



Encontros Bibli: revista eletrônica de biblioteconomia e
ciência da informação

ISSN: 1518-2924

Universidade Federal de Santa Catarina

Rodrigues, Débora Gomes Araújo; Dos Anjos, Renata Lemos; Dias, Guilherme Ataíde
TIPOLOGIA DOS DADOS DAS TESES BRASILEIRAS DA CIÊNCIA
DA INFORMAÇÃO: INVESTIGANDO SUA NATUREZA E ORIGEM

Encontros Bibli: revista eletrônica de biblioteconomia
e ciência da informação, vol. 27, 2022, pp. 1-25
Universidade Federal de Santa Catarina

DOI: <https://doi.org/10.5007/1518-2924.2022.e84729>

Disponível em: <https://www.redalyc.org/articulo.oa?id=14775278007>

- Como citar este artigo
- Número completo
- Mais informações do artigo
- Site da revista em [redalyc.org](https://www.redalyc.org)

UAEM [redalyc.org](https://www.redalyc.org)

Sistema de Informação Científica Redalyc
Rede de Revistas Científicas da América Latina e do Caribe, Espanha e Portugal
Sem fins lucrativos acadêmica projeto, desenvolvido no âmbito da iniciativa
acesso aberto

TIPOLOGIA DOS DADOS DAS TESES BRASILEIRAS DA CIÊNCIA DA INFORMAÇÃO: INVESTIGANDO SUA NATUREZA E ORIGEM

Data Typology from Brazilian Thesis in the Information Science: investigating its nature and origin

Débora Gomes Araújo Rodrigues

Doutora em Ciência da Informação
Universidade Federal da Paraíba, Programa de Pós-Graduação em Ciência da Informação, João Pessoa, Brasil
debora.g.de.araujo@gmail.com

<https://orcid.org/0000-0003-2283-0833> 

Renata Lemos dos Anjos

Mestre em Ciência da Informação
Universidade Federal da Paraíba, Programa de Pós-Graduação em Ciência da Informação, João Pessoa, Brasil
renata.anjos@academico.ufpb.br

<https://orcid.org/0000-0002-4599-3494> 

Guilherme Ataíde Dias

Professor titular
Universidade Federal da Paraíba, Departamento de Ciência da Informação, João Pessoa, Brasil
guilhermeataide@gmail.com

<https://orcid.org/0000-0001-6576-0017> 

A lista completa com informações dos autores está no final do artigo 

RESUMO

Objetivo: Classificar a tipologia dos dados científicos obtidos nas teses brasileiras, na área da Ciência da Informação, de acordo com a sua natureza e origem, no interstício de 2012 a 2020.

Método: Utilizou-se uma abordagem quali-quantitativa; caracterizou-se quanto aos meios como um estudo bibliográfico e quanto aos fins como uma pesquisa descritiva. A técnica de análise de conteúdo foi adotada para analisar os dados com o suporte do *software* QDAMiner.

Resultado: Verificou-se que os dados da área são diversificados em relação à natureza por serem dados textuais, visualizações, multimídia, números e *software*, além disso, esses dados também se diferenciam no que diz respeito à origem, pois foram identificados dados de registro, observacional, computacional e experimental. Sendo os dados de registros e observacionais predominantes.

Conclusões: Diante da realidade de que o presente estudo evidenciou a diversidade e a divisão desses dados no tocante a origem, especialmente pela presença dos dados observacionais, que são os mais difíceis de replicar, espera-se que a área da CI se conscientize sobre a importância de tratar adequadamente os dados de forma que possam ter condições de serem reutilizados em novas pesquisas no longo prazo.

PALAVRAS-CHAVE: Dados Científicos. Ciência da Informação. Natureza dos Dados. Origem dos Dados. Procedimentos Metodológicos.

ABSTRACT

Objective: To classify the scientific data typology obtained in the Brazilian Information Science theses, according to their nature and origin during the period between the years of 2012 and 2020.

Methods: A quali-quantitative approach was used; characterized as to the means as a bibliographic study and as for the purposes as descriptive research. The content analysis technique was adopted to analyze the data with the support of the QDAMiner software.

Results: The research shown that the data in the IC area are diversified in relation to its nature, as they are composed by textual data, visualizations, multimedia, numbers, and software. These data also differ about its origin, record, observational, computational, and experimental data were identified. Record and observational data were predominant.

Conclusions: Given the reality presented by the study, which demonstrated the data diversity and division regarding its origin, especially due to the observational data presence, which are the most difficult to replicate, it is

expected that the IC area will become aware of the importance of properly treating the data so that they can be reused in new long-term research.

KEYWORDS: Scientific Data. Information Science. Data Nature. Data Source. Methodological Procedures.

1 INTRODUÇÃO

Nos dias atuais, é crescente o debate a respeito da importância dos dados presentes nos trabalhos acadêmicos. A ideia é que os dados científicos sejam gerenciados de forma adequada para contribuir com a geração de conhecimento no âmbito da comunidade de pesquisadores. Nesse sentido, Curty e Aventurier (2017) revelam que esses dados correspondem a matérias-primas essenciais para a ecologia da ciência e são indispensáveis para a criação de novos ciclos do conhecimento científico.

A percepção de valor dos dados vem sendo obtida ao longo do tempo, embora efetivamente, se configure como algo novo. Entre as décadas de 1950 e 1960, a compreensão sobre os dados estava restrita a ser um resultado de uma pesquisa, passando posteriormente, a ser vistos como subprodutos. A partir do período de 1990 a 2010, a reutilização de dados ganhou importância para a promoção do desenvolvimento de novos estudos. Atualmente, os dados são considerados como elemento primordial para o avanço da ciência (SILVA, 2019).

A realidade é que essa temática passou a interessar os pesquisadores, as instituições acadêmicas e as agências de fomento à pesquisa, pelo entendimento de que esses dados, caso sejam bem preservados e gerenciados, compreendem uma fonte primordial de recursos informacionais que podem ser compartilhados e reutilizados como insumo para novas pesquisas (SALES; SAYÃO, 2019).

De acordo com Van den Eynden e Corti (2017), a publicação e o compartilhamento de dados científicos têm se destacado mundialmente, como resultado dos avanços tecnológicos, das necessidades de pesquisas e exigências de financiadores, editores e governos em relação à abertura, transparência e investimento eficiente em pesquisa. Nas ciências sociais, temos o exemplo do repositório digital UK Data Archive¹, que ao focar na curadoria, preservação e publicação dos dados tem procurado cada vez mais aprimorar as práticas de publicação de dados e aperfeiçoar as habilidades dos pesquisadores (VAN DEN EYNDEN; CORTI, 2017).

¹ Disponível em: <https://www.data-archive.ac.uk/>

Apesar dos dados impactarem o processo de geração de informação e, conseqüentemente do conhecimento, diversos pesquisadores obtêm os seus dados apenas para trazer respostas pontuais às suas próprias problemáticas de pesquisa e, em seguida, promovem o descarte de forma "consciente" ou "inconsciente", provocando o desperdício, pois impossibilita o reuso dos dados ao deixar de atender a novos estudos do próprio criador do conjunto de dados, ou de outros pesquisadores que não participaram da pesquisa original (RODRIGUES, 2021).

É importante que o pesquisador tenha presente que o processo de geração dos dados científicos geralmente requer do investimento de recursos como tempo e capital financeiro (BORGMAN, 2015). No entanto, muitos dados estão perdendo a potencialidade do seu valor por se encontrarem armazenados em computadores pessoais ou em outros dispositivos de armazenamento que não lhes permitem se tornarem bens públicos.

