Fabiana de Melo Amaral G. Pinto

PUBLICAÇÃO DE DADOS DE PESQUISA EM ENFERMAGEM: NOVAS PRÁTICAS EM CIÊNCIA ABERTA

Tese de doutorado
Junho de 2022

UNIVERSIDADE FEDERAL DO RIO DE JANEIRO
ESCOLA DE COMUNICAÇÃO
INSTITUTO BRASILEIRO DE INFORMAÇÃO EM CIÊNCIA E TECNOLOGIA
PROGRAMA DE PÓS GRADUAÇÃO EM CIÊNCIA DA INFORMAÇÃO

FABIANA DE MELO AMARAL GONÇALVES PINTO

Publicação de dados de pesquisa em Enfermagem: novas práticas em Ciência Aberta

Tese apresentada ao Programa de Pós-graduação em Ciência da Informação, convênio entre a Universidade Federal do Rio de Janeiro/Escola de Comunicação e o Instituto Brasileiro de Informação em Ciência e Tecnologia como parte do pré-requisito para a obtenção do grau de Doutor em Ciência da Informação.

Área de concentração: Informação e Mediações Sociais e Tecnológicas para o Conhecimento

Linha de pesquisa: Comunicação, Organização e Gestão da Informação e do Conhecimento

Orientadora: Profa. Dra. Lena Vania Ribeiro Pinheiro

Rio de Janeiro
2022

CIP - Catalogação na Publicação

P659p

Pinto, Fabiana de Melo Amaral Gonçalves
Publicação de dados de pesquisa em Enfermagem: novas práticas em Ciência Aberta / Fabiana de Melo Amaral Gonçalves Pinto. -- Rio de Janeiro, 2022.
112 f.

Orientador: Lena Vania Ribeiro Pinheiro.
Tese (doutorado) - Universidade Federal do Rio de Janeiro, Escola da Comunicação, Instituto Brasileiro de Informação em Ciência e Tecnologia, Programa de Pós-Graduação em Ciência da Informação, 2022.

1. Ciência da Informação. 2. Dados de pesquisa. 3. Ciência Aberta. 4. Comunicação Científica. 5. Enfermagem. I. Vania Ribeiro Pinheiro, Lena, orient. II. Título.

Elaborado pelo Sistema de Geração Automática da UFRJ com os dados fornecidos pelo(a) autor(a), sob a responsabilidade de Miguel Romeu Amorim Neto - CRB-7/6283.

UNIVERSIDADE FEDERAL DO RIO DE JANEIRO
ESCOLA DE COMUNICAÇÃO
INSTITUTO BRASILEIRO DE INFORMAÇÃO EM CIÊNCIA E TECNOLOGIA
PROGRAMA DE PÓS GRADUAÇÃO EM CIÊNCIA DA INFORMAÇÃO

FABIANA DE MELO AMARAL GONÇALVES PINTO

PUBLICAÇÃO DE DADOS DE PESQUISA EM ENFERMAGEM: NOVAS PRÁTICAS EM CIÊNCIA ABERTA

Tese apresentada ao Programa de Pós-graduação em Ciência da Informação, convênio entre a Universidade Federal do Rio de Janeiro/Escola de Comunicação e o Instituto Brasileiro de Informação em Ciência e Tecnologia como parte do pré-requisito para a obtenção do grau de Doutor em Ciência da Informação.

Área de concentração: Informação e Mediações Sociais e Tecnológicas para o Conhecimento.

Linha de pesquisa: Comunicação, Organização e Gestão da Informação e do Conhecimento

Rio de janeiro, 08 de junho de 2022.

BANCA EXAMINADORA

_____
Profa . Dra . Lena Vania Ribeiro Pinheiro (Orientadora)
Programa de Pós-Graduação em Ciência da Informação do convênio UFRJ/IBICT

_____
Profª. Dra. Eloísa da Conceição Príncipe de Oliveira (Membro titular interno)
Programa de Pós-Graduação em Ciência da Informação do convênio UFRJ/IBICT

_____
Prof. Dr. Luis Fernando Sayão (Membro titular interno)
Programa de Pós-Graduação em Ciência da Informação do convênio UFRJ/IBICT

_____
Profa. Dra. Silvana Aparecida Borsetti Gregorio Vidotti (Membro titular externo)
Programa de Pós-Graduação em Ciência da Informação da UNESP

_____
Profa. Dra. Simone da Rocha Weitzel (Membro titular externo)
Programa de Pós-Graduação em Biblioteconomia (PPGB) da UNIRIO

_____
Prof. Dr. Ricardo Medeiros Pimenta (Membro suplente interno)
Programa de Pós-Graduação em Ciência da Informação do convênio UFRJ/IBICT

_____
Profa. Dra. Dayanne da Silva Prudêncio (Membro suplente externo)
Departamento de Biblioteconomia da UNIRIO

Rio de Janeiro
2022

À minha família, em especial aos meus filhos Davi e Sofia e ao meu companheiro de todas as horas, Carlos. Amo vocês!

À Deus, por todas as coisas, as que eu compreendo e as que ainda não compreendo.

À minha orientadora, Profª Dra. Lena Vania Ribeiro Pinheiro, pela dedicação, superação e contribuição fundamentais a realização da pesquisa.

Às amigas de turma, Jannayne Carvalho do Amaral, Patrícia Romeu, Melina Brito dos Santos e Débora Nascentes, por nossas trocas em sala de aula e pelos ENANCIBs da vida!

Ao Programa de Pós-Graduação em Ciência da Informação IBICT/UFRJ, seus docentes e funcionários;

Aos colegas de trabalho, Inaya Gomes de Andrade, Suelen Cóquero e Renata Almeida pelo incentivo e companheirismo ao longo da jornada;

E a todos que direta e indiretamente colaboraram para a conclusão deste trabalho, meu MUITO OBRIGADA!!

RESUMO

A presente pesquisa investigou a prática de publicação de dados de pesquisa na área da Enfermagem, como parte das ações em prol ao movimento da Ciência Aberta. A metodologia utilizada caracterizou-se como pesquisa exploratória e descritiva com abordagem quantitativa para a análise dos dados e apresentação dos resultados. A pesquisa buscou identificar novas práticas de publicação e comunicação de dados de pesquisa no campo da Enfermagem. Foi discutida, por meio da TAR (Teoria Ator-Rede) de Bruno Latour, a mobilização de mundo oportunizados pela prática investigada. Os resultados demonstraram um alinhamento da área para a proposta do movimento da Ciência Aberta. Contudo, este mesmo panorama aponta para obstáculos a serem superados e para a necessidade de aprimoramento das práticas de dados abertos de pesquisa. De modo geral a Enfermagem está respondendo positivamente a proposta do movimento da Ciência Aberta no que se refere à publicação de dados de pesquisa. Tal constatação pode ser considerada como um caminhar rumo às boas práticas em Ciência Aberta e demostra que a área está no centro da discussão que se coloca.

Palavras-chave: Publicação de dados de pesquisa. Enfermagem. Ciência Aberta. Ciência da Informação. Comunicação Cientifica. Periódicos de dados de pesquisa. Repositório de dados de pesquisa.

ABSTRACT

The present research investigated the adherence to the practice of publishing research data in the area of Nursing, as part of the actions in favor of the Open Science movement. The methodology used was characterized as exploratory and descriptive research with a quantitative approach for data analysis and presentation of results. The research sought to identify new practices for publishing and communicating research data in the field of Nursing. It was discussed, through ANT (Actor-Network Theory) of Bruno Latour, the mobilization of the world made possible by the investigated practice. The results showed an alignment of the area to the proposal of the Open Science movement. However, this same scenario points to obstacles to be overcome and to the need to improve open research data practices. In general, Nursing is responding positively to the proposal of the Open Science movement regarding the publication of research data. This finding can be considered as a step towards good practices in Open Science and demonstrates that the area is at the center of the discussion that arises.

Keywords: Research data publish. Nursing. Open Science. Information Science. Scientific Communication. Data journals. Data repository.

LISTA DE FIGURAS

Figura 1 - Osso de Ishango 21

Figura 2 - Pinturas rupestres como dado de pesquisa (Parque Nacional da Serra da Capivara) 22

Figura 3 - Registros cuneiformes 23

Figura 4 - Registro de plantas medicinais publicados na obra “De Materia Medica” por Pedanius Dioscorides 255

Figura 5 - Fragmentos e primeiros rascunhos de Galileu Galilei sobre o Tratado "Das coisas que flutuam na água" 26

Figura 6 - Carta portulana do sudoeste do Oceano Atlântico com o Brasil 277

Figura 7 - Dados sistematizados sobre taxa de mortalidade do exército inglês (1837 a 1846) 299

Figura 8 - Caderno de pesquisa de Marie Curie 30

Figura 9 - Publicação: Observações sobre as contas de mortalidade de John Graunt (1662) 31

Figura 10 - Visibilidade dos processos e conjunto de dados de pesquisa 39

Figura 11 - Taxonomia para Ciência Aberta 46

Figura 12 - Representação do fluxo de trabalho para comunicação científica de dados de pesquisa 50

Figura 13 - A Publicação de dados de pesquisa e a mobilização de mundo 67

Figura 14 - Esquema de pontuação baseado em cinco estrelas para dados abertos 87

LISTA DE TABELAS

Tabela 1 - Publicação de dados de pesquisa de Enfermagem em publicações periódicas via *data paper* 17

Tabela 2 - Publicação de dados de pesquisa de Enfermagem em repositórios de dados via *dataset* 18

LISTA DE GRÁFICOS

Gráfico 1 – Prática de publicação de dados de pesquisa pela Enfermagem 71

Gráfico 2 - Aparecimento e crescimento da publicação de dados de pesquisa em Enfermagem via *data paper*..... 73

Gráfico 3 - Publicação de dados de pesquisa em Enfermagem por título de periódico 75

Gráfico 4 - Revisão por pares em *data paper* via periódico pela área de Enfermagem 77

Gráfico 5 - Acesso aos dados de pesquisa de Enfermagem via *data paper* 78

Gráfico 6 - Publicação de dados de pesquisa de Enfermagem via *dataset* 79

Gráfico 7 - Publicação de dados de pesquisa de Enfermagem em repositório por ano 81

Gráfico 8 - Visualização, downloads e citação de dados de pesquisa de Enfermagem 82

Gráfico 9 - Nível de abertura dos *datasets* de Enfermagem 83

Gráfico 10 - Licença de uso dos *datasets* 85

Gráfico 11 - Formato de dados publicado pela Enfermagem 86

Gráfico 12 - Arquivos complementares ou descrição dos metadados em *datasets* de Enfermagem 90

LISTA DE SIGLAS

AO	Open Acess
ABEC	Associação Brasileira de Editores Científicos
ABNT	Associação Brasileira de Normas Técnicas
ANDS	Australian National Data Service
CAPES	Coordenação de Aperfeiçoamento de Pessoal do Nível Superior
CNPq	Conselho Nacional de Desenvolvimento Científico e Tecnológico
COVID-19	Corona Vírus Disease
DANS	Data Archiving and Networked Services
DOAJ	Diretório de Revistas de Acesso Aberto
EMBRAPA	Empresa Brasileira de Pesquisa Agropecuária
FAIR	Findable, Accessible, Interoperable, Reusable
Fiocruz	Fundação Oswaldo Cruz
IBICT	Instituto Brasileiro de Informação em Ciência e Tecnologia
ISTA	Information Science and Technology Abstracts
JASIST	The Journal of the Association for Information Science and Technology
LISA	Library and Information Science Abstracts
LISTA	Library, Information Science & Technology Abstracts with Full Text
MCTIC	Ministério da Ciência, Tecnologia, Inovações e Comunicações
NBR	Norma Brasileira
OGP	Open Goverment Partnership
OKF	*Open Knowledge Foundation*
PGD	Plano de Gestão de Dados
RCAAP	Repositórios Científicos de Acesso Aberto de Portugal
RNP	Rede Nacional de Ensino e Pesquisa
SciELO	Scientific Electronic Library Online
TAR	Teoria Ator Rede
UNESP	Universidade Estadual Paulista Júlio de Mesquita Filho
UnB	Universidade de Brasília

SUMÁRIO

1 INTRODUÇÃO 10

2 PROCEDIMENTOS METODOLÓGICOS DA PESQUISA 15

2.1 ESTRATÉGIA DE COLETA DE DADOS NAS BASES BIBLIOGRÁFICAS 17

2.2 ESTRATÉGIA DE COLETA DE DADOS NOS REPOSITÓRIOS DE DADOS .. 18

3 TRILHA HISTÓRICA SOBRE PRÁTICA DA PUBLICAÇÃO DE DADOS DE PESQUISA 20

4 REFERENCIAL TEÓRICO: TRANSFORMAÇÕES DA COMUNICAÇÃO CIENTÍFICAPELO ACESSO LIVRE E DADOS DE PESQUISA 34

5 PANORAMA TERMINOLÓGICO-CONCEITUAL SOBRE PUBLICAÇÃO DE DADOS DE PESQUISA: UMA REVISÃO DE LITERATURA 43

5.1 ABORDAGEM BASEADA EM REPOSITÓRIO 51

5.2 ABORDAGEM BASEADA EM PERIÓDICO 533

6 A TEORIA ATOR REDE (TAR) NO CONTEXTO DA PUBLICAÇÃO DE DADOS DE PESQUISA: UMA POSSIBILIDADE DE MOBILIZAÇÃO DE MUNDO 63

7 PUBLICAÇÃO DE DADOS DE PESQUISA EM ENFERMAGEM: ASPECTOS QUANTITATIVOS 70

7.1 PRÁTICA DE PUBLICAÇÃO DE DADOS VIA PERIÓDICO 72

7.2 PRÁTICA DE PUBLICAÇÃO DE DADOS VIA REPOSITÓRIO 809

8 CONSIDERAÇÕES FINAIS 93

REFERÊNCIAS 998

1 INTRODUÇÃO

A órbita da presente pesquisa situa-se no campo da *e-science*[1], na qual novos cenários e práticas científicas se estabelecem e, ao mesmo tempo, se fundem às práticas já sedimentadas, aprimorando, oportunizando e ampliando novas relações e redes de trabalhos colaborativos em ambientes digitais.

Da introspecção dos arquivos dos pesquisadores à vitrine dos repositórios, os dados de pesquisas se veem diante de um protagonismo científico e ganham cada vez mais visibilidade e importância, oportunizando novos modos de interação e comunicação científicas.

Diante deste cenário, a pesquisa promoveu um diálogo entre duas áreas do conhecimento: a Enfermagem e a Ciência da Informação. Nele, a comunicação científica, cuja transversalidade perpassa todas as áreas da ciência, é representada pela prática da publicação de dados de pesquisa. A escolha pela Enfermagem deu-se por ser o *locus* da prática profissional da autora enquanto bibliotecária na Escola de Enfermagem Aurora de Afonso Costa da Universidade Federal Fluminense. A Ciência da Informação, por sua vez, vem dedicando inúmeros esforços em pesquisas para ampliar e disseminar estudos voltados à prática da gestão e à publicação de dados de pesquisa no âmbito do movimento da Ciência Aberta, e, de acordo com Borko (1968, p.1), a CI está interessada diretamente no “corpo de conhecimentos relativos à origem, coleção, organização, armazenagem, recuperação, interpretação, transmissão, transformação, e uso da informação”, investigando “as propriedades e o comportamento da informação, as forças que governam seu fluxo, e os meios de processá-la para otimizar sua acessibilidade e seu uso.”

Nesse sentido, o presente estudo considerou como objetivo principal identificar a prática de publicação de dados de pesquisa pela área da Enfermagem, enquanto novo fazer em Ciência Aberta. Para isso, buscamos responder duas questões norteadoras: 1) Se os pesquisadores da área da Enfermagem estão publicando os dados produzidos em pesquisas e, em caso positivo, 2) Quais canais de comunicação

[1]O termo *e-Science*, originário no Reino Unido e predominante no restante da Europa, foi cunhado por John Taylor, no ano de 1999, conforme Jankowski (2007). Traduzido para o português como e-Ciência, este termo adquiriu um significado que representa a potência da ciência melhorada com o uso intensivo das TICs e sua ampliação em torno de um esforço colaborativo (FERREIRA, 2018, p.15)

científica estão escolhendo para publicá-los? Para responder as questões anteriormente colocadas, consideramos os seguintes objetivos específicos: 1) Mapear a prática da publicação de dados de pesquisa via periódico científico; 2) Mapear a prática da publicação de dados via repositório de dados; 3) Identificar os canais de comunicação utilizados para a publicação de dados; 4) Apontar o potencial de mobilização de mundo proposto pela TAR (Teoria Ator Rede) de Bruno Latour em comparação à prática investigada, e por último, 5) Propor definições para os termos: Periódico de dados de pesquisa, Artigo de dados, Depósito de dados de pesquisa e Publicação de dados de pesquisa.

A investigação está amparada pela relevância e potencial que a prática pode exercer na engrenagem da ciência como um todo. A publicação do conjunto de dados viabiliza a transparência dos processos e métodos científicos, o que, consequentemente, potencializa e fomenta o avanço da ciência. Configura-se como uma prática que contribui para o avanço da gestão de dados de pesquisa da área e para o fortalecimento do movimento da Ciência Aberta de maneira geral.

Ao considerar o movimento como uma inciativa sem volta, cujas transformações envolvem principalmente a geração e tratamento dos dados de pesquisa, acreditamos que tal prática amplia o potencial de criação, colaboração e inovação. Além disso, a publicação e descrição desses dados desperta outros vieses para reflexão, tais como: aumento da visibilidade de pesquisas; ampliação da capacidade de respostas mais rápidas à sociedade, conforme vem sendo observado na pandemia do COVID-19; otimização de tempo gasto em alguma das etapas da pesquisa; colaboração, compartilhamento e reuso de dados entre os pares; estímulo à inovação e à descoberta científica; possibilidade de melhoria das práticas, métodos investigativos e experimentais de pesquisa; dentre outros.

Ao falarmos em publicação de dados, não é possível desviar da discussão que envolve a visibilidade e internacionalização da ciência. Assim sendo, a proposição de práticas colaborativas entre pares vem sendo cada vez mais estimulada e fomentada disciplinar e interdisciplinarmente na comunidade científica.

O compartilhamento e colaboração entre pares imprime maior agilidade na capacidade de respostas à sociedade, reafirmando o caráter social da ciência. Soma-se a esse aspecto, a possibilidade de agilidade no progresso das pesquisas, tendendo

a aumentar e, em alguns casos, resgatar a confiança nas instituições públicas de pesquisa.

A prática da publicação de dados de pesquisa descortina os dados, dando-lhe a transparência necessária e essencial para a melhoria das práticas e métodos investigativos e experimentais de pesquisa.

Busca-se garantir características pertinentes ao crédito e autoridade epistêmica atribuídos à ciência por meio de seus pesquisadores e demais atores que compõem o cenário científico, considerando principalmente os aspectos de: certeza, prova, evidência, conhecimento em detrimento aos ruídos que possam surgir a partir de opiniões, hipóteses, crenças, emoções e, mais recentemente, os dados *fake*.

Sendo assim, estruturamos o presente estudo em oito capítulos, considerando em alguns deles as subdivisões necessárias para a reflexão proposta, conforme descrição a seguir.

No primeiro capítulo, apresentamos uma breve contextualização sobre o objeto que viabiliza a prática investigada (os dados de pesquisa) no contexto do movimento da Ciência Aberta. Definimos a Enfermagem como área observada e sua interlocução com a Ciência da Informação, cujo interesse traduz-se, de modo geral, na investigação sobre as propriedades e o comportamento da informação. A comunicação científica ampara a discussão por promover a prática em questão.

No capítulo 2, descrevemos o caminho metodológico trilhado pelo estudo. Destacamos a natureza da pesquisa como exploratória e descritiva, com abordagem quantitativa para a análise dos dados e apresentação dos resultados. Tal método buscou identificar novas práticas em publicação e comunicação de dados de pesquisa no campo da Enfermagem.

Para amparar a reflexão pretendida pela pesquisa, propomos no capítulo 3 uma reconstrução da história da prática da publicação dos dados de pesquisa, que chamamos de “trilha histórica sobre prática da publicação de dados de pesquisa”. A intensão desta produção textual foi desmistificar a prática como um fazer contemporâneo. A prática como a conhecemos, à luz do movimento da Ciência Aberta, apresenta meios e tecnologias diferentes dos que apresentamos no capítulo, contudo, a natureza de sua essência permanece inalterada: o registro cujo propósito

seria o de tornar público dados e informações já utilizados e experimentados para dar conhecimento às gerações futuras.

