPA. E-mail: dmatnif@gmail.com. ORCID: <https://orcid.org/0000-0001-8468-462X>.
- b Doutor em Estudos Literários pela Universidade Estadual Paulista – UNESP. Professor na Universidade Estadual Paulista – UNESP. E-mail: joao.batista@unesp.br. ORCID: <https://orcid.org/0000-0002-8907-3593>.
- c Doutor em Ciência da Informação pela Universidade Estadual Paulista – UNESP. Professor na Universidade Federal do Pará – UFPA. E-mail: ggcandido@ufpa.br.
- d Doutor em Ciência da Informação pela Universidade Estadual Paulista – UNESP. Professor na Universidade Federal do Pará – UFPA. E-mail: deassis@ufpa.br. ORCID: <https://orcid.org/0000-0001-9634-1202>.

manifestadas em formas de representação tais como: a social, cultural, política e institucional. Nesse sentido o meme é descrito como sendo um elemento frequente percebido nas relações entre os indivíduos, onde podem ser considerados como gêneros discursivos, uma vez que são compostos de elementos textuais, enunciados, e de práticas como por exemplo as sociais (FERREIRA; VASCONCELOS, 2019).

A cultura dos eletrônicos foi um dos elementos que ajudaram a impulsionar a ocorrência de novos gêneros como os memes, motivada pelo uso de ferramentas tecnológicas como o computador (MARCUSCHI, 2020), acrescentando-se a isso o acesso à conexão de Internet, aos dispositivos móveis como smartphone e aos SRSO. Ressalta-se que o gênero apenas alcançou o reconhecimento que tem hoje graças a atuação participativa da “cultura digital” e do que se conseguiu produzir com as imitações, remixes e outras interações. Além disso, os memes têm um alcance mais rápido e de maior difusão em virtude dos SRSO (WIGGINS; BOWERS, 2015; RODRIGUES, 2017).

Recuero (2011, p. 122) reforça a ideia e acredita que a rede e a dinamicidade social tiveram grande influência nos memes e na sua intensificação tendo como exemplos os vídeos, jogos e as imagens.

O gênero denominado de meme foi atribuído por Dawkins (1976) quando propôs uma associação entre o termo Gene e meme, onde o primeiro serve para transmitir as informações de

conteúdo genético, e o outro tem como propósito a transmissão de informações culturais, ocorridas em espaços sociais e humanos.

O objetivo do estudo foi investigar os memes enquanto gênero discursivo e descrevê-los em uma página de um SRSO auxiliado pela classificação de Recuero (2011).

Utilizou-se como método o estudo descritivo com caráter exploratório tendo como locus de coleta de dados o Facebook, onde a escolha se justifica por se tratar de um dos SRSO mais utilizado (DIGITAL, 2019).

Foram observadas livremente páginas do Facebook em língua portuguesa que têm por finalidade fazer a divulgação de memes. A seleção para a realização da coleta de dados foi baseada na quantidade de curtidas e de compartilhamentos realizadas nas páginas chegando-se assim a página Capinaremos. Em março de 2021, a página Capinaremos foi classificada segundo seus idealizadores como de entretenimento com 1.513.476 seguidores, alcançando uma certa popularidade entre os internautas.

Para o desenvolvimento do estudo foram selecionados seis (6) memes baseado na classificação de postagens do ano de 2020, entre os dias 29 de novembro e 24 de dezembro de 2020. Na descrição dos memes, adotou-se a classificação de Recuero (2011), pois entende-se que sua pesquisa pode contribuir para a compreensão do gênero. Foram utilizados no estudo apenas os memes produzidos com um conjunto de imagens e textos/enunciados. Esta etapa da pesquisa vem abordar os

memes no contexto de fatos sociais, políticos, situações do cotidiano, de modo a descrevê-los como gênero discursivo. Para complementar a abordagem foram registradas algumas peculiaridades individuais de cada meme como número de compartilhamentos observados na página Capinaremos com a intenção de verificar o nível de alcance que o gênero obteve entre os internautas.

2. Os memes como gênero discursivo emergente

Os memes se tornaram populares e invadiram as redes sociais a partir da segunda década do século XXI. Segundo Ferreira e Vasconcelos (2019, p. 45) o meme é quase que “uma entidade onipresente”, ou seja, estão presentes em quase todos os lugares onde haja uma conexão ou onde possa acontecer uma interação virtual.

Para Marcuschi (2020) os gêneros textuais podem ser compreendidos como sendo eventos linguísticos, logo não podem ser observados como estruturas sem dinamicidade. Apesar dos gêneros textuais não possuírem traços linguísticos, são de fato atividades sociodiscursivas, e apesar de comporem os fenômenos sócio-históricos e culturalmente sensíveis possivelmente não exista uma relação de todos os gêneros textuais existentes.

Para Marcuschi (2020) a tecnologia fez surgir formas atualizadas de gêneros, não novas, mas que se apresentam de maneira diferente a sociedade. O e-mail é semelhante, por exemplo as cartas particulares, as comerciais e os bilhetes.

Portanto os gêneros não são novos, eles se reinventam e se transformam conforme as necessidades da sociedade por meio também dos avanços tecnológicos. Marcuschi (2020) acredita que se por um lado o e-mail não pode ser considerado como um novo gênero, as cartas eletrônicas por sua vez possuem características próprias o que as fazem serem consideradas como um novo gênero.

Neste sentido é válido fazer uma analogia entre os quadrinhos e os memes para verificar os aspectos que os aproximam ou os distanciam. Cagnin (1975) e Ramos (2009) dizem que os quadrinhos são considerados literatura de imagem fixa que reúnem vários gêneros em um só, chamado de hipergênero e caracterizados por possuir uma linguagem independente e constituída de elementos pictóricos e sem objetivos, embora consigam representar os elementos narrativos por meio de seus próprios mecanismos. Já os memes são uma mescla de composições com características um pouco diferentes dos quadrinhos, com objetivo bem definido, e que se disseminam por meio das plataformas de SRSO alcançando o público de maneira mais rápida (CALIXTO, 2017; FERREIRA; VASCONCELOS, 2019; RECUERO, 2011; WIGGINS; BOWERS; 2015).

Sobre os gêneros discursivo e emergentes, Ferreira e Vasconcelos (2019, p. 47) elucidam o valor na utilização do meme para a comunicação e para a cultura, onde o meme é um instrumento que permite a relação entre os indivíduos por meio

de textos e enunciados que se transformam em práticas sociais e por conseguinte em gênero discursivo. Recuero (2011) destaca que os memes têm relação com a questão da difusão da informação e com aquilo que é passado de pessoa para pessoa ou ainda com aquilo que pode sobreviver das ideias. Por outro lado, Calixto (2017) esclarece que os memes são entendidos como composições híbridas entre elementos virtuais e verbais, que são feitos a partir de mensagens e recortes.

Assim os gêneros são diferentes de textos e de enunciados, podendo atribuir a esses um sentido diferente daquele pensado em sua forma tradicional, servindo portanto de ferramenta de orientação, sendo capaz até mesmo de modificar a dinâmica cultural da humanidade, onde o meme, enquanto gênero, não pode ser aceito como algo pronto, como se fosse uma fórmula que os indivíduos utilizam para o ato comunicativo (WIGGINS; BOWERS, 2015). Os memes são complexos e fazem parte de um sistema moldado por motivações sociais e atividades culturais que resultam da comunicação e da vontade para que essa comunicação aconteça. Desta forma Recuero (2011, p.129) expõe que é preciso entender mais sobre os memes, como são constituídos, ou como se caracterizam, pois isso contribui para a compreensão de como eles são propagados.

Toda a construção elaborada sobre os memes enquanto gênero discursivo precisa ser melhor aprofundada. Assim, recorre-se às contribuições de Recuero (2011) como forma de evidenciar algumas características discutidas em seu estudo,

sintetizado no Quadro 1 trazendo as principais abordagens sobre gênero.

Quadro 1 – Características dos memes

Classificação	Tipos	Característica
Quanto a Fidelidade de cópia	Replicadores	Apresentam pouca variação em sua estrutura. São exatamente iguais à cópia original. Além disso, tem a função de serem informativos.
	Metamórficos ou Mutantes	Neste tipo de meme, as informações são discutidas e debatidas. Passam por alterações, mas sofrem também reinterpretações antes de serem passados adiante.
	Miméticos	Podem ser personalizados, entretanto mantêm a mesma estrutura e a ordem com que haviam sido determinadas anteriormente.
Quanto a longevidade	Persistentes	São do tipo que se replicam por um longo período mas depois acabam desaparecendo e reaparecendo em outro momento, voltando a serem replicados.
	Voláteis	Acontecem em um curto período de tempo, replicados em vários weblogs. No entanto, podem ser alterados ou esquecidos facilmente.
Quanto a fecundidade	Epidêmico	São do tipo que se disseminam facilmente em vários SRSO se comparando com uma epidemia.
	Fecundos	São difundidos em escala menor. Não são considerados epidêmicos.
Quanto ao Alcance	Globais	Possui grande alcance, com ligações por meio de nós fracos.
	Locais	Circulam entre os indivíduos próximos, ou seja, são memes restritos a nós da rede. Com o passar do tempo podem se transformar em globais.

Fonte: Autores adaptado de Recuero (2011).

Para elucidar este contexto, apresenta-se alguns exemplos de memes conhecidos dos internautas para facilitar o entendimento sobre alguns tipos de memes existentes. Um exemplo de meme epidêmico é o dos carregadores de caixão do país de Gana (Figura 1), noticiado globalmente devido à repercussão que teve durante a pandemia do COVID-19 em 2020. A imagem do grupo Dancing Pallbearers virou uma epidemia no sentido literal, tendo o meme replicado em SRSO tais como: TikTok e Facebook (CONHEÇA..., 2021).

Figura 1 – Exemplo de meme Epidêmico e Replicador



Fonte: Facebook. Disponível em:

<https://www.facebook.com/ApocalypseUnderLine/posts/3708249859223998>.

Acesso em: 09 mai. 2021.

O meme inspirado em uma vilã da telenovela Senhora do Destino (Figura 2) (OLIVEIRA, 2017) foi transformado em meme no ano de 2016 fazendo algum sucesso e circulando nos principais SRSO. Este meme trata de um tipo fecundo e

epidêmico e, ao mesmo tempo, global quanto ao nível de alcance, pois começou a circular entre usuários de outros países. O meme passou um tempo sem circular, permanecendo assim em uma fase de esquecimento, mas quando a novela é reprisada pela emissora volta a circular novamente.

O meme da Figura 3 tem a intenção de alertar os indivíduos sobre a importância de se consumir água. Esse tipo de meme não é tão compartilhado como os demais, mas vale ressaltar que ele tem um apelo mais no sentido de orientar do que divertir., por isso é considerado um meme informativo.

Figura 2 – Exemplo de meme “Persistente”



Fonte: Curtamais. Disponível em:

<https://www.curtamais.com.br/goiania/7-memes-de-nazare-tesesco-para-voce-entrar-no-clima-de-senhora-do-destino>. Acesso em 09 mai. 2021.

Figura 3 – Exemplo de meme Informativo



Fonte: Scielo. Disponível em:

<https://www.scielo.br/img/revistas/bak/v15n2//2176-4573-bak-15-02-0185-gf01.jpg>. Acesso em 09 mai. 2021.

Um exemplo de meme mutante e global é o caso do COVID-19 com a cantora de Cardi B (Figura 4), um dos primeiros surgidos durante a pandemia. A cantora do vídeo postado no YouTube teve a intenção de expressar sua preocupação com o vírus e com os problemas causados pela doença, mas o que se viu foi uma grande repercussão do vídeo e que logo virou meme, sendo modificado várias e várias vezes.

Para complementar o estudo dos memes como gênero discursivo, também é importante destacar a contribuição de alguns autores e o entendimento destes sobre o contexto em que os memes estão inseridos (Quadro 2).

Figura 4 – Exemplo de meme mutante e global



Fonte: G1. Disponível em: <https://g1.globo.com/pop-arte/musica/noticia/2020/03/13/cardi-b-e-coronavirus-como-a-cantora-pop-virou-simbolo-da-prevencao-da-doenca.ghml>. Acesso em: 09 mai. 2021.

Quadro 2 – Contexto dos memes a partir de uma seleção de autores

Autores	Ano	Contexto
Cândido	1970	Ferramenta de divulgação de ideias.
Chagas	2018	Veículo de propaganda política.
Brugger	2007	Propagação do Discurso de ódio.
Bertolini e Bravo	2001	Sociabilidade.
Blackmore; Dalkins	1999; 1982	Imitação.
Adar et al.; Recuero	2004; 2011	Replicação.
Recuero	2011	Estímulo à Interação, Propagação.
Wiggins e Bowers	2015	Humorístico ou socialmente crítico.

Fonte: Autores.

A partir dos elementos destacados, apresentamos os resultados sobre os memes analisadas na amostra desta pesquisa.

3. Resultados e discussões

A página Capinaremos registrada no Facebook apresenta um ranking semanal com os melhores memes produzidos, sintetizados no Quadro 3, contendo: autor, título, quantidade de reações, número de compartilhamentos, comentários, bem como a data de postagem.

Quadro 3 – Ranking dos melhores memes postados no Capinaremos

Posi.	Título	Autoria	Reações	Comp.	Come.	Data
1	Sim amigo, esses jornais retratam a realidade exatamente como ela é	Luís Eduardo	1.573	239	53	03/12/2020
2	É tudo invenção da imprensa	Daniel Pires	1.451	278	56	29/11/2020
3	Saudades da merenda	Giuliano Luigi Biondi Madioto	1.196	241	18	24/12/2020
4	Baita argumento	Dudu D'Avila Marques	944	358	57	05/12/2020
5	Vai lá campeão, confia no teu potencial	Heriko Rocha	899	199	155	03/12/2020
6	E se você falar “eu te avisei” ainda falam que você torce contra.	Guilherme Claudino	885	129	54	02/12/2020

Fonte: Autores.

O primeiro meme (Figura 5) refere-se a três programas jornalísticos de emissoras de televisão brasileira. No meme, os programas de noticiários são criticados pelos internautas por se aproveitarem de situações do cotidiano consideradas dramáticas

para elevarem a audiência, tanto que foi o meme que mais obteve reações (como ilustrado no Quadro 3).

Figura 5 – Meme da página Capinaremos, primeira posição no ranking



Fonte: Página Capinaremos – Facebook. Disponível em:
<https://www.facebook.com/capinaremos/posts/3560797260634976>.
Acesso em 09 mai. 2021.

O título atribuído “Sim amigo, esses jornais retratam a realidade exatamente como ela é.” é bastante sugestivo, uma vez que os noticiários trazem muitas notícias de crimes, assaltos, acidentes automobilísticos, feminicídios, tudo para chamar a atenção da sociedade para a realidade de insegurança e de medo vivida pelos brasileiros.

No texto foi utilizada a expressão gado que se refere aos seguidores do Presidente do Brasil (Jair Messias Bolsonaro) e o termo da COVID-19 que é uma das preocupações globais atualmente, isso porque a pandemia afetou a vida dos indivíduos

em todos os seguimentos tais como a saúde e o econômico,0 principais centros de polêmicas políticas. Para enfatizar a polêmica, o autor do meme utilizou a figura de três apresentadores de telejornais brasileiros para criticar a forma apelativa do jornalismo realizado para se referir a temática principal.

Considera-se o meme como socialmente crítico, pois indica uma situação restrita ao contexto brasileiro. Em relação à fidelidade de cópia não são considerados replicadores e nem miméticos. São do tipo voláteis quanto a longevidade dado que o meme ocupa o 1º lugar no ranking em número de reações e de compartilhamentos. Este meme é ainda do tipo fecundo e local, já que se propagam de maneira mais discreta nos SRSO.

O segundo meme (Figura 6) foi considerado replicador e mimético, pois mantém a mesma estrutura de uma imagem anterior original (o meme já existia com outras frases), sendo apenas personalizado para um novo contexto.

Quanto à longevidade é do tipo volátil apesar de estar em segundo lugar no ranking em número de reações e de compartilhamentos. Percebe-se que é um meme fecundo e local, além disso, foi um meme que não permaneceu tanto tempo circulando nos SRSO.