No que se refere aos dados das pesquisas sociais, Borgman (2015) esclarece que esses podem ser recursos de acesso comum, através de arquivos abertos para alguns pesquisadores, bem como, bens de equipe, quando limitados aos membros de grupos de pesquisa. Também podem ser particulares, quando partilhados apenas entre os pesquisadores individuais ou têm a possibilidade de ser bens públicos, ao serem compartilhados de forma aberta, sem restrição de uso. Neste contexto, a aquisição de dados científicos pode ser um processo oneroso, tanto no sentido da necessidade de financiamento por agências de fomento à pesquisa, quanto no dispêndio de pessoal qualificado, tempo e capital. No entanto, a reutilização pode economizar recursos como dinheiro, tempo e experiência para expandir a diversidade de insumos disponíveis voltados à pesquisa. Além disso, os pesquisadores normalmente, fazem uso de recursos públicos para a obtenção dos seus dados científicos, o que seria outra razão para disponibilizá-los publicamente (SILVA, 2019).

A falta de acesso aos dados traz impactos à ciência, podendo provocar o retrabalho e o atraso de novas descobertas, deixando de atender prontamente às demandas emergentes de pesquisas. Observa-se que, o "engavetamento" de dados traz prejuízos incalculáveis quando o seu não acesso e disponibilização entravam a possibilidade de respostas científicas pontuais (RODRIGUES, 2021).

Sant'Ana (2016) afirma que a Ciência da Informação (CI) pode e deve contribuir para que este cenário de acesso e uso intensivo de dados desenvolva-se da melhor

maneira por meio da identificação e estudo dos fatores e características que maximizam a otimização do acesso e uso dos dados.

Araújo (2018) mostra que a CI atualmente investiga temáticas como: a Produção e Comunicação Científica com ênfase no Acesso Livre ao Conhecimento e a Curadoria Digital; discute a Representação e Organização da Informação, no tocante, por exemplo, à exploração de assuntos como Ontologias, *Web Semântica* e Recuperação da informação; inclui ainda os estudos sobre os Sujeitos da Informação, dentre outros temas relacionados de forma direta ou indireta com a gestão dos dados científicos.

Sales e Sayão (2019) alertam que para uma boa curadoria e gestão efetiva dos dados é necessário levar em consideração as peculiaridades do domínio científico, assim como as características dos dados gerados nesse domínio. Para os autores, o estudo sobre a natureza dos dados evidencia a diversidade e heterogeneidade de tipos de dados que podem ser originados no ambiente de pesquisa no tocante a formatos, mídias, suportes, expressões, arcabouços tecnológicos, etc. Já em relação à divisão dos dados na categoria origem, provavelmente, é o princípio de divisão mais significativa para o tratamento dos dados, por possibilitar o estabelecimento do nível e a abrangência da gestão e curadoria que necessita ser aplicado em cada categoria.

Diante dessas considerações iniciais, o objetivo deste trabalho foi classificar a tipologia dos dados científicos obtidos nas teses brasileiras, na área da CI, de acordo com a sua natureza e origem, no interstício compreendido entre os anos de 2012 a 2020.

2 A NATUREZA DIVERSIFICADA DOS DADOS CIENTÍFICOS

Na busca por apresentar um conceito amplo para atender as necessidades dos dados oriundos da pesquisa, Sales e Sayão (2019, p. 36) propuseram a seguinte definição: “[...] é todo e qualquer tipo de registro coletado, observado, gerado ou usado pela pesquisa científica, tratado e aceito como necessário para validar os resultados da pesquisa pela comunidade científica”. Para Curty e Aventurier (2017), os dados científicos são fontes de insumos para um processo iterativo no ciclo de vida da pesquisa, possibilitando o avanço da ciência e o desenvolvimento tecnológico.

De acordo com Silva (2019), dados científicos são definidos como informações registradas ou produzidas através de alguma forma ou meio, no decorrer de uma pesquisa, além do que, também são evidências necessárias ao pesquisador para

validar suas conclusões depois de ter realizado um estudo. O autor esclarece que podem ser atribuídas diversas definições aos dados científicos, pelo fato deles envolverem todos os domínios do conhecimento. Nesse sentido, Assante *et al.* (2016) coadunam com a ideia ao enfatizarem que não tem uma definição partilhada sobre esses dados. O termo normalmente se refere aos diversos materiais produzidos nas pesquisas.

No contexto da investigação científica, Schöpfel *et al.* (2016) mostram 2 (duas) funções dos dados: dados como materiais (insumos) em um primeiro estágio do processo de investigação – os dados são coletados e analisados de diversas fontes e de distintas formas e formatos; dados como resultado (saída) – referem-se aos dados produzidos no decorrer do processo de pesquisa e no final, juntamente com publicações, como resultados de pesquisa.

O relatório do *National Science Board* (NSB) de 2005, intitulado *Long-Lived Digital Data Collections*, ao se referir ao termo “dados” mostra que ele pode se diferenciar em relação a sua natureza, os dados podem ser textos, números, imagens, vídeos, áudios, *software*, algoritmos, equações, animações, modelos e simulações. A *Organisation for Economic Co-operation and Development* – Organização para Cooperação e Desenvolvimento Econômico (OECD) (2007, p. 13, tradução nossa) apoia essa classificação ao conceituar dados de pesquisa como: “[...] registros factuais (pontuações numéricas, registros textuais, imagens e sons) usados como fontes primárias de pesquisa científica e que são comumente aceitos na comunidade científica como necessários para validar resultados da pesquisa.”²

De acordo com um estudo empírico realizado por Sales e Sayão (2019), na área das ciências nucleares, os autores agruparam os dados no tocante a natureza em algumas categorias, tais como: números, multimídias, *software*, visualizações, textuais, artefatos, processos e outros, conforme ilustrado no Quadro 1.

Quadro 1: Tipo de dados encontrados na área das Ciências Nucleares

Natureza dos dados	Características
Números	Medidas, resultados de levantamentos, resultados de experimentos, fórmulas, equações e algoritmos.
Multimídias	Imagens, vídeo, fotografias, áudios, animações e filmes.
Software	Bases de dados, simulações e códigos nucleares.

² [...] *factual records (numerical scores, textual records, images and sounds) used as primary sources for scientific research, and that are commonly accepted in the scientific community as necessary to validate research findings.*

Visualizações	Tabelas, diagramas, modelos em 3D, modelos reduzidos, desenhos.
Textuais	Metadados, questionários, entrevistas anotações, normas, padrões, certificados, caderno de laboratório, transcrição, correspondência, diário de campo.
Artefatos	Espécimes, amostras e maquete.
Processos	Procedimentos operacionais padronizados, workflows, protocolos, testes.
Outros	Phanton ou manequim.

Fonte: Adaptado de Sales e Sayão (2019).

Desta forma, esses autores foram além das definições apresentadas pela NSB (2005) e OECD (2007) ao identificarem no estudo empírico novos tipos de dados, fato que remete à diversidade. A partir de então, os dados foram agrupados de acordo com as similaridades correspondentes as características da natureza dos dados.

A heterogeneidade inerente aos dados produzidos na pesquisa científica implica na necessidade de estabelecer políticas e estratégias de gestão amplas que envolvam os diversos tipos de dados coletados e analisados por uma instituição ou comunidade acadêmica. Reconhecer essas diferenças é fundamental para a adoção de múltiplas ações no escopo da gestão de dados científicos e do ciclo de vida da curadoria, visto que cada tipo de dado requer processos distintos de gestão e de profundidade de ações de curadoria, como por exemplo, arquivamento de longo prazo (SAYÃO; SALES, 2020).

Essa realidade envolve a necessidade de incluir os dados considerados da “cauda longa” da ciência, composto por diversos projetos de pesquisa desenvolvidas em pequenos laboratórios, pois cada vez mais esses dados estão sendo gerados e analisados, à proporção que novos métodos e instrumentação possibilitam aos pesquisadores individuais e pequenas equipes coletarem um vasto volume de dados. Nesse cenário, os cientistas sociais estão analisando grandes quantidades de dados provenientes das estatísticas governamentais, levantamentos *on-line* e modelos comportamentais (BORGMAN, 2015; SAYÃO; SALES, 2020).

A partir dessa reflexão, é possível observar que o valor dos dados não se limita aos insumos necessários para a realização de uma investigação científica ou os resultados que se podem obter a partir deles. Além disso, os dados trazem legitimidade a uma pesquisa e podem ser reutilizados em novos estudos, desde que passem por uma gestão adequada, pois por envolverem todos os domínios do conhecimento, os dados são diversificados, conforme visualizado com os dados das ciências nucleares do estudo realizado por Sales e Sayão (2019).