No capítulo 4, apresentamos o referencial teórico adotado pelo estudo. Nele, apresentamos autores e suas contribuições que corroboram a reflexão no contexto da Ciência Aberta, uma vez que a prática observada nasce nesse contexto. Destacamos as transformações ocorridas no modo de comunicar a ciência impulsionadas pelo movimento. Os autores discutem a invisibilidade dos dados produzidos em pesquisas e a contribuição que a prática observada pode apresentar para o alargamento e transparência da ciência.

A proposta do capítulo 5 consistiu em apresentar como a literatura científica vem considerando a publicação de dados de pesquisa na contemporaneidade. Desse modo, propomos, por meio de uma revisão de literatura, um panorama terminológico-conceitual sobre o termo “publicação de dados de pesquisa”. Sendo assim, foi possível identificar uma certa divergência conceitual para o termo investigado, o que oportunizou uma reflexão baseada em duas vertentes: repositórios e periódicos.

No capítulo 6, propomos um diálogo entre a prática observada e a Teoria Ator Rede de Bruno Latour no tocante a mobilização de mundo. Consideramos para tal, a característica volátil dos conjuntos de dados de pesquisa associada à transparência e à comunicabilidade proporcionada pela publicação via periódico ou via repositório, bem como sua possibilidade de compartilhamento e colaboração entre pares.

No capítulo 7, verticalizamos a reflexão baseada nos dados produzidos na pesquisa e apresentamos a prática observada no contexto da Enfermagem. A exploração dos dados possibilitou desmembrar a discussão sob dois aspectos: publicação de dados via periódico e publicação de dados via repositório. Tais aspectos revelam a importância da gestão dos dados e o formato com que estes devem ser apresentados em cada um dos canais de comunicação científica explorados. Expomos por meio de gráficos uma análise quantitativa sobre a prática observada.

As considerações finais da pesquisa apontam para novas possibilidades de pesquisa que envolve a prática da publicação de dados de pesquisa na Enfermagem e nas demais áreas do conhecimento. Coube-nos aqui apresentar os resultados, considerando os objetivos propostos, a cientificidade da área observada e a

proposição de novas práticas científicas impulsionadas pelo Movimento da Ciência Aberta.

Encerramos esta introdução ansiosos por apresentar, nos capítulos que se seguem, tudo o que vimos e refletimos sobre a temática escolhida pela pesquisa e certos de que este é um tema bastante recente e que ainda trará muitas ramificações em estudos futuros.

No próximo capítulo, descrevemos o caminho metodológico trilhado por esta pesquisa.

2 PROCEDIMENTOS METODOLÓGICOS DA PESQUISA

A pesquisa proposta caracteriza-se como exploratória e descritiva, com abordagem quantitativa para a análise dos dados e apresentação dos resultados. Busca identificar novas práticas em publicação e comunicação de dados de pesquisa no campo da Enfermagem.

A metodologia utilizada pela pesquisa se dividiu em duas etapas. A primeira, consistiu em uma exaustiva revisão de literatura sobre o tema: publicação de dados de pesquisa, com o propósito de entender o universo terminológico e conceitual do objeto e da prática investigada. Para tal, foram realizadas buscas nas seguintes bases de dados, bibliotecas digitais e repositórios: *Scopus, Web of Science, Science Direct, Information Science and Technology Abstracts – ISTA, Library, Information Science & Technology Abstracts with Full Text – LISTA, Library and Information Science Abstracts – LISA, Springer Link,* Literatura Latino-Americana e do Caribe em Ciências da Saúde – LILACS*, MEDLINE, CINAHL, Scielo*, Biblioteca Digital Brasileira de Teses e Dissertações - BDTD do IBICT e o Repositórios Científicos de Acesso Aberto de Portugal – RCAAP.

A estratégia de busca utilizada nas bases de dados e repositórios considerou os seguintes termos e operadores booleanos: ("data journal" OR "data publish" OR "data publication" AND "science communication" OR "scholarship communication" AND "open science").

Cabe ressaltar a limitação dos campos de busca dos repositórios utilizados na pesquisa, tais como: Biblioteca Digital Brasileira de Teses e Dissertações - BDTD do IBICT e Repositórios Científicos de Acesso Aberto de Portugal – RCAAP, ainda que não tenham retornado publicações relevantes para a pesquisa. Nestes casos, a *string* de busca teve que ser fragmentada para que pudesse ter algum resultado.

Os critérios de inclusão adotados na pesquisa consideraram estudos que estabelecessem relação entre a prática da publicação de dados de pesquisa e sua participação no *workflow* da comunicação científica. Além disso, buscamos documentos que propusessem uma definição terminológica e conceitual para o objeto investigado (a prática da publicação de dados de pesquisa). Os critérios de exclusão referem-se a documentos que não explorassem a prática da publicação de dados de

pesquisa no âmbito da Ciência Aberta. Desse modo, foram recuperados um total 546 itens, entre artigos e comunicações de eventos. Após a aplicação dos critérios de inclusão e exclusão e a retirada de resultados duplicados, foram eliminados 439 itens, participando da análise e reflexão do estudo 107 artigos. Utilizamos o software *Endnoteweb* para o gerenciamento bibliográfico dos itens recuperados nas bases de dados e posterior classificação e gerenciamento por temas explorados na pesquisa.

Os estudos selecionados na revisão de literatura oportunizaram a composição do capítulo que descreve o panorama terminológico para o conceito de publicação de dados de pesquisa e seus desdobramentos.

A segunda etapa da pesquisa consistiu na exploração das bases e repositórios de dados, com o objetivo de investigar a prática da publicação de dados pelo campo da Enfermagem. Utilizamos como fonte para investigar a prática da publicação de dados via periódico as seguintes bases de dados: *WEB OF SCIENCE, SIENCE DIRECT e SCOPUS,* por serem fontes multidisciplinares, além das bases *LILACS, MEDLINE e CINAHL*, por se tratarem de fontes de informação específicas do campo da saúde.

Para a investigação da prática de publicação de dados via repositório, recorremos ao diretório *Re3Data*[2]*,* com o propósito de selecionar os repositórios de dados mais adequados ao recorte temático do estudo. Na ausência de repositórios de dados específicos para o campo da Enfermagem, optamos pelos repositórios de dados multidisciplinares: *Figshare*, *Dryade* e *Zenodo*.

Cabe ressaltar que, antes de recorrermos às fontes anteriormente citadas, foi feita uma busca preliminar na base Ulrich, considerada como um dos diretórios mais completos e detalhados sobre publicações periódicas do mundo todo. Infelizmente, o tipo de busca proposta (por tipo de publicação: publicação de dados) não obteve um retorno favorável, na medida em que o sistema de recuperação da informação do diretório apresentou inconsistências em sua recuperação. Desse modo, optamos por não utilizá-lo como fonte de extração de dados para evitar inconsistências nos resultados das buscas.

[2] O *Re3Data* é um diretório de repositórios de pesquisa e uma ferramenta de ciência aberta que oferece a pesquisadores, organizações financiadoras, bibliotecas e editores uma visão geral dos repositórios internacionais existentes para dados de pesquisa. Fonte: https://en.wikipedia.org/wiki/Registry_of_Research_Data_Repositories

2.1 ESTRATÉGIA DE COLETA DE DADOS NAS BASES BIBLIOGRÁFICAS

Com o objetivo de identificar a prática da publicação de dados de pesquisa através da publicação periódica pela Enfermagem, buscamos nas bases de dados bibliográficas *WEB OF SICENCE, SCIENCE DIRECT, SCOPUS, LILACS, MEDLINE e CINAHL* artigos de dados (*data papers*) indexados pelo assunto Enfermagem. Aplicamos esse critério em todas as fontes consultadas com a finalidade de obter um parâmetro homogêneo nas buscas.

Assunto: Nursing

Filtro: Tipo de publicação: *data paper*

Tabela 1 - Publicação de dados de pesquisa de Enfermagem em publicações periódicas via data paper[3]

Base de dados	*Tipo de publicação: data paper*	*Período*
Web of Science	75	2015-2021
Science Direct	86	2015-2021
Scopus	16	2015-2021
Lilacs	0	-
Medline	0	-
Cinahl	0	-
Total	177	2015-2021

Fonte: A autora, 2021.

Nas bases de dados específicas do campo da saúde: *MEDLINE, CINAHL e LILACS,* não foram obtidos resultados para a busca em publicação do tipo *datapaper*. O período descrito na Tabela 1 refere-se ao tempo coberto pelo aparecimento da tipologia *datapaper* publicado pela área da Enfermagem. Não houve um recorte temporal pré-estabelecido, tendo em vista a juventude deste tipo de publicação.

[3] Pesquisa rodada em 08/11/2021.

2.2 ESTRATÉGIA DE COLETA DE DADOS NOS REPOSITÓRIOS DE DADOS

Para a seleção dos repositórios de dados, recorremos ao buscador *Repository finder,* disponível no diretório *Re3Data*, para selecionar repositórios por área do conhecimento. O *Repository finder* é uma ferramenta de busca que auxilia na identificação de repositórios por assunto. Na ausência de repositório de dados específicos para o campo da Enfermagem, seguimos a sugestão da ferramenta de busca e optamos em utilizar os repositórios multidisciplinares *Figshare*, *Dryad* e *Zenodo*. Aplicamos o mesmo critério de busca utilizado nas bases bibliográficas: por assunto “Nursing” associado ao filtro tipo de publicação: *dataset* com o intuito de obter um parâmetro homogêneo nas buscas.

Tabela 2– Publicação de dados de pesquisa de Enfermagem em repositórios de dados via dataset

Repositórios de dados	*Tipo de publicação; data sets*	*Data de depósito*
Figshare	224	2013-2021
Dryad	134	2017-2021
Zenodo	103	2017-2021
Total	461	2013-2021

Fonte: A autora, 2021.

Aplicou-se o mesmo critério temporal para a pesquisa em repositório. O período descrito na Tabela 2 refere-se ao tempo coberto pelo aparecimento da tipologia *dataset* publicado pela área da Enfermagem. Não houve um recorte temporal pré-estabelecido, tendo em vista a juventude deste tipo de publicação.

A sistematização dos resultados obtidos a partir da revisão de literatura e da exploração das bases e repositórios de dados proporcionou a composição dos próximos capítulos, onde apresentaremos, respectivamente: a trilha histórica sobre a prática da publicação de dados de pesquisa, o referencial teórico adotado pela pesquisa, o panorama terminológico que ampara a prática da publicação de dados de pesquisa e, na sequência, os aspectos quantitativos sobre a prática enquanto exercícios de Ciência Aberta pela área da Enfermagem.

A seguir, numa tentativa de contextualizar a prática investigada sob a perspectiva de produção e publicização do conhecimento ao longo da história da

humanidade, produzimos um lastro histórico do que chamamos de “trilhas” da prática de publicação de dados. Por meio deste resgate, pretendeu-se contar a “história da prática da publicação de dados”, buscando entender as tecnologias e formas de comunicar os dados produzidos ao longo dos tempos.

3 TRILHA HISTÓRICA SOBRE PRÁTICA DA PUBLICAÇÃO DE DADOS DE PESQUISA

A história da publicação dos dados de pesquisa está intimamente relacionada à história da escrita humana. A transição da humanidade, desde a pré-história, passando pela idade antiga, idade média, idade moderna e contemporânea, fez com que as formas de registros e inscrições fossem sendo modificadas ao longo dos tempos. Tais registros tornaram-se ferramenta de sobrevivência, além de ser uma forma de salvaguardar a memória para a posteridade.

Com o passar dos anos, o desenvolvimento da escrita, enquanto evolução tecnológica, ampliou a maneira de registrar dados e informação para comunicação, memória e possível reutilização futura.

Para contar um pouco do que acreditamos ter sido “os primórdios” da publicação de dados, optamos por adotar uma definição bastante ampla para nos orientar nesta jornada. Desse modo, consideramos, para fins dessa reconstrução histórica, a prática da publicação de dados como sendo toda forma de registro deixado pelo homem ao longo dos tempos, desde a pré-história até a modernidade, com o propósito de tornar público dados e informações já utilizados e experimentados para dar conhecimento às gerações futuras.

Salientamos que o termo “publicação de dados de pesquisa” é uma expressão deste século, e, como tal, não poderia ser assim identificada anteriormente. Contudo, a prática em questão é identificada ao longo dos tempos por diversos pesquisadores através dos esforços de arqueólogos, historiadores e demais cientistas, de todas as partes do mundo, em tornar os dados fruto de pesquisas o mais público possível.

Para tal, adotamos o recorte temporal com base nos principais períodos da história da humanidade (pré-história, idade antiga, idade média, idade moderna e contemporânea), associando-os às suas formas de registros de dados e informação.

Iniciamos nossa trilha histórica a partir do artigo “Uma breve história do big data desde 18.000 a.C”, publicado pelo Instituto *Information Management*[4]. O texto nos

[4] Cf. https://docmanagement.com.br/04/22/2015/uma-breve-historia-do-big-data-desde-18-000-a-c/.

remete ao período pré-histórico, remontando a história do registro de dados a 18.000 a.C, com evidências de que no período paleolítico

> nossos ancestrais tinham um sistema de arquivo de dados conhecido como osso de Ishango. O artefato citado consiste em ser uma fíbula de um babuíno, que era usada com uma marcação de traços e servia como um sistema matemático simples." (INSTITUTO INFORMATION MANAGEMENT, 2015, online).

Almeida (2002) confirma o entendimento anterior e nos ajuda a reconstruir essa história ao afirmar que a partir da descoberta do osso de Ishango "os homens estavam assim inventando os rudimentos da contabilidade escrita, traçando algarismos nos ossos (ou madeira), no sistema numérico mais rudimentar que existiu." (ALMEIDA, 2002, p. 120). Verifica-se, desde então, o registro e a publicação de dados em objetos e/ou superfícies que serviriam de orientação para a posteridade.

Figura 1 - Osso de Ishango

Fonte: *Natural Science*

As inscrições (o registro dos dados) presentes no osso de Ishango e em outros artefatos de mesma natureza ilustram e ao mesmo tempo reforçam a máxima que diz: "os números não falam por si só". Embora tais instrumentos sejam a base para o desenvolvimento da matemática[5], e, ainda, considerando toda a pesquisa histórica-arqueológica já realizada sobre tais artefatos, acredita-se que tais objetos isolados e sem descrições histórico-contextuais não revelariam exatamente em que condições e propósitos os dados foram produzidos e registrados.

[5] Cf. ALMEIDA, 2016; SANTOS, 2019; CAMPOS, 2017.

Outra figura arqueológica pertencente ao mesmo período e que contribuiu para a construção da trilha histórica da publicação de dados de pesquisa é a pintura rupestre. A abordagem voltada para a descoberta sobre o significado, propósito e condições de produção dos registros em pedras e demais superfícies ressignificam a arte contida nas pinturas rupestres em dados de pesquisa. Segundo Viana et al. (2016), tais registros:

> respondem pela primeira tentativa de sistematização dos dados fornecidos pela arte rupestre europeia, assegurando que suas figuras formavam "composições" que, embora tratadas até então como meras acumulações casuais de imagens independentes, eram portadoras de significados complexos para as sociedades que as conceberam (VIANA et al., 2016, p.3).

No Brasil, temos vários sítios arqueológicos que contribuem para demonstrar como a pintura rupestre pode ser considerada um "*dataset*". Destacamos aqui os dados registrados nas paredes das cavernas no Parque Nacional da Serra da Capivara, no Piauí, região nordeste do Brasil. Não obstante a natureza artística do painel, é inegável que o modo de registro produzido pelos povos deste período revela a riqueza e a versatilidade de formas de representação de dados ao longo dos tempos.

Figura 2 - Pinturas rupestres como dado de pesquisa (Parque Nacional da Serra da Capivara)

Fonte: Educa mais Brasil[6]

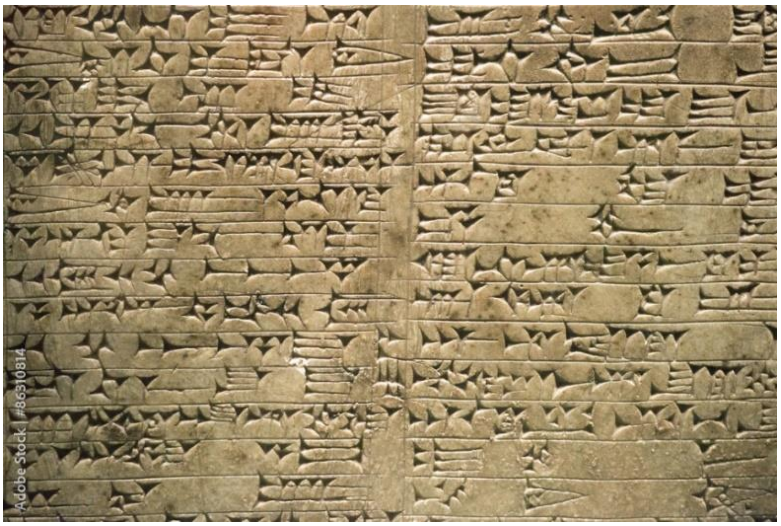
[6] https://www.educamaisbrasil.com.br/enem/artes/arte-rupestre

Percebe-se que neste período o propósito de organização, registro, administração, publicidade e memória é bastante evidente. Ao considerarmos os registros nas paredes das cavernas e demais artefatos pré-históricos, como “suporte de publicação de dados”, respeitando a compatibilidade tecnológica do período, podemos de fato dar início à construção de uma trilha histórica sobre a publicação dos dados de pesquisa.

A arte representada nas pinturas das cavernas revela também um outro olhar sobre dados. Os pigmentos das cores e materiais utilizados para cada textura abre portas para estudos no campo da química e áreas afins. As descobertas arqueológicas revelam a fertilidade do período e a possibilidade de reuso dos dados ali registrados por vários campos do conhecimento.

Adiante, temos, na passagem do período pré-histórico para a idade antiga, o desenvolvimento da escrita cuneiforme, conforme Figura 3.

Figura 3 - Registros cuneiformes



Fonte: Banco de Imagem[7]

[7] Disponível em: Freepick.(https://as1.ftcdn.net/v2/jpg/00/86/31/08/1000_F_86310814_atunc799M6S2RJHA8FCx3CfQeBCwqJRH.jpg)

A primeira indagação que nos vem à mente ao olharmos para essa placa de argila é: que dados e informações se escondem por detrás desse bloco de pedra talhado em escrita cuneiforme? Apenas especialistas neste tipo de escrita são capazes de decifrar. O registro feito em pedra, considerando-a como um suporte aberto (uma parede de pedra em espaço público), não garante o acesso à informação ali registrada.

Supomos que o registro desses dados revela informações importantes sobre o período (tempo), a civilização, costumes entre outros, contudo, apenas alguém que saiba decifrar esses códigos terá condições de fazer uma leitura completa e adequada do conteúdo descrito neste suporte e revelar que tipo de dados e informação foram ali publicadas.

O exemplo da escrita cuneiforme permite trazer para a discussão a “limitação” de descrição e contextualização ainda presente na prática da publicação de dados de pesquisa da contemporaneidade. O fato de publicar o conjunto de dados nem sempre garante seu entendimento e possibilidade de reuso em novas pesquisas. A ausência de uma descrição cuidadosa dos dados, por meio de sua gestão, é um fator que impacta diretamente na usabilidade do reuso de dados em novas pesquisas.

De acordo com Reis (2019), a maioria das descrições cuneiformes referem-se a dados contábeis e administrativos

> Com a escrita cuneiforme, a literatura teve início, os textos literários mais antigos do mundo apareceram em tabuletas sumerianas, em forma de poemas e narrativas, no entanto, a grande maioria das inscrições cuneiformes reveladas na Mesopotâmia são de registros contábeis e administrativos (REIS, 2019, p. 12).

Ao avançarmos para a idade média, observamos o desenvolvimento e aprimoramento de diversas formas de escrita em diferentes partes do mundo. O Alfabeto como o conhecemos nasce na Fenícia e é aprimorado pelos gregos. Contudo, a escrita, a leitura e a compreensão dos registros ainda eram restritas a pequenos grupos “letrados”.

Deste ponto em diante, a trilha histórica sobre a prática da publicação de dados de pesquisa começa a ficar mais próximas aos nossos olhos e ouvidos, uma vez que grandes pesquisadores e cientistas, bem como seus registros e descobertas, se

destacam neste período. A partir desse momento, nossa tentativa em organizar a trajetória histórica da publicação de dados confunde-se com a própria história da comunicação científica.