Figura 6 – Meme do Capinaremos, segunda posição no ranking



Fonte: Página Capinaremos – Facebook. Disponível em: <https://www.facebook.com/capinaremos/photos/a.239235636124505/3551066088274760/>. Acesso em: 09 mai. 2021.

O terceiro meme (Figura 7) é socialmente crítico, pois expõe a questão da corrupção na alimentação escolar, onde foi atribuída uma sigla de partido político à situação, onde os alunos da rede estadual de ensino da cidade de São Paulo ficaram sem merenda escolar. O meme estimula os leitores a pensar sobre desvios de verbas e problemas com a merenda escolar, segundo o autor do meme.

Figura 7 – Meme do Capinaremos, terceira posição no ranking



Fonte: Página Capinaremos – Facebook. Disponível em:

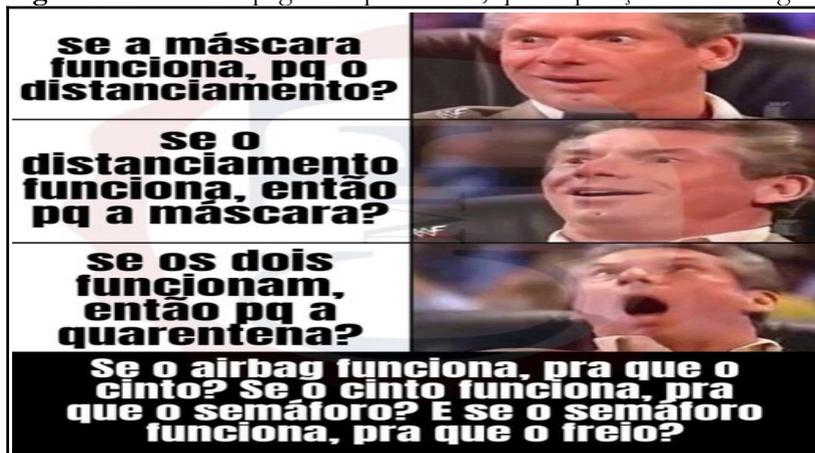
<https://www.facebook.com/capinaremos/posts/3558191197562249>.

Acesso em 09 mai. 2021.

Observou-se que o meme não se enquadra no quesito fidelidade de cópia, é considerado volátil quanto a longevidade e fecundo já que não obteve grande repercussão entre os usuários de SRSO. Não se tornou uma epidemia, sendo disseminado apenas no Facebook. É um meme considerado local quanto ao alcance, circulando apenas entre os indivíduos de redes sociais próximas.

O quarto meme (Figura 8) faz menção a ideia da crítica de algumas pessoas sobre o uso de medidas preventivas contra a propagação do COVID-19. Neste meme é utilizado uma sequência de texto que culmina com uso de um quarto texto que descaracteriza a ideia principal e mostra que se os indivíduos forem nesta linha de argumento então tudo pode ser questionável.

Figura 8 – Meme da página Capinaremos, quarta posição do ranking



Fonte: Página Capinaremos – Facebook. Disponível em: <https://www.facebook.com/capinaremos/posts/3567191219995580>. Acesso em: 09 mai. 2021.

O meme é do tipo humorístico, com estímulo à propagação tanto que foi o mais compartilhado. Quanto à fidelidade de cópia, é do tipo metamórfico pois se percebe que há debates sendo construídos sobre as informações relacionadas ao assunto proposto. Quanto à longevidade é do tipo volátil, pois o meme não ganhou tanta repercussão entre SRSO, sendo limitado apenas ao compartilhamento no Facebook. O meme é também do tipo fecundo e é considerado local, já que não obteve grande alcance.

O quinto meme (Figura 9) trata da questão das mudanças de regras sobre o Estatuto do Armamento. O meme traz tanto texto quanto imagem em um formato interessante, em que a figura central é colocada em tamanho maior e acima das outras imagens

dando uma dimensão de superioridade. Isso se faz necessário para mostrar aos leitores que a força maior de armamento discutida no Brasil conforme o autor do meme está supostamente sob o domínio dos bandidos, enquanto que os brasileiros ficam com aquilo que condiz com a sua realidade.

Figura 9 – Meme da página Capinaremos, quinta posição do ranking



Fonte: Página Capinaremos – Facebook. Disponível em: <https://www.facebook.com/capinaremos/posts/3560343990680303>. Acesso em: 09 mai. 2021.

O debate que se coloca por meio desse meme é considerado de grande valor informacional em função daquilo que se discute sobre a questão da segurança pública sendo portanto considerado como um meme socialmente crítico. É do tipo mimético, sendo fiel a cópia original, modificado apenas na inserção de algumas imagens e textos para se adaptar a uma dada situação ou contexto, podendo ser utilizado em outra circunstância. São

voláteis, fecundos e locais quanto a longevidade, fecundidade e alcance, respectivamente.

O sexto meme (Figura 10) trata sobre um problema frequente vivenciado em regiões metropolitanas e grandes cidades que são as enchentes. O autor do meme indica que há uma relação entre os problemas decorrentes das chuvas e os gestores municipais, neste caso os prefeitos. Segundo o autor, enquanto a cidade (no caso, São Paulo) está imersa na água e sofre com todos os transtornos decorrentes dela, o prefeito viajava durante as férias.

Figura 10 – Meme do Capinaremos, sexta posição no ranking



Fonte: Página Capinaremos – Facebook. Disponível em:

<https://www.facebook.com/capinaremos/posts/3558169854231050>.

Acesso em: 09 mai. 2021.

Trata-se de um meme de crítica social, replicador e mimético, pois, segundo a fidelidade de cópia, imita a cópia original. Não

houve modificação em sua estrutura original, cabendo mudanças apenas no texto, dando uma segurança ao autor, pois reduz erros possíveis no encaminhamento das informações e de suas interpretações. É volátil, fecundo e local, pois é um meme que não tem capacidade de permanecer muito tempo em circulação. Além disso, não ocorreu indícios de transmissão em outro SRSO, tendo também a característica de serem não epidêmicos.

4. Considerações finais

O estudo demonstrou que os memes além de serem gêneros discursivos, são um tipo aprimorado do gênero de quadrinhos, moldados em diversos suportes, tais como: em vídeo, em imagem e em texto. São gêneros que são potencializados pela internet e especialmente pelos SRSO, produzindo reflexões sobre o contexto de realidades ou simplesmente de um filme, série ou quaisquer outros produtos de audiovisual.

A pesquisa permitiu compreender o gênero a partir do seu tempo de permanência na internet (o alcance) e das modificações que sofreram até chegarem ao público. Além disso, a classificação apresentada por Recuero (2011) serve ainda para se discutir e entender valores, tais como os sociais, que são construídos com auxílio e pretensão de difusão de informações pelos SRSO.

Assim espera que a pesquisa possa contribuir tanto com aspectos relacionados à utilização dos memes quanto à construção de contextos bem como do entendimento sobre sua influência na vida dos indivíduos, dado que eles refletem muitas

vezes uma realidade revelada de forma humorística, mas não deixando de evidenciar também uma importante criticidade.

Referências

ALEXA. The top 500 sites on the web. **ALEXA**, 2021.

Disponível em: <https://www.alexa.com/topsites>. Acesso em: 20 mai. 2021.

CAGNIN, A. L. **Os quadrinhos**. São Paulo: Ática, 1975.

Disponível em:

<https://pt.scribd.com/document/327790205/Os-Quadrinhos-Antonio-Luiz-Cagnin>. Acesso em: 12 dez. de 2020.

CALIXTO, D. de. O. **Memes na internet**: entrelaçamentos entre educomunicação, cibercultura e a ‘zoeira’ de estudantes nas redes sociais. 2017. Dissertação (Mestrado em Interfaces Sociais da Comunicação) - Escola de Comunicações e Artes, Universidade de São Paulo, São Paulo, 2017. Disponível em:

<https://www.teses.usp.br/teses/disponiveis/27/27154/tde-01112017-102256/publico/>

[DOUGLASDEOLIVEIRACALIXTO.pdf](#). Acesso em: 11 mai. de 2021.

CONHEÇA a história por trás do meme dos “carregadores de caixão”. **Mundo Conectado**. Disponível em:

<<https://mundoconectado.com.br/noticias/v/13180/conheca-a-historia-por-tras-do-meme-dos-carregadores-de-caixao>>.

Acesso em: 2 mai. 2021.

DAWKINS, Richard. **O gene egoísta**. Universidade de Harvard, 1976. Disponível em:

https://www2.unifap.br/alexandresantiago/files/2014/05/Richard_Dawkins_O_Gene_Egoista.pdf. Acesso em: 7 jun. 2020.

DIGITAL 2019: Global Internet Use Accelerates – We Are Social. Disponível em:

<https://wearesocial.com/blog/2019/01/digital-2019-global-internet-use-accelerates>. Acesso em: 30 jun. 2020.

FERREIRA, D. M. M.; VASCONCELLOS, M. A. Discurso de memes: (Des)memetizando ideologia antifeminista. Bakhtiniana, **Rev. Estudo Discurso**, v.14, n. 2 São Paulo, apr./june 2019, Epub Apr 15, 2019. Disponível em:

<https://www.scielo.br/pdf/bak/v14n2/2176-4573-bak-14-02-0044.pdf>. doi.org/10.1590/2176-457339504. Acesso em: 11 dez. 2020.

MARCUSCHI, L. A. **Gêneros textuais: definição e funcionalidade**. 2020. Disponível em:

https://edisciplinas.usp.br/pluginfile.php/133018/mod_resource/content/3/Art_Marcuschi_G%C3%AAneros_textuais_defini%C3%A7%C3%B5es_funcionalidade.pdf. Acesso em: 5 dez. de 2020.

OLIVEIRA, P. Criador de memes da Nazaré, de “Senhora do Destino”, explica perfil nas redes sociais. **Gshow**, 2017.

Disponível em:

<https://gshow.globo.com/Bastidores/noticia/criador-de-memes->

[da-nazare-de-senhora-do-destino-explica-perfil-nas-redes-sociais.ghtml](#). Acesso em: 27 mai. 2021.

RECUERO, R. **Redes sociais na internet**. 2011. Disponível em: <https://www.researchgate.net/publication/259328435>. Acesso em: 12 dez. de 2020.

RODRIGUES, F. de. A. **Coleta de dados em redes sociais: privacidade de dados pessoais no acesso via Application Programming Interface**. Universidade Estadual Paulista (UNESP), 2017. Disponível em: <https://repositorio.unesp.br/handle/11449/149768>. Acesso em: 07 jun. 2020.

WIGGINS, B. E.; BOWERS, G. B. memes as genre: a structural analysis of the memescape. **New media & society**, v.17, n.11, p. 1887-1906, 2015.

A desinformação sob a ótica da qualidade da informação

João Pedro Silva de Albuquerque^a e Valdirene Aparecida Pascoal^b

1. Introdução

Em tempos de mídias digitais muito se fala em desinformação. No entanto, pouco se reflete sobre o conceito que tanto preocupa as sociedades contemporâneas. O que é a desinformação? Quais são seus impactos na conduta individual e coletiva? Neste trabalho, tem-se o intuito de investigar a desinformação sob a perspectiva da Qualidade de Informação. Para tanto, a questão que guia a presente investigação é assim formulada: As dimensões de Qualidade de Informação poderiam auxiliar no entendimento da estrutura de conteúdos desinformativos? Para tanto, tem-se como objetivo refletir sobre o conceito de informação genuína e degenerada, o conceito de desinformação e os parâmetros que norteiam a Qualidade de Informação.

Desse modo, explora-se o conceito de informação genuína e degenerada elaborado por De Tienne(2006) para iniciar o debate sobre o conceito de desinformação. Em seguida, é discutido como a desinformação também é informação, com o propósito

a Doutorando em Ciência da Informação pela Universidade Estadual Paulista – UNESP. E-mail: joao.albuquerque@unesp.br.

b Doutoranda em Ciência da Informação pela Universidade Estadual Paulista – UNESP. E-mail: valdirene.pascoal@unesp.br. ORCID: <https://orcid.org/0000-0002-3695-6560>.

de ilustrar as relações existentes entre informação degenerada, desinformação e Qualidade de Informação. Por fim é apresentada uma breve análise de conteúdos desinformativos que circularam em mídias sociais a partir das dimensões de Qualidade da Informação, propostas por Batini e Scannapieco (2016) como forma de exemplificar que as dimensões da Qualidade de Informação podem ser princípio-guia para a compreensão da desinformação na contemporaneidade.

2. Informação genuína e informação degenerada

Em consonância com as dimensões de Qualidade de Informação, propõe-se uma relação com a teoria fundamentada por De Tienne (2006) acerca da abordagem de informação Peirceana. O autor elabora uma distinção entre informação genuína e degenerada, a qual propõe cinco características para mensurar a genuinidade de uma informação: verdade, novidade, familiaridade, geração de interpretantes e antecipação. Para que uma proposição seja capaz de veicular informação genuína, ela deve assumir as cinco propriedades a seguir:

- Verdade: a proposição deve transmitir verdade sobre o objeto, possibilitando uma “conexão com a realidade”. O autor ressalta que “conexão com a realidade” implica em “[...] parte da informação, reflete ou relata sobre a condição de uma situação que foi alterada a partir de uma reação direta com a observação original da fonte de informação” (2006, p.8, tradução nossa). Neste contexto,

a proposição será verdadeira mesmo que alguém negue sua veracidade.

- **Novidade:** Além da verdade, De Tienne (2006) considera que uma proposição deve veicular conteúdos que ainda não são conhecidos pelo intérprete. De Tienne ressalta que “A verdade transmitida deve ser algo para alguma mente que ainda não a experienciou, e assim ainda não foi modificada por aquela verdade” (2006, p.8, tradução nossa).
- **Familiaridade:** A familiaridade presente em uma proposição se dá por meio da conexão de elementos de experiências passadas, e, também, alinhadas ao futuro. De acordo com De Tienne: “A verdade transmitida deve referir-se a um universo que efetivamente envolve à mente que a escuta. Tal interesse deve estar enraizado numa parte da experiência passada da mente que ainda não tenha se tornado obsoleta, mas que ainda permaneça conectada ao futuro” (2006, p.8, tradução nossa).
- **Gerar interpretantes:** Na cadeia semiótica peirceana, o interpretante é o efeito, que a “informação recebida” causa na conduta de um intérprete. Ou seja, o interpretante é o elemento semiótico que direciona a ação de um agente. O autor destaca que a geração de interpretantes é característica fundamental para a veiculação de informação: “[...] a semiose que está ocorrendo deve ser bem objetiva, com um poder real para

gerar interpretantes efetivos que serão eles mesmos férteis” (2006, p.8, tradução nossa). Dessa forma a geração de interpretantes fundamenta a informação genuína, garantindo a funcionalidade do sistema.

- Antecipação: Uma proposição, para ser informativa, precisa antecipar fatos, contribuindo assim para que um agente tenha informação sobre acontecimentos futuros e a partir disso molde sua ação. De Tienne pontua: “Informação antecipa fatos quando: não propicia falsos medos ou falsas esperanças, mas simplesmente fornece condições que auxiliem uma negociação bem-sucedida em relação aos eventos futuros” (2006, p.9, tradução nossa).

Em síntese, De Tienne (2006) compreende que para uma proposição veicular informação de modo genuíno ela deve cumprir os cinco requisitos elencados acima, caso essa máxima não se cumpra, a proposição veiculará informação degenerada, dificultando a tomada de decisão do receptor. Entende-se que uma informação degenerada pode ser analisada no âmbito da desinformação, com potencial de auxiliar nas investigações acerca dos impactos da desinformação na conduta. No que segue, apresenta-se uma análise da desinformação a partir de elementos que sustentam uma abordagem de análise baseada na Qualidade da Informação.

3. Desinformação e qualidade da informação

O que se chama de desinformação pode ser entendido como informações não acuradas, falsas ou manipuladas capazes de

influenciar a opinião de alguém mediante a deturpação da verdade (BRITO; PINHEIRO, 2015) e informações, erradas, fora de contexto ou propositalmente falsas (KARLOVA; HA LEE, 2011). Fallis (2015) propõe que a desinformação deve ser analisada a partir de três características básicas: (i) desinformação é informação; (ii) desinformação é uma informação enganosa; (iii) desinformação não é uma informação acidentalmente enganosa. A partir do que os autores falam é possível entender que desinformação é informação, porém a sua finalidade é o engano e manipulação da realidade. Neste sentido, estruturalmente, a desinformação pode ser vista sob alente do conceito de informação, então, da mesma forma que a informação, ela é um conjunto estruturado de representações mentais e emocionais codificadas (signos e símbolos) e modeladas pela interação social, passíveis de serem registradas em qualquer suporte e comunicadas de forma assíncrona e multidirecionada (SILVA; RIBEIRO, 2002).