3 A ORIGEM DOS DADOS E A RELAÇÃO COM A SUA PRESERVAÇÃO

O relatório do NSB (2005) mostra que os dados podem se distinguir conforme a sua origem, a saber: observacional, computacional e experimental. Borgman (2015) acrescenta uma quarta categoria de origem dos dados, a de registros. A compreensão sobre essa diferença é primordial para as decisões acerca do arquivamento e preservação.

Segundo o relatório do NSB (2005), os dados de origem observacional são registros históricos que não podem ser replicados, logo podem necessitar de um arquivamento indefinido. A proveniência desses dados está relacionada a observações de fenômenos e eventos que só podem ocorrer uma vez, e, sendo assim, precisam ser arquivados de forma fidedigna, autêntica e íntegra em repositórios confiáveis que promovam uma preservação adequada no longo prazo (SALES; SAYÃO, 2019).

Borgman (2015) revela que os dados observacionais são considerados os mais importantes para preservar, por serem os menos reprodutíveis. Silva (2019) exemplifica que em ciências sociais as observações podem estar associadas ao comportamento de vários grupos ou indivíduos. Desta forma, pela razão que os dados desse tipo de origem não podem ser obtidos mais de uma vez, é primordial aprofundar as práticas de curadoria para posterior reutilização desses dados.

No que se refere aos fundamentos da metodologia científica no campo das ciências sociais, Marconi e Lakatos (2019) revelam que a observação direta intensiva ocorre por meio das técnicas de coleta de dados: observação e entrevista. Neste contexto, a observação pode ser assistemática, sistemática, não participante, participante, individual, em equipe, em laboratório, na vida real ou naturalista. Já a entrevista pode ser padronizada ou estruturada, não padronizada, semi estruturada ou livre. A observação extensiva acontece através de questionário, formulário, medidas de opinião e atitudes e técnicas mercadológicas.

Dias, Anjos e Araújo (2019) realizaram uma pesquisa envolvendo os pesquisadores brasileiros da CI, cujo domínio do conhecimento pertence a grande área das ciências sociais aplicadas. Neste contexto, o estudo investigou a tipologia dos dados científicos, no tocante a serem observacionais, computacionais, experimentais e uma combinação desses. Foi constatado que os dados

observacionais foram os prevalentes. Desta forma, é possível visualizar uma associação desses tipos de dados com a área, fato que chama a atenção para a adoção de medidas de preservação.

Em relação aos dados computacionais, o relatório NSB (2005) revela que correspondem a resultados da execução de um modelo ou simulação de computador (envolve uma documentação robusta do *hardware*, *software* e dados de entrada). A preservação dos dados de saída pode não ser essencial, pois os dados podem ser reproduzidos, todavia, é necessário o arquivamento do próprio modelo e de um conjunto completo de metadados sobre ele. Ressalta-se que, podem ocorrer casos em que o *hardware* e o *software* utilizados podem tornar-se obsoletos e/ou findados, acentuando a necessidade da preservação. Borgman (2015) complementa que normalmente eles são encontrados nas ciências físicas e da vida, porém também podem ser identificados nas ciências sociais e humanas.

No que se refere aos dados experimentais (exemplo das taxas de reações químicas), conforme o relatório NSB (2005), pode não ser possível reproduzir de forma fidedigna todas as condições experimentais, pela falta de conhecimento de algumas variáveis e quando os custos da reprodução do experimento são proibitivos. Desta forma, as questões de reprodutibilidade e custos são primordiais para a implementação de políticas para preservação de dados experimentais.

Segundo Gil (2019, p. 5), “o experimento em investigações sociais é bem pouco utilizado, visto que, de modo geral, o cientista não possui o poder de introduzir modificações nos fenômenos que pretende pesquisar.” Apesar da experimentação não ser comum nas Ciências Sociais, ela pode ser identificada em estudos que buscam medir com variáveis controladas a reação de uma população a determinado fenômeno.

Borgman (2015) enfatiza que a categoria de registro inclui formas de dados que não se adequam facilmente nas categorias de observação, experimentação e computação ou que sejam resultado de qualquer uma destas categorias. Para a autora, registros de qualquer fenômeno ou atividade humana podem ser tratados como dados para a pesquisa. Eles podem incluir a documentação de governos, empresas, públicas e atividades privadas; livros e outros textos; materiais de arquivo; documentação na forma de áudio e gravações de vídeo, placas de vidro, papiros, cuneiformes, bambu, entre outros.

Diante da relação da origem dos dados e a sua preservação, Sayão e Sales (2019) apresentam o arquivamento em repositórios confiáveis como um caminho para que os dados se mantenham preservados ao longo do tempo, de forma fidedigna, autêntica e íntegra. Nesse sentido, para Monteiro (2017), os repositórios de dados científicos possuem uma participação significativa no processo de gestão de dados. Eles tornam viável a representação, o processamento, o gerenciamento, a disseminação e o acesso aos dados que neles foram depositados, com os métodos de curadoria, fortalecem a colaboração científica internacional ao propiciar o compartilhamento, acesso, uso e reuso dos dados.

Para atender a necessidade do cenário atual, alguns tipos de repositórios foram identificados na literatura, a saber: repositório de dados específicos da área, repositório de dados institucionais, repositórios genéricos (BEZJAK et al., 2018). No diretório global de repositórios, o OpenDoar³, é possível localizar repositórios em acesso aberto que publicam conjuntos de dados de diferentes tipos.

Van den Eynden e Corti (2017) apresentam alguns benefícios em preservar os dados em um repositório de dados específicos, como: a garantia da qualidade de que os dados estão alinhados com os padrões estabelecidos; preservação de dados no longo prazo em formatos de arquivos acessíveis; o depósito de dados em um local seguro que possibilita, quando preciso, o controle do acesso; a descoberta de recursos *online* e acesso a dados por meio de catálogos de dados; suporte ao usuário; e oportunidades promocionais e de treinamentos com foco na coleta de dados, possibilitando uma maior visibilidade dos dados.

A partir dessa relação entre a origem dos dados e a necessidade de preservação, é possível verificar que os dados, em todas as suas origens, devem ser preservados, caso contrário, se limitarão a atender a necessidade do momento do(s) criador(es) dos conjuntos de dados, além de que, não poderão ser reutilizados pelo próprio pesquisador nem por outros possíveis usuários/pesquisadores terceiros.

4 PROCEDIMENTOS METODOLÓGICOS

O presente estudo é resultante do trabalho de tese de Rodrigues (2021), intitulado de “*Elementos de ciclos de vida dos dados no percurso metodológico das teses brasileiras da área de ciência da informação: um estudo diagnóstico*”. Desta

³ Disponível em: https://v2.sherpa.ac.uk/open_doar/.

forma, o caminho percorrido foi realizado com o levantamento das teses brasileiras da CI, disponibilizadas nos *sítes* dos seus respectivos programas, nos repositórios institucionais e na biblioteca virtual da Universidade de São Paulo, visto que, o Programa de Pós-Graduação em CI da referida instituição, disponibiliza suas teses nessa plataforma diferente dos outros que utilizam seus próprios *sítes* ou repositórios institucionais.

Na sequência, os procedimentos metodológicos das teses foram separados para analisar os tipos de dados em relação à natureza e origem que são utilizados pelos pesquisadores da área, no âmbito das teses.

A presente pesquisa teve como finalidade descrever características das teses da CI, no tocante a origem e a natureza dos dados científicos. A pesquisa bibliográfica foi utilizada para a coleta de dados, visto que foram investigadas as teses da CI. Além disso, foi adotada para aprofundar as definições conceituais sobre a temática, para isso, realizou-se uma investigação na literatura pertinente (MARCONI; LAKATOS, 2019).