O período em questão é marcado por grandes nomes da ciência, disputas e conflitos teóricos e religiosos. Os manuscritos, suas reproduções por escribas e tradutores são recursos cada vez mais utilizados como forma de registro, memória e salvaguarda de dados e informações produzidos pelos pesquisadores e cientistas do período.

O registro de dados a partir da observação da natureza e de experimentos é relatado em manuscritos e cartas como forma de comunicação e troca entre os pares. O espírito de memória científica é fortemente identificado a partir desse período.

A Figura 4 apresenta um registro de dados (descrições) encontrados em uma publicação da idade média, escrito pelo botânico e médico grego Pedanius Dioscorides, do exército romano, sobre plantas medicinais. Sua pesquisa deu origem a enciclopédia “De matéria Medica”, escrita originalmente em grego e posteriormente traduzido para o árabe e outros idiomas.

Figura 4 - Registro de plantas medicinais publicados na obra “De Materia Medica” por Pedanius Dioscorides

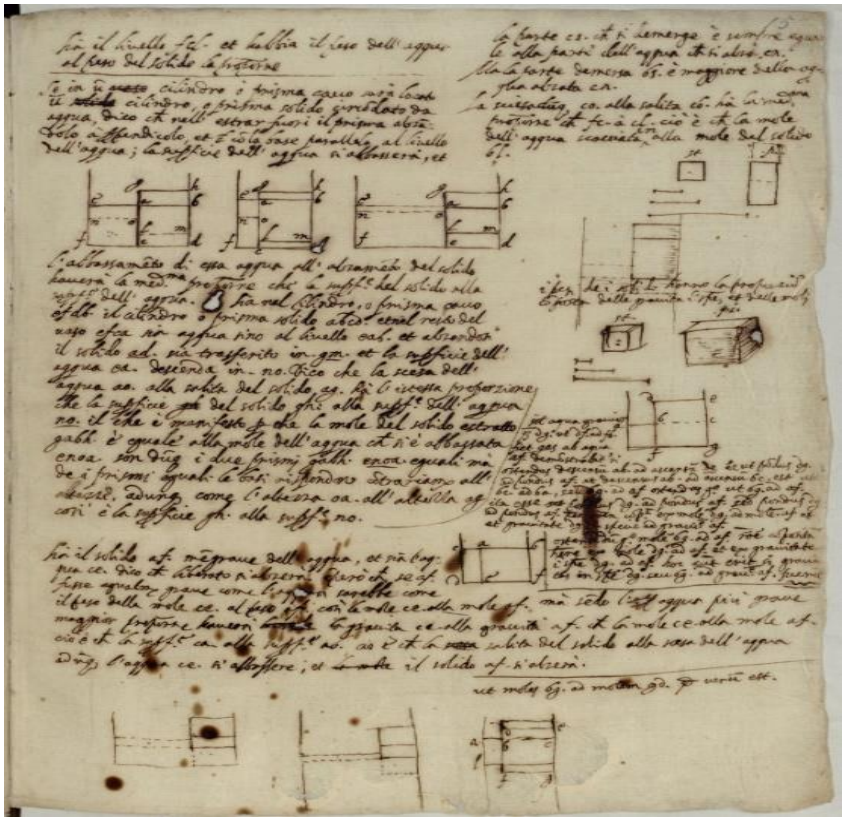
Fonte: Biblioteca Digital Mundial[8].

[8] Disponível em: https://www.wdl.org/pt/item/10690/view/1/87/

Os detalhes das ilustrações, bem como as informações sobre cada uma delas, refletem o zelo na gestão dos dados ali registrados. A prática de anotações a partir de observação e experimentos, bem como a descrição dos dados, condições e características intrínsecas e extrínsecas de cada exemplar de planta, revelam o empenho e a intensão do pesquisador em registrar e tornar público tais conteúdos, para que os dados e conhecimento até então produzidos não se perdessem ao longo do tempo.

Antes de avançarmos para o período moderno, consideramos relevante destacar o indiscutível legado da obra de Galileu Galilei no curso deste trajeto histórico. Os dados produzidos e registrados por suas investigações e experimentações, e que foram preservados até hoje, servem de base para pesquisas recentes em diversos campos do conhecimento.

Figura 5 - Fragmentos e primeiros rascunhos de Galileu Galilei sobre o Tratado "Das coisas que flutuam na água"



Fonte: Biblioteca Mundial Digital.[9]

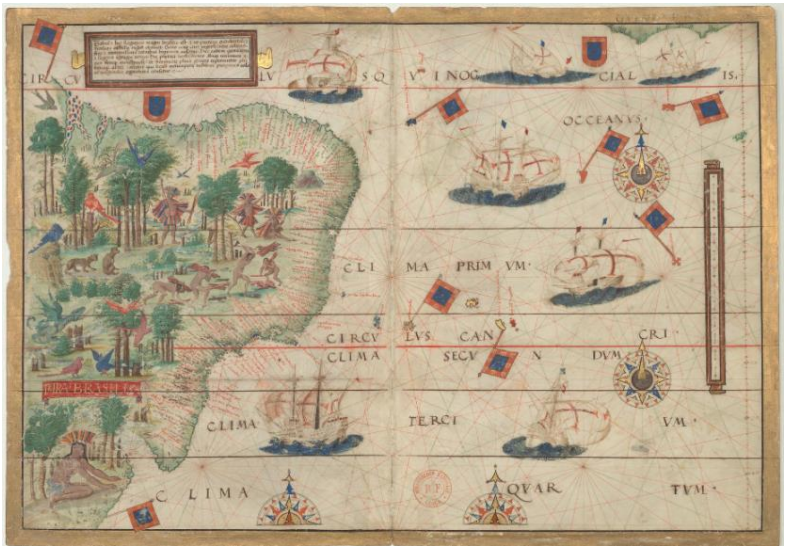
[9] Disponível em: https://dl.wdl.org/4182_1_38.png

Embora esteja em formato de rascunho, tais dados foram publicados sob o formato de livro e estão disponíveis na Biblioteca Mundial Digital[10]. Consegue-se perceber nitidamente a preocupação do cientista com o registro dos dados, seus cálculos e conclusões acerca dos experimentos realizados.

Seguimos para a idade moderna, na qual buscamos reconstruir, através de registros e publicação de dados ligados aos fatos históricos da época, a composição da trilha histórica aqui pretendida. É neste período que temos as grandes expedições marítimas e a produção de dados sobre astronomia, oceanografia e geografia dos novos territórios explorados, além de dados sobre a fauna e a flora. Dados e informações sobre fenômenos da natureza e seu impacto sobre as embarcações e as rotas de navegação também ganharam volume.

As cartas náuticas, chamadas à época de “cartas portulanas[11]”, são belíssimas publicações de dados de pesquisas. Nelas, estão registrados com riqueza de detalhes os dados para o desenvolvimento e exploração de rotas marítimas e expansão territorial, bem como a instalação e ampliação de relações comerciais. Os dados ali registrados foram produzidos a partir de observações e experimentações empíricas realizadas por navegadores, cientistas e demais estudiosos.

Figura 6 - Carta portulana do sudoeste do Oceano Atlântico com o Brasil



Fonte: Atlas Miller. Biblioteca Digital Mundial.[12]

[10] Cf. https://www.wdl.org/pt/item/4182/view/1/5/
[11] De acordo com a Biblioteca Digital Mundial, seu objetivo principal era retratar com a maior precisão possível os litorais e os portos, razão pela qual eram chamadas de portulanas. Cf. https://www.wdl.org/pt/item/18177/
[12] Disponível em: https://www.wdl.org/pt/item/18561/view/1/1/

As cartas náuticas produzidas a partir da idade moderna são excelentes documentos ou suportes tecnológicos da época, que podem nos ajudar a contar a história e evolução do registro e publicação de dados de pesquisa. Não obstante ao propósito ao qual foram criadas, hoje, ao olharmos para trás, conseguimos enxergar o quão rico e valioso é este tipo de suporte onde os dados foram registrados.

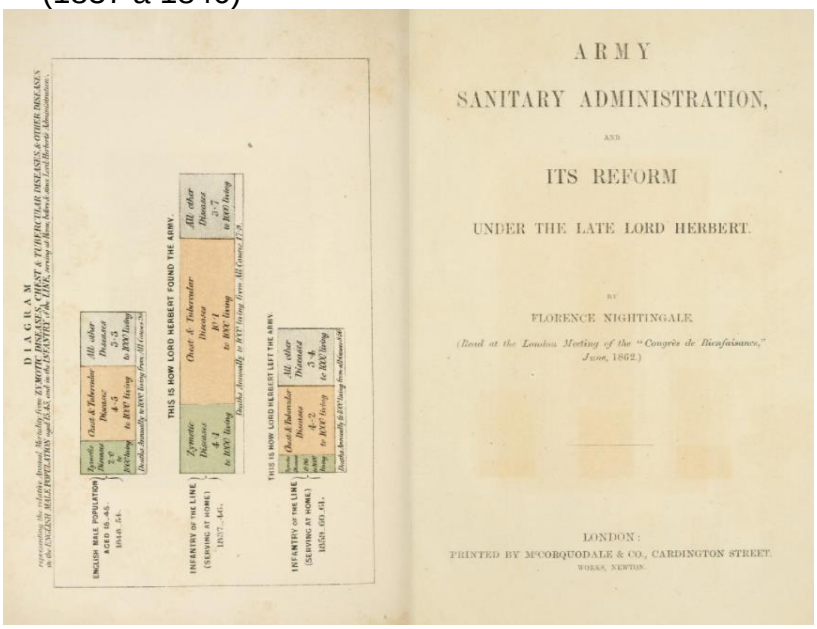
Foi no bojo da Revolução Francesa que diversas pesquisas científicas se desenvolveram e avançaram. Os dados de pesquisa produzidos neste período impulsionaram a ciência em diversos seguimentos. Antoine Lavosier na Europa e suas pesquisas em química emergem nas buscas por dados científicos produzidos no período.

Contudo, optamos por destacar como registro e publicação de dados do período em questão a publicação "*Army sanitary administration and its reform under the late Lord Herbert*" (1862), de Florence Nightingale. Considerada a fundadora da Enfermagem Moderna em todo o mundo. Ela foi uma enfermeira pesquisadora precursora na produção e gestão de dados produzidos no período pelo campo da Enfermagem. Provavelmente devido a sua formação em matemática, que antecedeu a Enfermagem, deixou um legado amplamente reconhecido em todo o domínio da saúde, principalmente no que se refere à sistematização dos dados observados e produzidos na prática da Enfermagem. Sua experiência em hospitais da Europa e na Guerra da Criméia (1853 a 1856) colaborou para o desenvolvimento e proposição de estudos quantitativos e para a prática científica da Enfermagem e saúde de modo geral.

A importância de seus registros é tamanha que, de acordo com artigo publicado no jornal *The New York Time*, "Ela escreveu sobre saúde pública, saneamento, [...], e por 50 anos ela esteve por trás da maior parte da legislação relacionada à saúde na Inglaterra." (SWEET, 2014, on-line).

A Figura 7 apresenta a publicação de dados sistematizados por Naghtingale (1862) por meio de um diagrama que representa a taxa de mortalidade anual de soldados do exército inglês.

Figura 7 - Dados sistematizados sobre taxa de mortalidade do exército inglês (1837 a 1846)



Fonte: Archive.org[13]

De acordo com a autora, o diagrama representa a taxa de mortalidade da população masculina inglesa da infantaria, com idade entre quinze e quarenta e cinco anos, nos anos de 1837 a 1846. Inclui, ainda, uma terceira divisão no diagrama, que representa a taxa de mortalidade da mesma infantaria durante três anos após a introdução de melhorias sanitárias por ela proposta, entre os anos de 1859 a 1861. (NIGHTINGALE, 1862). Consideramos a obra como uma autêntica publicação de dados da época, que revela o potencial interdisciplinar e de reuso dos dados produzidos em pesquisas.

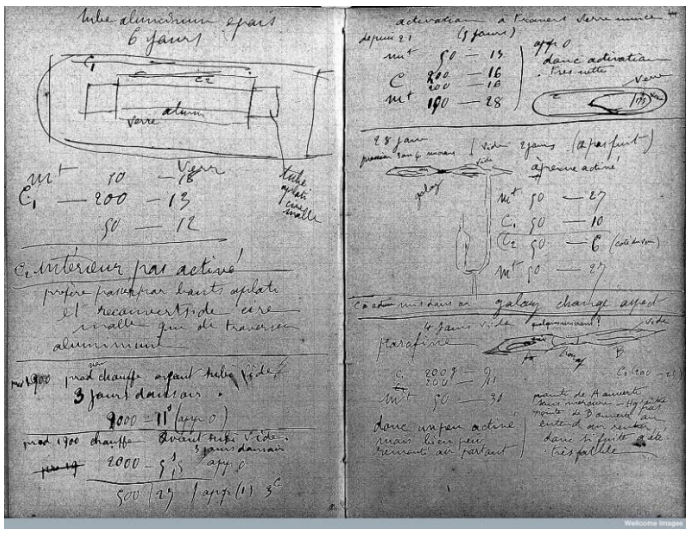
Ainda no mesmo período, não poderíamos deixar de mencionar as contribuições dos pesquisadores Marie Curie e Pierre Curie para a física e química modernas. Nas palavras de Tasch (2015),

> Curie, a primeira e única mulher a ganhar o Prêmio Nobel em duas áreas diferentes (física e química), deu continuidade às pesquisas do físico francês Henri Becquerel, que em 1896 descobriu que o elemento urânio emite raios. Ao lado de seu marido, físico francês, Pierre Curie, a brilhante dupla científica descobriu um novo elemento radioativo em 1898. A dupla batizou o elemento

[13] Disponível em: https://archive.org/details/armysanitaryadmi00nigh/page/n3/mode/2up.

polônio, em homenagem à Polônia, país natal de Marie (TASCH, 2015, *on-line,* tradução nossa[14]).

Figura 8 - Caderno de pesquisa de Marie Curie



Fonte: Science Alert.[15]

O suporte tecnológico utilizado pelos pesquisadores Marie e Pierre para registrar os dados produzidos em seus experimentos foi o caderno de pesquisa. Nele, é possível visualizar esquemas, dados numéricos e descrições textuais dos experimentos. Revela-se, assim, um arquivo de dados que reúne em uma mesma plataforma dados do tipo: imagem, número e texto.

Inúmeros cientistas deste período mereceriam destaque na construção do que chamamos de trilhas da história da publicação de dados. Porém, a proposta deste capítulo não consiste na exploração exaustiva sobre o assunto. Nosso objetivo foi contextualizar a prática e identificá-la, ao longo dos anos, sob diversas formas e condições de registros.

O período contemporâneo é fortemente marcado pelo avanço da tecnologia computacional e seus diversos suportes de produção e de registro de dados e informação. As tecnologias de informação e comunicação (TICs) nos levam a percorrer, cada vez mais, vias de mão única, onde não há como retornar ao antigo modus operandi de produção e registro de dados de pesquisa.

A *e-science* e o *big data* são o presente. A maneira com que pesquisadores contemporâneos trabalham em seus laboratórios, a infraestrutura que dá suporte aos

[14] Curie, the first and only woman to win a Nobel Prize in two different fields (physics and chemistry), furthered the research of French physicist Henri Becquerel, who in 1896 discovered that the element uranium emits rays.
[15]Disponível em: https://www.sciencealert.com/these-personal-effects-of-marie curie-will-be-radioactive-for-another-1-500-years

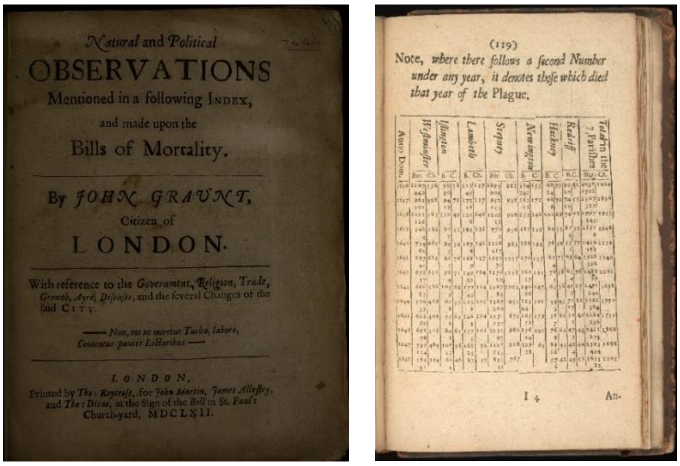
experimentos e a produção massiva de dados demandam cada vez mais conhecimento técnico, infraestrutura tecnológica e investimento. Na contemporaneidade, os repositórios de dados estabelecem-se, cada vez mais, como locais apropriados para o armazenamento e publicação de dados de pesquisa.

Stephanie Simms (2012), do Centro de Curadoria da Universidade da Califórnia (UC3), colabora com a nossa "trilha histórica da publicação de dados" ao remontar a história da publicação de dados de pesquisa há meio século, embora só tenha recebido os holofotes recentemente no bojo do movimento da Ciência Aberta. Segundo a autora, a publicação de dados, conduzida de uma maneira que seria reconhecida pelos estudiosos de hoje, vem ocorrendo há quase meio milênio. (SIMMS, 2012, on-line)

Foote (2017), em um breve artigo sobre a história do *big data*, cita a pesquisa de John Graunt, no século XVII, como exemplo de registro e reutilização de dados a partir de dados demográficos já publicados: De acordo com a autora,

> Em 1663, John Graunt conduziu uma das primeiras pesquisas registradas envolvendo análise de dados. Segundo o autor: "Ele estudou os registros de óbitos mantidos pelas paróquias de Londres. A partir disso, ele foi capaz de fazer observações sobre as taxas de mortalidade variáveis entre os sexos e até mesmo prever a expectativa de vida (FOOTE, 2017, *on-line,* tradução nossa[16]).

Figura 9 - Publicação: Observações sobre as contas de mortalidade de John Graunt (1662)



[16] In 1663, John Graunt dealt with "overwhelming amounts of information" as well, while he studied the bubonic plague, which was currently ravaging Europe. Graunt used statistics and is credited with being the first person to use statistical data analysis.

Fonte: Wellcomecollection.org. [17]

A reutilização dos dados demográficos por John Graunt e demais iniciativas de mesma natureza demonstram a importância da gestão e publicação de dados para reutilização futura. Abre-se um parêntese que nos ocorreu ao longo da pesquisa: será que Naghtingale não reutilizou a metodologia de Graunt para coletar e analisar os dados de mortalidade por ela observados? Sua formação em matemática possibilita tal elocubração e torna plausível tal hipótese sobre o assunto.

De fato, vimos que a cultura da publicação de dados vem sendo construída de forma distinta entre as áreas do conhecimento, e, de acordo com relatório publicado por Swan e Brown (2008), que aborda a visão dos pesquisadores sobre a publicação de dados:

> Algumas disciplinas estão bem à frente de órgãos de financiamento no sentido de que eles têm uma cultura de compartilhamento de dados por um longo tempo e desenvolveram uma infraestrutura e métodos para o fazer. Em outras disciplinas, o compartilhamento de dados não é comum e portanto, as políticas do financiador podem implicar em modificações significativas nas atitudes e no comportamento dos pesquisadores[18] (SWAN; BROWN, 2008, p.12, tradução nossa).

CRONIN (2013, p. 435) percebe a diversidade cultural que permeia a prática da publicação de dados ao compreender que

> Diferentes campos têm convenções mais ou menos bem estabelecidas para contabilizar o uso de dados, desde a captura e rotulagem até publicação (incluindo o processo de revisão por pares) preservação, acesso, compartilhamento e redirecionamento. [...] Mas muitos campos não têm tipos de dados consistentes ou repositórios específicos para promover prática padrões. Geralmente em domínios de natureza altamente interdisciplinares, poliparadigmáticos, cujas definições e condições de contorno geralmente são confusas (CRONIN. 2013, p. 435, tradução nossa).

Nos últimos anos, novos fluxos de trabalho envolvendo os dados de pesquisa foram sendo implementados, permitindo assim que novas modalidades de

[17] Disponível em: https://wellcomecollection.org/works/ear22f7e/items?canvas

[18] . Some disciplines are well ahead of funding bodies in that they have had a culture of sharing data for a long time and have developed the infrastructures and methods for doing this. In other disciplines, data sharing is not commonplace and therefore funder policies may imply significant modifications to researchers.

comunicação na ciência e disseminação da informação fossem exploradas pela comunidade científica.