De acordo com Fallis (2014), a desinformação cumpre sua função de engano por meio de diferentes estratégias como, fazer com que alguém continue a ter uma crença falsa existente, fazer com que alguém adquira uma crença falsa, desista de uma crença verdadeira ou simplesmente não adquira uma nova crença verdadeira. Para tanto, segundo o autor, a desinformação pode ser construída a partir da utilização de fontes de informação falsas, utilizar estratégias que descredibilizam fontes de informações verdadeiras, ocultação, restrição ou polarização do acesso a

determinados tipos de informações, informações inventadas, ou até mesmo informações verdadeiras aplicadas de formas vagas, descontextualizadas, ou representadas de forma que sua visualização cause confusão.

Isso implica em uma variedade de dimensões que podem influenciar na estruturação de um conteúdo desinformativo e na forma como ele vai gerar engano. Isto torna possível a reflexão sobre como os estudos sobre Qualidade da Informação podem ajudar na detecção e na avaliação de conteúdos desinformativos, pois a Qualidade da Informação refere-se ao estabelecimento e definição de papéis, responsabilidades, políticas e procedimentos relativos à aquisição, manutenção, representação e disseminação de informações (BOTEGA et al., 2016).

Illari e Floridi (2014) argumentam que informação de qualidade é informação adequada para um propósito e os usuários das informações possuem diferentes critérios para determinar sua qualidade. Ademais Fallis (2014) baseado em Soo Young Rieh (2002) coloca que a Qualidade da Informação é um “conceito multidimensional” formado por dimensões como precisão, integridade, atualidade, autoridade cognitiva, acessibilidade, entre outras. Estas dimensões permitem entender o significado e a relevância da informação em relação a um contexto ou propósito. Portanto, a manipulação dessas dimensões pode dificultar o entendimento da realidade ou deturpá-la, isto faria com o que passa a ser transmitido seja desinformação em vez de informação.

Ademais, Illari e Floridi (2014) falam que para entender a Qualidade da Informação é necessário entender o usuário, as informações devem ser ajustadas a seus parâmetros e necessidades, para tanto eles falam em dimensões como interpretabilidade, moeda, volatilidade, oportunidade, precisão, integridade e credibilidade, então uma informação para ter aderência deve levar em conta estas dimensões. Já um conteúdo desinformativo para ter aderência usaria estas dimensões para criar uma noção paralela da realidade de forma a se adequar aos vieses e ideologias do usuário, onde a criação do conhecimento dá espaço para distorção ou enganação e a desinformação ganha valor.

A partir disso uma abordagem que avalia a desinformação presente na web sob a ótica de dimensões de Qualidade de Informação pode trazer perspectivas sobre fatores relativos à sua exatidão, atualidade, confiabilidade de suas fontes e os principais fatores que a fazem confundir e serem tidas como verdade por aqueles que a consome.

4. Análise de conteúdos desinformativos online

Nesta seção serão analisados três exemplos de desinformação que circularam em mídias sociais para que se possa ter uma visualização de como as dimensões existentes em Qualidade da Informação se relacionam com a desinformação. Os exemplos foram retirados do site da agência de checagem de informações falsas, denominada ‘aosfatos.org’.

Optou-se por utilizar a classificação de Qualidade da Informação proposta por Batini e Scannapieco (2016) na qual é apresentada uma classificação das dimensões da Qualidade da Informação. Na classificação dos autores as dimensões são incluídas em um mesmo aglomerado de acordo com sua similaridade o que permite avaliar diferentes tipos de informação. São elas: (i) Exatidão correção, validade e precisão enfocam a aderência a uma dada realidade de interesse; (ii) Completude: pertinência e relevância referem-se à capacidade de representar todos e somente os aspectos relevantes da realidade de interesse; (iii) Redundância: minimalidade, compactação e concisão referem-se à capacidade de representar os aspectos da realidade de interesse com o uso mínimo de recursos informativos; (iv) Legibilidade: compreensão, clareza e simplicidade referem-se à facilidade de compreensão e fruição das informações pelos usuários; (v) Acessibilidade: leva em conta disponibilidade e está relacionada à capacidade do usuário de acessar informações de sua cultura, estado físico/funções e tecnologias disponíveis; (vi) Consistência: coesão e coerência referem-se à capacidade das informações de cumprir sem contradições com todas as propriedades da realidade de interesse, conforme especificado em termos de restrições de integridade, edições de dados, regras de negócios e outros formalismos; (vii) Utilidade: relacionada à vantagem que o usuário obtém com o uso da informação; (viii) Confiança: inclui credibilidade, confiabilidade e reputação, capturando quantas informações derivam de uma fonte

autorizada. O aglomerado de confiança abrange também questões relacionadas à segurança.

Para a análise, foram atribuídos os seguintes parâmetros a cada dimensão de Qualidade da Informação: (i) Exatidão: apresentação de dados ou argumentos falsos em relação ao objeto; (ii) Completude: ausência de informações em relação ao objeto; (iii) Redundância: indução a uma resposta; (iv) Legibilidade: complexidade da apresentação dos argumentos; (v) Acessibilidade: tipo de suporte utilizado para compartilhar a desinformação; (vi) Consistência: Apresentação de informações que aproximem o argumento da realidade e evitem contradições; (vii) Utilidade: intencionalidade em relação ao que está sendo apresentado de acordo com um contexto; (viii): Confiança: nível de credibilidade das fontes ou ausência delas.

Caso 1 – O ex-presidente Lula doou terreno para o Hamas construir base no Brasil.

Circulou em mídias sociais um vídeo no qual um homem desconhecido afirmava que o ex-presidente brasileiro Luiz Inácio “Lula” da Silva teria doado um espaço em Brasília que foi usado pela organização palestina Hamas para construção de um quartel-general. Na realidade, o prédio mostrado no vídeo corresponde à embaixada da Palestina em Brasília e sua representação é feita pela ANP (Autoridade Nacional Palestina), liderada pelo Fatah, grupo que se opõe ao Hamas. No vídeo também é falado que o

ex-presidente doou 25 milhões de reais ao Hamas, quando na verdade foi ao Fatah para reconstrução da Faixa de Gaza após o cessar-fogo entre Israel e o Hamas em 2007. Por fim, o vídeo tem origem no ano de 2018, mas foi ressuscitado após os novos conflitos entre Israel e o Hamas iniciados em maio de 2021.

Este exemplo oferece um baixo nível relacionado à dimensão confiança, pois o conteúdo transmitido parte de uma fonte que não se identifica e não cita fontes que colaboram com suas afirmações. Tem uma baixa exatidão para atrelar a embaixada da Palestina a um quartel-general do Hamas, isto é reforçado por uma manipulação da completude que aponta apenas para as características de uma arquitetura relacionada ao oriente médio do prédio. O conteúdo passado não possui consistência por não apresentar nenhuma evidência de ligação do prédio apresentado com o Hamas. Contudo o conteúdo aproveita da legibilidade e acessibilidade para ganhar aderência, pois é um vídeo gravado de forma simples que mostra o objeto de que se fala e o interlocutor utiliza uma linguagem cotidiana. Sua utilidade parece estar relacionada a uma campanha de difamação.

Caso 2 – A suíça teve aumento de Mortes por COVID-19 após suspender hidroxiquina

Uma ex-candidata a deputada federal no Brasil afirmou em uma entrevista dada a um “jornalista especialista em saúde” que a decisão da Suíça em suspender o uso da hidroxiquina no

tratamento da COVID-19 aumentou a taxa de mortes causada pela doença no país europeu entre maio e junho de 2020 de 3% para 11%. A ex-candidata disse que a decisão de banimento do medicamento foi devido a revisão de um artigo científico lançado na revista *The Lancet*. Contudo, no período mencionado a taxa de mortes era de 0,01 para cada 100.000 habitantes, a fonte usada pela ex-candidata é um jornal francês que vem recebendo críticas por divulgar informações incorretas, o artigo citado por ela tinha extraído os dados de um erro de uma base de dados suíça e diversas pesquisas atestam a ineficiência da hidroxicloroquina no tratamento da COVID-19.

Neste exemplo também existe o problema na dimensão confiança pelo fato das fontes citadas serem pouco confiáveis e terem dados errados. Por outro lado, tenta-se criar uma falsa confiabilidade no que dito pelo formato de entrevista a um suposto “jornalista da área médica”. Destaca-se aqui também a legibilidade e acessibilidade pelo conteúdo em formato de vídeo online e possuir linguagem simples. O conteúdo também passa longe da exatidão por usar dados falsos. Tem problemas de consistência, pois não são apresentados outros estudos que colaborem com o argumento. Também há problemas de completude relacionados ao argumento de eficácia da hidroxicloroquina. Existe o apelo a uma redundância para dizer que o número de mortes aumentou, mesmo não sendo verdade, devido apenas a suspensão da cloroquina. Sua utilidade aparenta ser para a criação de uma narrativa em defesa de um

medicamento que não tem comprovação científica contra a COVID-19.

Caso 3 – Fotos de fugitivos de prisão no RN são atribuídas nas redes a mortos no Jacarezinho

Foi compartilhada nas mídias sociais uma imagem de foragidos de um presídio do estado brasileiro Rio Grande do Norte (RN) no ano de 2017 atribuída às pessoas mortas durante uma operação policial na favela do Jacarezinho no Rio de Janeiro (RJ). A imagem foi retirada de um anúncio realizado pela Coordenação de Administração Penitenciária do Rio Grande do Norte (Coape-RN) em 25 de maio de 2017 após uma fuga de 91 presidiários da Penitenciária estadual de Parnamirim, inclusive o nome das pessoas na imagem não corresponde ao de ninguém morto na favela do Jacarezinho. A operação na favela do Jacarezinho foi marcada pelo grande número de mortes, 28 ao todo, também foi alvo de denúncias relacionadas ao excesso de violência policial e a realização de execuções de pessoas dentro de casas e desarmadas.

Aqui o problema na confiança se dá no fato da fonte da informação ser de outro contexto e ser ocultada para que se possa dar credibilidade a informação falsa. A informação traz uma falsa exatidão quanto a identidade das pessoas. A completude foi manipulada ao mostrar apenas imagens de criminosos e houve um apelo à redundância ao tentar trazer a

ideia de que pelo fato da imagem apresentar criminosos logo eles seriam as pessoas que foram mortas na favela do Jacarezinho. Existe um problema de consistência, pois não existe nada na imagem que mostre que as pessoas são da favela do Jacarezinho. O conteúdo se vale de acessibilidade e legibilidade por ser uma fotografia apoiada por postagens que afirmam que as fotos são dos mortos da favela do Jacarezinho. A utilidade se dá na tentativa de justificar as mortes causadas pela ação policial diante das denúncias.

Figura 1 – Dimensões da qualidade da informação relacionadas a desinformação



Fonte: Autores.

A partir da análise desses três casos de conteúdos desinformativos que foram compartilhados em mídias sociais é possível iniciar uma reflexão sobre como as dimensões de Qualidade da Informação, propostas por Batini e Scannapieco (2016) permitem notar as estruturas que dão forma à desinformação. A Figura 1 apresenta como as dimensões de Qualidade da Informação evidenciam partes que compõem a desinformação.

5. Considerações finais

Ao retomar a questão que motivou o presente trabalho “As dimensões de Qualidade da Informação poderiam auxiliar no entendimento da estrutura de conteúdos desinformativos?”, identificou-se análises e definições possíveis para o conceito de desinformação, tendo como parâmetro as dimensões da Qualidade de Informação. Ao investigar o conceito de desinformação, é necessário refletir sobre o conceito de informação, para tanto apresentou-se o conceito de informação genuína elaborada por De Tienne (2006), que em linhas gerais propõe uma definição fundamentada em cinco pilares: verdade, novidade, familiaridade, geração de interpretantes e antecipação. Ainda que não seja uma definição reducionista, auxilia na identificação da informação degenerada e relaciona-se com os parâmetros estruturados na Qualidade de Informação.

Focalizou-se o conceito de desinformação, que de acordo com Fallis (2014), cumpre sua função enganadora a partir de

diversas estratégias que são construídas utilizando fontes de informação falsa ou degenerada. Além disso, apresentou-se as dimensões da Qualidade da Informação enquanto um recurso relevante para a compreensão da informação, seus contextos e estruturas. Bem como, seu potencial para identificar a desinformação transmitida no lugar da informação. Por fim, houve um estudo acerca de exemplos de desinformação veiculados em mídias digitais. Os exemplos elencados possibilitaram identificar a desinformação a partir da estrutura da Qualidade de Informação. Espera-se com a presente discussão contribuir para a reflexão dos impactos que a desinformação pode vir desempenhar na conduta.

Referências

ALLCOTT, H.; GENTZKOW, M. Social Media and Fake News in the 2016 Election. **Journal Of Economic Perspectives**, [S.L.], v. 31, n. 2, p. 211-236, 1 maio 2017. American Economic Association. Disponível em: <https://web.stanford.edu/~gentzkow/research/fakenews.pdf>. Acesso em: 23 jul. 2020.

BATINI, C.; SCANNAPIECO, M. **Data and Information Quality: dimensions, principles and techniques**. [S. I.]: Springer, 2016.

BOTEGA, L. C. *et al.* Methodology for Data and Information Quality Assessment in the Context of Emergency Situational Awareness. **Universal Access In The Information Society**,

[S.L.], v. 16, n. 4, p. 889-902, 3 jun. 2016. Springer Science and Business Media LLC. Disponível em:
<http://dx.doi.org/10.1007/s10209-016-0473-0>. Acesso em: 20 maio 2020.

BRITO, V. de. Pa.; PINHEIRO, M. M. K. PODER INFORMACIONAL E DESINFORMAÇÃO. **Tendências da Pesquisa Brasileira em Ciência da Informação**, João Pessoa, v. 8, n. 2, p. 144-164, dez. 2015.

DE TIENNE, A. Informação em Formação: Uma Perspectiva Peirciana. **Cognitio**: Revista de Filosofia, 6(2), 149-165, 2013. Disponível em:
<https://revistas.pucsp.br/index.php/cognitiofilosofia/article/view/13602/10109>. Acesso em: 24 mai 2021.

FALLIS, D. What Is Disinformation? **Library Trends**, Illinois, v. 63, n. 3, p. 401-426, out.2015.

FALLIS, D. The Varieties of Disinformation. In: FLORIDI, Luciano; ILLARI, Phyllis. **The Philosophy of Information Quality**. Londres: Springer, 2014. p. 1-303.

ILLARI, P.; FLORIDI, Luciano. Information Quality, Data and Philosophy. In: FLORIDI, Luciano; ILLARI, Phyllis. **The Philosophy of Information Quality**. Londres: Springer, 2014. p. 1-303.

KARLOVA, N. A.; LEE, J H. Notes from the underground city of disinformation: a conceptual investigation. **Proceedings Of The American Society For Information Science And Technology**, [S.L.], v. 48, n. 1, p. 1-9, 2011. Wiley.

SARACEVIC, T. Ciência da informação: origem, evolução e relações. **Perspec. Ci. Inf**, Belo Horizonte, v. 1, n. 1, p. 41-62, jan/jun 1996. Disponível em: <http://portaldeperiodicos.eci.ufmg.br/index.php/pci/article/view/235>. Acesso em: 13 jul. 2020.

SILVA, A. M. da.; RIBEIRO, F. **Das “Ciências” documentais à ciência da informação**: ensaio epistemológico para um novo modelo curricular. Porto: Edições Afrontamento, 2002.