Houve a articulação das abordagens quantitativas e qualitativas com o objetivo de obter um olhar duplo para o objeto de estudo, já que quando os dois métodos são usados de forma complementar, eles se nutrem de forma recíproca (DIETRICH; LOISON; ROUPNEL, 2015).

O *corpus* trabalhado na pesquisa envolve as teses dos programas de pós-graduação brasileiros da CI que foram desenvolvidas quando a área começou a discutir efetivamente a temática dos dados científicos no país. Fato este que remete às publicações dos artigos “*Compartilhamento de dados e e-Science: explorando um novo conceito para a comunicação científica*”, dos autores Medeiros e Caregnato (2012); e “*Curadoria digital: um novo patamar para preservação de dados digitais de pesquisa*”, dos autores Sayão e Sales (2012). Assim sendo, o recorte da pesquisa inclui as teses da CI do período de 2012-2020.

Através de uma busca preliminar no *site* da Plataforma Sucupira, os programas de pós-graduação da CI no Brasil, com curso de doutorado, foram encontrados, ao clicar no ícone “cursos avaliados e reconhecidos” e posteriormente, na opção “área de avaliação” – o que levou ao campo cursos avaliados e oferecidos pela Plataforma. Na sequência, foi selecionada a opção de “Comunicação e Informação” e, por fim, o curso “Ciência da Informação”.

A partir da busca, foram identificados 13 programas de pós-graduação com curso de Doutorado, na área da CI, referentes às seguintes instituições: Universidade de Brasília (UnB), Universidade de São Paulo (USP), Universidade Estadual de Londrina (UEL), Universidade Estadual Paulista Júlio de Mesquita Filho (UNESP-MAR), Universidade Federal da Bahia (UFBA), Universidade Federal da Paraíba (UFPB-JP), Universidade Federal de Minas Gerais (UFMG), Universidade Federal de Pernambuco (UFPE), Universidade Federal de Santa Catarina (UFSC), Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT)/Universidade Federal do Rio de Janeiro (UFRJ), Universidade Federal Fluminense (UFF) e da Universidade Fumec (FUMEC).

A população do estudo foi levantada no período de dezembro de 2020 a janeiro de 2021, correspondendo a 565 teses de doutorado brasileiras em CI, defendidas em quaisquer anos durante o período de 2012 a 2020 disponível em acesso aberto.

Foi realizada uma amostragem aleatória simples sem repetição. Para determinar o tamanho amostral (n), foi usado o cálculo amostral para populações finitas (LOHR, 2009), levando em consideração uma proporção de $p = 0,5$, erro amostral de $e = 0,05$ sob o nível de confiança de 95%. Segue a fórmula utilizada para o cálculo do tamanho da amostra:

$$\text{Tamanho da amostra} = \frac{\frac{z^2 \times p(1-p)}{e^2}}{1 + \left(\frac{z^2 \times p(1-p)}{e^2 N} \right)}$$

Onde, N se refere ao tamanho populacional e z é o valor tabelado da distribuição normal padrão que corresponde ao nível de 95% de confiança. Desta forma, obteve-se o tamanho amostral de 229 teses.

As teses disponíveis entre 2012 e 2020 foram baixadas e enumeradas de 1 (um) a 565 para a obtenção da amostragem aleatória simples, que ocorreu através de um sorteio no *software* Excel da empresa Microsoft. Na sequência, utilizou-se a função =ALEATÓRIOENTRE(\$A\$1;\$A\$565)+ALEATÓRIO() e, posteriormente, o sorteio foi ordenado através da função =ORDEM.EQ(B1;\$B\$1:\$B\$565). Logo, foram selecionadas as primeiras 229 teses sorteadas para a composição da amostra analisada.

A técnica de pesquisa de análise de conteúdo foi adotada para analisar e interpretar os percursos metodológicos das teses brasileiras da CI, à luz da tipologia

dos dados quanto à natureza e a origem, com base no referencial teórico. Desta forma, foram quantificadas as teses através das suas respectivas metodologias quanto a sua natureza, tais como: números, imagens, textos, vídeos, filmes, áudios, *software*, algoritmos, equações, animações, modelos e simulações. A obtenção desta informação é relevante para identificação dos tipos de dados que predominam na CI. A classificação dos dados usados nas respectivas teses incluiu também uma análise quali-quantitativa segundo as suas origens, seja observacional, computacional, experimental e registro ou a partir de uma correlação entre estas origens. A obtenção da informação foi fundamental para que os pesquisadores da área possam adotar medidas de preservação dos seus dados.

Para realização desse processo de categorização e codificação, foi utilizado o QDA Miner (PROVALIS RESEARCH, 2021), que é um *software* qualitativo para análise de dados. Por meio dele, foram geradas codificações de textos para análise das metodologias.

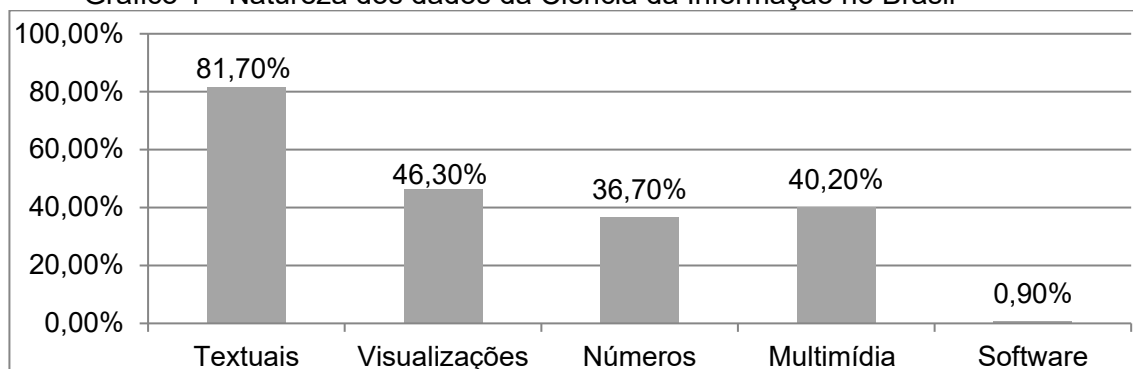
5 ANÁLISE E DISCUSSÃO DOS RESULTADOS

A análise e discussão dos resultados foi dividida em 2 (duas) subseções. Na primeira foi realizada uma análise sobre a natureza dos dados identificados nos procedimentos metodológicos das teses da CI e na segunda foram discutidos os resultados levantados no tocante à origem desses dados.

5.1 Natureza dos Dados

A compreensão acerca da natureza dos dados é de suma importância para a criação de modelos e estratégias, para dimensionar o aparato tecnológico e organizacional pertinente à sua gestão, além de possibilitar a determinação do investimento que será necessário a preservação dos dados ao longo do tempo (SAYÃO; SALES, 2020). Diante desse cenário, no Gráfico 1 está ilustrada a natureza dos dados da CI que foram levantados nos textos analisados.