A propagação de uma cultura voltada para a gestão e prática da publicação de dados de pesquisa requer inúmeros esforços tecnológicos e humanos, além de conceber a ciência de forma aberta e acessível a todos. Empreende um conceito livre e colaborativo para as descobertas científicas, com propósitos e aplicações para a sociedade como um todo.

A finalização desta trilha histórica reafirmou o caráter de transformação e evolução da prática de publicação de dados ao longo dos tempos. Permitiu-nos atribuir ao pesquisador ou ao produtor do dado o caráter inovador da prática, considerando as condições de registro do período. Descartou o caráter de novidade da prática, atribuída por muitos autores ao período contemporâneo. Verificou que as formas de registros e inscrições foram sendo modificadas e melhoradas de acordo com a tecnologia correspondente ao período, e, por último, reafirmou a prática enquanto instrumento de sobrevivência e forma de salvaguardar a memória para a posteridade.

A seguir, retomamos o curso da pesquisa e apresentaremos o referencial teórico que amparou a investigação proposta.

4 REFERENCIAL TEÓRICO: TRANSFORMAÇÔES DA COMUNICAÇÃO CIENTÍFICA PELO ACESSO LIVRE E DADOS DE PESQUISA

A prática de publicação de dados de pesquisa pode ser considerada uma atividade bastante recente em diversos campos do conhecimento, e até mesmo não implementada por alguns. No capítulo anterior, buscamos desmistificar este entendimento por meio de um resgate histórico de antigas práticas de publicização de dados “de pesquisa”. Tal prática vem se estabelecendo como uma modalidade de comunicação científica, e, como tal, começa a ser reconhecida e demandada como boa prática científica em diversos campos do conhecimento.

Sabe-se que processo de comunicar a ciência vem sendo estudado por muitos autores e, ao longo de pelo menos quatro séculos, considerando a publicação do primeiro periódico por volta do século XVII, foram observadas modificações em vários de seus aspectos, ainda que suas características essenciais permaneçam as mesmas.

Um dos aspectos mais significativos dessas transformações, e considerado como “divisor de águas”, pode ser atribuído à internet/web. A integração da rede mundial de computadores ao circuito da comunicação científica colaborou significativamente para os moldes atuais de disponibilização e acesso à informação científica, e, na contemporaneidade, avança rumo ao acesso aos dados de pesquisa.

Torna-se oportuno destacar que, há quase 20 anos, Schwartzman (2001, p. 5) já mencionava em seu livro sobre a formação da comunidade científica no Brasil: “temos mais ciência do que nunca, nunca dependemos tanto dela como agora, e nunca tivemos tanta clareza sobre seus problemas, limites e também possibilidades”. A atualidade de tal declaração encontra interlocução com o presente estudo, uma vez que a prática da publicação dos dados de pesquisa demanda, sobretudo, uma mudança no fazer dos atores que compõem a comunidade científica, principalmente no exercício da gestão dos dados de pesquisa.

No curso das transformações ocorridas no campo científico, no início dos anos 2000, na Europa, inicia-se o movimento em prol ao acesso livre[19] à informação

[19] Inicialmente discute-se o acesso livre, e, posteriormente, acesso aberto, conforme consagrado na Ciência da informação.

científica. Tal iniciativa tem como objetivo promover discussões e propor ações que garantam o acesso aberto à informação científica nas diversas áreas de conhecimento. Com o passar do tempo, o movimento consolida-se na literatura como Acesso Aberto (AO – Open Acess), e tem como marcos fundadores as seguintes declarações: Budapeste (2002), Bethesda (2003) e Berlim (2003). (PINTO; PINHEIRO, 2019, p.2)

Na sequência das ações em prol da ampliação ao acesso à informação científica, ocorre o movimento da abertura da ciência, conhecido mundialmente como Ciência Aberta ou Open Science. O referido movimento empreende uma postura mais ampliada para as ações que colaboram para maior transparência dos métodos, processos e ações no exercício da ciência. Albagli, Clinio e Raychtock (2014) apontam alguma das ações que compõem o movimento, a saber: acesso aberto a publicações científicas, educação aberta e recursos educacionais abertos, dados científicos abertos, ferramentas e materiais científicos abertos, ciência cidadã e os cadernos de pesquisa abertos. Especificamente para o universo da presente pesquisa, a open science inaugura um novo modus operandi na maneira de pensar, executar, e, sobretudo, registrar e comunicar a ciência. De acordo com Albagli, Clinio e Raychtock:

> a produção e comunicação científicas são processos indissociáveis [...] Daí a relevância das novas formas de registro, de documentação e de abertura das etapas da pesquisa, e não apenas o relato de “casos de sucesso” ou de resultados consolidados no formato de artigo científicos. (ALBAGLI, CLINIO E RAYCHTOCK, 2014, p.438-445)

No Brasil, diversas instituições de pesquisa científica e de governo aderiram ao compromisso proposto pela Open Science e atualmente compõem o movimento rumo a ampliação das ações pró Ciência Aberta. Dentre estas, podemos citar: o Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT); a Empresa Brasileira de Pesquisa Agropecuária (EMBRAPA); o Ministério da Ciência, Tecnologia e Inovações (MCTI); a Coordenação de Aperfeiçoamento de Pessoal do Nível Superior (CAPES); o Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq); a Fundação Oswaldo Cruz (Fiocruz); a Universidade de Brasília (UnB); SciELO, a Rede Nacional de Ensino e Pesquisa (RNP); Fundação de Amparo à Pesquisa de São Paulo (FAPESP) entre outras.

Considera-se importante destacar o protagonismo do IBICT nesse contexto ao lançar, em 2005, o Manifesto de Acesso Livre à Informação Científica e, em 2016, o Manifesto de Acesso Aberto a Dados da Pesquisa Brasileira para Ciência Cidadã. No primeiro, o IBICT propõe à comunidade científica a adesão ao movimento, contudo, sinaliza para a necessidade de criação de política nacional de acesso livre à informação científica para que a aderência seja ampla e irrestrita (IBICT, 2005).

Mais adiante, o Instituto estende a sua visão sobre o acesso aberto e reconhece os dados de pesquisa como um recurso imprescindível para as ações de Ciência Aberta, Ciência para todos, Ciência Cidadã na convicção de que o conhecimento é um bem público e a "inclusão cognitiva" é fundamental para a justiça social (IBICT, 2016).

Em 2017, sensível às transformações na Comunicação Científica, o IBICT publica o Plano de Dados Abertos 2017-2018, cujo objetivo é a promoção, a ampliação e o aprimoramento da abertura de dados da instituição, garantindo a sua publicidade, transparência e eficiência na administração pública, além de facilitar o intercâmbio de dados com outros órgãos para a melhoria da gestão pública (IBICT, 2017).

Neste cenário, o Brasil é um dos cofundadores da parceria para governo aberto OGP (*Open Goverment Partnership*[20]) e, como tal, apresenta em seu 4º Plano de Ação em Governo Aberto o compromisso 3, cuja relevância nota-se a partir da previsão de mecanismos de governança de dados científicos para o avanço da ciência aberta no Brasil. Destacamos a seguir o compromisso 3 do plano que estabelece

> O compromisso pretende avançar nos processos relacionados à disponibilização de dados abertos de pesquisa científica por meio do aprimoramento de instrumentos de governança. Objetivo - Agenda 2030: Fortalecer a pesquisa científica, melhorar as capacidades tecnológicas de setores industriais em todos os países, particularmente os países em desenvolvimento, inclusive, até 2030, incentivando a inovação e aumentando substancialmente o número de trabalhadores de pesquisa e desenvolvimento por milhão de pessoas e os gastos público e privado em pesquisa e desenvolvimento (BRASIL, 2018, p.25).

[20] A Parceria para Governo Aberto (*Open Government Partnership* – OGP) foi criada em setembro de 2011 e conta, atualmente, com a adesão de mais de 70 países e de 15 governos subnacionais. O Brasil, que é um dos co-fundadores da iniciativa, tem trabalhado intensamente para fortalecer seus princípios e práticas e, com base neles, avançar rumo à construção de um Estado mais aberto.

Neste compromisso, verifica-se a necessidade de desenvolvimento de infraestrutura e de competências informacionais e técnicas para a gestão, a guarda, a curadoria, a preservação, a descrição e a disponibilização dos dados produzidos em pesquisas científicas. O documento destaca também a relevância do compromisso ao reforçar a importância da transparência das pesquisas e da utilidade dos dados para o reuso em novas investigações.

O compromisso 3 impulsiona ações rumo à implementação de publicações de dados de pesquisa no Brasil, aumentando as possibilidades de colaboração e compartilhamento de dados, bem como o exercício de uma ciência aberta brasileira.

Uma das ações concretas que visam ao fortalecimento do movimento rumo às boas práticas em Ciência Aberta e busca atender o compromisso acima descrito pode ser observado por meio do acordo firmado entre o CNPQ e o IBICT, onde, segundo informação publicada na página do Governo Federal (2020),

> [...] resultará na criação de uma plataforma de armazenamento de dados científicos, o Lattes Data. A iniciativa é uma expansão da Plataforma Lattes com o objetivo de armazenar e permitir acesso aos dados oriundos dos projetos fomentados pelo CNPq, permitindo compartilhamento e reuso pela comunidade científica e acompanhamento pela sociedade, além de múltiplas oportunidades de inovação (BRASIL, 2020, *on-line*).

Recentemente, a Liga das Universidades de Pesquisas Europeias (LERU) publicou outra declaração onde reafirma a importância da transparência dos dados de pesquisa para fins de compartilhamento e colaboração científica. Denominada "*Sorbonne declaration on research data rights*" convida a comunidade acadêmica do mundo inteiro a se unir para, dentre outras ações, "garantir que as publicações avaliadas pelos pares sejam apoiadas pelos conjuntos necessários de dados segundo os princípios FAIR, pois os resultados da pesquisa devem ser acessíveis, verificáveis e replicáveis" (DECLARAÇÂO, 2020, on-line).

É oportuno destacar que a trajetória histórica pela publicização e acessibilidade à informação e aos dados de pesquisa tem sido protagonizada pelos pesquisadores em seus laboratórios e projetos de pesquisa. Surge como uma inovação no exercício da investigação científica, com intuito de dar transparência, confiabilidade, além de

otimizar os resultados e custos. Ao longo dos anos, vem ganhando novos aliados e é apoiada por diversos atores que compõem a engrenagem da comunicação científica.

Na maré das ações pró Ciência Aberta e infraestrutura essencial para a composição de um cenário favorável à implementação da prática da publicação de dados de pesquisa no Brasil, destacamos o lançamento da rede de Repositórios de Dados Científicos do Estado de São Paulo[21.]. Tal iniciativa fortalece e expande a participação do país nas discussões sobre ciência aberta e reforça as ações que visem ao compartilhamento e reuso de dados de pesquisa.

Portanto, é sob a égide da Ciência Aberta que a prática de publicar dados de pesquisa nasce e dá os primeiros passos em direção a uma ciência na qual o conjunto de dados produzidos em investigações científicas ganham transparência, contextualidade e mobilidade. Seu objetivo, enquanto exercício de comunicação científica, é alcançar, além de seus pares, a sociedade como um todo, tendo em vista a transparência, a colaboração e o desenvolvimento científico.

A prática propõe uma mudança cultural no fazer científico e na forma de comunicar a ciência, conforme constatou Blaise Cronin (2013) em editorial do periódico JASIST (*The Journal of the Association for Information Science and Technology*) sobre o compartilhamento e reutilização de dados de pesquisa:

> Todas as esferas da academia estão sendo afetadas, ainda que em proporções diferentes. Esses desenvolvimentos têm implicações significativas para o sistema de comunicação acadêmica em geral (incluindo JASIST) e também para o bem público.[22] (CRONIN, 2013, p.435, tradução nossa).

Tal mudança resgata a natureza “livre” e “aberta” da ciência e de sua forma de comunicá-la. A prática em questão vem assumindo novos moldes, e, em diversos

[21] A rede disponibiliza em uma plataforma aberta de dados associados a pesquisas científicas desenvolvidas em todas as áreas de conhecimento por instituições de ensino superior e pesquisa públicas no Estado de São Paulo. O desenvolvimento da rede, contou com fomento da FAPESP e com a parceria das seis universidades públicas do Estado de São Paulo – USP, Universidade Estadual de Campinas (Unicamp), Universidade Estadual Paulista (Unesp), Universidade Federal de São Carlos (UFSCar), Universidade Federal do ABC (UFABC) e Universidade Federal de São Paulo (Unifesp) –, o Instituto Tecnológico de Aeronáutica (ITA) e a Embrapa Informática Agropecuária (CNPTIA/Embrapa). https://metabuscador.uspdigital.usp.br/

[22] All corners of the academy are being affected, albeit at differential rates. These developments have significant entail-ments for the scholarly communication system in general (including JASIST) and also for the public good.

domínios, se rende ao apelo da abertura da ciência e, consequentemente, à transcrição e ao compartilhamento dos dados de pesquisa.

A ciência moderna inaugurou um certo cerceamento ao conhecimento científico, na medida em que manteve oculto e, talvez até como descreveram Austin e colaboradores (2017), “sequestrou”[23] os dados produzidos em pesquisa. Via de regra, a comunicação das descobertas e intentos científicos acontecem de maneira descritiva por meio de artigos convencionais. A matéria prima da pesquisa, ou seja, o(s) conjunto(s) de dados produzidos, por muito tempo ficou oculta e pouco acessível aos pares, o que por vezes esterilizou o potencial para colaboração científica e até mesmo a internacionalização da ciência, conforme ilustrou Curty (2017) na Figura 10:

Figura 10 – Visibilidade dos processos e conjunto de dados de pesquisa

Fonte: Curty (2018) adaptado pela autora (2020).

Curty (2017) ilustra através da Figura 10, a submersão dos processos e do conjunto de dados produzidos em pesquisas. A prática da publicação de dados quer emergir essa informação, possibilitando um alargamento e transparência para a pesquisa.

Ainda sobre a invisibilidade dos dados de pesquisa, Chavan e Penev (2011, p. 2) reforçam a mensagem representada pela Figura 10, ao afirmarem que “os conjuntos de dados estão presos invisivelmente nos armários institucionais e

[23] O termo “sequestro de dados” foi utilizado por Austin e colaboradores (2017) em seu artigo: *Key components of data publishing: using current best practices to develop a reference model for data publishing* e se refere ao apagamento/ocultamento do conjunto de dados que deu suporte a pesquisa.

individuais, computadores e discos. Este pode ser considerado um obstáculo à interdisciplinaridade e internacionalidade".

Ao longo do tempo, a comunicação na ciência vem se moldando às novas demandas sociais e científicas. As transformações, principalmente tecnológicas, ocorridas no decurso de quatro séculos demonstraram o quão modificado foi o workflow da comunicação cientifica. A transição da versão impressa dos periódicos para a versão digital e, em alguns casos, a coexistência dessas versões, exigiu esforços da comunidade científica no que se refere ao conhecimento em tecnologia e financiamento para implementação e acesso ao conteúdo no formato impresso, eletrônico e digital. Esse passo complexo, e nada trivial, potencializou ainda mais o alcance na maneira de comunicar a ciência.

O espírito inovador dos cientistas, associado às possibilidades tecnológicas sempre em desenvolvimento, oportuniza mais uma vez a transformação do modo de comunicar a ciência. Os dados produzidos em pesquisas alcançam status e valor estratégico, podendo ser utilizados e reutilizados não somente por seus produtores, mas também por pares que desenvolvem pesquisas afins. Ao invés de ficarem armazenados em arquivos particulares (cauda longa), por vezes inacessíveis após a pesquisa, passam a ser visíveis e conhecidos pela comunidade científica.

O conceito "cauda longa da ciência", de acordo com Wyborn; Lehnert (2016), vem sendo atribuído a pequenas coleções de dados produzidas por um número significativo de núcleos de pesquisa e estão sendo consideradas como ativos informacionais. Sayão e Sales (2019) complementam a noção de cauda longa ao afirmarem que "os dados da cauda longa representam a maior parcela de dados produzida pela ciência e constituem um território de constante criatividade e inovação que precisam ser revelados, integrados e compartilhados." (2019, p. 6). Ao serem dispostos em "vitrines", o conjunto de dados ganha potencial mobilidade, podendo ser reaproveitado em outras pesquisas que tragam nova utilidade científica.

Uma parcela da comunidade científica, imbuída do sentimento e necessidade cada vez mais urgente da abertura para o compartilhamento e reuso dos dados científicos, reconhece a urgência de se adaptar às novas demandas da ciência e ampliar a maneira de comunicá-la. Assim, no início dos anos 2000, surgem os periódicos de dados de pesquisa. Sua composição se dá, essencialmente, por artigos

de dados. Tais artigos configuram-se como uma nova tipologia documental e têm por finalidade descrever (por meio de metadados) o conjunto de dados produzidos por uma pesquisa, bem como seu método, modelos e condições de sua produção.

Soma-se a essas características a indicação, via identificador persistente, do acesso ao conjunto de dados que deverão estar armazenados em um repositório de dados para fins de preservação e reutilização em novas pesquisas.

Cabe ressaltar que, em estudos anteriores a este, autores como García-García; López-Borrull e Peset (2015) e Schöpfel, Farace, Prost e Zane (2019) sinalizaram o surgimento dos periódicos de dados a partir dos anos de 1956. Contudo, entendemos que o periódico de dados em conformidade com os moldes propostos pelo movimento da ciência aberta, de fato, só começa a ser delineado a partir dos anos de 1982, com a revista *Data Science Journal*.

O periódico de dados de pesquisa, comumente conhecido na literatura internacional como *data journal ou research data journal*, tem como função precípua reunir, comunicar e dar visibilidade a informações sobre os conjuntos de dados produzidos em pesquisas por meio de artigos de dados (*data paper*). Outras variações para o termo foram identificadas ao longo da pesquisa e serão apresentadas mais adiante, uma vez que compõem o arcabouço terminológico e conceitual que envolve o objeto do estudo.

No cenário brasileiro, até o presente momento da pesquisa, não se tem registro de um periódico de dados de pesquisa, contudo, em seu volume 25 de maio de 2020, a revista Encontros Bibli: revista eletrônica de Biblioteconomia e Ciência da Informação, da Universidade Federal de Santa Catarina, inaugurou uma nova página na história da comunicação científica brasileira via descrição de dados de pesquisa. Ao lançar a modalidade de publicação “artigo de dados”, compactua e contribui com o movimento de acesso aberto a dados de pesquisa, além de potencializar o compartilhamento e reuso dos dados em novas pesquisas. Com isso, nota-se que a discussão sobre gestão e compartilhamento de dados de pesquisa vem avançando, ganhando robustez e visibilidade no cenário científico brasileiro.

A publicação de dados de pesquisa demanda novas habilidades e conhecimentos sobre o fluxo de trabalho da comunicação e publicação científica. Principalmente no que se refere à descrição detalhada do conjunto de dados. O fato

de o conjunto de dados estar sob a curadoria de um repositório de dados, não garante automaticamente a sua comunicação, ou seja, necessita de informação e descrição contextualizada sobre condições e propósitos de sua produção e uso.

No âmbito da comunicação cientifica, a literatura mostrou que, ao nos referirmos à “publicação de um conjunto de dados”, é necessário distingui-los sob duas abordagens: a da publicação periódica e a do repositório de dados. Publicar um conjunto de dados em um periódico científico não significa o mesmo que publicar um conjunto de dados em repositórios de dados de pesquisa. Observou-se que o termo “publicação de dados” ou “*data publishing*” é utilizado indiscriminadamente na literatura, o que desencadeia diversas interpretações sobre o assunto.

No próximo capítulo, com base em uma revisão de literatura sobre a terminologia “publicação de dados de pesquisa”, apresentaremos como a comunidade científica vem adotado a terminologia na literatura.

5 PANORAMA TERMINOLÓGICO-CONCEITUAL SOBRE PUBLICAÇÃO DE DADOS DE PESQUISA: UMA REVISÃO DE LITERATURA

Antes de apresentar e discutir as definições encontradas na literatura, à luz do movimento da Ciência Aberta, considerou-se oportuno iniciar a reflexão a partir da contribuição da pesquisadora Ingetraut Dlahberg (1978), sobre o conceito e suas relações. Embora seja um trabalho de mais de 40 anos, é um clássico e reconhecido mundialmente como fundamentação teórica para a estruturação de um conceito.