O trabalho colaborativo entre Ciência da Informação e Assistência Técnica como fator mitigante da assimetria informacional na agricultura familiar

Luciana Sekito de Freitas Zambelli^a e Cristiane Pantoja de Moraes^b

1. Introdução

Na virada do século XXI, mudanças tecnológicas ocorreram na sociedade causando um fenômeno denominado Big Data em que grandes volumes de dados heterogêneos eram gerados, processados e disponibilizados em tempo real e em velocidade na rede de computadores. O Big Data causou transformações profundas nos âmbitos corporativo, governamental, organizacional e científico permitindo análise e compreensão automatizadas de variados eventos não-identificáveis pelo humano, oriundos de diversas bases de dados. Estas análises e compreensões permitem a definição de padrões e comportamentos dos usuários norteadores da tomada de decisão de entidades para diferentes fins (CONEGLIAN; SEGUNDO; SANT'ANA, 2017).

A assimetria informacional pode ser definida como um fenômeno em que alguns agentes econômicos têm mais

a Mestranda em Zootecnia pela Universidade Estadual Paulista – UNESP. E-mail: luciana.sekito@unesp.br.

b Mestranda em Ciência da Informação pela Universidade Estadual Paulista – UNESP. E-mail: cristiane.moraes@unesp.br. ORCID: <https://orcid.org/0000-0002-3872-2717>.

informações do que outros (BELO; BRASIL, 2006). Já Nascimento e Reginato (2008) citam que, quando se trata de informação seja ela para a tomada de decisão dos gestores, os responsáveis por armazenar e tratar os dados sejam eles físicos, econômicos ou financeiros, sempre disponibilizarão ao usuário um número reduzido de informações do que os profissionais que a disponibilizam, caracterizando assim a assimetria informacional.

Na percepção da Ciência da Informação (CI) a assimetria informacional é um desequilíbrio das densidades entre as dimensões *input* e *output* no ambiente virtual em que os atores da sociedade com maior vantagem competitiva são aqueles que têm maior quantidade de dados acessados e processados lhe conferindo maior poder de uso dos dados para interesse próprio (SANT'ANA, 2019).

Novos potenciais de uso, processos e análise em um ambiente de alta disponibilidade de dados podem trazer consequências aos usuários e à sociedade em geral. A crescente assimetria informacional é uma das consequências e aponta para criação de interações entre a origem e destino dos dados do ponto de vista do usuário com suporte tecnológico profundo e complexo. Este cenário hodierno indica a necessidade de novas visões e reflexões sobre o processo de acesso, uso, análise e visualização de dados disponíveis à sociedade. A Ciência da Informação pode contribuir para a eficiência dos processos dos dados e disseminação do conhecimento facilitando as interações entre os usuários e os dados (SANT'ANA, 2016).

Uma das áreas da sociedade com limitações significativas no acesso às tecnologias de informação e no conhecimento sobre acesso e uso de dados e elemento alvo essencial na mitigação da assimetria informacional é a Agricultura Familiar (SANT'ANA, 2017).

A relevância socioeconômica da Agricultura Familiar reside na sua dispersão geográfica de 4,3 milhões de propriedades rurais representando 84% do total de estabelecimentos agropecuários no país que permitem o abastecimento de alimentos convencionais ou orgânicos de qualidade para a sociedade brasileira. Esta classe contribui com cerca de 12,3 milhões de empregos, representando 74% do total de pessoas que exercem algum tipo de atividade rural, gerando renda de 54,3 bilhões de reais equivalente a 38% do total gerado pelo agronegócio brasileiro, além de implantar ações sustentáveis demonstrando sua responsabilidade socioambiental. Por sua importância socioeconômica, os pequenos produtores merecem receber devida atenção para a manutenção do abastecimento interno alimentar alinhado às demandas da população de produção de alimentos de qualidade de forma sustentável e do bem-estar animal (CODAF, 2021).

A assimetria informacional origina-se da dificuldade de acesso e entendimento do ambiente digital pelo agricultor familiar. Iniciativas públicas efetivas como o projeto Competências Digitais para Agricultura Familiar (CoDAF) buscam minimizar a assimetria informacional proporcionando

fácil acesso e uso dos dados para a otimização da atividade pelos produtores familiares (AFFONSO; PERRONI, 2016).

Gomes (2018), em uma análise dos resultados de eficiência e rentabilidade na produção de leite de um programa de assistência técnica de iniciativa público-privada no Estado de Minas Gerais, demonstrou a existência de um ciclo virtuoso no processo contínuo e dinâmico de assistência técnica de qualidade nas propriedades, significando que produtores que recebiam assistência técnica há mais tempo mostraram-se tecnicamente mais eficientes. Estes produtores tiveram maior capacidade de expansão da renda, maior capacidade de reinvestimento na atividade em decorrência do aumento de produção com eficiência atingido com suporte da Assistência Técnica. A Assistência Técnica contribui na melhoria da produtividade através da implantação e aprendizado de tecnologias para aumento da rentabilidade da propriedade familiar (GOMES et al., 2018).

Os esforços de vários atores da sociedade, como governos, empresas de desenvolvimento de aplicações e instituições de ensino são necessários a fim de reduzir a assimetria de informação nas atividades agropecuárias (AFFONSO; PERRONI, 2016).

Neste contexto, o objetivo deste estudo foi desenvolver uma narrativa relacionando conceitos teóricos relevantes, como Ciclo de Vida dos Dados, Campo e Transduções Informacionais, Qualidade dos dados, complementares às competências da Ciência da Informação e que podem ser aplicados no contexto do

trabalho conjunto à Assistência Técnica para mitigar a assimetria informacional na agricultura familiar. O objetivo secundário foi criar um ambiente estimulador à reflexão da CI sobre as possibilidades na disseminação e aplicação do conhecimento amplo e relevante da acessibilidade e gestão dos dados tendo o papel de facilitador, executor e, principalmente, gestor em trabalho colaborativo com profissionais da Assistência Técnica.

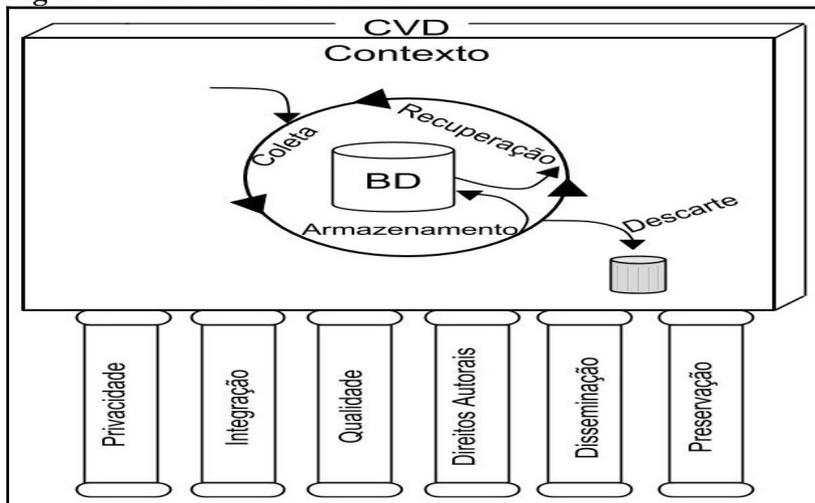
2. Procedimentos metodológicos

A metodologia se constituiu em um estudo teórico, de caráter qualitativo através de análise e interpretação do conteúdo de aulas expositivas virtuais ministradas nos meses de abril e maio de 2021 na Disciplina de Dados, Tecnologia e Informação na Ciência da Informação (DTI-CI) do Programa de Pós-Graduação da Faculdade de Filosofia e Ciências (PPGCI – FFC) da UNESP de Marília/SP. Foram consideradas as respostas de um questionário no formato Google dos participantes da aula sobre transdução informacional de um exercício denominado “De volta para o futuro”. A pesquisa bibliográfica no contexto da agricultura familiar e assistência técnica correlatas ao projeto complementam o estudo. A coleta de artigos complementares foi realizada através de pesquisa de palavras-chave como “agricultura familiar” e “importância da assistência técnica” no Google Acadêmico e de textos, artigos e links disponibilizados no site da CODAF.

3. Ciclo de Vida dos Dados (CVD)

O Ciclo de Vida de Dados (CVD) proposto a partir da perspectiva da CI apresenta-se como uma estrutura de referência cíclica orientadora dos estudos, reflexões e definições envolvidos no complexo processo de acesso, uso e manutenção dos dados. O conhecimento e a aplicabilidade do CVD permitem à CI propor soluções em todo processo de gestão dos dados. O CVD compreende as fases de coleta, armazenamento, recuperação e descarte (Figura 1) (SANT'ANA, 2016).

Figura 1 – Ciclo de Vida dos Dados



Fonte: Sant'Ana (2016).

As fases são codependentes e interrelacionam-se de forma concomitante ou não, tornando o processo cíclico com representação de todos os momentos em que os dados são envolvidos. Importante ressaltar que todas as fases deste ciclo são permeadas por todos os fatores em maior ou menor grau e que durante o planejamento do projeto ou processo devem ser considerados, funcionando como elos entre as fases. Os fatores são: privacidade, integração, qualidade, direitos autorais, disseminação e preservação (SANT'ANA, 2019).

A coleta e recuperação têm maior aderência à CI, enquanto as fases de armazenamento e descarte apresentam maior aderência ao trabalho da Ciência da Computação (SANT'ANA, 2016). A exposição e discussão sobre o CVD na sua integralidade não se alinha ao âmbito do estudo. Um enfoque maior nas fases de coleta e recuperação dos dados está mais alinhado à temática do estudo.

- a) Coleta: nesta fase em que se objetiva a obtenção dos dados deve incluir em seu escopo a definição, escolha, localização e avaliação dos dados necessários, assim como os mecanismos, metodologias e ferramentas para a consecução dos dados. A coleta pode ser processual em que a aquisição dos dados é dinâmica e contínua e a cadência da coleta deve ser definida através da identificação do tempo entre as tomadas de medidas ou obtenção dos valores. A cadência da coleta está diretamente ligada à granularidade dos dados, o intervalo

menor com maior precisão gerará um volume maior de dados obtidos, enquanto o intervalo maior gerará volume menor com menor granularidade. A dinâmica desta coleta permite que o processo coexista com as demais fases. Por outro lado, a coleta pode ser pontual com estrutura de projeto com início e fim pré-definidos, e cada procedimento pode ter configurações distintas. A coleta pontual pode apontar a necessidade de atualizações dos dados coletados em um movimento cíclico com definição de blocos e subconjuntos de dados dentro do escopo da coleta, por exemplo, no entanto sem apresentar caráter contínuo pelas suas características iniciais intrínsecas. Nesta fase, são necessárias competências específicas e o nível de conhecimento dependerá da necessidade informacional que direciona a coleta. Percebe-se uma fase em que o trabalho em conjunto da CI, usuário e Ciência da Computação pode tornar o processo mais eficiente (SANT'ANA, 2016).

- b) Armazenamento: nesta fase tem grande enfoque tecnológico e uma atuação maior da Ciência da Computação que define o processo como persistência de dados. Há exigência de um planejamento detalhado de ações para assegurar a reutilização dos dados através de especificações físicas e lógicas para registro em um suporte, além de novos processos de análise direta ou com interação com outras bases de dados. Nesta fase, há

potencial de participação da CI e o usuário fica mais distante colaborando na validação dos modelos de estruturas para os dados solicitados (SANT'ANA, 2016).

- c) Recuperação: para Sant'Ana (2016), é notável o quanto a tecnologia vem se desenvolvendo e com isso a aquisição e manutenção dos equipamentos digitais se tornam cada vez mais acessíveis facilitando o armazenamento e disponibilização das informações. Manter e fornecer os conteúdos pode gerar inquietações, principalmente quanto ao armazenamento, pois não se sabe quanto e quais dados estarão disponíveis e em quais suportes físicos serão armazenados para garantir a permanência da informação. Além disso, entra o questionamento sobre a privacidade, integridade e veracidade das informações. Todos esses aspectos levam a somente um conceito que está diretamente ligado à encontrabilidade ou recuperação dos dados armazenados, sugerindo a necessidade de colaboração entre a instrumentalização da Ciência da Computação, a visão de acesso aos dados da CI e aqueles que tenham conhecimento das necessidades do público-alvo ou potenciais usuários. Moreira, et al (2020) em seu estudo sobre dados de políticas públicas para o setor agrícola na internet é compreensível que o ato de compartilhar dados depende de uma certa qualidade podendo ser beneficiários em potencial os pesquisadores que necessitam de fontes de dados para análise e, em

geral, para a sociedade, pois envolve o exercício da transparência governamental e da cidadania. Entender como a dimensão da qualidade se reflete no fornecimento de dados governamentais pode ajudar aqueles que possuem o banco de dados e fornecer conteúdo para adoção de estratégias de comunicação mais eficazes. O governo deve desempenhar um papel na concepção e distribuição dos dados, gerando condições que melhor se adéquem ao aproveitamento e reutilização dos conteúdos, levando sempre em consideração a Lei de Acesso à Informação e o princípio da transparência ativa. Com isso, aumentará a possibilidade de novos requisitos para o acesso à informação e a participação da sociedade nas políticas públicas.

- d) Descarte: a evolução dos recursos digitais proporciona custos de aquisição, manutenção e armazenamento de dados cada vez mais viáveis. A decisão de não disponibilizar parte ou todo dos dados por questões como privacidade, direitos autorais ou espaço de armazenamento é persistente. Esta fase é identificada como descarte e é demandante de atividades e competências específicas para o planejamento e execução das ações da Ciência da Computação. A CI exerce papel importante e os usuários contribuem com a validação e acompanhamento do processo durante esta fase (SANT'ANA, 2016).

As fases são interdependentes apresentando forte vinculação entre si exigindo que no planejamento inicial considere-se os elementos contextuais de cada uma das fases ocorrendo de forma simultânea ou não. Assim, os dados coletados podem ser armazenados, recuperados ou descartados, mas não obrigatoriamente e somente, nessa ordem, o que configura o caráter cíclico e passível de sincronicidade do processo resultando em fluxos informacionais (SANT'ANA, 2019).

4. Campo e assimetria informacional

A resultante dos fluxos informacionais em um ambiente virtual com alta disponibilidade de conectividade possibilita a reflexão sobre a existência de um campo informacional representado pela interação de Ciclos de Vida dos Dados sujeito a fatores como: competência do elemento alvo ou agente, características culturais e forma de controle sobre recursos necessários para o processo.

O elemento alvo ou agente pode ser identificado como aquele que está sendo submetido, com ou sem consentimento ou ciência, a um conjunto de CVDs na fase de coleta ou de recuperação sendo indivíduo ou instituição. Assim, considera-se a existência da dimensão *output* do campo informacional caracterizada pela densidade e gradiente resultantes do volume e variedade de dados gerados pelas interações dos conjuntos de CVDs.

Analogamente, podemos considerar a existência da dimensão *input* caracterizada pela densidade resultante do conjunto de

processos de coleta dos dados os quais os elementos alvo detêm. Dada a ciclicidade do processo de coleta de dados a partir dos elementos alvo ou de outros CVDs, consideramos a possível ocorrência do reúso dos dados.

As dimensões resultantes podem, em uma primeira instância, serem mensuradas em bits/tempo, ou seja, o volume de dados transacionais nos fluxos em todos CVDs gerados em um determinado momento. Neste contexto, as análises das dimensões *output* e *input* permitem a concepção de um novo espaço multidimensional composto por ambientes virtuais com características e limitações indefinidas, mas que gera desconforto ao sair do âmbito físico. Como resultado, é possível percebermos a existência de assimetria do elemento alvo no campo informacional por meio do desequilíbrio entre a quantidade de dados que fluem entre as ambas as dimensões configurando desigualdade entre os elementos da sociedade. A proporção entre as dimensões definirá qual elemento detém maior vantagem e o poder de uso dos dados (SANT'ANA, 2019).

5. Transdução informacional

Seguindo no entendimento um pouco mais profundo deste novo espaço multidimensional e as estruturas sobre as quais os processos se alicerçam, consideramos a constituição das camadas de abstração derivadas da interpretação da transdução informacional. Na visão da CI, a transdução informacional do conteúdo registrado em suportes digitais é definida como processo de transformação de um tipo de sinal ou estímulo de

energia (elétrica, química, física) em um outro tipo de sinal ou produto. A transdução pode ser medida através de uma representação deste processo transdutor do conteúdo digital em bits persistidos nos suportes agregando mudanças semânticas consecutivas até o formato resultante nas interfaces dos usuários finais (SANT'ANA, 2019).