Gráfico 1 - Natureza dos dados da Ciência da Informação no Brasil

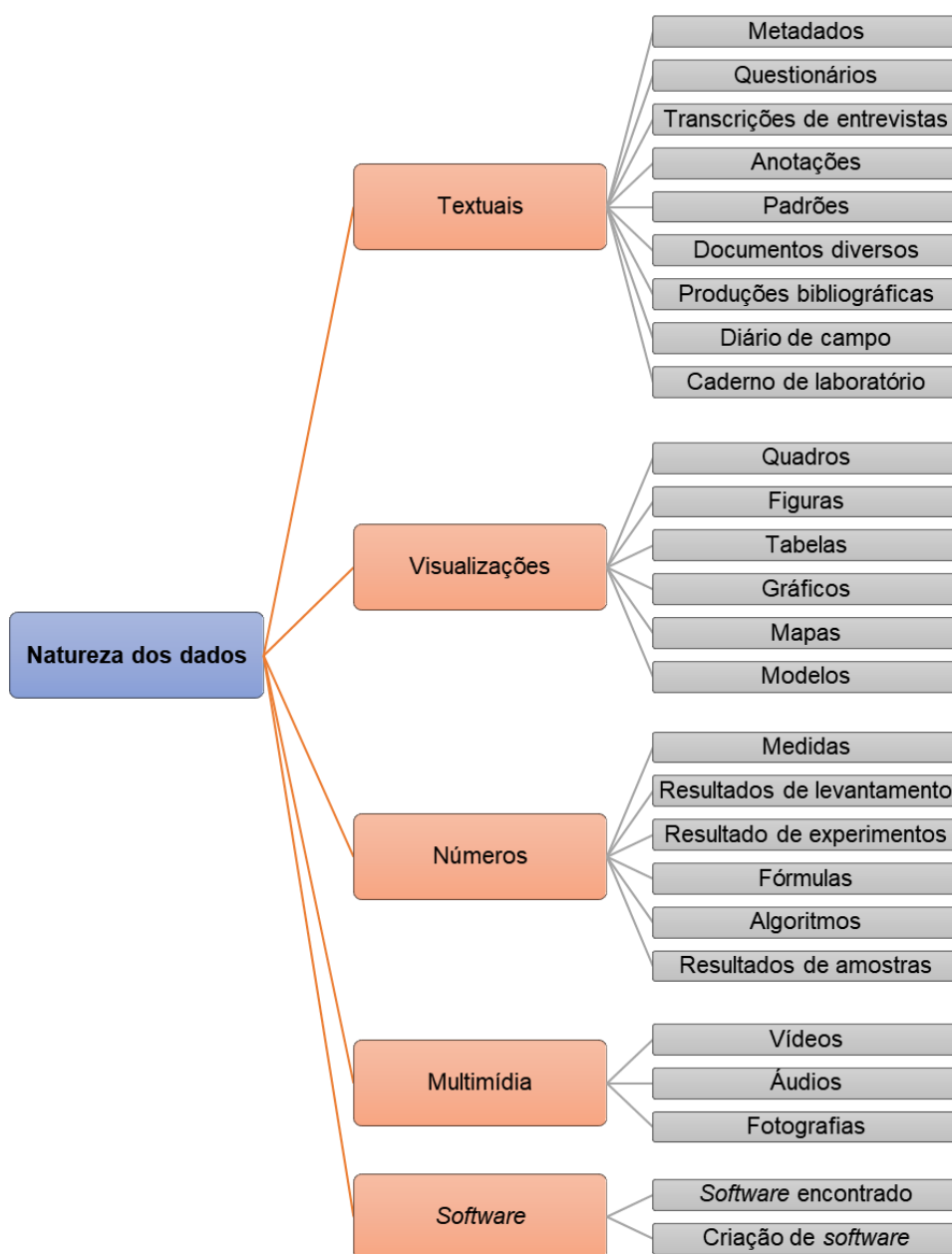


Fonte: Rodrigues (2021).

Por meio dos resultados foi possível constatar que em um mesmo trabalho houve a presença de dados de naturezas diferentes. Assim sendo, apesar dos dados textuais serem os mais comuns nas metodologias da CI estudadas, foram identificados diferentes conjuntos de tipos de dados nas teses.

Com base no estudo desenvolvido por Sales e Sayão (2019), voltado à área das Ciências Nucleares, no que se refere a natureza dos dados, foram reunidos em um mesmo grupo vários tipos de textos, diversas formas de visualização dos dados, distintos tipos de representações numéricas, alguns recursos multimídia e diferentes dados como *software*. Nesse caso, houve uma adaptação para a CI, conforme demonstrado na Figura 1.

Figura 1 – Natureza dos dados nas teses brasileiras da Ciência da Informação



Fonte: Elaboração própria com base nos dados da pesquisa

Neste sentido, os dados textuais encontrados referem-se aos metadados que foram importados de bases de dados com a descrição de documentos. Questionários e formulários foram preenchidos pela população pesquisada. As transcrições de entrevistas, diversas anotações que foram usadas como fonte de dados, padrões de metadados, vários documentos, a exemplo, leis, memorandos, cartas, atas relatórios, entre outros. Foram considerados ainda, os dados provenientes de produções bibliográficas científicas como livros, artigos, material apresentado em evento, etc. Também fizeram parte do mesmo contexto, os diários de campo, pois neles foram

registradas as percepções dos pesquisadores no trabalho de campo e, por fim, foram identificados cadernos de laboratórios. Assim sendo, constatou-se que os dados de um mesmo grupo são heterogêneos, o que requer cada vez mais práticas específicas de gestão e curadoria dos dados por meio aplicação das etapas de um ciclo de vida dos dados que mais se adeque ao projeto de cada pesquisador, com ações voltadas, por exemplo, à coleta, armazenamento, preservação e recuperação dos dados.

Outra diversidade de dados identificada está relacionada ao grupo das visualizações. Neste foram agrupados os dados escolhidos de forma preliminar para compor os critérios de definição das amostras, como também representaram os dados validados para posterior análise. Segundo Gil (2019), a disponibilização de dados em tabelas, gráficos e diagramas é primordial para o processo de interpretação e análise dos dados.

A pesquisa demonstrou a existência de dados em gráficos de nuvens de palavra e pelo método Reinert⁴, além de outros exemplos de tipos de dados, a exemplo de gráficos de superfície e do tipo diagrama de caixa (*box plot*), além disso, o mapa conceitual e o temático foram considerados como forma de visualização dos dados.

Os dados referentes a modelos corresponderam a estudos realizados nesse tipo de visualizações, assim como a própria construção deles, como SRDigital⁵, modelo-síntese, conceitual, extração de relações semântica, analíticos, dentre outros. A construção de modelos pode ser considerada um dado de saída de um trabalho, ou seja, refere-se aos dados produzidos na própria pesquisa (SCHÖPFEL *et al.*, 2016).

De acordo com o estudo realizado por Sales e Sayão (2019), foram reunidos no mesmo grupo numérico, segmentos dos textos referentes a medidas, resultados de levantamentos, resultados de experimentos, fórmulas, algoritmos e incluímos dados de amostras que foram expressas quantitativamente nos casos analisados.

No grupo dos números também foram consideradas medidas referentes a cálculos estatísticos como moda, mediana e frequência. Michel (2015) revela a importância do dado estatístico para contribuir com uma análise qualitativa mais

⁴ Ou Classificação Hierárquica Descendente (CHD) técnica para análise léxica automatizada, inferindo a correlação entre termos em um mesmo *corpus* textual.

⁵ O SRDigital constitui-se em um modelo baseado no processo de referência de Denis Grogan (2001) para o ambiente digital que depende da atuação de atores importantes, tais como: usuário e os agentes computacionais (SOUZA, 2012, p. 97).

robusta. De modo geral, neste momento da pesquisa, os segmentos codificados expressaram a quantificação dos dados.

No grupo multimídia, os vídeos, os áudios e as fotografias, identificadas nos textos, corresponderam a materiais registrados que foram coletados para análise, como também produzidos na própria pesquisa, quando os vídeos e áudios foram gravados no momento da realização de entrevistas presenciais e/ou por *Skype*⁶, assim como as fotografias que foram tiradas do ambiente investigado. Em algumas metodologias, constatou-se uma combinação desses recursos multimídia, porém os áudios predominaram.

Menos de 1% dos casos analisados apresentaram dados como *software*, oriundos das pesquisas que estudaram ou criaram algum produto desse tipo.

A partir dos resultados obtidos referentes à natureza dos dados da CI, foi possível constatar que essa área faz parte do grande número de projetos científicos que geram pequenas quantidades de dados heterogêneos. Cenário esse, discutido por Sayão e Sales (2019, 2020) e Borgman (2015) no contexto dos dados da cauda longa.

Ressalta-se novamente a necessidade da preservação dos dados encontrados no presente estudo diante da sua natureza, sobretudo, no que tange à possibilidade de contenção de gastos e tempo que seriam destinados a outras investigações. Evidencia-se que, a preservação desses dados deve ser realizada objetivando a sua descoberta, interpretação e, sobretudo o seu reuso, envolvendo estratégias e processos adequados como a elaboração e atualização de um Plano de Gestão de Dados - PGD, realizando curadoria dos dados de pesquisa durante todo o seu ciclo de vida.