As ideias de Dalberg (1978) foram decisivas para a elaboração da definição de um conceito, o que possibilitou ampliar o escopo da discussão sobre as definições propostas para a terminologia referente à publicação de dados de pesquisa. De acordo com Dlahberg (1978, p. 106), “as definições são pressupostos indispensáveis na argumentação e nas comunicações verbais e se constituem elementos necessários na construção de sistemas científicos”. Ou seja, estabelecer uma definição homogênea e sistêmica sobre um determinado conceito é fundamental para o amadurecimento e construção de um campo, prática ou objeto científico.

A mesma autora reforça o entendimento anteriormente exposto, tendo em vista a internacionalização da ciência, ao afirmar que

> [...] a importância das definições evidencia-se também quando se tem em vista a comunicação internacional do conhecimento. É pelo domínio perfeito das estruturas dos conceitos que será possível obter também perfeita equivalência verbal (DAHLBERG, 1978, p. 106).

Contudo, tanto a terminologia encontrada da literatura quanto os conceitos atribuídos à publicação de dados de pesquisa revelam a juventude da temática investigada, ainda muito incipiente na Ciência da Informação. Essas circunstâncias levam, algumas vezes, a um entendimento equivocado sobre o que entendemos por “publicação de dados de pesquisa”. Tal constatação decorre do fato de a ação de depositar ou armazenar um conjunto de dados em um repositório ser considerada completamente diferente do fluxo de trabalho da publicação em um periódico científico.

Campos (2001) também colaborou com a presente pesquisa ao salientar que, de acordo com a Teoria Geral da Terminologia, de Eugen Wuester, o trabalho terminológico estabelece a fixação de conceitos, visando à elaboração de definições orgânicas, além de estabelecer princípios para a criação de novos termos e possibilitar a comunicação mais precisa entre especialistas de diversas áreas do conhecimento no âmbito da Ciência e da Tecnologia, em nível nacional e internacional (CAMPOS, 2001, p.71). Autores como Parsons e Fox (2013), Borgman (2015), Li, Greenberg e Dunic (2020) se referiram a essa inconsistência terminológica como uma “metáfora de publicação de dados”.

Sendo assim, entende-se que termos e conceitos devem estar logicamente sincronizados, para que produzam um entendimento harmonioso sobre um campo, prática ou objeto informacional, ainda que estes transitem no domínio da ciência como um todo, como é o caso do movimento da Ciência Aberta, e seu objeto se aplique à diversos domínios, como é o caso dos dados de pesquisa. A harmonia conceitual pode favorecer o desenvolvimento de estudos interdisciplinares e a colaboração entre pares.

Para subsidiar a proposição terminológica e conceitual referente à prática da publicação de dados de pesquisa e suas variações, adotou-se o conceito de Sales e Sayão (2019) sobre dados de pesquisa, baseado na vasta literatura que estudaram. De acordo com os autores: “Dado de pesquisa 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” (SALES; SAYÃO, 2019, p. 36).

Cronin (2013), como pesquisador e editor científico no campo da Ciência da Informação, reforça a importância da acessibilidade aos dados de pesquisa no contexto da comunicação científica ao afirmar que:

> [...] os dados não são simplesmente adendos ou artefatos de segunda ordem; ao contrário, eles são o coração de boa parte da literatura narrativa, o material proteico que permite inferência, interpretação, construção de teoria, inovação e invenção [...].[24] (CRONIN, 2013, p. 435, tradução nossa).

[24] Data are not simply addenda or second-order artifacts;rather, they are the heart of much of the narrative literature,the protean stuff that allows for inference, interpretation,theory building, innovation, and invention.

A palavra publicar significa, em linhas gerais, dar publicidade; ato de tornar público; registrar; comunicar. No âmbito da ciência, isso se aplica a dar ciência e transparência aos pares e à sociedade como um todo às pesquisas científicas e seus métodos, por meio de um canal formal de comunicação (livros, periódicos, eventos e seus *proceedings*). Meadows (1999) e demais autores chancelam essa afirmação ao apresentarem o periódico científico como instrumento legitimado pela comunidade científica para comunicar e tornar público as pesquisas concluídas ou em andamento, na medida em que foram aprovadas e avaliadas pelos pares.

É oportuno lembrar que no inglês as duas palavras "*publish*" e "*edit*" têm conotações distintas. Nos Estados Unidos, há clara diferença entre o "editor", responsável intelectualmente pela obra e o "*publisher*", responsável pela sua publicação ou editoração, enquanto no Brasil, para ambas as funções, é usada a palavra editor".

Contudo, no contexto da Ciência Aberta e sua proposta de abertura dos dados de pesquisa, novas maneiras de oferecer transparência à ciência surgem, ampliando o arcabouço terminológico e conceitual do campo científico. Ao nos referirmos à "publicação de dados de pesquisa", inúmeras interpretações, aplicações e definições são apresentadas na literatura para o termo, o que ousamos considerar como uma inconsistência terminológica inerente a uma nova prática de comunicação científica.

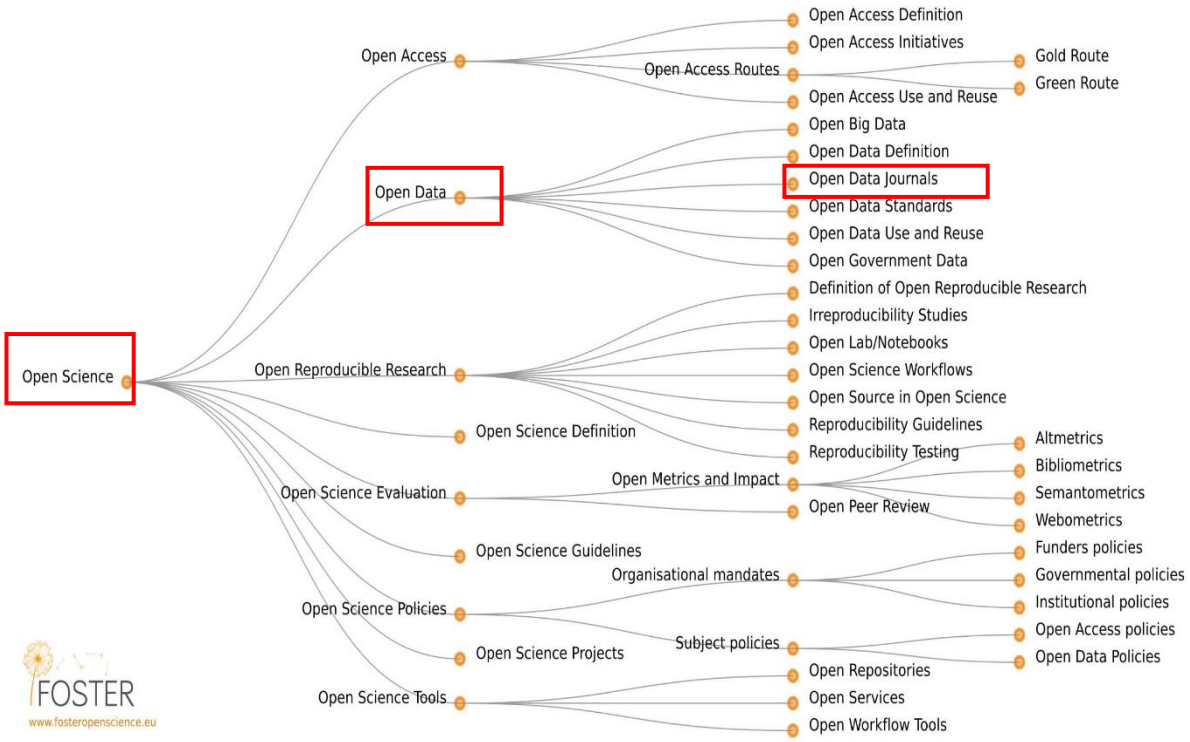
Diversos autores contribuem para a pluralidade conceitual e terminológica do campo, mostrando claramente ser essa uma definição em processo de construção, consolidação e amadurecimento. Ao mesmo tempo, revela características interdisciplinares em suas relações ao trazer para a discussão colaborações de múltiplos campos do conhecimento.

Compilamos, a seguir, as definições encontradas na literatura, sobretudo internacionais, que se referem à publicação de dados de pesquisa e que possibilitaram um entendimento mais completo e objetivo sobre a prática investigada. As contribuições aqui apresentadas permitiram-nos identificar características distintas aplicadas aos termos: "**periódico de dados de pesquisa**", "**publicação de dados de pesquisa**", "**documento de dados de pesquisa"** e seus correlatos em língua inglesa.

Sendo assim, iniciamos nossa reflexão a partir da representação ontológica do objeto "periódico de dados de pesquisa" no contexto do movimento da Ciência Aberta.

Ao apresentarmos a taxonomia desenvolvida pelo projeto *Foster Open Science* (Figura 11), foi possível inferir relações no âmbito da comunicação científica e no campo científico de modo geral.

Figura 11- Taxonomia para Ciência Aberta



Fonte: Foster Plus - Foster Open Science, 2020.[25]

Na taxonomia representada pela Figura 11, os periódicos de dados aparecem diretamente ligados aos dados abertos, e estes, por sua vez, representam uma das abas do guarda-chuva da Ciência Aberta. Tais periódicos configuram-se como um braço do movimento, além de ser local para escoamento para publicação de dados.

A *FOSTER Plus* é uma iniciativa da União Europeia para a implementação e prática de Ciência Aberta *no Horizont2020*[26]. Seu objetivo é contribuir para a mudança real e duradoura no comportamento dos pesquisadores europeus para garantir que a Ciência Aberta se torne uma realidade normativa.

[25] Disponível em: https://www.fosteropenscience.eu/themes/fosterstrap/images/taxonomies/os_taxonomy.png
[26] O Programa Horizonte 2020 é um programa de pesquisa e inovação da União Europeia (UE). Ele promove a transferência de tecnologia e inovação entre Academia e Indústria. https://www.fapema.br/wp-content/uploads/2018/06/Brochure-H2020-Brasil.pdf

Nesta iniciativa, os periódicos de dados, denominados na literatura internacional como “*data journals*” ou “*open data journals*”, são publicações revisadas por pares que reúnem artigos de dados, também conhecidos como documentos de dados ou “*data papers*” (*FOSTER OPEN SCIENCE*, 2019, *on-line*).

Os periódicos de dados nascem no âmbito do movimento da Ciência Aberta e sua matéria prima, os dados de pesquisa, são protagonistas de inúmeras abordagens de pesquisa científica. Representam uma parcela da literatura que concebem a publicação de dados de pesquisa pelo viés tradicional da comunicação científica, ou seja, o periódico científico.

Desse modo, a construção do panorama terminológico teve início com o trabalho de Lawrence e colaboradores, em 2011, quando entendem ser a

> publicação de dados uma ponte entre humanos e computadores, na medida em que preenche a lacuna entre dados, que são consumidos diretamente por um computador, e informações, que são produzidas a partir de dados, mas consumidas por seres humanos.[27] (LAWRENCE et al., 2011, p. 15, tradução nossa)

Para estes autores, uma publicação de dados compõe o fluxo de trabalho da comunicação científica, sob os seguintes vieses: 1. Publicação de dados s*tand-alone.* 2. Publicação de dados por *proxy*. 3. *Appendix data*. 4. *Journal driven data archival*. 5. *Overlay.*

Na publicação de dados do tipo *stand-alone*, o conjunto de dados é publicado (depositado) em um repositório sem a necessidade de um documento de dados que o descreva.

Na publicação de dados por *proxy*, a publicação dos dados (depósito em repositório) está atrelada à publicação do artigo convencional, porém, apresenta uma desvantagem. A política que rege a preservação a longo prazo (no repositório) do conjunto de dados é independente do documento convencional e limitada pelas políticas e financiamento da instituição que gerencia o repositório. Ou seja, ao longo do tempo, o acesso ao conjunto de dados pode ser descontinuado.

[27] Data publication provides a bridge between human and computer in that it bridges the gap between data, which is consumed by a computer directly, and information, which is produced from data, but consumed by humans.

No tipo de publicação de dados "*Appendix data*", o conjunto de dados é publicado (arquivado) como material suplementar na própria revista.

Na opção de publicação de dados do tipo "*Journal driven data archival*", os autores devem reservar uma seção do artigo convencional para fazer a descrição do conjunto de dados produzidos na pesquisa. A descrição dos dados está contida no artigo convencional, portanto, passa pela avaliação dos pares. A seção de descrição dos dados está condicionada à indicação do depósito do conjunto de dados em um repositório de dados. Este modelo estabelece uma relação bilateral em seu fluxo de trabalho, entre a revista e um repositório de dados.

O modelo de publicação *overlay* prevê uma publicação de dados específica, ou seja, um artigo de dados em periódico de dados cuja proposta é publicar, via descrição por metadados, o conjunto de dados produzidos por uma pesquisa. De acordo com Lawrence (2011)

> o conteúdo principal seria um documento de descrição de dados, que adiciona todo o conteúdo que pode estar faltando no arquivo primário: por exemplo, afirmações adicionais sobre a importância do conjunto de dados e sua proveniência. (LAWRENCE , 2011, p.22)

Assim como no modelo *journal driven data archival,* o fluxo de trabalho para a publicação do tipo *overlay* também inclui revisão por pares, tanto para o artigo quanto para o conjunto de dados. A curadoria do "pacote de dados" presente no repositório também é avaliada, e os metadados criados pelo curador também são passíveis de ser modificados em resposta à revisão, uma vez que há repositórios que se limitam a metadados básicos para descrever o conjunto de dados.

Autores como Dillo; Hodson; Sierman e demais colaboradores (2014) seguiram a mesma trilha de Lawrence et al. (2011) por entenderam o periódico de dados também sob duas perspectivas: a publicação *stand-alone* do conjunto de dados (em repositórios) e a publicação de dados em um periódico de dados.

> No modelo de publicação *standalone* de dados, os dados são publicados como um conjunto de dados. Em outras palavras, os dados são uma publicação por si só, sem a necessidade de uma coexistência artigo de revista padrão descrevendo os dados. O arquivo de dados fornece sistemas, que

> fornecem um documento de descrição de dados como o item citável e os dados são obtidos eletronicamente.
> O modelo de publicação “diário de dados" é baseado em "periódico de dados". Publicação especializada na publicação de "documentos de dados". Um documento de dados é uma publicação de periódico cujo objetivo principal é descrever dados (fornecer informações sobre o quê, onde, por que, como e quem dos dados), em vez de relatar uma investigação de pesquisa.[28] (DILLO; HODSON; SIERMAN et al., 2014, p.1, tradução nossa).

Esse olhar plural para o termo publicação de dados de pesquisa também foi apresentado por García-García, López-Borrull, Peset (2015), que interpretaram o termo sob três possibilidades:

> 1) Publicação como objeto de informação independente em um repositório de dados de pesquisa.
> 2) Publicação de dados de pesquisa na forma de artigo de dados em um periódico de dados.
> 3) Publicação dos dados da pesquisa juntamente com o artigo, em um formulário de publicação ampliada.[29] (GARCÍA-GARCÍA, LÓPEZ-BORRULL, PESET, 2015, p. 846, tradução nossa).

Ao considerarem a publicação de dados um dos principais pilares para uma ciência aberta, sustentável, confiável e moderna, Bloom e colaboradores (2015) contribuíram com o estudo ao entenderem que termo “publicação de dados” aplica-se tanto para repositório quanto para periódicos, e, dessa forma, fomenta entendimentos conceituais diversos que envolvem o termo e seus correlatos. De acordo com os autores:

> Os principais processos da publicação em periódicos de dados consistem na revisão por pares científicos e divulgação dos conjuntos de dados. Naturalmente, para a equipe de revisão, isso geralmente inclui acesso a pré-publicação ao conjunto de dados, que estaria disponível em algum repositório de dados, e também exige soluções explícitas de controle de versão para conjuntos de dados e documentos de dados. Os passos finais

[28] In the “stand-alone publication” model data are published as a stand alone dataset. In other words, the data are a publication in their own right, with no requirement for a coexisting standard journal article describing the data. The data archive provides systems, which provide a data description document as the citable item, and the data is obtainable electronically. The “data journal publication” model is based on “data journals” specialized in publishing “data papers”. A data paper is a journal publication whose primary purpose is to describe data (providing information on the what, where, why, how and who of the data), rather than to report a research investigation

[29] 1) Publicación como objeto de información independiente en un repositorio de datos de investigación; 2) Publicación de datos de investigación en forma de data paper en un data jornal; 3) Publicación de datos de investigación junto al artículo, en la forma de enriched / enhanced publication.

> para a publicação no periódico de dados estão alinhados com o processo tradicional de publicação[30]. (BLOOM et al., 2015, p.11, tradução nossa)
>
> Os periódicos de dados geralmente dependem de repositórios de dados externos para lidar com a gestão de dados. Isso requer uma forte colaboração entre as equipes do periódico e do repositório, além de confiar que o repositório buscará gerenciamento e ingestão de dados de acordo com procedimentos padrão aceitáveis. Revistas e repositórios de dados são incentivados a fazer tais acordos (por exemplo, acordos de nível de serviço) públicos e transparentes para os usuários.[31] (BLOOM et al., 2015, p.14, tradução nossa).

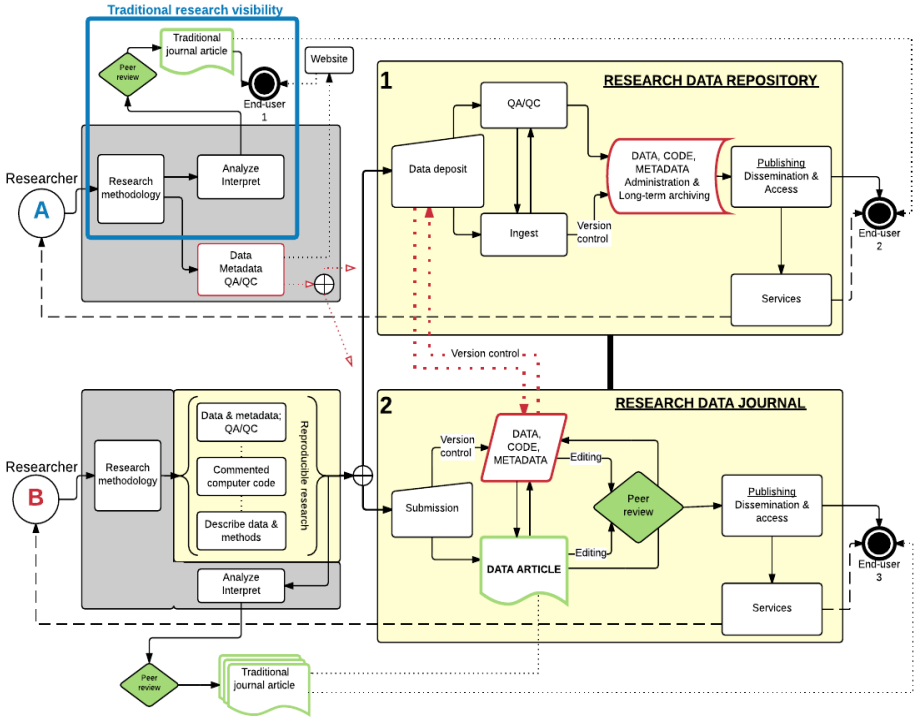
A diferenciação conceitual estabelecida pelos autores corrobora a reflexão conceitual proposta pela tese, na medida em que sustenta que a publicação de dados e o depósito de dados configuram-se em práticas distintas, porém, interligadas por um elo fundamental (o conjunto de dados), o que amplia as possibilidades de transparência e abertura da ciência.

O esquema de Bloom e colaboradores (2015), representado na Figura 12, demonstra a diferença entre a publicação dos dados pelo viés do repositório de dados e a publicação de dados pelo viés do periódico científico:

[30] Data journal publishing's core processes consist of scientific peer review and dissemination of the datasets. Naturally, for the review team, this usually comprises pre-publication access to the dataset, which would be available in some data repository, and also demands explicit version control solutions for datasets and data papers. The final steps of data journal publishing are aligned with the traditional publishing process.

[31] The journal workflows we examined typically straddled the dual processes of reviewing the dataset itself and the data papers, which were carried out separately and then checked to ensure that the relationship between the two was valid. Such QA/QC workflows for data journals demand a strong collaboration with the research community and their peer reviewers.