Um exercício em aula foi proposto aos alunos para que realizassem a identificação, descrição e análise comparativa de uma determinada atividade relacionada ao acesso, uso ou transmissão de informação envolvendo dados, desempenhada nos dias de hoje e há 30 anos. Os resultados demonstraram variabilidade nas interpretações e descrições das transduções e sugeriram que a prática de identificar e descrevê-las pode ser influenciada pelo nível de formação e conhecimento técnico de cada indivíduo (dados não mostrados).

O entendimento pelo leitor desta descrição ficou passível de sofrer uma possível influência pela sua concepção, conhecimento e experiência de vida. Os alunos relataram diferentes níveis de dificuldade e/ou facilidade para a realização das tarefas, seja no âmbito da escolha da atividade, da comparação entre o tempo atual e há 30 anos (devido à heterogeneidade de faixas etárias entre os alunos) ou na descrição e profundidade das transduções informacionais que podem requerer conhecimentos específicos como o domínio sobre os algoritmos envolvidos nas fases de acesso a dados e as diferentes e pervasivas interfaces e suas características relacionadas aos usuários finais.

6. Qualidade dos dados

O aprimoramento digital das técnicas praticadas que permitem o acesso contínuo a dados digitais confiáveis e de qualidade no contexto da pesquisa com autenticidade verificada, demonstram que os registros de arquivos são válidos e que eles podem ser usados como prova legal no futuro. É comum que em diferentes conjuntos de dados fornecidos pela coleta, a criação de oportunidades de pesquisa seja de forma colaborativa, do ponto de vista financeiro, gerando compartilhamento, reutilização e posterior análise desses dados benéficos aos usuários (SAYÃO; SALES, 2012).

A qualidade da representação dos dados publicados de forma legível e entendível por analistas humanos ou máquinas para gerar conhecimento através de interpretações mais complexas evitando impactos negativos nas tomadas de decisão é essencial. Uma informação de qualidade é aquela que atende aos requisitos para seu uso específico e considerando a complexidade dos dados pela característica pervasiva e heterogênea disponíveis em processos dinâmicos em volumes, em diferentes âmbitos, fontes e acessados nas redes por diversas arquiteturas. Entende-se que a CI possui a responsabilidade e competências para, por exemplo, desenvolver metodologias de verificação que garanta a qualidade dos dados (BOTEGA et al., 2019, 2017).

7. Considerações finais

Como resultado, obteve-se uma narrativa didática relacionando os conceitos teóricos sobre o CVD e suas interações resultantes em campo informacional, a complexidade das transduções informacionais e a importância da qualidade dos dados gerados em processos de mitigação da assimetria informacional na Agricultura Familiar.

O agricultor familiar busca a melhora da sua produtividade, a fim de manter-se na atividade de forma eficiente e garantir a renda familiar. Porém, no cenário atual, encontra dificuldades no acesso e atendimento das políticas públicas, no acesso a informações técnicas relevantes ao aumento da produtividade agrícola e no acesso ao financiamento agrícola, por exemplo.

O trabalho colaborativo entre a Ciência da Informação e a Assistência Técnica torna-se uma reflexão necessária a fim de se buscar continuamente a mitigação da assimetria informacional, através da organização dos esforços, da implementação de ações benéficas ao agricultor familiar considerando a complementaridade das competências de cada profissional atuante no processo como um todo.

Referências

AFFONSO, E. P.; PERRONI, V. Tecnologias da Informação e Comunicação na Agricultura Familiar: um olhar na produção científica de Ricardo César Gonçalves Sant'Ana. **Revista Eletrônica Competências Digitais para Agricultura Familiar**, v. 2, n. 1, p. 20–40, 3 nov. 2016. Disponível em:

<http://200.145.54.28:8082/index.php/recodaf/article/view/20>. Acesso em: 10 maio 2021.

BELO, N. M.; BRASIL, H. G. Assimetria informacional e eficiência semiforte do mercado. **Rev. adm. empres.**, São Paulo, v. 46, n. spe, p. 48-57, 2006. Disponível em: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0034-75902006000500004&lng=en&nrm=iso. Acesso em: 17 maio 2021.

BOTEGA, L.C. et al. Gestão da qualidade de dados conectados: fundamentos e metodologia. In: OLIVEIRA H; VIDOTTI, S. (Orgs.). **Informação e tecnologias: desenhando fronteiras científicas**. Editora UFPB – João Pessoa. 2018. p. 113–134. Disponível em: <http://www.editora.ufpb.br/sistema/press5/index.php/UFPB/catalog/book/246>. Acesso em: 04 maio 2021.

_____. Methodology for Data and Information Quality Assessment in the Context of Emergency Situational Awareness. **Universal Access in the Information Society**, v. 16, n. 4, p. 889–902, 1 nov. 2017. Disponível em: [https://www.researchgate.net/publication/303793713 Methodology for Data and Information Quality Assessment in the Context of Emergency Situational Awareness](https://www.researchgate.net/publication/303793713_Methodology_for_Data_and_Information_Quality_Assessment_in_the_Context_of_Emergency_Situational_Awareness). Acesso em: 04 maio 2021.

CoDAF – Competências Digitais para Agricultura Familiar.

Disponível em: <https://codaf.tupa.unesp.br/>. Acesso em: 16 maio 2021.

CONEGLIAN, C. S.; SEGUNDO, J. E. S.; SANT’ANA, R. C. G.

Big Data: fatores potencialmente discriminatórios em análise de dados. **Em Questão**, v. 23, n. 1, p. 62–86, 1 jan. 2017. Disponível em:

<https://www.seer.ufrgs.br/EmQuestao/article/view/62122>.

Acesso em: 21 mar. 2021.

GOMES, A. P. et al. Assistência técnica, eficiência e rentabilidade na produção de leite. **Revista de Política Agrícola**, v. 27, n. 2, p. 79, 26 dez. 2018. Disponível em:

<https://seer.sede.embrapa.br/index.php/RPA/article/view/1414>

. Acesso em: 02 maio 2021.

MOREIRA, F. M. et al. A qualidade na recuperação de dados governamentais: um estudo sobre dados de políticas públicas na internet. **Perspectivas em Ciência da Informação**, v. 25, n. 2, p. 103-132, 2020. Disponível em:

[https://www.scielo.br/scielo.php?pid=S1413-](https://www.scielo.br/scielo.php?pid=S1413-99362020000200103&script=sci_arttext&tlng=pt)

[99362020000200103&script=sci_arttext&tlng=pt](https://www.scielo.br/scielo.php?pid=S1413-99362020000200103&script=sci_arttext&tlng=pt). Acesso em: 21

mar. 2021.

NASCIMENTO, A. M.; REGINATO, L. Divulgação da informação contábil, governança corporativa e controle organizacional: uma relação necessária. **Revista Universo**

Contábil, v. 4, n. 3, p. 25-47, 2008. Disponível em:

<http://www.spell.org.br/documentos/ver/6165/divulgacao-da->

[informacao-contabil--governanca-corporativa-e-controle-organizacional--uma-relacao-necessaria](#). Acesso em: 18 maio 2021.

SAYÃO, L. F. et al. Curadoria digital: um novo patamar para preservação de dados digitais de pesquisa. **Informação & Sociedade: Estudos**, v. 22, n. 3, p. 179-191, 2012. Disponível em: https://www.researchgate.net/profile/Luana-Sales-2/publication/235334882_CURADORIA_DIGITALum_novo_patamar_para_preservacao_de_dados_digitais_de_pesquisa/links/09e41510fb129a419c000000/CURADORIA-DIGITAL-um-novo-patamar-para-preservacao-de-dados-digitais-de-pesquisa.pdf. Acesso em: 18 maio 2021.

SANT'ANA, R. C. G. Campo Informacional Resultante da Interação de Ciclos de Vida dos Dados. In: DIAS, G.; FREIRE, B. (Orgs). **Dados Científicos: perspectivas e desafios**. Editora UFPB – João Pessoa. 2019. p.5-19.

_____. Ciclo de vida dos dados: uma perspectiva a partir da ciência da informação. **Informação & Informação**, v. 21, n. 2, p. 116, 20 dez. 2016. Disponível em: <https://brapci.inf.br/index.php/article/view/0000022210/6eb47c5863b44510764cd65ea0ffefc5>. Acesso em: 08 abril 2021.

_____. Estudo sobre acesso a dados de pesquisa sobre Agricultura Familiar. **Revista Eletrônica Competências Digitais para Agricultura Familiar**, v. 3, n. 1, p. 24–37, 26 jul. 2017. Disponível em:

<http://200.145.54.28:8082/index.php/recodaf/article/view/46>.
Acesso em: 02 maio 2021.

_____. Transdução Informacional: impactos do controle sobre os dados. In: MARTÍNEZ-ÁVILA, D; SOUZA, E.A.; GONZALEZ, M. E. Q. (Orgs.). **Informação, conhecimento, ação autônoma e big data**: continuidade ou revolução? Cultura Acadêmica – FiloCzar – São Paulo, 2019. p.117-128. Disponível em: <http://books.scielo.org/id/gfrbh/pdf/martinez-9788572490559.pdf>. Acesso em: 08 abril 2021.

Impacto da qualidade da informação no prontuário do paciente: Reflexões sobre a fase de coleta

Fabrizio Amadeu Gualdani^a, Nathália Adrielle de Lima^b e Leonardo Castro Botega^c

1. Introdução

No campo da saúde, assim como em outras áreas do conhecimento, as tecnologias de informação estão presentes nos mais diversos setores, seja na parte administrativa e de gestão hospitalar, na produção de medicamentos, no ensino e pesquisa, ou na construção de sistemas de informação dedicados aos prontuários visando favorecer o acesso e a recuperação das informações referentes aos pacientes. Aprimorar cada vez mais os serviços de diagnóstico e oferecer um tratamento adequado às necessidades do paciente é o papel principal das instituições e dos sistemas de saúde.

O prontuário é um dos recursos fundamentais para a manutenção de um sistema de saúde e é utilizado como um objeto de intermédio entre o profissional da saúde e o paciente. Ou seja, é um documento multidisciplinar, temporal e abrangente

a Mestrando em Ciência da Informação pela Universidade Estadual Paulista – UNESP. E-mail: fabriziogualdani@gmail.com.

b Mestranda em Ciência da Informação pela Universidade Estadual Paulista – UNESP. E-mail: nathalia.adrielle@unesp.br.

c Doutor em Ciência da Computação pela Universidade Federal de São Carlos – UFSCar. Professor na Universidade Estadual Paulista – UNESP. E-mail: <https://orcid.org/0000-0003-1495-5935>. ORCID: <https://orcid.org/0000-0003-1495-5935>.

devido aos diversos tipos de profissionais que registram informações neste documento e acessam seus conteúdos informacionais (CARVALHO, 2018).

Quanto ao acesso ao conteúdo desse documento, o suporte papel carrega limitações para se encontrar uma informação específica sobre um dado do paciente, pois depende da disponibilidade humana em ter acesso físico ao prontuário. Cabe à instituição de saúde planejar, organizar e executar um sistema de acesso físico a essa documentação (GALVÃO; RICARTE, 2012).

O prontuário eletrônico, com o conteúdo integralmente no formato digital, abre novas possibilidades de uso dessa importante fonte de informação, para a assistência ao paciente, para a pesquisa, para a gestão e integração de novas fontes de informação. (GALVÃO; RICARTE, 2012). No entanto, não é a mera existência e utilização do prontuário em formato eletrônico que irá automaticamente melhorar a qualidade do preenchimento das informações inseridas neste documento.

No momento em que um profissional da saúde for realizar o preenchimento dos dados no prontuário, torna-se necessário que ele desenvolva uma consciência desse processo como uma fase de obtenção, isso é, uma coleta dos dados durante o processo de anamnese, para se organizar a informação (seja no suporte analógico ou eletrônico) e dessa forma facilitar a visualização como um todo desses dados, sendo possível para o profissional da saúde tanto chegar a um diagnóstico como realizar uma melhor tomada de decisão direcionada ao paciente.

Para Sant'Ana (2016) ao se coletar dados é necessário compreender quais são as necessidades informacionais, assim como definir se serão necessárias novas realizações do procedimento de coleta, gerando um processo contínuo de atualização dos dados. Da mesma forma, uma característica essencial ao se coletar dados é a definição e a garantia de elementos que permitam a percepção da qualidade desses dados coletados, portanto, elementos como a procedência, mecanismos de coleta e garantias de integridade física e lógica representam apenas alguns dos aspectos a serem considerados. A confiabilidade dos dados é uma condição prévia para que um dado seja útil para um possível acesso futuro, buscando identificar elementos contextuais dos dados que possam favorecer sua localização e interpretação na fase de recuperação.

Os prontuários do paciente possuem informações registradas em campos textuais em um formato de dados não estruturados, outro tipo de situação que prejudica muito a recuperação desses dados clínicos (SOUZA; ALMEIDA, 2019).

Diante dessas situações, recuperar as informações em todos os seus formatos, sejam elas estruturadas, não estruturadas (informações que se fazem presentes em textos livres), trata-se de uma atividade necessária.

Feitas essas observações, nasce uma preocupação com a qualidade de dados e informações internas, que caso apresentarem um baixo desempenho, poderão proporcionar maiores falhas, incertezas e erros quanto aos procedimentos

executados. Gerando uma série de problemas e dificuldades perante o processo de recuperação da informação, problemas esses relacionados a qualidade das informações preenchidas nos prontuários eletrônicos que dificultam o processo de recuperação da informação, tais como: erros de digitação, falta de revisão no texto final, erros ortográficos, gramaticais, remoção de sufixos e prefixos, sinônimos, abreviações, siglas, símbolos, pontuações, termos e jargões médicos (CARVALHO, 2018).

Desta maneira, este artigo possui como objetivo demonstrar como que o fator da qualidade de dados pode interferir no processo de coleta em prontuários do paciente, através de uma revisão de literatura tanto nacional como internacional, compreendendo a quais resultados e conclusões os autores têm obtido por meio de suas pesquisas. Isso é, seguindo os procedimentos metodológicos de uma pesquisa básica e exploratória, visando compreender e comparar os resultados obtidos, assim como as conclusões de diferentes autores quanto aos principais problemas de qualidade da informação, voltados para o momento da coleta de dados para o prontuário do paciente que conseqüentemente necessitarão de uma recuperação desses dados.

A realização dessa pesquisa se justifica pelo fato de que antes de propor soluções tecnológicas ou até mesmo aperfeiçoar os métodos já existentes sobre uma boa coleta de dados direcionadas aos prontuários do paciente (seja no formato analógico ou eletrônico), torna-se necessário compreender e

delimitar com clareza como a qualidade da informação vem sendo tratada nesse processo. Dessa forma, feita a apresentação da temática deste artigo, a próxima seção apresenta a metodologia utilizada para o desenvolvimento deste estudo.

2. Procedimentos metodológicos

Os procedimentos metodológicos adotados nesta pesquisa, consistem do ponto de vista da sua natureza em uma pesquisa básica e considerando seus objetivos, trata-se de uma pesquisa exploratória. No que diz respeito aos procedimentos técnicos, este trabalho compõe-se de uma pesquisa bibliográfica que incluiu artigos científicos publicados em periódicos, sendo que as principais áreas que dialogam com a abordagem proposta, foram nas áreas de ciência da informação, ciência da computação e ciências da saúde.

Logo, o presente artigo pode ser configurado como uma revisão sistemática de literatura, pois permite a realização de uma análise de artigos que respondem a determinados critérios de inclusão e exclusão, possibilitando, desta forma, encontrar melhores evidências que fundamentem conclusões (FONTES-PEREIRA, 2017).

A busca por artigos que compõem esta revisão se iniciou no período de março até junho de 2021 e foram inseridos os termos [*Recuperação da Informação Prontuário Eletrônico do Paciente* *Qualidade Informação Prontuário do Paciente* *Data Quality Electronic Health Records* *Information Retrieval Electronic Health Record*], em bases de dados como a BRAPCI,

SCIELO, ACM, ScienceDirect, PubMed, Archivaria e IBICT. Foram inseridos artigos nacionais e artigos internacionais, publicados nos anos de 2003 até 2021.