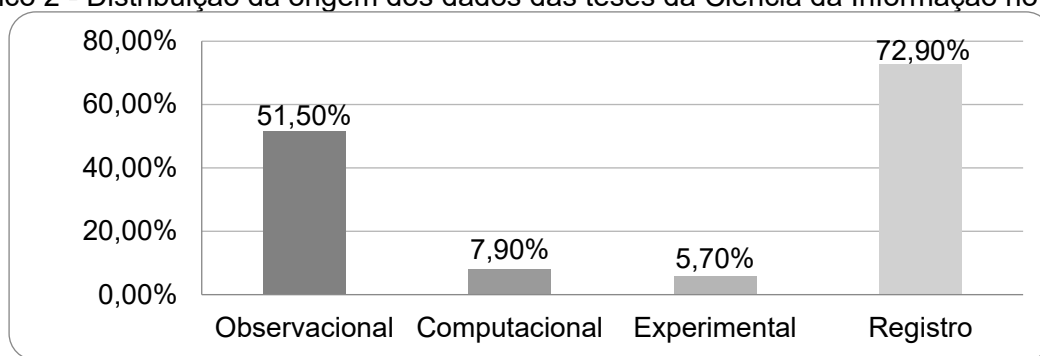
Para Curty (2019, p. 191), “o reuso é uma prática que requer estratégias e incentivos para que se materialize de forma comprometida com os princípios da ciência.” Assim sendo, esses dados necessitam de uma curadoria adequada para serem reutilizados em novos estudos.

⁶ Ferramenta de comunicação para chats e chamadas gratuitas. Disponível em: <https://www.skype.com/pt-br/>.

5.2 Origem dos dados

Conforme observado na literatura, os dados científicos podem ser de origem observacional, computacional, experimental e de registro (BORGMAN, 2015; NATIONAL SCIENCE BOARD, 2005). Por meio do presente estudo e de acordo com os percursos metodológicos das teses analisadas, verificou-se que os dados da CI podem ser provenientes dessas 4 (quatro) categorias, conforme pode ser verificado no Gráfico 2.

Gráfico 2 - Distribuição da origem dos dados das teses da Ciência da Informação no Brasil



Fonte: Rodrigues (2021).

Existe uma diferença significativa entre os percentuais levantados nas categorias de origens dos dados (Gráfico 2). Grande parte dos dados encontrados nas teses é de registro, pois essa categoria foi identificada em 167 teses, correspondente a 72,9% das 229 teses estudadas. Na sequência, destacaram-se os dados de origem observacional, que foram descobertos em 118 casos, logo, dizem respeito a 51,5% do total da amostra, ou seja, mais da metade dos trabalhos analisados.

Os dados computacionais foram verificados em 18 casos, ou seja, em apenas 7,9% dos casos estudados. A origem experimental dos dados se deu ainda em menor participação, sendo descoberta em 13 metodologias, correspondente a 5,7% de participação. Este dado corrobora com o exposto por Gil (2019) que a experimentação é pouco utilizada nas pesquisas sociais.

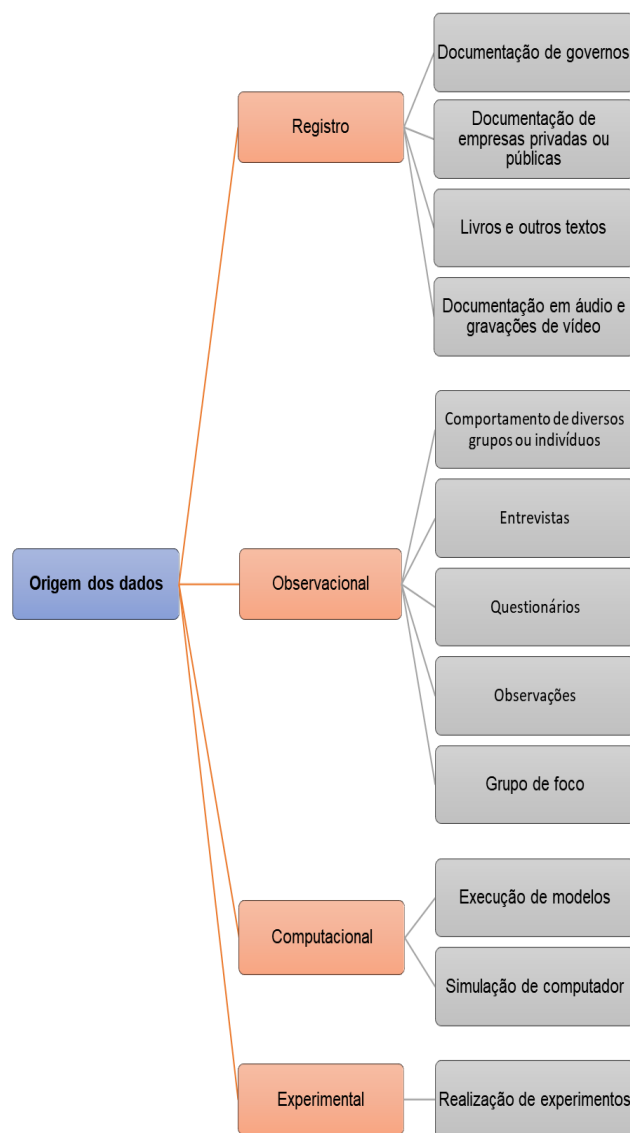
De acordo com uma pesquisa realizada por Dias, Anjos e Araújo (2019), que investigou a origem dos dados dos pesquisadores vinculados aos programas de pós-graduação em CI no Brasil, em relação a ser observacionais, computacionais, experimentais e/ou uma combinação dessas categorias, foi constatado que a maioria dos dados desses pesquisadores é de origem observacional ou observacional

correlacionada com às demais categorias. Levando em consideração que o estudo dos autores não contemplou a categoria registro, é possível identificar um alinhamento com os resultados do presente trabalho. Desta forma, fica evidenciada a necessidade de atentar para as medidas de preservação dos dados da área da CI, principalmente pela presença expressiva dos dados observacionais. Esse fato não exclui a necessidade de preservação dos dados das demais origens.

Houve a presença de mais de uma categoria de origem dos dados em um mesmo trabalho. A partir de uma análise de correlação entre essas categorias, realizada no *software* QDA Miner, 4 (quatro) casos são de origem computacional e experimental, 4 (quatro) observacional e computacional, 11 de registro e computacional, 5 (cinco) experimental e observacional, 4 (quatro) experimental e de registro, e um número considerável de 77 casos, corresponde às categorias observacional e de registro, que são as que mais originaram os dados nas teses.

É fundamental saber a origem dos dados para possibilitar a compreensão sobre em que medida deve-se adotar a curadoria dos dados voltada a cada categoria. Nesse sentido, a Figura 2 representa os resultados obtidos no âmbito das teses brasileiras CI, no tocante à origem dos dados.

Figura 2 – Origem dos dados nas teses brasileiras da Ciência da Informação



Fonte: Elaboração própria com base nos dados da pesquisa

Segundo Borgman (2015), o registro de qualquer fenômeno ou atividade humana pode estar relacionado a dados para a pesquisa. Partindo dessa reflexão, foram identificados dados dessa categoria como: documentação de governos, empresas públicas e atividades privadas; livros e outros textos; materiais de arquivo; documentação na forma de áudio e gravações de vídeo (Figura 2).

De modo geral, os dados que foram considerados como de registros, foram coletados através de atas, relatórios, legislação, memorandos, organogramas, planos, dados estatísticos, arquivos, tabelas cedidas ou descobertas em algum *site*, fontes históricas, projetos políticos pedagógicos, padrões, catálogos de bibliotecas ou de universidades, matrizes curriculares, ementas, programas de disciplinas,

documentação que pode ser considerada de governo e/ou documentação de empresas públicas ou privadas. Também foram incluídos a obtenção de dados através de livros e artigos, cujo levantamento aconteceu utilizando bases de dados nacionais e internacionais, e os dados coletados em fontes como enciclopédias, dicionários, manuais e dados de outros estudos. Essa coleta está de acordo com Borgman (2015), que apresenta dados de registro como livros e outros textos.