Figura 12– Representação do fluxo de trabalho para comunicação científica de dados de pesquisa



Fonte: (BLOOM, 2015)

O fluxo de trabalho esquematizado na Figura 12 expõe a diferença entre atores e processos envolvidos, além das especificidades que cada uma das etapas de publicação de dados exige da comunidade científica.

Exemplificando com dois pesquisadores (A e B) e três tipos de usuários finais, os autores analisaram três possíveis modos de acesso à informação sobre o conjunto de dados produzidos em uma pesquisa, bem como seus níveis de acesso à informação científica por estes produzidos. Ao optar somente pela publicação tradicional, o pesquisador A limita o acesso do usuário final 1 somente ao artigo tradicional derivado da pesquisa. Já o pesquisador B, ao depositar os dados de sua investigação em um repositório de dados de pesquisa, dá a possibilidade ao usuário final 2 de acessar tanto o artigo tradicional quanto o conjunto de dados que possibilitou a reflexão e publicação deste. O mesmo pesquisador B pode, ainda, ampliar o alcance de sua pesquisa por meio da descrição detalhada das condições sob as quais o conjunto de dados foi gerado, via artigo de dados.

Desse modo, foi possível observar que a terminologia utilizada na literatura para “publicação de dados” **em repositórios** se refere ao depósito (*upload*) do

conjunto de dados, o que não garante, necessariamente, a publicidade, comunicação, citação e provável reutilização dos mesmos. Já a “publicação de dados” **em periódico**, cuja ação envolve o fluxo de trabalho de um periódico científico tradicional, nos parece garantir a representação e descrição das características do conjunto de dados via metadados, revisão por pares, além da indexação para recuperação e citação adequada.

No entanto, tal entendimento não é consenso entre estudiosos do assunto e pôde ser verificado a partir das seguintes contribuições identificadas na literatura:

5.1 ABORDAGEM BASEADA EM REPOSITÓRIO

Na abordagem baseada em repositório, a publicação de dados refere-se ao depósito do conjunto de dados em um repositório, garantindo sua curadoria, preservação e potencial para compartilhamento e reutilização em novas pesquisas. De acordo com essa abordagem, é no repositório de dados que ocorre a publicação, e não no periódico. Encontramos na literatura alguns autores que corroboram essa visão tecnicista da publicação de dados, conforme veremos a seguir:

Callaghan e colaboradores (2013) contribuem para essa interpretação de publicação de dados de maneira mais consistente, ao argumentarem que:

> A publicação formal de dados fornece um serviço além do simples ato de postar um conjunto de dados em um site, na medida em que inclui uma série de verificações técnicas no conjunto de dados (formato, metadados) ou de natureza mais científica. A publicação formal de dados também fornece ao usuário de dados certas garantias sobre a persistência dos dados e fornece um fórum para que o conjunto de dados seja encontrado e avaliado - uma parte essencial do processo científico.[32] (CALLAGHAN et al., 2013, p. 194, tradução nossa).

Em artigo que explora o desenvolvimento de competências para a publicação de dados de pesquisa, Stoker, Melzack e McLean (2019, p.175) se referem à

[32] Formal publication of data provides a service over and above the simple act of posting a dataset on a website, in that it includes a series of checks on the dataset of either a technical (format, metadata) or a more scientific (is the data scientifically meaningful?) nature. Formal data publication also provides the data user with certain assurances about data persistence, and provides a forum for the dataset to be found and evaluated – an essential part of the scientific process

publicação de dados “como o processo de disponibilizar um conjunto de dados por meio de um repositório”. Esse artigo levou-nos até a equipe de Curadoria de Dados Digitais da Biblioteca da Universidade de Sidney, na Austrália, onde a equipe desenvolveu um programa de apoio a pesquisadores para publicação de dados de pesquisa. Ao descrever o serviço em uma das páginas da biblioteca, a equipe considera que “a melhor maneira de disponibilizar seus dados de pesquisa é publicando-os em um repositório de dados”, ou seja, a publicação de dados pelo viés do depósito em repositório, e não pela publicação em revista. Os integrantes da equipe complementam essa orientação em seu “Guia de publicação de dados” ao destacarem os benefícios de uma publicação de dados:

> A publicação de seus dados e conjuntos de dados de pesquisa em um repositório de dados pode permitir que seus dados sejam reutilizados e citados em outros trabalhos. Alguns repositórios podem até fornecer ferramentas para rastrear como as pessoas usam seus dados com a altmetria. Os conjuntos de dados podem ser publicados por si mesmos, sem um trabalho de pesquisa associado, o que permite obter reconhecimento por trabalhos que podem não se qualificar para autoria em um periódico de pesquisa tradicional. [33](UNIVERSITY OF SIDNEY, *online*, tradução nossa)

Observa-se claramente, por meio desta definição, a proposta de depósito ou armazenamento do conjunto de dados, quando o autor se refere à preservação dos dados a longo prazo.

5.2 ABORDAGEM BASEADA EM PERIÓDICO

Na abordagem baseada em periódico, ocorre, por vezes, uma variação terminológica entre “documento de dados” e “periódico de dados”. A exploração da literatura permitiu-nos observar que o documento de dados pode referir-se tanto a um artigo de dados quanto a um periódico de dados. Nossa interpretação leva-nos a crer que a atribuição do termo “documento de dados” se aplica ao conceito proposto para “artigo de dados”, tanto por suas definições quanto por sua terminologia, por ser um

[33] Publishing your research data and datasets in a data repository can allow for your data to be reused and cited in other works. Some repositories may even provide tools to track how people use your data, such as Altmetrics. Datasets can be published in their own right, without an associated research paper, which can allow you to gain recognition for work that may not qualify for authorship in a traditional research journal.

documento, algo que descreve um estado, condição e/ou características de um conjunto de dados.

Costas e colaboradores (2013) definiram um periódico de dados como

> uma publicação que reúne artigos de dados especializados, cuja finalidade principal é descrever dados (fornecendo informações sobre o quê, onde, por que, como e que tipo de dados), em vez de relatar uma investigação de pesquisa. Como tal, contém fatos sobre dados, não hipóteses e argumentos em apoio a essas hipóteses baseadas em dados, como encontrado em um artigo de pesquisa convencional. Seus objetivos são três: fornecer uma publicação de periódicos citáveis que leve crédito acadêmico a editores de dados; descrever os dados de forma estruturada e legível por humanos; e trazer a existência dos dados à atenção da comunidade acadêmica.[34] (COSTAS et al., 2013, p.13).

Hrynaszkiewicz e Shintani (2014) compartilharam seu entendimento sobre o periódico de dados ao descreverem características e benefícios deste tipo de publicação. De acordo com os autores, o periódico de dados:

> possibilita melhor descoberta de dados, através da indexação de artigos em base de dados (por exemplo, Scopus, PubMed); permite crédito via artigo de revista; e maior confiabilidade dos dados, graças a revisão por pares. O periódico de dados também pode fornecer recomendações sobre formatação e deposição de dados. O artigo de dados permite publicação de informações contextuais e metodológicas detalhadas, permitindo validação independente e reprodução de dados. Os periódicos de dados também podem fornecer uma saída para conjuntos de dados e experiências não publicados ou sem prévia publicação. E, ao fazer parceria com repositórios, os periódicos de dados também podem melhorar a acessibilidade e a integração de dados com artigos de periódicos e aprimorar a experiência dos leitores (HRYNASZKIEWICZ, SHINTANI, 2014, p. 4, tradução nossa).

Apesar de não contribuírem autoralmente com o conceito de periódico de dados de pesquisa em artigo publicado sobre consensos e controvérsias sobre a publicação de dados, Kratz e Strasser (2014) apresentam características que norteiam este tipo de publicação e a conduzem à ascensão à “primeira classe” dos produtos de pesquisa, no contexto da comunicação científica. De acordo com os autores

[34] The “data journal publication” model is based on “data journals” specialized in publishing “data papers”. A data paper is a journal publication whose primary purpose is to describe data (providing information on the what, where, why, how and who of the data), rather than to report a research investigation. As such, it contains facts about data, not hypotheses and arguments in support of those hypotheses based on data, as found in a conventional research article. Its purposes are threefold: to provide a citable journal publication that brings scholarly credit to data publishers; to describe the data in a structured human-readable form; and to bring the existence of the data to the attention of the scholarly community

> A força de um documento de dados está no fornecimento de documentação rica, o que é especialmente útil para dados de pesquisa únicos e heterogêneos de “cauda longa” da ciência. O comprimento e a estrutura do papel de dados variam entre os periódicos, mas a tendência é para um formato curto e bem estruturado. Todas as revistas requerem um resumo, métodos de coleta e uma descrição do conjunto de dados; alguns incentivam os autores a sugerir possíveis usos para os dados (por exemplo, *Internet Archaeology, e Open Health Data*) [...] Os documentos de dados são definidos com maior nitidez, não pela presença de qualquer informação em particular, mas pela ausência de análises ou conclusões.[35] (KRATZ; STRASSER, 2014, p.5, tradução nossa).

O termo concebido por Candela e colaboradores (2015) generaliza um **documento de dados** como objeto, que pode fornecer ao consumidor de dados os detalhes que facilitam o entendimento e possível reutilização dos dados.

Autores como García-García, López-Borrull e Peset (2015) destacam alguma das vantagens da publicação de dados sob o viés do periódico científico e afirmam que este tipo de publicação torna a pesquisa científica mais confiável. De acordo com os autores

> as revistas de dados se concentram em dados abertos de pesquisa, o que que incentiva a reutilização, uma vez que descrevem a metodologia de coleta e análise de dados – com profundidade semelhante à exigida nas patentes. Em muitas revistas convencionais, a descrição da coleção de dados é reduzida a etapa de resultados e sua discussão. Os artigos de dados trazem à luz conjuntos de dados que por várias razões não foram publicadas em artigos pesquisa.[36] (GARCÍA-GARCÍA, LÓPEZ-BORRULL, PESET, 2015, p. 852).

Reforçam o entendimento do cenário proposto para a publicação de dados de pesquisa Assante e colaboradores (2016) ao destacarem três principais atores necessariamente envolvidos na produção de uma publicação de dados de pesquisa: autores (pesquisadores), editores e gestores de repositório de dados de pesquisa.

---

[35] Data paper length and structure varies between journals, but the tendency is toward a short, tightly structured format. All journals require an abstract, collection methods, and a description of the dataset; a few encourage authors to suggest potential uses for the data (e.g., Internet Archaeology, and Open Health Data). Some journals supplement this general framework with field-specific sections. (e.g., Internet Archaeology and the Journal of Open Archaeology Data each include a section for temporal and geographic scope). Data papers are most sharply defined not by the presence of any particular information, but by the absence of analysis or conclusions.

[36] Las revistas de datos es que se centran en los datos de investigación abiertos, lo cual fomenta la reutilización ya que con ellos se describe la metodología de obtención y análisis de los datos –con una profundidad similar a la que se exige en las patentes-. En muchos artículos convencionales, la descripción de la reco-gida de datos se reduce para ceder el protagonismo a los resultados y su discusión. Los artículos de datos hacen salir a la luz conjuntos de datos que por distintos motivos no fueron publicados en artícu-los de investigación

Destacam o repositório como uma ferramenta de apoio à publicação de dados ao afirmarem que: “Os repositórios de dados científicos são frequentemente propostos como instrumentos para apoiar a publicação de dados, pois fornecem instalações para todos os diferentes atores envolvidos nesse processo” (ASSANTE, 2016, p.2, tradução nossa).

Embora compreendam os diferentes papéis assumidos, tanto pelo periódico quanto pelo repositório, os autores utilizam o termo publicação de dados ambiguamente ao se referirem a fluxos de trabalho distintos no contexto da comunicação científica. Destacam a importância da descrição e documentação do conjunto de dados por meio de uma publicação de dados ao afirmarem que:

> O objetivo final da publicação de dados é disponibilizar conjuntos de dados para validação e reutilização, tanto pela comunidade científica que a produziu e, mais amplamente, em outras comunidades. Para atingir esse objetivo, um conjunto de dados produzido por uma comunidade de prática deve receber informações auxiliares e contextuais sobre o conjunto de dados, tais como os metadados descritivos, e também, como foram obtidos, ou seja, informações sobre sua proveniência. Esta documentação, em geral influencia profundamente a descoberta, a compreensibilidade, a verificação, a citação e a reutilização prática do conjunto de dados.[37] (ASSANTE, 2016, p.8, tradução nossa)

Além disso, os autores supracitados definem um documento de dados a partir de suas características: “um documento de dados contém informações sobre os conjuntos de dados e como foram produzidos, facilitando sua melhor compreensão.” (ASSANTE, 2016, p. 18).

Outra definição relevante para a pesquisa foi a proposta por Pauline Ward (2016), assistente de serviço de dados de pesquisa na Universidade de Edimburgo. A autora descreve um periódico de dados como

> um novo formato de publicação que se concentra no conjunto de dados e não no artigo de pesquisa convencional. Estes são vistos como um passo significativo na alteração das comunicações acadêmicas, para que os criadores de dados obtenham crédito por seu trabalho. Para que isso

[37] The ultimate goal of data publishing is to make datasets available for validation and reuse both within the scientific community that has produced it and, more widely, within other communities. This documentation at large, deeply influences dataset discoverability, understandability, verification and practical re-use.

> aconteça, um conjunto de dados deve ser citável e revisado por pares, como as publicações tradicionais.[38] (WARD, 2016, *on-line*, tradução nossa)

Esta visão é importante porque, na comunicação científica, a natureza de cada área defini seu comportamento quanto à forma de produzir, disseminar e comunicar a ciência, não existindo um padrão geral.

Em artigo publicado *no International Journal on Digital Libraries*, Austin e colaboradores (2017) contribuíram para o estudo a partir das seguintes definições:

> Um **periódico de dados** é um diário (invariavelmente acesso aberto) que publica artigos de dados. O periódico de dados geralmente fornece modelos para a descrição dos dados e oferece aos pesquisadores orientações sobre onde depositar e como descrever e apresentar seus dados. Dependendo do periódico, esses modelos podem ser genéricos ou focados na disciplina. Alguns periódicos ou seus editores mantêm seus próprios repositórios. Além de apoiar a vinculação bidirecional entre um artigo de dados e o (s) seu (s) conjunto (s) de dados correspondente (s) e facilitar práticas de identificação persistentes, os periódicos de dados fornecem fluxos de trabalho para garantir a qualidade (ou seja, revisão por pares de dados).[39] (AUSTIN et al., 2017, p. 82, tradução nossa).

O compartilhamento de dados como condição de financiamento de pesquisa foi apresentado por Eynden e Corti em artigo publicado em 2017. Destaca-se a colaboração de tais autores para o presente estudo por meio de sua investigação sobre a promoção das práticas de publicação de dados de pesquisa no domínio das ciências sociais. Embora utilizem o termo “publicação de dados” tanto para o depósito do conjunto de dados quanto para a publicação em periódico, tais autores se referem ao periódico de dados como um “diário de dados” e o descrevem da seguinte maneira:

> Cada vez mais conjuntos de dados de ciências sociais também são publicados em diários de dados, como *Scientific Data, Research Data Journal for the Humanities and Social Sciences, ou Journal of open psychology data*, nos quais um documento descreve a metodologia de geração dos dados, proveniência e potencial de reutilização de um conjunto de dados

[38] Data journals are a new format of publication that focuses on the dataset rather than the conventional research article. They are seen as a significant step in changing scholarly communications so that data creators get credit for their work. For this to happen, a dataset must be both citable and peer reviewed, like traditional publications.

[39] A data journal is a journal (invariably open access) that publishes data articles. The data journal usually provides templates for data description and offers researchers guidance on where to deposit and how to describe and present their data. Depending on the journal, such templates can be generic or discipline focused. Some journals or their publishers maintain their own repositories. As well as supporting bi-directional linking between a data article and its corresponding dataset(s), and facilitating persistente identification practices, data journals provide work-flows for quality assurance(i.e. data peer review) and should also provide editorial guidelines on data quality assessment.

> apresentado em um repositório em detalhes.[40] (EYNDEN; CORTI, 2017, p.116, tradução nossa).

Abrimos um parêntese neste ponto da pesquisa para ressaltar a diferença das áreas e de seus comportamentos no que tange a comunicação científica. Com base nas leituras realizadas ao longo da pesquisa, acreditamos que a comunicação cientifica via publicação dos dados produzidos em pesquisas se apresentará de diversas maneiras, podendo ser "nomeada" de diversas formas, contudo, estará sempre em sintonia com o comportamento e características diversificadas de cada área do conhecimento. Tal comportamento demonstra a juventude da prática e de sua absorção nas áreas e por atores que compõem o cenário da comunicação científica, oportunizando estudos de áreas de acordo com seus padrões de informação e dados.

Retomamos a reflexão sobre a publicação de dados de pesquisa na abordagem baseada em periódico ao trazer a proposta Hrynaszkiewicz e Shintani (2014). Tais pesquisadores veem no periódico de dados um veículo legitimado pela ciência para comunicar o conjunto de dados de uma pesquisa, além de ser um instrumento para fomentar a reprodutibilidade científica e a confiabilidade nos dados abertos. Para os autores, os dados devem estar disponíveis e compreensíveis aos pares. Segundo eles:

> Os periódicos de dados ajudam a incentivar os cientistas a publicar dados e ajudá-los a atender aos requisitos de agências de fomento e instituições de pesquisa, fornecendo uma publicação citável vinculada ou publicada ao lado de seus conjuntos de dados. [41] (HRYNASZKIEWICZ; SHINTANI, 2014, p. 3, tradução nossa).

Em artigo recente publicado no periódico *Journal of the Association for Information Science and Technology*, Li, Greenberg e Dunic (2020) consideram o documento de dados um gênero científico emergente e o definem como:

> publicação acadêmica de metadados pesquisáveis, documento que descreve um conjunto de dados acessível on-line, ou um grupo de conjuntos de dados, publicado de acordo com a norma práticas acadêmicas. (LI, GREENBERG e DUNIC, 2020, p. 172)

---

[40] Increasingly social sciences data sets are also published indata journals, such as Scientific Data, Research Data JournalfortheHumanitiesandSocialSciences,ortheJournalofopenpsychology data, whereby a data paper describes the datageneration methodology, provenance, and reuse potential fora data set lodged in a repository in detail.

[41] Data journals help incentivize scientists to publish data, and help them meet the requirements of funding agencies and research institutions, by providing a citable publication linked to or published alongside their dataset(s).

De todas as definições encontradas até o momento, a que mais se aproxima do nosso entendimento sobre o que é um periódico de foi a adotada pelo *Australian National Data Service* (ANDS)

> As revistas de dados são publicações cujo objetivo principal é expor conjuntos de dados. Estas permitem que o autor se concentre nos dados em si, em vez de produzir uma análise extensiva dos dados como ocorre no modelo de periódico tradicional. Fundamentalmente, os periódicos de dados buscam: promover acreditação científica e reutilização, melhorar a transparência do método científico e dos resultados, apoiar boas práticas de gerenciamento de dados e fornecer uma rota acessível, permanente e solucionável para o conjunto de dados. (AUSTRALIAN NATIONAL DATA SERVICE, *on-line,* tradução nossa).

Para impulsionar ainda mais a flutuação terminológica[42] referente ao objeto investigado, acrescentamos a este panorama o conceito de artigo de dados para complementar o entendimento sobre o objeto investigado. De maneira sucinta, Newman e Corke (2009) consideram um item de dados como um novo tipo de comunicação científica. Corroboram o entendimento anterior e complementam a definição de autores como Chavan e Penev (2011) ao afirmarem que um

> documento de dados é uma publicação de periódico cujo objetivo principal é descrever dados, em vez de relatar uma investigação de pesquisa. Como tal, contém fatos sobre dados, não hipóteses e argumentos que apoiam essas hipóteses baseadas em dados, conforme encontrado em um artigo de pesquisa convencional. Seus objetivos são triplos: fornecer uma publicação de periódico citável que ofereça crédito acadêmico aos editores de dados; descrever os dados de forma estruturada e legível por humanos; e levar a existência dos dados à atenção da comunidade acadêmica.[43] (CHAVAN; PENEV, 2011, p.3, tradução nossa).

Outra variação terminológica encontrada na literatura que envolve a publicação de dados foi identificada em artigo já citado no decorrer desta pesquisa de Hrynaszkiewicz e Shintani (2014), quando discutem a importância do periódico de dados como um novo tipo de publicação para ajudar a resolver problemas de reprodutibilidade e confiabilidade científica dos dados de pesquisa. De acordo com os

[42] Termo utilizado a partir do que Medows (1999) chamou de “flutuação de significado das palavras".