Os critérios de inclusão dos artigos para esse estudo, compõem-se de pesquisas que trabalharam diretamente com o desenvolvimento, aplicação, análise, avaliação ou revisão de literatura a respeito seja exclusivamente da questão da qualidade do preenchimento das informações no prontuário eletrônico, da recuperação em si da informação presente neste documento, como na qualidade da informação disponibilizada no processo de recuperação de dados no prontuário eletrônico.

A partir da coleta dos artigos procedeu-se a leitura classificatória, na qual foram construídos resumos que destacaram as conclusões de cada artigo. Foram selecionados aqueles que abordaram, direta ou indiretamente, os critérios de coleta e qualidade de dados, assim como a recuperação da informação. Esses resumos serão apresentados na sequência.

3. Resultados e discussões

Martha, Barra e Campos (2004) realizaram um levantamento a respeito de recuperação de informações médicas em textos livres de prontuários eletrônicos do paciente. A metodologia empregada, consistiu no desenvolvimento de um artigo dissertativo. Algumas dificuldades podem ocorrer na recuperação de informações como erros de digitação pela falta de corretor ortográfico, falta de revisão do texto final. Além disso, profissionais da saúde possuem um estilo próprio na inserção de

textos e podem usar palavras sinônimas, abreviações, jargões médicos.

Dorileo et al. (2006) apresentaram uma proposta de estruturação do registro de evolução clínica, objetivando sua integração ao prontuário eletrônico do paciente. Esta estruturação proporciona um nível de organização que otimizará a recuperação da informação, com conseqüente agilização da prática clínica, representando mais um passo em direção à informatização eficiente do prontuário médico. Apesar de serem bem preenchidos, durante o levantamento de informações realizado a partir da leitura de prontuários médicos foram encontrados diversos problemas, tais como: registros ilegíveis, falta de padronização de termos e uso de abreviações não padronizadas.

Galvão et al. (2008) mapearam e sistematizaram as superestruturas das tipologias textuais mais frequentes em áreas específicas do setor da saúde, pretendendo com isto aprimorar a análise e recuperação dos conteúdos informacionais contidos nestes textos, buscando disponibilizar conhecimentos básicos para a gestão eletrônica dos documentos institucionais. Neste esforço, devem estar presentes não apenas os profissionais da informática, bem como, os profissionais da ciência da informação, da ciência da terminologia, da linguística e os profissionais provenientes de diferentes campos. Finalmente, a gestão da informação deve ser não apenas projetiva (considerar a criação da informação do presente para o futuro), mas devem ser

também retroativos (ou seja, devem se atentar para a organização e recuperação da informação já existente e armazenada).

Dixon et al. (2011) afirmam que a qualidade dos dados clínicos não é homogênea entre as fontes. Existem dados de baixa qualidade em sistemas de informação clínica, o que representa um desafio para os interessados em estudos secundários sobre as diversas formas de utilização de dados de registros eletrônicos de saúde. Além disso, o processo de coletar dados deve ser melhor compreendido por todas as partes interessadas no intercâmbio de informações sobre saúde. Novos métodos para coletar esses dados ou reaproveitá-los devem ser explorados, aperfeiçoando, assim, o fluxo de trabalho.

Pavão et al. (2011) avaliaram a qualidade da informação dos prontuários de três hospitais de ensino do estado do Rio de Janeiro. A qualidade dos prontuários foi considerada ruim. Conclui-se que informações indispensáveis ao cuidado dos pacientes apresentaram baixo registro. Por fim, os autores ressaltam a importância da elaboração de medidas que visem melhorias na qualidade do prontuário, que refletirão na qualidade da assistência ao paciente.

Feder (2017) demonstra que a avaliação da qualidade dos dados é um empreendimento significativo para qualquer estudo que pretenda usar dados do registro eletrônico de saúde. Dados de alta qualidade aprimoram a confiabilidade dos resultados do estudo e são um componente essencial desses registros.

Souza e Almeida (2019) analisaram os dados clínicos textuais do prontuário eletrônico do paciente para descrever uma forma de conexão com as terminologias médicas padronizadas (ontologias, vocabulários controlados) para possibilitar a recuperação desses dados. Entretanto a maioria das informações contidas em um prontuário eletrônico do paciente são disponibilizadas como dados não estruturados, o que dificulta a utilização dos dados clínicos para fins de pesquisa. Os dados não estruturados, como registrados por médicos, apresentam uma grande variedade de sinônimos, acrônimos, e idiossincrasias que não corresponde a terminologias médicas padronizadas, resultando em dificuldades para a recuperação de informação.

Hanauer et al. (2020) afirmam que registros médicos eletrônicos podem ser uma ferramenta valiosa como apoio para pesquisas direcionadas a doença de câncer, bem como outros domínios clínicos. Isto é, uma ferramenta de autoatendimento simples de operar que é poderosa, escalável e generalizável entre os casos de uso, permitindo equipes de vários campos aumentar sua produtividade e obter acesso a dados precisos do paciente que normalmente exigiriam uma abordagem manual para identificação.

Sengan et al. (2020) em cooperação com profissionais de saúde aplicou modelos de aprendizado de máquina baseado em Fuzzy em sistemas de recuperação de informações médicas para problemas clínicos específicos. Conclui-se que o uso de métodos de aprendizado de máquina para implementações clínicas têm um

enorme potencial para melhorar a atual oferta de serviços de saúde.

Oliveira e Favaretto (2021) aplicaram um modelo adaptado de avaliação da qualidade da informação do prontuário eletrônico do paciente do hospital de clínicas de Itajubá com o objetivo de propor melhorias na qualidade dos dados. O caminho metodológico permitiu desenvolver um projeto de qualidade de informação tendo como resultado as ações necessárias para melhoria contínua da informação. Foi possível identificar as dimensões da reputação, acessibilidade e valor agregado como as que mais impactam na qualidade da informação do prontuário eletrônico.

Os resultados obtidos através do desenvolvimento da revisão sistemática de literatura desse estudo demonstraram que muitos dos dados oriundos do prontuário independentemente do suporte, tanto físico como eletrônico se apresentaram em um formato não estruturado, apresentando uma grande variedade de sinônimos, acrônimos, e idiossincrasias que não correspondem a terminologias médicas padronizadas, resultando em dificuldades para a recuperação de informação. Existem dados de baixa qualidade em sistemas de informação clínica, portanto, realizar a avaliação da qualidade de dados é uma atividade necessária ao se trabalhar com dados pertencentes a prontuários do paciente, devido ao fato de que dados de alta qualidade aprimoram a confiabilidade dos estudos. Além disso, o processo de coletar dados em registros eletrônicos de saúde deve ser melhor

compreendido por todos partes interessadas no intercâmbio de informações sobre saúde.

Novos métodos para coletar esses dados ou reaproveitá-los aperfeiçoando assim o fluxo de trabalho devem ser explorados. Esta estruturação proporciona um nível de organização que otimizará a recuperação da informação, com consequente agilização da prática clínica, representando mais um passo em direção à informatização eficiente do prontuário médico.

4. Considerações finais

A partir dos resultados e discussões apresentadas, conclui-se que há necessidade da compreensão de que trabalhar com a qualidade de dados envolve uma série de critérios direcionados, inclusive, com o ciclo de vida dos dados (coleta, armazenamento, recuperação e descarte) que emerge de um campo informacional.

Diante deste contexto e corroborando com Sant'Ana (2016), a Ciência da Informação pode e deve contribuir em ambientes que contam com a presença de acesso e uso intensivo de dados, buscando elementos que possibilitam a construção de estruturas de referências que permitam identificar características em contextos específicos. Com isto, acredita-se que a presença da Ciência da Informação em estudos referentes à qualidade de dados pode possibilitar a identificação, o reconhecimento de fatores e características que propiciem uma ampliação da harmonia entre os profissionais envolvidos no processo e na performance da utilização desses dados, favorecendo assim, o desenvolvimento da área.

Graças a natureza interdisciplinar e colaborativa da Ciência da Informação, foi possível lidar com questões de qualidade de dados pertencentes tanto a prontuários físicos como eletrônicos, isto é, conforme afirma Sant’Ana (2019), esta área lida diretamente com a camada informacional de determinado campo, sendo capaz de atender as especificidades que o acesso a dados acrescenta.

Referências

CARVALHO, R. C. Aplicação de mineração de dados em informações oriundas de prontuários de paciente. **Informação em Pauta**, Fortaleza, v. 3, p. 161-181, 2018. Disponível em: <http://www.brapci.inf.br/index.php/res/v/106569>. Acesso em: 05 jun. 2021.

DE A. OLIVEIRA, S; FAVARETTO, F. Qualidade da Informação do Prontuário Eletrônico do Paciente no Processo de Apoio à Decisão Clínica. **Journal of Health Informatics**, v. 13, n. 1, 2021. Disponível em: <http://www.jhi-sbis.saude.ws/ojs-jhi/index.php/jhi-sbis/article/view/767>. Acesso em: 05 jun. 2021.

DIXON, B. E.; MCGOWAN, J. J.; GRANNIS, S. J. Electronic laboratory data quality and the value of a health information exchange to support public health reporting processes. In: **AMIA annual symposium proceedings**. American Medical Informatics Association, 2011. p. 322. Disponível em: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3243173>. Acesso em: 05 jun. 2021.

DORILEO, E. A.; PONCIANO, M.; COSTA, T. Estruturação da evolução clínica para o prontuário eletrônico do paciente. In: **CBIS-X Congresso Brasileiro de Informática em Saúde, Florianópolis**. 2006. Disponível em: https://www.researchgate.net/profile/Joaquim-Felipe/publication/238075164_Estruturação_da_Evolução_Clinica_para_o_Prontuario_Eletronico_do_Paciente/links/548985850cf2ef3447929a03/Estruturação-da-Evolução-Clinica-para-o-Prontuario-Eletronico-do-Paciente.pdf. Acesso em: 05 jun. 2021.

FEDER, S. L. Data quality in electronic health records research: quality domains and assessment methods. **Western journal of nursing research**, v. 40, n. 5, p. 753-766, 2018. Disponível em: <https://journals.sagepub.com/doi/abs/10.1177/0193945916689084>. Acesso em: 05 jun. 2021.

FONTES-PEREIRA, A. Revisão Sistemática da Literatura: Como Escrever um Artigo Científico em 72 Horas. Rio de Janeiro: **Edição do Kindle**., 2017.

GALVÃO, M. C. B. et al. **Linguagens empregadas em prontuários do paciente frente aos processos de organização e recuperação da informação no contexto da saúde**. 2008. Disponível em: <http://repositorios.questoesemrede.uff.br/repositorios/handle/123456789/804>. Acesso em: 05 jun. 2021.

GALVÃO, M. C. B.; RICARTE, I. L. M. Prontuário do paciente. Rio de Janeiro: **Guanabara Koogan**, 2012.

HANAUER, D. A. et al. Electronic Medical Record Search Engine (EMERSE): An information retrieval tool for supporting cancer research. **JCO clinical cancer informatics**, v. 4, p. 454-463, 2020. Disponível em: <https://ascopubs.org/doi/full/10.1200/CCI.19.00134>. Acesso em: 05 jun. 2021.

MARTHA, A. S.; BARRA, P. S. C.; CAMPOS, C. J. R. Recuperação de Informações em Textos Livres de Prontuários do Paciente. In: **CBIS-IX Congresso Brasileiro de Informática em Saúde**. 2004. Disponível em: <http://telemedicina.unifesp.br/pub/sbis/cbis2004/trabalhos/arquivos/636.pdf>. Acesso em: 05 jun. 2021.

PAVÃO, A. L. B. et al. Estudo de incidência de eventos adversos hospitalares, Rio de Janeiro, Brasil: avaliação da qualidade do prontuário do paciente. **Revista Brasileira de Epidemiologia**, v. 14, p. 651-661, 2011. Disponível em: <https://www.scielo.org/article/rbepid/2011.v14n4/651-661/>. Acesso em: 05 jun. 2021.

SANT'ANA, R. C. G. Ciclo de vida dos dados: uma perspectiva a partir da ciência da informação. **Informação & Informação, Londrina**, v. 21, n. 2, p. 116-142, 2016. Disponível em: <https://www.brapi.inf.br/index.php/article/view/0000022210/6eb47c5863b44510764cd65ea0ffefc5>. Acesso em: 05 jun. 2021.

SENGAN, S. et al. Medical information retrieval systems for e-Health care records using fuzzy based machine learning model. **Microprocessors and Microsystems**, p. 103344, 2020. Disponível em: <https://www.sciencedirect.com/science/article/abs/pii/S0141933120305032>. Acesso em: 05 jun. 2021.

SOUZA, A. D.; DE ALMEIDA, M. B. Análise de dados clínicos textuais de Prontuários Eletrônicos do Paciente para integração com terminologias médicas padronizadas. Disponível em: https://www.researchgate.net/profile/Mauricio-Almeida/publication/337658059_Analise_de_dados_clinicos_textuais_de_Prontuarios_Eletronicos_do_Paciente_para_integracao_com_terminologias_medicadas_padronizadas/links/5de3e415299bf10bc33749f8/Analise-de-dados-clinicos-textuais-de-Prontuarios-Eletronicos-do-Paciente-para-integracao-com-terminologias-medicadas-padronizadas.pdf. Acesso em: 05 jun. 2021.

TIAN, Q. et al. Application of openEHR archetypes to automate data quality rules for electronic health records: a case study. **BMC medical informatics and decision making**, v. 21, n. 1, p. 1-11, 2021. Disponível em: <https://bmcmmedinformdecismak.biomedcentral.com/articles/10.1186/s12911-021-01481-2>. Acesso em: 05 jun. 2021.

Diagrama estrutural para teses e dissertações: uma proposta didática

Ricardo César Gonçalves Sant'Ana^a

1. Introdução

Existem muitas regras de formatação para textos acadêmicos, tais como as normas da ABNT⁹ e das próprias instituições mantenedoras dos programas de pós-graduação, que oferecem todo o suporte necessário para que os autores possam estruturar seus textos. No entanto, quando pensamos em textos resultantes de cursos como os de pós-graduação, temos além das necessidades vinculadas à publicação dos resultados das pesquisas as questões relacionadas aos processos de acompanhamento e avaliação.

Bancas compostas, muitas vezes por membros que nunca tiveram contato com as pesquisas a serem avaliadas (ou validadas) recebem a difícil incumbência de estabelecer um processo avaliativo que muitas vezes fica prejudicado pelo tempo disponível para tal verificação, não só para a leitura crítica do

-
- a Doutor em Ciência da Informação e Livre-docente em Sistemas de Informações Gerenciais pela Universidade Estadual Paulista – UNESP. Professor Associado na Universidade Estadual Paulista – UNESP. E-mail: ricardo.santana@unesp.br. ORCID: <https://orcid.org/0000-0003-1387-4519>.
- 9 Associação Brasileira de Normas Técnicas – que apresentam normas como: NBR 14724:2011: Trabalhos acadêmicos; NBR 6023:2002: Referência; NBR 6024:2012: Numeração progressiva das seções de um documento; NBR 6027:2012: Sumário; NBR 6028:2003: Resumo; NBR 12225:2004: Lombada; NBR 10520:2002: Citações, entre outras.

texto como também para compreender elementos como escopo, justificativa, coerência, aplicação e até mesmo validade das metodologias e citações utilizadas.

Textos resultantes de cursos de pós-graduação, em especial os definidos como acadêmicos, sejam eles formatados como artigo ou como relatórios finais, apresentam sempre elementos comuns tais como: problema de pesquisa, objetivo geral, objetivos específicos, metodologia, resultados e considerações finais. Entende-se que existem ainda outros elementos e também variações em como estes elementos são tratados e apresentados no texto, mas considerando estes elementos fundamentais já se torna possível estabelecer uma perspectiva em que a estrutura explicitamente apresentada destes elementos permitem uma espécie de visualização gráfica do todo, uma dissecação dos elementos e suas relações. Para os que estão escrevendo o texto passa a ser uma oportunidade de identificar eventuais problemas de coesão e coerência da pesquisa, que muitas vezes, quando não aplicadas técnicas como as aqui apresentadas podem levar a interpretação de que o problema está nas respostas quando em muitos dos casos pode estar na própria pergunta.