Os dados observacionais são registros históricos que não podem ser replicados, oriundos de eventos singulares, logo, precisam ser preservados, de forma que possam ser recuperados (BORGMAN, 2015; NATIONAL SCIENCE BOARD, 2005; SAYÃO; SALES, 2019). Esse fato chama a atenção da área da CI, uma vez que foi verificado que os dados observacionais fazem parte de forma significativa das metodologias das teses brasileiras desse campo de estudo, ao serem identificados em 118 trabalhos, o que corresponde a um percentual de 51,5%.

Para a preservação adequada dos dados, Sayão e Sales (2019) indicam o arquivamento em repositórios confiáveis, que permitam que no decorrer do tempo os dados se mantenham preservados de forma fidedigna, autêntica e íntegra. Van den Eynden e Corti (2017) apresentam como alternativa para as ciências sociais, o repositório de dados específico Reshare⁷, de grande relevância no Reino Unido.

Para categorizar os segmentos dos textos como observacional, alguns critérios foram definidos com base na literatura. Codificou-se os segmentos de textos referentes ao comportamento de vários grupos ou indivíduos (SILVA, 2019); os trechos relacionados a observação direta e intensiva, com a identificação das técnicas de coleta de dados: observação, entrevista e grupo de foco e partes dos textos inerentes a observação extensiva, quando utilizados questionário, formulário, medidas de opinião e atitudes e técnicas mercadológicas (MARCONI; LAKATOS, 2019).

As sentenças dos textos referentes à categoria observacional foram categorizadas a partir da discussão dos autores Marconi e Lakatos (2019), Michel (2015) e Silva (2019).

Assim sendo, foram identificados como dados observacionais os que procederam de entrevistas, questionários e de vários tipos de observação que foram encontrados nos textos das metodologias pesquisadas: assistemática, sistemática, descritiva, comportamental, observação em laboratórios, entre outros.

⁷ Disponível: <https://reshare.ukdataservice.ac.uk/>

Os dados provenientes da categoria computacional são mais presentes nas ciências físicas e da vida, porém, também podem ser descobertos nas ciências sociais e humanas (BORGMAN, 2015). Esta realidade foi verificada no presente estudo, pois mesmo que a presença desse tipo de dados tenha sido pouco expressiva, ele pôde ser encontrado na CI, que é uma área do conhecimento pertencente a grande área das ciências sociais.

Os trechos das metodologias foram categorizados como sendo de origem computacional com base no relatório NSB (2005), ao demonstrar que eles são resultados da execução de um modelo ou simulação de computador (Figura 2). Desta forma, os exemplos extraídos dessa categoria utilizaram ferramentas computacionais para originar os dados.

No que se refere aos dados pertencentes a categoria experimental, conforme exposto por Gil (2019), o experimento é pouco explorado nas ciências sociais, em detrimento do cientista dessa área do conhecimento não possuir o poder de introduzir modificações nos fenômenos que pretende investigar. Realidade essa que foi observada na presente pesquisa, visto que os dados de origem experimental foram os menos expressivos nos trabalhos analisados.

Apesar dessa categoria não ter sido comum nas metodologias analisadas, as poucas vezes que foi aplicado o método experimental, buscou-se determinar a relação de causa e efeito. Esses trabalhos envolveram questões computacionais, teóricas, humanas e documentais (Figura 2).

Diante da diversidade de dados identificados no âmbito das metodologias das teses brasileiras da CI e a significativa presença dos dados observacionais, é necessário que a área procure desenvolver entre os seus pesquisadores uma cultura voltada à preservação dos dados científicos, a fim de possibilitar dados acessíveis e contextualizados, que contribuam com o desenvolvimento de novos estudos e a otimização dos recursos investidos nas pesquisas da área. Esse debate pode ser enfatizado na disciplina de metodologia científica, já que os dados normalmente são discutidos nesse contexto.

6 CONSIDERAÇÕES FINAIS

Por meio da investigação realizada, foi possível classificar a tipologia dos dados científicos obtidos nas teses brasileiras, na área da CI, de acordo com a sua natureza e origem, no interstício de 2012 a 2020. O tipo de dados no tocante à natureza, foram

classificados, majoritariamente, como dados textuais, seguidos dos dados de visualizações (quadros, figuras, tabelas, gráficos, mapas e modelos). Também foram constatados dados de natureza multimídia (vídeos, áudios e fotografias), dados numéricos (medidas, resultados de levantamentos, resultados de experimentos, fórmulas, algoritmos e resultados de amostras) e uma parte pouco significativa de dados como *software*. Desta forma, a pesquisa mostrou que os dados da área no âmbito das teses são diversos. Essa informação é fundamental por evidenciar a necessidade de adotar ações necessárias para cada tipo de dado, como o arquivamento, de modo a contribuir com novas pesquisas.

Os dados das teses foram classificados de forma prevalente como sendo de origem de registro e observacional e em uma menor proporção como computacional e experimental. Devido à presença significativa dos dados da categoria observacional, sendo os mais difíceis de replicar, é preciso que a área aplique medidas de preservação e curadoria aos seus dados para assegurar que possam ser reutilizados pelos próprios criadores dos dados em novos projetos, assim como por outros usuários. Essa realidade mostra a importância de aprofundar ações de gestão dos dados que envolvam aspectos como a descrição do contexto dos dados, a escolha de um repositório confiável para o depósito dos dados, a fim de possibilitar um arquivamento de longo prazo.

Diante do estudo proposto, espera-se que a área da CI, ao discutir assuntos relacionados com a temática dos dados, possa ser referência para outros domínios do conhecimento científico ao desenvolver uma cultura voltada à preservação dos dados objetivando o seu acesso, uso e reúso.

A discussão sobre os dados normalmente é evidenciada na etapa dos procedimentos metodológicos, quando os pesquisadores especificam, por exemplo, a coleta e análise dos seus dados, sendo o momento de especificar o caminho da pesquisa. Logo, essa etapa é uma ocasião oportuna para aprofundar o percurso dos dados na investigação científica. A partir dessa realidade, sugere-se que outros aspectos relacionados à curadoria dos dados como descrição, arquivamento, preservação e reutilização dos dados possam fazer parte desse contexto.

REFERÊNCIAS

ARAUJO, C. A. A. **O que é ciência da informação**. Belo Horizonte: KMA, 2018.



ASSANTE, M.; CANDELA, L.; CASTELLI, D.; TANI, A. Are scientific data repositories coping with research data publishing?. **Data Science Journal**, v.15, p.6, 2016. Disponível em: <http://doi.org/10.5334/dsj-2016-006>. Acesso em: 19 jun. 2020.

BARDIN, L. **Análise de conteúdo**. Tradução: Luis Antero Reto, Augusto Pinheiro. Lisboa: Edições 70, LDA, 2011.

BEZJAK, S.; CLYBURN-SHERIN, A.; CONZETT, P.; FERNANDES, P.; GÖRÖGH, E.; HELBIG, K.; KRAMER, B.; LABASTIDA, I.; NIEMEYER, K.; PSOMOPOULOS, F.; ROSS-HELLAUER, T.; SCHNEIDER, R.; TENNANT, J.; VERBAKEL, E.; BRINKEN, H.; HELLER, L. **Open Science Training Handbook**. Hannover: German National Library of Science and Technology, 2018, 204 p. Disponível em: <https://www.fosteropenscience.eu/content/open-science-training-handbook/>. Acesso em: 10 dez. 2019.

BORGMAN, C. L. **Big data, little data, no data: scholarship in the networked world**. London: The MIT Press, 2015.

BORKO, H. Ciência da Informação: o que é isto. **American Documentation**, v. 19, n. 1, p. 3-5, 1968.