[43] A data paper is a journal publication whose primary purpose is to describe data, rather than to report a research investigation. As such, it contains facts about data, not hypotheses and arguments in support of those hypotheses based on data, as found in a conventional research article. Its purposes are threefold: to provide a citable journal publication that brings scholarly credit to data publishers; to describe the data in a structured human-readable form; and to bring the existence of the data to the attention of the scholarly Community.

autores, um artigo de dados também é denominado de “descritores de dados”, e como tal:

> Incluem descrições detalhadas dos métodos usados para coletar os dados e descrever análises técnicas que suportam a qualidade das medições. Os descritores de dados não contêm testes de novas hipóteses científicas, análises extensivas destinadas a fornecer novos conhecimentos científicos ou descrições de novos métodos científicos.[44] (HRYNASZKIEWICZ, SHINTANI, 2014, p. 8, tradução nossa)

O panorama terminológico que vem sendo apresentado ao longo deste capítulo para referir-se à “publicação de dados de pesquisa” encontrou autores como Austin e colaboradores (2017), que propõem um conceito abrangente, que envolve incentivos para os pesquisadores compartilharem seus dados, fornecendo resultados citáveis, favorecendo para uma mudança de paradigma emergente. Tais autores compreendem o artigo de dados como um

> produto de publicação de dados, que pode aparecer em um periódico de dados ou em qualquer outro periódico. Quando os editores se referem à "publicação de dados", geralmente eles aditam os dados comuns ao conjunto de dados subjacente. Os artigos de dados concentram-se em tornar os dados detectáveis, interpretáveis e reutilizáveis, em vez de testar hipóteses ou apresentar novas interpretações (em contraste com os artigos tradicionais de periódicos). O comprimento desses artigos pode variar de micropapers (focados em uma tabela ou gráfico) a uma apresentação detalhada de conjuntos de dados complexos.[45] (AUSTIN et al., 2017, p. 82, tradução nossa).

No Brasil, embora sejam poucos os autores que abordem a temática, apresentamos a seguir a contribuição de pesquisadores da Universidade Estadual Paulista Júlio de Mesquita Filho - UNESP sobre o assunto. Em recente pesquisa para

---

[44] Data Descriptors include detailed descriptions of the methods used to collect the data and describe technical analyses that support the quality of the measurements. Data descriptors do not contain tests of new scientific hypotheses, extensive analyses aimed at providing new scientific insights, or descriptions of fundamentally new scientific methods.

[45] A data article is a ‘data-publishing’ product, also known as a ‘data descriptor’ that may appear in a data journal or any other journal. When publishers refer to ‘data publishing’, they usually mean a data article rather than the underlying dataset. Data articles focus on making data discoverable, interpretable, and reusable, rather than testing hypotheses or presenting new interpretations (by contrast with traditional journal articles). Whether linked to a dataset in a separate repository, or submitted in tandem with the data, the aim of the data article is to provide a formal route to data sharing. The parent journal may choose whether or how standards of curation, formatting, availability, persistence, or peer review of the dataset are described. By definition, the data article provides a vehicle to describe these qualities, as well as some incentive to do so. The length of such articles can vary from micro papers (focused on one table or plot) to very detailed presentation of complex datasets.

a proposição de uma estrutura comum para o artigo de dados de pesquisa, Roa-Martínez, Vidotti e Santana (2017) definem um artigo de dados como

> um novo tipo de comunicação científica que faz parte das classes de documentação resultantes do processo Publicação de dados. Esse tipo de artigo descreve um conjunto de dados de pesquisa acessível por outros pesquisadores junto com informações próprio, ou seja, com metadados relacionados ao processo a partir do qual os dados foram fornecidos os métodos e considerações de coleta, formatos, condições específicas e autores entre outras (ROA-MARTÍNEZ; VIDOTTI e SANTANA, 2017, p. 5).

Curty e Aventurier (2017) discutem novos formatos de publicação que estão em ascensão no contexto da Ciência Aberta e que vem gradativamente conquistando seus espaços entre os membros da comunidade científica. Para os autores,

> os artigos de dados elevam os dados científicos à condição de protagonistas, pois se dedicam a descrever exaustivamente a coleção de dados, acompanhados de descrições do contexto, do percurso metodológico e dos aspectos procedimentais da pesquisa, e das possíveis aplicações dos dados. Essa abordagem pode se materializar por meio da publicação dos artigos de dados em periódicos científicos híbridos, receptivos aos data papers, ou periódicos dedicados à publicação de dados. Em geral os artigos de dados são submetidos ao escrutínio dos pares, validando os dados e tornando-os mais transparentes e críveis perante a comunidade científica. Os periódicos de dados também trazem à superfície para comunidades de interesse, coleções de dados com maior potencial de reuso, tendo em vista a garantia de maior detalhamento documentação, e melhor endossamento. Por preservar características dos periódicos de dados e de cristalizada aceitação e de ampla aquiescência entre os membros da comunidade científica, como peer-review, periodicidade e corpo editorial, esta abordagem tem maior potencial de receptividade no meio científico. (CURTY; AVENTURIER, 2017, p. 18).

Em trabalho apresentado no *International Conference on Grey Literature*, Farace, Frantzen e Smith (2017, p.27) discutem o documento de dados como um novo tipo de literatura cinza, além de entendê-lo como um dos meios mais tangíveis de implementação dos princípios de dados FAIR.

Ao proporem uma discussão onde a publicação de dados é vista sob a ótica da organização do conhecimento, Schöpfel, Farace, Prost e Zane (2019) concebem o documento de dados da seguinte maneira:

> Ao contrário dos documentos de pesquisa comuns, o principal objetivo dos documentos de dados é descrever conjuntos de dados, incluindo as

> condições e o contexto de sua aquisição e sua utilidade potencial, em vez de relatar e discutir resultados. Além disso, é geralmente assumido que papéis de dados são papéis curtos com até 4 páginas.[46] (SCHÖPFEL; FARACE; PROST; ZANE, 2019, p.3).

Com base nas contribuições dos autores que compõem a discussão e que não esgotam a investigação sobre o tema, e, ainda, ancorados na trajetória histórica e questões epistemológicas quem envolvem a prática da publicação de dados e sua sustentação terminológica no campo científico, propomos, a seguir, definições para os termos: **Periódico de dados de pesquisa, Artigo de dados, Depósito de dados de pesquisa e Publicação de dados de pesquisa**. Tais proposições terminológicas e conceituais buscam auxiliar o entendimento e o papel desses objetos e práticas no contexto da infraestrutura informacional da comunicação científica, gestão e compartilhamento de dados de pesquisa.

Periódico de dados de pesquisa – publicação periódica científica revisada por pares que segue os trâmites tradicionais do *workflow* da comunicação científica. Reúne artigos de dados que contêm a descrição do conjunto de dados de uma pesquisa.

Artigo de dados – documento cujo objetivo é descrever de forma detalhada e por meio de metadados as condições técnicas (procedimentos), tecnológicas, metodológicas e contextuais em que o conjunto de dados foi produzido, e, assim, possibilitar uma interpretação adequada para fins de compartilhamento e reuso em novas pesquisas. Ao serem publicados em um periódico científico revisado por pares, os artigos de dados ganham visibilidade, citabilidade e credibilidade científica.

Depósito de dados de pesquisa – processo de armazenamento ou depósito de um conjunto de dados em repositórios de dados, podendo estar ou não em acesso aberto.

Publicação de dados de pesquisa – ato de comunicar e dar publicidade aos dados produzidos por uma pesquisa, quer seja por meio de periódico científico, quer seja por meio de repositório de dados ou outro canal de comunicação científica.

[46] Unlike usual research papers, the main purpose of data papers is to describe datasets, including the conditions and context of their acquisition and their potential utility, rather than to report and discuss results. Also, it is generally assumed that data papers are short papers with up to 4 pages.

Refere-se a representação dos dados e não necessariamente aos dados propriamente ditos.

Tais conceitos buscam compreensão desse universo complexo, por vezes novo, e quase sempre ambíguo, que envolve os dados de pesquisa e sua interoperabilidade semântica e tecnológica.

A reflexão proposta por este capítulo não encerra a possibilidade de discussão sobre a prática investigada, contudo, amparou a elaboração do próximo capítulo, onde discutiremos a prática da publicação de dados de pesquisa à luz da Teoria Ator Rede - TAR de Bruno Latour e a mobilização de mundo por esta proposta.

6 A TEORIA ATOR REDE (TAR) NO CONTEXTO DA PUBLICAÇÃO DE DADOS DE PESQUISA: UMA POSSIBILIDADE DE MOBILIZAÇÃO DE MUNDO

O referencial teórico apresentado pela pesquisa considerou principalmente a característica volátil da prática da publicação dos conjuntos de dados de pesquisa associada à transparência e comunicabilidade proporcionada via periódico ou via repositório, bem como sua possibilidade de compartilhamento e colaboração entre pares.

A reflexão proposta buscou respaldo teórico na ideia de mobilização de mundo presente na Teoria Ator Rede (TAR) de Bruno Latour e sua aplicação no contexto da comunicação científica.

De acordo com Latour (2000) a mobilização de mundo compreende o deslocamento ou "mobilização" de uma "coleção", representada na pesquisa por conjunto de dados, para realidades e ambientes distintos aos quais deram origem aquela coleção. Segundo o autor:

> Estes podem ser retirados do contexto em que estão e levados embora durante as expedições. Portanto, a história da ciência é em grande parte a história da mobilização de qualquer coisa que possa ser levada a mover-se... (LATOUR, 2000, p. 364)

Acredita-se que, por meio da descrição detalhada dos dados de pesquisa e comunicada por meio de *data papers* ou *datasets*, o compartilhamento e reuso do conjunto de dados possam ser potencializados. A observação da ciência em ação proposta por Latour (2000) nos permite, aqui também, acompanhar um modelo de comunicação científica emergente em ação.

Consideramos a prática da publicação de dados de pesquisa uma janela oportuna para a observação da ciência em transformação. A ideia de ciência em transformação é bem ilustrada pela figura tanto do periódico de dados quanto do repositório de dados, uma vez que sua chegada recente ao campo da comunicação científica e, mais especificamente na área da Enfermagem, nos coloca diante de oportunidades e desafios inerentes a este novo modo de descrever e comunicar a ciência. Enquanto está estabelecendo-se, firmando-se no campo da comunicação

científica, necessita de experimentação por parte de seus principais colaboradores (autores, pesquisadores, editores, revisores, bibliotecários, entre outros), num movimento constante entre os laboratórios (que Latour chama de centros de cálculo) e a periferia, que podemos considerar nesta pesquisa como o acesso à descrição dos dados por seus potenciais reutilizadores.

A discussão proposta a partir da Teoria Ator Rede considerou as relações entre ciência, tecnologia e sociedade e os respectivos atores envolvidos nesse processo. Classificou como actantes humanos os pesquisadores e demais agentes envolvidos no cenário da investigação científica e que são capazes de descrever detalhadamente as características e aspectos que compõem um conjunto de dados. Os actantes não-humanos são caracterizados pelos periódicos, repositórios de dados e infraestruturas tecnológicas que dão suporte a este formato de comunicação científica.

De acordo com o ecossistema de comunicação científica, nos baseamos na ideia de que o conteúdo presente em tais conjuntos ou pacotes de dados devam ser trabalhados como em laboratórios, onde a informação vinda do centro de cálculo em formato de dados (produzida por actantes humanos ou não-humanos) ganha forma, valor e transparência para possibilitar seu reuso por outros pesquisadores. Nesse ecossistema, os reutilizadores são considerados em suas periferias (outros laboratórios) e estabelecem trocas (colaboração) entre si ao fazerem uso distinto do inicial e, assim, promover novas abordagens de investigação para o campo científico.

A reflexão sobre a pratica da publicação de dados, amparada pela TAR no circuito da comunicação científica e sob a égide da *e-science,* permitirá entendê-la como uma prática de visibilidade interdisciplinar, ampliando a malha da comunicação científica na área da Enfermagem e, consequentemente, as possibilidades incursivas da ciência, numa espécie de composição científica. A ideia de composição é representada por Latour ao trabalhá-la sob a ótica do composicionismo, onde as ciências trabalham de forma integrada e não mais separada em cantões disciplinares (DIAS, SZTUTMAN, MARRAS, 2014, p.500).

De acordo com Latour (2012), a TAR busca superar três obstáculos: **dispersão, destruição e desconstrução** das relações tecnocientíficas. Sendo assim, entendemos que, no que se refere ao obstáculo “dispersão”, consideramos que a prática da publicação de dados possibilita uma diminuição da dispersão dos dados

produzidos em pesquisas, promovendo o registro e a comunicação dos dados produzidos, que, por diversas vezes, ficaram ocultos e dispersos na cauda longa da ciência[47].

Para superarmos o obstáculo “destruição”, compreendemos a ausência de informações sobre o conjunto de dados como um fator de “perecimento” desses dados. A superação desse obstáculo é proposta por meio da representação dos dados via metadados, associados ao seu identificador persistente. Os repositórios de dados aparecem como espaços de preservação, onde os *datasets* estarão “a salvo”. Tais medidas prometem garantir “vida longa” ao conjunto de dados e possibilitar sua reutilização pelos pares.

A superação do obstáculo “desconstrução**”** pode ser entendida no âmbito do *datapapers* como uma espécie de abertura da “caixa preta”. Tal abertura requer uma mudança cultural na gestão dos dados de pesquisa, e, ao mesmo tempo, propõe uma desconstrução do “apego” aos dados. Ao descrever o conjunto de dados, o sentimento de posse dos dados é substituído pelo desejo de transparência, compartilhamento e reuso.

Desse modo, entendemos que a prática da publicação de dados de pesquisa contribui para eliminar os obstáculos elencados na TAR, uma vez que envolve a publicação via artigo e via repositório. Os autores Hrynaszkiewicz e Shintani (2014), citados anteriormente, reforçam o entendimento de que a publicação de dados surge como uma nova maneira de ajudar a resolver problemas de compreensão, confiabilidade e reutilização dos dados de pesquisa.

A transparência e acessibilidade promovidas pela publicação do conjunto de dados fomenta a reutilização dos dados produzidos em pesquisas e reafirma a “mobilização” de realidades ou mundos, conforme descrito por Latour, em seu livro Ciência em ação (2000).

O viés tecnosocial oportunizado pela prática traduz-se na capacidade de entendê-la como uma mudança cultural a partir dos elos, recursos e associações epistêmicas que pode proporcionar. Ao aumentar o número de associações

[47] Segundo Sales e Sayão (2018, p. 4183), o termo “cauda longa da ciência” refere-se a uma parcela considerável do trabalho científico que não está visível nem para a sociedade em termos de benefícios e qualidade de vida, nem para os pares no contexto da dinâmica de uma comunidade científica.

(colaborações em produções, colaboração em reuso, dentre outras), recruta novos atores humanos e não humanos a participarem do empreendimento científico, agregando e reagregando as relações estabelecidas no campo da ciência de um modo geral.

A prática da publicação de dados, na perspectiva da comunicação científica, é compreendida não somente como objeto investigado, mas também como um instrumento de laboratório. Segundo Latour (2000, p. 115), “um instrumento, seja qual for sua natureza, é o que nos leva do artigo ao que nos dá sustentação ao artigo”. Ou seja, nesse caso, consideramos a prática em questão como instrumento que irá proporcionar a descrição e a publicação dos dados. Além disso, irá indicar o acesso ao conjunto de dados de pesquisa e, por conseguinte, sustentar a pesquisa e possibilitar a composição de novos arranjos e produções científicos. Desse modo, a reflexão sobre a prática da publicação de dados pela área da Enfermagem pode ser interpretada a partir da mobilização de mundo e de atores (humanos e não-humanos) participantes do ecossistema da comunicação científica.

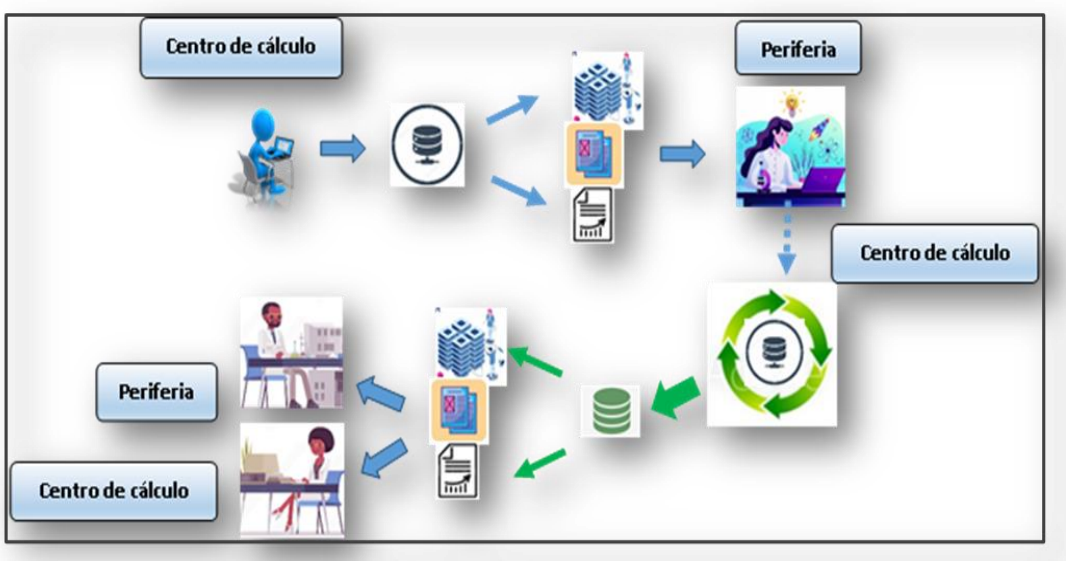
O *modus operandi* de uma publicação de dados por si só já demanda controvérsias para o campo científico, conforme pode ser observado através da revisão de literatura sobre o termo “publicação de dados de pesquisa”. A Ciência da Informação, na qualidade de território de observação desse processo, propõe uma leitura interdisciplinar da prática. A mobilização de mundo proposta fomenta o estabelecimento, quase que obrigatório, de relações interdisciplinares entre domínios tecnocientíficos. Uma das características mais marcantes da Ciência da Informação, a interdisciplinaridade, também colabora com a mobilização de mundo envolvidos e proporcionados no processo de comunicação científica.

A partir do entendimento de que “os dados não falam por si só”, é fundamental que sua descrição seja adequadamente elaborada, como se houvesse um porta-voz. Assim, podemos dizer que a prática da publicação de dados reúne os porta-vozes dos dados de uma pesquisa, ou seja, *datapapers* e *datasets*. Esse entendimento leva-nos à reflexão sobre uma prática que conduz a mobilização de mundos, e, como tal, agente de inovação, colaboração e desenvolvimento científico, uma vez que, ao descrever a informação contida no conjunto de dados, permite compreender as condições em que os dados foram produzidos, descortinando a janela do laboratório.

Enquanto objeto/instrumento de um laboratório, provoca reações no campo científico quando comunica e compartilha com a comunidade científica a descrição do conjunto de dados.

Os actantes humanos e não humanos que participam desse processo compõem o ecossistema da comunicação científica e desta se beneficiam, graças às trocas e colaborações possibilitadas por esse *habitat* científico, conforme ilustrado na Figura 13.

Figura 13 – A publicação de dados de pesquisa e a mobilização de mundo



Fonte: A autora, 2020

A materialização da mobilização de mundo proposta por Latour (2000) e vislumbrada por este estudo é demonstrada na Figura 14. Nela, a publicação via periódico ou repositório de dados de pesquisa é compreendida tanto como canal de comunicação científica, ao reunir e publicar artigos de dados, quanto como fonte de informação para se chegar ao conjunto de dados original. Essa via de mão dupla mobiliza saberes e mundos, em movimentos que podem ser ora disciplinares, ora interdisciplinares. Pinheiro (1999, p. 175) já destacava essa via de mão dupla, característica da Ciência da Informação na qualidade de campo de proposições de estudos: “A Ciência da Informação tem dupla raiz: de um lado a Documentação e, de outro, a Recuperação da Informação. Na primeira, o que importa é o registro do conhecimento científico [...] e, no segundo, as tecnologias de informação”.