Este relato propõe o compartilhamento de práticas adotadas em orientações realizadas pelo autor bem como requisitos derivados de participações em bancas. Não se trata de uma proposta de mais formalizações para textos e sim sugestões de técnicas que podem ser utilizadas para facilitar não só a construção do texto por parte dos pesquisadores como, ainda,

colaborar com o processo de acompanhamento por parte dos orientadores e, principalmente, para compreensão de leitores que tenham a tarefa de validar e avaliar tais textos, não sendo descartada ainda a possibilidade de uso de tais recursos pelos futuros leitores de tais textos, permitindo uma leitura estruturada de tais textos, que por sua natureza, apresentam uma estrutura intrínseca.

Destaca-se que neste texto, a adoção de técnicas de explicitação de elementos estruturais e suas relações com os textos não tem motivação estética, mas essencialmente pragmática, visando a ampliação da visibilidade das ‘entranhas’ do texto.

1.1 Escopo e delimitação

Não se pretende com este texto apresentar elementos para o auxílio no desenvolvimento do projeto e nem mesmo ao desenvolvimento de pesquisas, mas, antes, a estruturação de elementos adicionais ao texto resultante, no entanto, reconhece-se que tais elementos podem contribuir para a proposta e mesmo para a fase de desenvolvimento se estes elementos passarem a ser construídos previamente, contribuindo assim, inclusive, com a manutenção da coerência entre proposta, elaboração e resultados.

Outra questão relevante na definição do escopo deste texto reside na heterogeneidade das áreas da ciência e seu impacto nos textos, o que impõe limitações ao detalhamento que se aplica a proposta aqui apresentada, exigindo que os exemplos sejam

menos específicos, sem, no entanto, perder a possibilidade de interpretação e adaptações necessárias a cada área.

2. Elementos estruturais escolhidos

Perceba que a escolha do termo “estrutural” em detrimento a “estruturante” já aponta para o caminho de que os elementos escolhidos para esta proposta não são novos ou não serão agentes promotores de novas formas de organização dos textos mas, antes, elementos já existentes e que simplesmente passam a ser explicitados em uma formatação ampliada e que facilite não só sua visualização como também suas relações.

Parte-se do pressuposto de que mesmo não estando definidos em seções próprias no texto estarão presentes de alguma forma na construção do texto, cabendo ainda maior importância nestes casos a sua explicitação.

Assim, inicialmente apresenta-se os seguintes elementos descritos no Quadro 1. Perceba que na coluna Cardinalidade do Quadro 1 estão identificadas as cardinalidades dos elementos do diagrama de tal forma a apontar se tais elementos podem conter mais de uma instância na pesquisa, ou seja, podem ser quebrados em mais de uma ocorrência. No entanto, é importante destacar que mesmo em elementos como motivação, por exemplo, que aparece como elemento único, sua própria vinculação com os elementos de relevância já deixam claro que podem haver múltiplos argumentos, no entanto, no diagrama recomenda-se que se tenha em mente a importância de destacar a motivação de maior peso, e se for o caso, na própria descrição do elemento

(dentro do retângulo que o representa) podem ser descritos mais de um argumento. Vale lembrar que muitas vezes o excesso de informações pode prejudicar a percepção de quem interpreta, diluindo a visibilidade do que realmente busca-se representar. Essa mesma questão de cardinalidade se aplica a outros elementos identificados no Quadro 1 como único, tais como objetivo geral, justificativa, problema de pesquisa entre outros.

Quadro 1 – Elementos do diagrama estrutural da pesquisa

Elemento	Vínculos	Cardinalidade
Título	(Todo o diagrama)	Único
Motivação	Relevância Social, Relevância Científica, Justificativa, Problema de Pesquisa	Único
Relevância Social	Motivação	Único
Relevância Científica	Motivação	Único
Justificativa	Motivação	Único
Delimitações	Problema de Pesquisa	Único
Problema de Pesquisa (ou Questão de Pesquisa)	Motivação, Delimitações	Único
Hipótese	Problema de Pesquisa	Único
Tese	Hipótese	Único
Objetivo Geral	Objetivos Específicos	Único
Objetivo Específico N	Objetivo Geral, Metodologia, Resultado	Múltiplo
Metodologia N	Objetivo Específico	Múltiplo
Resultado N	Objetivo Específico, Conceito	Múltiplo
Conceito N	Resultado, Referência	Múltiplo
Referência N	Conceito	Múltiplo

Fonte: Autor.

Com relação ao conjunto das vinculações entre os elementos representadas no diagrama, cabe lembrar que em essência todos

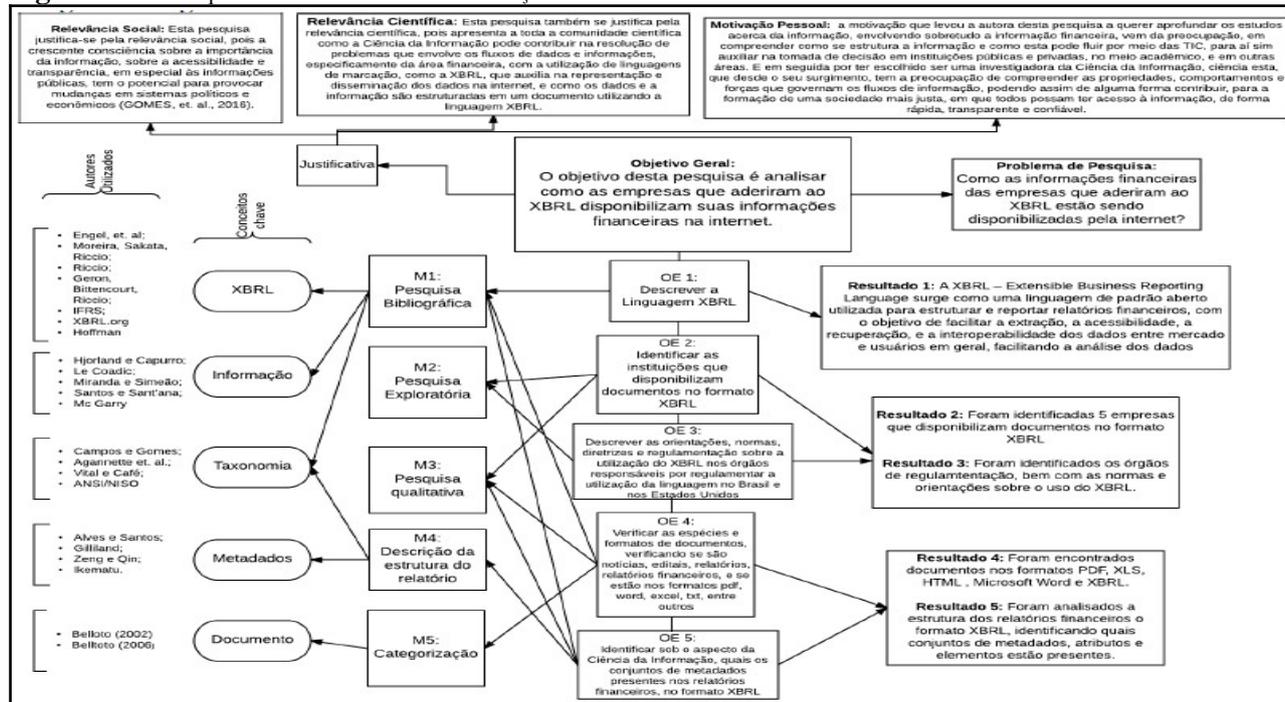
os elementos encontram-se de alguma forma relacionados entre si, no entanto, no diagrama as vinculações mais relevantes para a interpretação são destacadas, de tal forma a facilitar a leitura.

2.1 Exemplos de Implementação

Esta proposta de diagrama estrutural já vem sendo utilizada desde o meu início como orientador e, mesmo sem a formalização, vem sendo apresentada de forma muito similar ao que está descrito neste texto, no entanto, conforme descrito na introdução, a utilização em cada projeto de pesquisa precisa ser ajustado para as especificidades que se fossem atendidas pelo modelo tornaria sua implementação ‘cara’ demais para o pesquisador, assim, como proposto, a sugestão que o diagrama tenha o papel de direcionar a representação estrutural da tese ou dissertação.

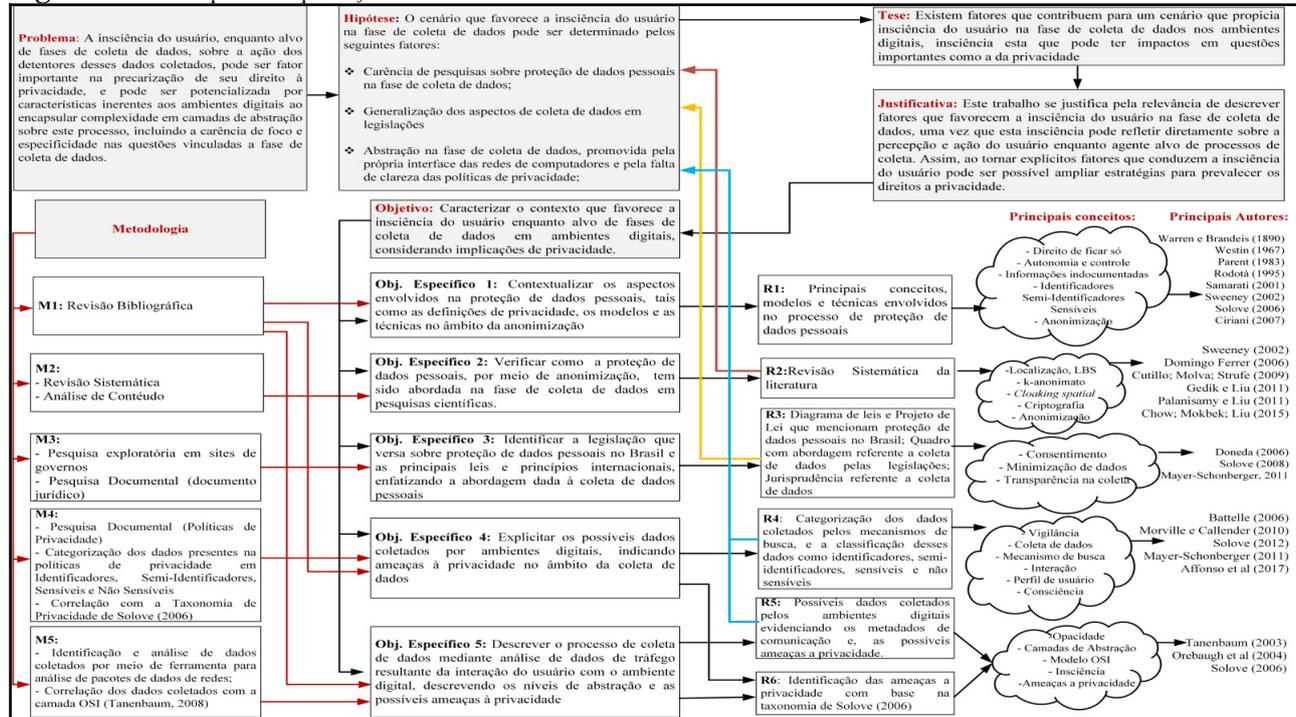
Neste sentido, são apresentados nas Figuras 1 a 4 exemplos de aplicação do diagrama estrutural. Importante perceber nos exemplos de aplicação a potencial de adaptação que o esquema proposto apresenta, mesmo considerando que se trata de exemplos de uma mesma área e, mais ainda, do mesmo programa, mesma linha e com o mesmo orientador, e da mesma proposta. Vale ainda ressaltar que mesmo com as variações aplicadas pelos então alunos percebe-se que a essência do diagrama permanece.

Figura 1 – Exemplo de uso em uma dissertação de mestrado



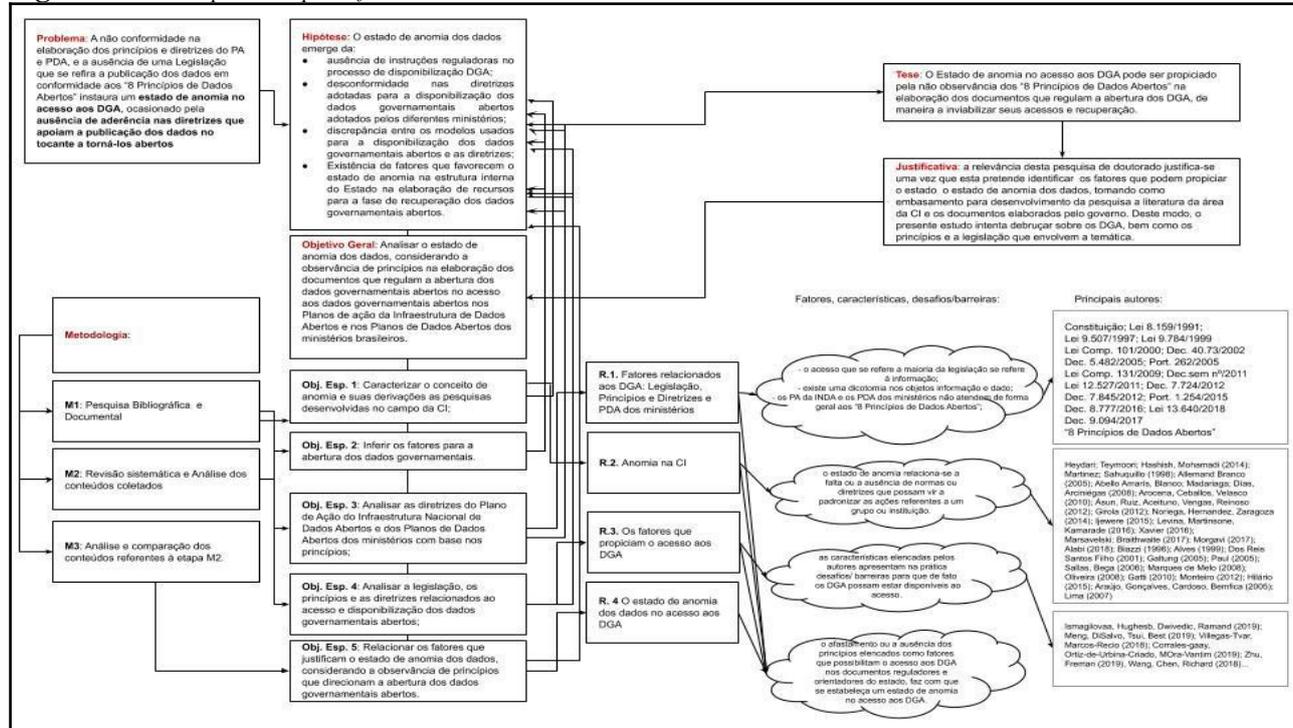
Fonte: Nagai (2017).

Figura 2 – Exemplo de aplicação em uma tese de doutorado



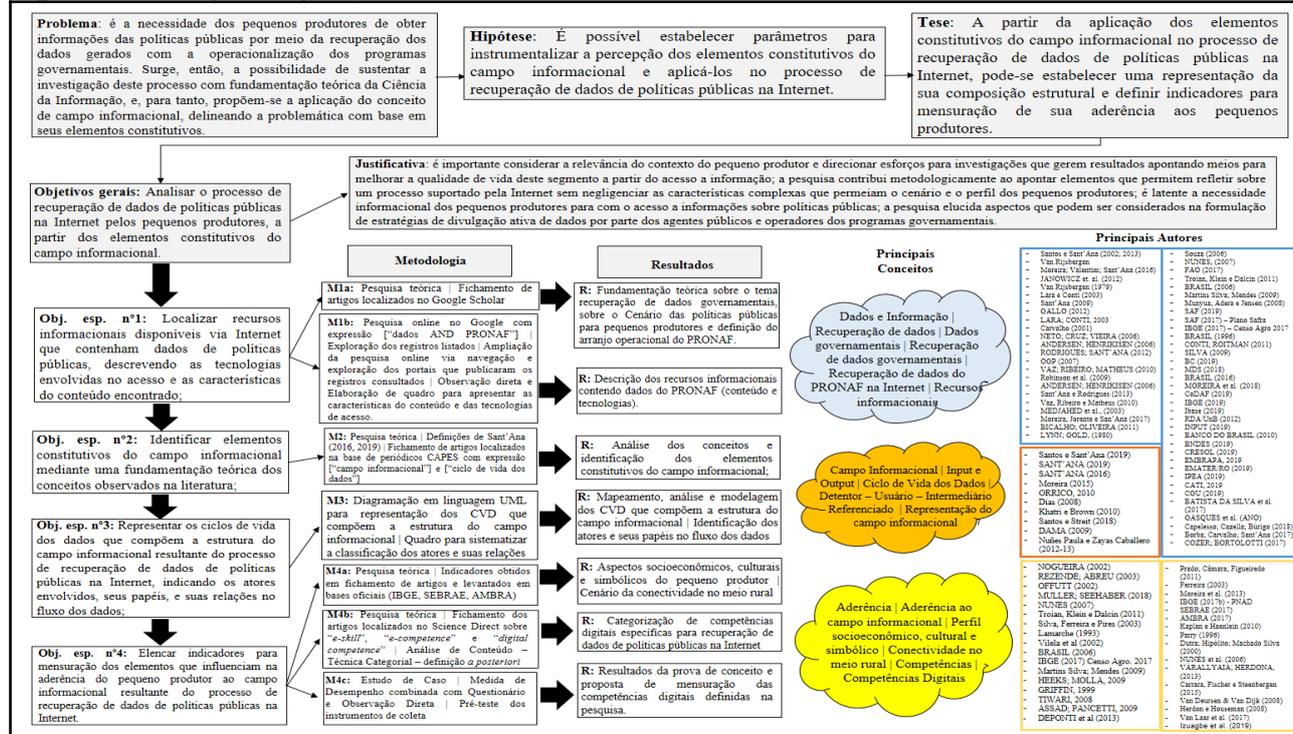
Fonte: Affonso (2018).