CURTY, R. Abordagens de reúso e a questão da reusabilidade dos dados científicos. **Liinc Em Revista**, v. 15, n.2, 2019. Disponível em: <https://doi.org/10.18617/liinc.v15i2.4777>. Acesso em: 15 mai. 2021.

CURTY, R. G; AVENTURIER, P. O paradigma da publicação de dados e suas diferentes abordagens. *In*: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 18., 2017. **Anais [...]**, Marília: UNESP, 2017. Disponível em: <https://www.brapci.inf.br/index.php/res/v/105144>. Acesso em: 8 jan. 2020.

DIAS, G. A.; ANJOS, R. L.; ARAUJO, D. G. A gestão dos dados de pesquisa no âmbito da comunidade dos pesquisadores vinculados aos programas de pós-graduação brasileiros na área da ciência da informação: desvendando as práticas e percepções associadas ao uso e reúso de dados. **Liinc em Revista**, Rio de Janeiro, v.15, n.2, p. 5-31, nov. 2019. Disponível em: <http://revista.ibict.br/liinc/article/view/4683>. Acesso em: 15 jan.2020.

DIETRICH, P; LOISON, M; ROUPNEL, M. Articular as abordagens quantitativa e qualitativa. *In*: PAUGAM, S. **Pesquisa Sociológica**. Petrópolis: Vozes, 2015. p. 171-182.

GIL, A. C. **Métodos e técnicas de pesquisa social**. 7. ed. São Paulo: Atlas, 2019.

LOHR, S. L. **Sampling: design and analysis**. Nelson Education, 2009.

MARCONI, M. A.; LAKATOS, E. M. **Fundamentos de metodologia científica**. 8. ed. São Paulo: Atlas, 2019.

MICHEL, M. H. **Metodologia e pesquisa científica em ciências sociais**: um guia prático para acompanhamento da disciplina e elaboração de trabalhos monográficos. 3. ed. São Paulo: Atlas, 2015.

MONTEIRO, E. C. S. A. **Direitos autorais nos repositórios de dados científicos**: análise sobre os planos de gerenciamento dos dados. Orientador: Ricardo César Gonçalves Sant'Ana. 2017. 115 f. Dissertação (Mestrado em Ciência da Informação) – Faculdade de filosofia e Ciências, Universidade Estadual Paulista, Marília, 2017. Disponível em: <https://repositorio.unesp.br/handle/11449/149748>. Acesso em: 5 jan. 2020.

NATIONAL SCIENCE BOARD (NSB). **Long-lived digital data collections**: enabling research and education in the 21st century. 2005. Disponível em: <https://www.nsf.gov/geo/geo-data-policies/nsb-0540-1.pdf>. Acesso em: 30 mar. 2020.

ORGANIZATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT. **OECD principles and guidelines for access to research data from public funding**. Paris: OECD, 2007. Disponível em: <https://www.oecd.org/sti/inno/38500813.pdf>. Acesso em: 16 jul. 2020.

PLATAFORMA SUCUPIRA. 2021. Disponível em: <https://sucupira.capes.gov.br/sucupira/public/index.xhtml>. Acesso em: 20 maio 2021.

PROVALIS RESEARCH. **QDA MINER**. Disponível em: <https://provalisresearch.com/products/qualitative-data-analysis-software/>. Acesso em: 15 mar. 2021.

RODRIGUES, D.G.A. **Elementos de ciclos de vida dos dados no percurso metodológico das teses brasileiras da área de ciência da informação**: um estudo diagnóstico. 2021. Tese (Doutorado em Ciência da Informação) – Universidade Federal da Paraíba, João Pessoa, 2021.

SALES, L. F.; SAYÃO, L. F. Uma proposta de taxonomia para dados de pesquisa. **Revista Conhecimento em Ação**, v. 4, n. 1, jan./jun. 2019. Disponível em: <https://revistas.ufrj.br/index.php/rca/article/view/26337>. Acesso em: 10 maio. 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, maio/ago. 2016. Disponível em: <http://www.uel.br/revistas/uel/index.php/informacao/article/view/27940/20124>. Acesso em: 20 jun. 2018.

SAYÃO, L.F.; SALES, L.F. A ciência invisível: os dados da cauda longa da pesquisa científica. In: DIAS, G. A; OLIVEIRA, B. M.J.F. (org.). *Dados científicos: perspectivas e desafios*. João Pessoa: UFPB, 2019. p. 33-52. Disponível em: <http://www.editora.ufpb.br/sistema/press5/index.php/UFPB/catalog/view/359/508/2949-1>. Acesso em: 20 out. 2021.

SAYÃO, L. F.; SALES, L. F. Afinal, o que é dado de pesquisa?. *Biblos*, v.34, n.2, 2020. Disponível em: <https://www.seer.furg.br/biblos/article/view/11875>. Acesso em: 20 abr. 2021.

SCHÖPFEL, J.; PROST, H.; MALLERET, C.; JUZNIC, P.; CESAREK, A.; KOLER-POVH, T. Dissertations and data. *Grey Journal (TGJ)*, [s. l.], v. 12, n. 3, p. 126–148, 2016. Disponível em: <http://search-ebscohost-com.ez15.periodicos.capes.gov.br/login.aspx?direct=true&db=lih&AN=118841427&lang=pt-br&site=ehost-live>. Acesso em: 16 mar. 2020.

SILVA, F. C. C. **Gestão de dados científicos**. 1. ed. Rio de Janeiro: Interciência, 2019.

VAN DEN EYNDEN, V., CORTI, L. Advancing research data publishing practices for the social sciences: from archive activity to empowering researchers. *International Journal on Digital Libraries*, v. 18, p. 113–121, 2017. Disponível em: <https://doi.org/10.1007/s00799-016-0177-3>. Acesso em: 25 out. 2020.

CONTRIBUIÇÃO DE AUTORIA

Concepção e elaboração do manuscrito: D. G. A. Rodrigues, R. L. Anjos, G. A. Dias

Coleta de dados: D. G. A. Rodrigues

Análise de dados: D. G. A. Rodrigues

Discussão dos resultados: D. G. A. Rodrigues, R. L. Anjos, G. A. Dias

Revisão e aprovação: D. G. A. Rodrigues, R. L. Anjos, G. A. Dias

CONJUNTO DE DADOS DE PESQUISA

Escolha uma das opções e apague as demais.

- 1) O conjunto de dados que dá suporte aos resultados deste estudo não está disponível publicamente.

FINANCIAMENTO

Guilherme Ataíde Dias

Bolsista de Produtividade em Pesquisa PQ – Conselho Nacional de Desenvolvimento Científico e Tecnológico
Número do projeto/processo: 311563/2018-0

Renata Lemos dos Anjos

Bolsista CAPES – Programa de Demanda Social – Coordenação de Aperfeiçoamento de Pessoal de Nível Superior
Número do processo: 88887.503749/2020-00

LICENÇA DE USO

Os autores cedem à **Encontros Bibli** os direitos exclusivos de primeira publicação, com o trabalho simultaneamente licenciado sob a [Licença Creative Commons Attribution](#) (CC BY) 4.0 International. Esta licença permite que **terceiros** remixem, adaptem e criem a partir do trabalho publicado, atribuindo o devido crédito de autoria e publicação inicial neste periódico. Os **autores** têm autorização para assumir contratos adicionais separadamente, para distribuição não exclusiva da versão do trabalho publicada neste periódico (ex.: publicar em repositório institucional, em site pessoal, publicar uma tradução, ou como capítulo de livro), com reconhecimento de autoria e publicação inicial neste periódico.

PUBLISHER

Universidade Federal de Santa Catarina. Programa de Pós-graduação em Ciência da Informação. Publicação no [Portal de Periódicos UFSC](#). As ideias expressadas neste artigo são de responsabilidade de seus autores, não representando, necessariamente, a opinião dos editores ou da universidade.

EDITORES

Edgar Bisset Alvarez, Ana Clara Cândido, Patrícia Neubert e Genilson Geraldo.

HISTÓRICO

Recebido em: 05-11-2021 – Aprovado em: 16-03-2021 - Publicado em: 25-03-2022.