Figura 3 – Exemplo de aplicação em uma tese de doutorado



Fonte: Aleixo (2020).

Figura 4 – Exemplo de aplicação em uma tese de doutorado



Fonte: Moreira (2020).

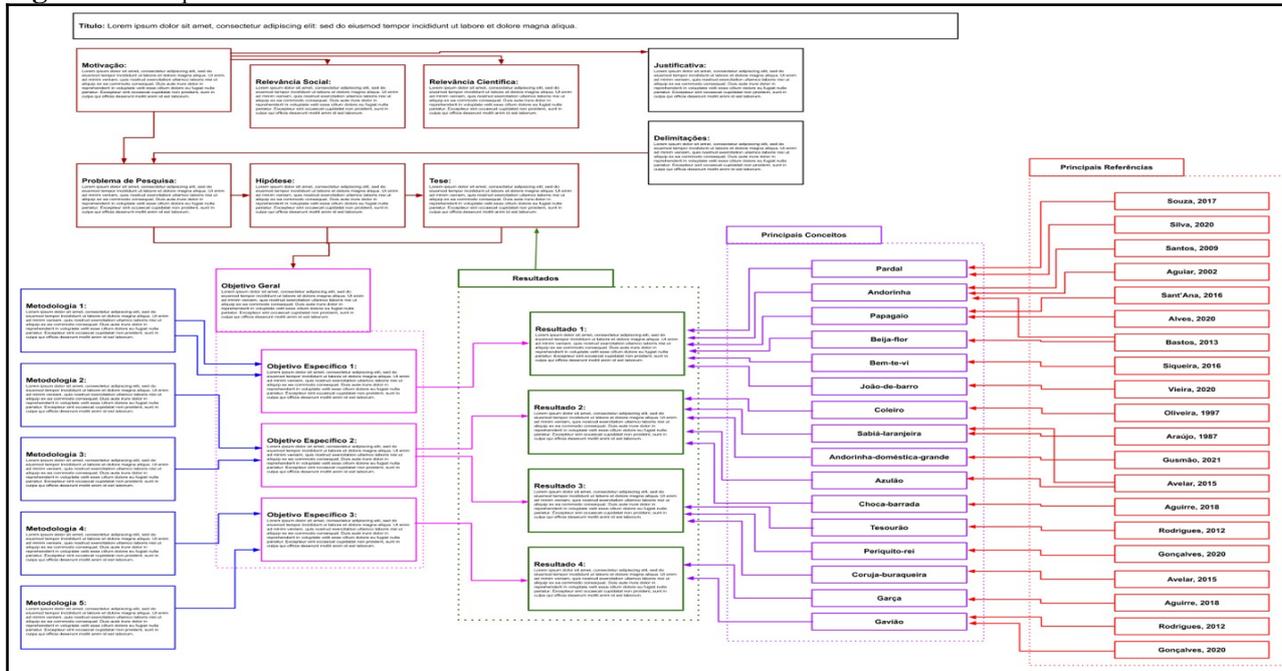
3. Diagrama Estrutural para Teses e Dissertações

Como forma de facilitar a compreensão e colaborar com futuras implementações do diagrama, são apresentados na figura 5 os elementos propostos para o diagrama, com dados fictícios e número de ocorrências dos elementos múltiplos composto de forma a exemplificar didaticamente a presença de mais de um elemento no diagrama. Na descrição dos elementos foi utilizado o texto padrão em latim (*lorem ipsum*) para preencher os espaços de texto, tal como utilizado em produções gráficas, para testar e ajustar aspectos visuais antes do seu uso em situações reais.

Observando a Figura 5 é possível perceber a distribuição dos elementos estruturais do texto, suas vinculações e organizados de tal forma que a interpretação da figura leve a construção de uma percepção geral da pesquisa.

Elementos como os objetivos específicos são apresentados em número de três, mas devem ser adequados à realidade de cada situação, aumentando ou diminuindo o número de elementos. O mesmo vale para os elementos identificados como metodologias, resultados, conceitos e referências, que na proposta aparecem em número de 5, 4, 16 e 20 respectivamente apenas como exemplo, cabendo as devidas adequações às situações específicas de cada texto.

Figura 5 – Proposta de desenho de estrutura do texto



Fonte: Autor.

Com relação ao tamanho, o desenho foi projetado pensando-se em uma folha A2 e, portanto, com dimensões de 59 centímetros de largura por 42 centímetros de altura e orientação horizontal (paisagem). As relações entre os elementos são representadas por setas e também devem ser refeitas conforme a estrutura estiver sendo construída. Ainda sobre as relações, quando se pensa na vinculação dos conceitos aos resultados, e entre as referências e os conceitos, vale lembrar que elas podem ser um para muitos, muitos para um ou ainda muitos para muitos (cardinalidade), o que confere ao modelo a flexibilidade necessária para quaisquer situações. Recomenda-se, para efeito de simplificação, que a relação entre referências e conceitos seja feita com base no principal autor adotado na definição de cada conceito, mas, novamente, trata-se somente de uma recomendação.

4. Considerações

Espera-se que esta proposta auxilie na construção de formas de leitura (ou de avaliação) que propiciem a superação das limitações das palavras e seus significados e permitindo que sejam estudados as teorias e os problemas que elas colocam e buscam responder.

Espera-se que a compreensão sobre o texto e, por consequência, sobre a pesquisa realizada seja auxiliada pelo diagrama estrutural já que a explicitação dos elementos estruturais pode contribuir para a ‘navegação’ do leitor pelo texto.

Destaca-se ainda a identificação de elementos que não aparecem diretamente nesta proposta mas que podem emergir mais facilmente no processo de compreensão auxiliado pelo diagrama, evidenciando elementos como a concepção filosófica adotada (Pós-positivista, Construção social, Reivindicatória / participativa, Pragmática) e suas relações com as estratégias de investigação adotadas (qualitativas, quantitativas, mistas).

Referências

AFFONSO, Elaine Parra, A INSCIÊNCIA DO USUÁRIO NA FASE DE COLETA DE DADOS: PRIVACIDADE EM FOCO, 325 f, **Tese** (Doutorado), Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, 2018.

ALEIXO, Diana Vilas Boas Souto, O ESTADO DE ANOMIA DOS DADOS NO ACESSO AOS DADOS GOVERNAMENTAIS ABERTOS NO BRASIL, 260 f, **Tese** (Doutorado), Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, 2020.

ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. ABNT. **NBR 6023**: informação e documentação: referências: elaboração. Rio de Janeiro, 2002a. 24 p.

_____. **NBR 10520**: informação e documentação: citações em documentos: apresentação. Rio de Janeiro, 2002b. 7 p.

_____. **NBR 6028**: informação e documentação: resumo: apresentação. Rio de Janeiro, 2003. 2 p.

_____. **NBR 12225**: informação e documentação: lombada: apresentação. Rio de Janeiro, 2004. 3 p.

_____. **NBR 14724**: informação e documentação: trabalhos acadêmicos: apresentação. Rio de Janeiro, 2011. 11 p.

_____. **NBR 6024**: informação e documentação: numeração progressiva das seções de um documento. Rio de Janeiro, 2012a. 4 p.

_____. **NBR 6027**: informação e documentação: sumário: apresentação. Rio de Janeiro, 2012b. 3 p.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. **Normas de apresentação tabular**. 3. ed. Rio de Janeiro, 1993. Disponível em: <http://biblioteca.ibge.gov.br/visualizacao/livros/liv23907.pdf>. Acesso em: 02 dez. 2021.

MOREIRA, Fabio Mosso, RECUPERAÇÃO DE DADOS DE POLÍTICAS PÚBLICAS NA INTERNET PELO PEQUENO PRODUTOR: ELEMENTOS CONSTITUTIVOS DO CAMPO INFORMACIONAL, 167 f, **Tese** (Doutorado), Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, 2020.

NAGAI, Cristina Toyoko Hashimoto. DIVULGAÇÃO DE DADOS FINANCEIROS NA INTERNET: UM ESTUDO SOBRE O EXTENSIBLE BUSINESS REPORTING LANGUAGE XBRL. 93 f. **Dissertação** (Mestrado) - Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília, 2017.

Organizadores do livro



Fábio Mosso Moreira

Doutor e Mestre pelo Programa de Pós-Graduação em Ciência da Informação (UNESP/Marília). Bacharel Administração de Empresas (UNESP/Tupã). Atualmente atua como professor no Departamento de Ciência da Informação da UNESP/Marília. Membro do Grupo de Pesquisa Tecnologia de Acesso a Dados (GPTAD) e do Projeto Competências Digitais para Agricultura Familiar (CoDAF). Editor de Conteúdo da Revista Eletrônica Competências Digitais para Agricultura Familiar (RECoDAF).



Leonardo Castro Botega

Doutor em Ciência da Computação pela Universidade Federal de São Carlos - UFSCar com Pós-doutorado pela Universidade de São Paulo – USP. Membro Permanente do Programa de Pós-graduação em Ciência da Informação da UNESP-Marília. Membro Colaborador do Programa de Pós-graduação em Ciência da Computação da UNESP-Bauru/Prudente Pesquisador Colaborador do Instituto de Computação da UNICAMP. Data Product Manager na empresa PISMO. Líder do Grupo de Interação Humano-Computador (GIHC) - UNESP. Revisor de periódicos nas áreas de fusão de dados, sistemas críticos de tomada de decisão, web-semântica e sistemas de informação. Tem experiência acadêmica e profissional nos seguintes temas: Fusão de Dados e Informações, Mineração de Dados, Qualidade de Dados e Informações, Websemântica, Gestão de Dados Críticos e Sistemas Críticos de

Tomada de Decisão. Obteve diversas publicações em eventos e periódicos nacionais e internacionais, além de orientar diversos trabalhos de graduação, mestrado e doutorado com bolsa CAPES, CNPq e FAPESP.



José Eduardo Santarem Segundo

Livre Docente em Informação e Tecnologia pela Universidade de São Paulo (USP), 2020. Pós-Doutorado pela Faculdade de Engenharia da Computação da Western University/Canadá, 2018. Doutor e Mestre em Ciência da Informação pela Universidade Estadual Paulista UNESP-Marília/SP; Professor Doutor no Departamento de Educação, Informação e Comunicação, da Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto, da Universidade de São Paulo (USP); Docente do Programa de Pós-Graduação em Ciência da Informação da UNESP/Marília na linha de Informação e Tecnologia. Bolsista de Produtividade em Pesquisa PQ-2 do CNPq. Coordenador do GT8 – Informação e Tecnologia, da Associação Nacional de Pesquisa e Pós-Graduação em Ciência da Informação (ANCIB). Atua na linha de pesquisa – Ambientes Digitais e Tecnologias Aplicadas a Informação e Comunicação, com ênfase em Web Semântica, Linked Data, Big Data, Aprendizado de Máquina (Machine Learning), Dados Abertos e Acervos Digitais. Líder do NEWSDA – Núcleo de Estudos em Web Semântica e Dados Abertos. Recebeu o Prêmio de Melhor Tese pela Associação Nacional de Pesquisa e Pós-Graduação em Ciência da Informação (ANCIB) - Ano 2011. Recebeu também Menção Honrosa no Prêmio Capes de Teses – 2011 (Ciências Sociais Aplicadas).



Ricardo César Gonçalves Sant'Ana

Professor Associado da UNESP/Tupã, em regime de dedicação exclusiva. Professor do Programa de Pós-Graduação em Ciência da Informação da UNESP/Marília. Graduado em Matemática e Pedagogia, Mestrado em Ciência da Informação (2002), Doutorado em Ciência da Informação (2008) e Livre-Docente em Sistemas de Informações Gerenciais pela UNESP (2017). Possui especializações em Orientação a Objetos (1996) e Gestão de Sistemas de Informação (1998). Parecerista ad hoc de periódicos e de agências de fomento. Líder do Grupo de Pesquisa – Tecnologias de Acesso a Dados (GPTAD) e membro do Grupo de Pesquisa – Novas Tecnologias em Informação GPNTI. Membro da ANCIB. Tem experiência na área de Ciência da Computação, atualmente realiza pesquisas com foco em: ciência da informação e tecnologia da informação, investigando temas ligados ao Ciclo de Vida dos Dados, Transparência e ao Fluxo Informacional em Cadeias Produtivas. Foi Presidente da primeira composição da Comissão de Acompanhamento e Avaliação dos Cursos de Graduação – CAACG da UNESP, entre 2018 e 2020. Atuou como professor na Faccat Faculdade de Ciências Contábeis e Administração de Tupã, onde coordenou curso de Administração com Habilitação em Análise de Sistemas por dez anos e o curso de Licenciatura em Computação. Atuou no setor privado como consultor, integrador e pesquisador de novas tecnologias informacionais de 1988 a 2011.

Comissão Organizadora do IV DTI

 Fáblio Mosso MOREIRA	UNESP – Universidade Estadual Paulista
 Leonardo Castro BOTEGA	UNESP – Universidade Estadual Paulista
 José Eduardo SANTAREM SEGUNDO	UNESP – Universidade Estadual Paulista
 Ricardo C. G. SANT ^ª ANA	UNESP – Universidade Estadual Paulista

Comissão Científica do IV DTI

 Cristian BERRIO-ZAPATA	UFPA – Universidade Federal do Pará
 Daiane Marcela PICCOLO	FATEC – Faculdade de Tecnologia de São Paulo
 Diana Vilas Boas Souto ALEIXO	UEL – Universidade Estadual de Londrina
 Elaine Parra AFFONSO	FATEC – Faculdade de Tecnologia de São Paulo
 Elizabete C. de S. A. MONTEIRO	UNESP – Universidade Estadual Paulista
 Fáblio Mosso MOREIRA	UNESP – Universidade Estadual Paulista
 Fernando de Assis RODRIGUES	UFPA – Universidade Federal do Pará
 Jacquelin T. CAMPEROS-REYES	UNESP – Universidade Estadual Paulista
 José Eduardo SANTAREM SEGUNDO	UNESP – Universidade Estadual Paulista
 Leonardo Castro BOTEGA	UNESP – Universidade Estadual Paulista
 Ricardo C. G. SANT ^ª ANA	UNESP – Universidade Estadual Paulista

Transmissão do IV DTI



Legenda: Gravação da transmissão do IV DTI, realizado no dia 16/07/2021 e transmitido via Canal do Youtube do GPTAD. Playlist Disponível em: https://www.youtube.com/watch?v=RcHRnAfUzyQ&list=PLYxyY03EkjN3mKFJta0QzIH0Wwr2AHK_U

Apoio

Organizacional

Grupo de Pesquisa Tecnologia de Acesso a Dados –
GPTAD/UNESP

Programa de Pós-Graduação em Ciência da Informação –
PPGCI/UNESP

Institucional

Faculdade de Filosofia e Ciências-FFC, UNESP-Marília

Faculdade de Ciências e Engenharia-FCE, UNESP-Tupã