A documentação dos dados (sua descrição) por aquele que produziu ou por aquele que programou/calibrou a máquina que produziu, bem como sua comunicação, oferece credibilidade e transparência ao processo científico, numa espécie de “bula de remédio”, onde os dados são devidamente descritos e contextualizados. Isso possibilita seu reuso em novas produções científicas, num constante movimento progressivo e colaborativo.

A representação gráfica proposta pela Figura 14 demonstra a identificação de mobilização de mundo percebida pelo estudo. A publicação dos dados permite que a periferia assuma o papel ora de consumidor de dados, ora de produtor, possibilitando uma dinâmica que agrega a comunidade científica e potencializa a capacidade de pesquisas e descobertas. A disseminação desta cultura no circuito da comunicação científica brasileira insere o país em uma tecnociência, entendida por Latour (2000) como uma rede onde os pontos fortes e fracos se aliam, como uma malha, cobrindo a comunicação científica nacional e internacional.

Tornar este instrumento de comunicação e descrição de dados um ponto de passagem no workflow da comunicação científica parece reduzir a distância entre pesquisadores e comunidade científica. De acordo com Ziman[48] (1979, p. 50): “a publicação dos resultados de uma experiência é algo que fica muito distante, na verdade, daquela batalha direta e extenuante que o pesquisador trava no início, individualmente, com a natureza bruta, em seu laboratório”. Essa alusão ao distanciamento afirmada há 41 anos, entre os dados produzidos em laboratório e a publicação dos resultados da pesquisa em um artigo convencional, ilustra a lacuna identificada pela presente pesquisa e realça a necessidade de aproximar os dados de pesquisa dos pares afins.

Ao considerarmos o protagonismo atribuído aos dados de pesquisa no contexto da Ciência Aberta, bem como a descrição de como se chegou a estes, podemos nos juntar novamente a Ziman (1979) e reafirmar a importância do registro dos detalhes da experiência, que nem sempre é detectável em um artigo convencional.

Segundo o autor, cujas obras são clássicas da história da ciência e na Ciência da Informação, especialmente na comunicação científica, o registro de um

[48] Jonh Ziman foi um físico teórico cujo trabalho se caracterizou por sua clareza e simplicidade. Seu interesse pela responsabilidade social na ciência o levou a lecionar disciplinas sobre ensino e aprendizagem em ciência e sociedade. Historiador da ciência, cujos últimos trabalhos concentrara-se na compreensão pública da ciência.

experimento e de sua circunstância é fundamental para o progresso científico, bem como seus desdobramentos tecnológicos e inovadores.

> [...] um experimento é simplesmente uma observação passível de ser reproduzida. Sua essência está em que a experiência se tenha registrado sob circunstâncias determinadas, de maneira que possa ser repetida pelo pesquisador original ou por qualquer outra pessoa. Um único observador, por mais honesto que seja pode enganar-se; é essencial, a bem da precisão científica, que outros também vejam o que está ocorrendo. Através da repetição do experimento, eles ratificam os seus resultados e, incidentalmente, convencem a si próprios, de forma satisfatória, da sua validade. Assim, a prova feita através da experiência passa a ser parte do conhecimento público, sendo totalmente digna de crédito (ZIMAN, 1979, p.47).

Esse registo não se refere às discussões e aos resultados encontrados, mas, sobretudo, à descrição e detalhamento que deram origem e substância à discussão científica em questão. A publicação dos dados garante a característica de transparência atribuída ao método científico, potencializando-se em uma ciência mais aberta.

A rede científica se expande, ampliando a malha que compõe a ciência. Verifica-se também um maior cuidado na gestão e planejamento dos dados. A controvérsia apresenta-se como salutar, na medida em que exige do pesquisador e de seu grupo de pesquisa cuidados necessários para garantir a reputação e integridade dos dados sob sua autoria. Acreditamos nos benefícios que esta prática apresenta para a ciência, uma vez que se desvenda a caixa preta presente na maioria dos conjuntos de dados de pesquisa. Abre-se ao mundo científico inúmeras possibilidades de debates, aquecendo os fóruns científicos de pesquisa.

No capítulo a seguir apresentaremos a análise quantitativa sobre a prática observada. Abordaremos a prática da publicação de dados de pesquisa (via repositório e via periódico) pela área da Enfermagem sobre os vieses identificados na pesquisa.

7 PUBLICAÇÃO DE DADOS DE PESQUISA EM ENFERMAGEM: ASPECTOS QUANTITATIVOS

As pesquisas no domínio da Enfermagem produzem dados de natureza quantitativa e qualitativa decorrente de investigações observacionais, experimentais, de campo, entre outras.

No contexto brasileiro, de acordo com o diretório dos grupos de pesquisa no Brasil do CNPQ – Conselho Nacional de Desenvolvimento Científico e Tecnológico, existem 852 grupos na área de Enfermagem[49]. Tais números apontam para um universo significativo de produção de dados de pesquisa a serem explorados no cenário científico.

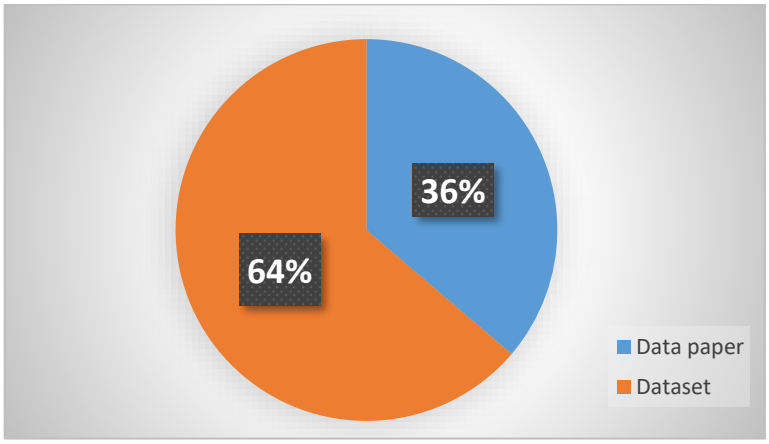
Paralelo à produção de dados oriundo de pesquisas, estão os dados produzidos no âmbito dos sistemas de saúde. Estes, embora não tenham sido gerados com um propósito investigativo, podem vir a ser um fecundo campo de pesquisa para a área em questão e áreas afins, configurando-se como um rico universo de dados a ser explorado.

No contexto da Ciência Aberta, esses dados passam a ser vistos de forma mais cuidadosa. O zelo na produção, documentação e gestão, bem como a possibilidade de reutilização dos mesmos, é cada vez mais discutido. Tais dados saem de suas planilhas e documentos invisibilizados e ganham espaços apropriados para fins de preservação e comunicação científica.

Diante desta vasta possibilidade de exploração de dados de pesquisa, associada aos critérios metodológicos adotados pelo estudo, iniciamos nossa reflexão apresentando, através do Gráfico 1, o panorama da prática da publicação de dados de pesquisa pela área da Enfermagem.

[49] Busca realizada na base corrente do Diretório dos Grupos de Pesquisa no Brasil (CNPQ) em 11/11/2021.

Gráfico 1- Prática da publicação de dados de pesquisa pela Enfermagem



Fonte: A autora, 2021.

Para representar a prática da publicação de dados de pesquisa pela área da Enfermagem, utilizamos os dados sistematizados nas tabelas 1 e 2 (anteriormente apresentadas no capítulo 4).

De acordo com o Gráfico 1, a prática da publicação via repositório é superior à publicação via periódico. A publicação em repositório foi representada por 64% dos dados recuperados na pesquisa, ainda que o número de bases de dados bibliográficas incluídas no estudo tenha sido superior ao número de repositórios explorados.

A publicação em periódico representou 36% dos dados recuperados o que aponta para um desequilíbrio entre as duas práticas observadas. Ainda que as fontes bibliográficas consultadas para identificação da publicação via periódico tenham sido superiores ao número de repositórios explorados, evidenciou-se uma “preferência” por este tipo de publicação de dados pela área observada.

Considera-se importante esclarecer, que as práticas representadas no gráfico 1 não significam que sejam ações excludentes. As práticas, tanto de depósito em repositório, quanto publicação em periódico configuram-se como ações que devam acontecer de formar concomitante.

Embora pré-matura, essa constatação ajudou-nos a começar a esboçar um perfil do comportamento da área. Desse modo, para melhor entendermos a prática investigada, consideramos oportuno segmentar a análise, conforme apresentaremos a seguir, sob a lente das bases e repositórios de dados de pesquisa.

7.1 PRÁTICA DE PUBLICAÇÃO DE DADOS VIA PERIÓDICO

A prática da publicação de dados de pesquisa em periódico científico pela área de Enfermagem foi evidenciada a partir da exploração nas bases de dados: Web of Science, Science Direct e Scopus. Contudo, antes de apresentá-la, consideramos relevante destacar, a título de contextualização histórica, a juventude da produção científica da área investigada. Segundo Avelino; Oliveira; Silva; Santos; Ferreira (2010, p. 1)

> O primeiro local para divulgação da pesquisa cientifica na Enfermagem brasileira foi a Revista Brasileira de Enfermagem, criada em 1932, denominada Annaes de Enfermagem, sendo considerado um evento bastante significativo para a divulgação e estímulo à pesquisa.

Santos e Gomes (2007, *online*) ampliam o panorama histórico de amadurecimento científico da área ao afirmarem que:

> O investimento das enfermeiras brasileiras no aprimoramento profissional mediante cursos de pós-graduação teve seu início já na década de 20, quando cerca de dezessete enfermeiras diplomadas pela Escola de Enfermagem Anna Nery realizaram cursos de pós-graduação nos Estados Unidos, com bolsa de estudos da Fundação Rockfeller. Também nos anos 40 e 50, o espectro das brasileiras em cursos de pós-graduação no exterior foi amplo.
> No que tange às primeiras iniciativas de criação de cursos de pós-graduação no Brasil, o Curso de Especialização em Enfermagem e Obstetrícia começou a ser oferecido em 1943, em São Paulo.
> No que tange a pós-graduação stricto sensu teve início em 1972, com a criação do curso de Mestrado da Escola de Enfermagem Anna Nery. Nesta mesma década foram implantados mais sete cursos de mestrado: quatro na região sudeste, dois na região nordeste e um na região sul. (SANTOS e GOMES, 2007, *online*)

Na contemporaneidade, Salles e Barreira (2010) colaboram para o entendimento sobre o comportamento da Enfermagem quanto no contexto da Ciência Aberta, ao resgatarem a cientificidade do campo no artigo “Formação da comunidade científica de Enfermagem no Brasil”. As autoras complementam Avelino et al (2010) ao apresentarem como se formou o que chamaram de “*habitus científico”* da área a partir dos anos de 1950.

No contexto internacional, a pesquisa no campo da Enfermagem e a construção de sua cientificidade é marcada inicialmente pelas investigações de natureza quantitativa, conforme nos mostra McEwen; Wills (2009)

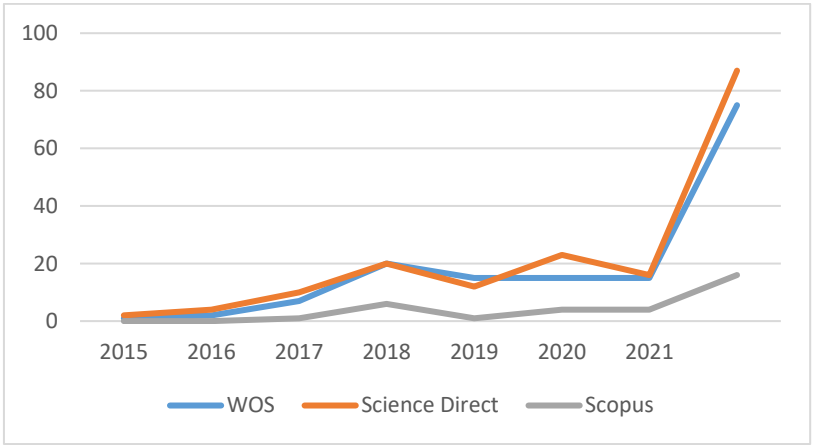
> Bastante influenciados pelo empirismo lógico, à medida que a Enfermagem começou a desenvolver-se como disciplina científica em meados de 1900, os métodos quantitativos eram usados quase exclusivamente na pesquisa. Nas décadas de 1960 e 1970, as escolas de Enfermagem alinharam a pesquisa de Enfermagem com a pesquisa científica no intuito de trazer reconhecimento ao ambiente acadêmico, e os enfermeiros pesquisadores (MCEWEN; WILLS, 2009, p.16).

Esta brevíssima história da construção científica da área colaborou com a pesquisa permitindo uma melhor análise sobre o comportamento da área no tocante à publicação de dados de pesquisa.

Com base nos dados obtidos a partir das buscas nas bases de dados selecionadas, iniciaremos nossa reflexão sobre a prática da publicação de dados de pesquisa via periódico pela Enfermagem.

O Gráfico 2 ilustra a adoção do *datapaper* enquanto canal de comunicação científica pela Enfermagem. A publicação de dados de pesquisa por meio deste instrumento demonstrou uma aceitabilidade por parte dos pesquisadores a esta tipologia de escrita para tornar público os dados produzidos em pesquisas.

Gráfico 2 – Aparecimento e crescimento da publicação de dados de pesquisa em Enfermagem via data paper[50]



Fonte: A autora, 2021.

[50] Pesquisa rodada em 08/11/2021.

De acordo com os dados apresentados no Gráfico 2, foi possível identificar quando a Enfermagem começou a aderir a este tipo de publicação[51]. Ao compararmos com o início do movimento da Ciência Aberta, por volta dos anos 2000, associado a propagação da cultura de acesso aberto à informação científica e abertura dos dados, vimos que o primeiro *data paper* foi publicado 15 anos após a deflagração do movimento. Se confrontarmos este dado com a cientificidade da área no contexto nacional e internacional (século XX), percebemos uma receptividade da Enfermagem quanto à proposta do movimento no cenário científico global, principalmente ao que se refere às novas práticas de publicação e comunicação científicas.

O crescimento exponencial para este tipo de publicação, visualizado através do Gráfico 2, revela novos comportamentos e práticas por parte da comunidade científica. Acreditamos que tal crescimento esteja ligado não apenas, mas principalmente, às políticas de fomento à pesquisa.

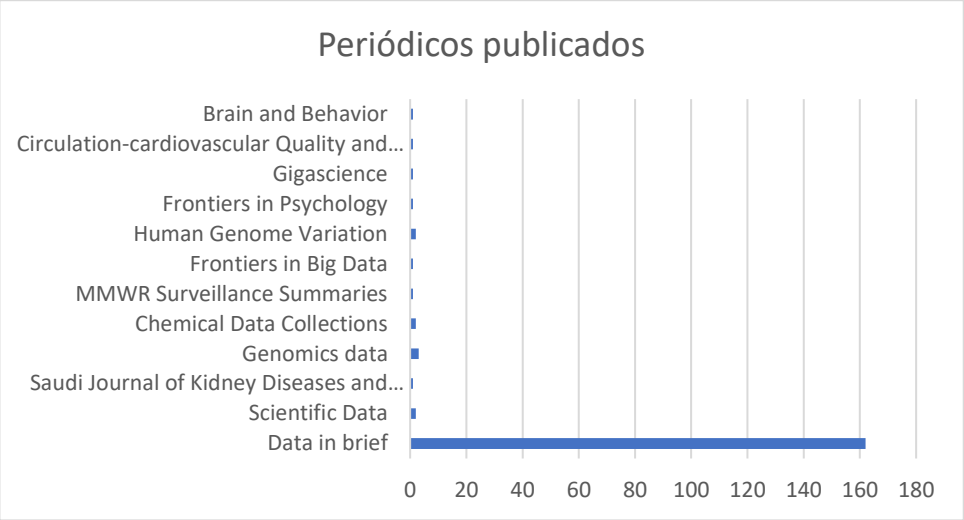
Se analisarmos a prática de publicação de dados via *data paper* em números absolutos, totalizando 178[52] artigos em 6 anos (2015-2021), vimos o desenvolvimento de uma transformação cultural científica em curso. A composição desta tipologia de artigo requer, antes de sua produção, uma mudança cultural no modo de gerar e tratar os dados produzidos em pesquisa. Não é apenas uma nova forma de escrita científica ou um novo *template*, é, sobretudo, uma nova forma de fazer ciência. O cuidado com os dados que o antecedem é condição *sine qua non* para sua reutilização e citação em pesquisas futuras.

Outro aspecto observado a partir dos dados coletados e considerado relevante para a prática investigada, bem como para o ecossistema da ciência como um todo, foi a escolha do periódico para o escoamento da produção da área. Para nossa surpresa, a maioria, quase que absoluta, dos *data papers* foram publicados em periódicos de escopo multidisciplinar, e não em periódicos especializados ou disciplinares.

O gráfico a seguir demonstra a discrepância entre os títulos de periódicos eleitos para a publicação dos dados:

[51] Informação baseada na exploração das bases citadas na metodologia.
[52] Cf. Conjunto de dados da pesquisa em https://doi.org/10.6084/m9.figshare.19790425.v1

Gráfico 3 - Publicação de dados de pesquisa em Enfermagem por título de periódico[53]



Fonte: A autora, 2021.

Nossa hipótese é a de que tal desequilíbrio possa estar relacionado a não disponibilização dessa tipologia de artigo em periódicos do escopo da Enfermagem. A falta de periódicos especializados que ofereça ao pesquisador a possibilidade de comunicar sua pesquisa via *data paper* pode estar redesenhando o escoamento da produção científica da área investigada. Outro fator que possa ter influenciado neste resultado são as pesquisas de natureza interdisciplinar e/ou de colaboração multidisciplinar, onde os resultados podem ser publicados em fontes de escopo amplo, como por exemplo a Nature.

Outro aspecto que pode ser levado em consideração é a juventude do campo, já contextualizada historicamente neste capítulo, além da mudança de práticas científicas propostas pelo movimento da Ciência Aberta, principalmente no que se refere a gestão dos dados de pesquisa.

Este aspecto observado pela pesquisa implica, também, além da análise que envolve a disponibilização de *data paper* para a publicação, a avaliação do fator de impacto do periódico de acordo com a área do conhecimento.

A revista em destaque (*Data in Brief*), da editora Elsevier, é responsável, em números absolutos, por 162[54] artigos de dados identificados pela pesquisa de um total

[53] Pesquisa rodada em 08/11/2021.
[54] Cf. Conjunto de dados da pesquisa em https://doi.org/10.6084/m9.figshare.19790425.v1

de 178. Caracteriza-se como um periódico multidisciplinar, de acesso aberto e revisado por pares, que publica artigos curtos, além de descrever e fornecer acesso aos dados de pesquisa. Contudo, a escolha em publicar neste periódico implica em custos que giram em torno de U$500 por artigo, além dos impostos envolvidos em publicações internacionais, o que poderia apresentar-se como uma barreira para a escolha por este tipo de publicação científica.

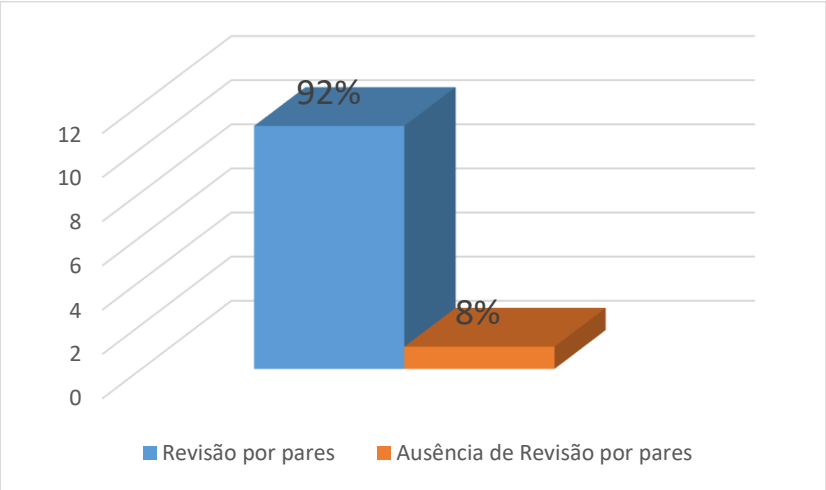
No entanto, com base nos dados representados no Gráfico 2, a pesquisa evidenciou que o fator econômico não se apresentou como barreira para a escolha por este tipo de publicação. Nossa hipótese é a de que a pouca oferta por esta tipologia de documento possa estar promovendo esse desvio de opção de publicação da área investigada.

Quanto aos demais periódicos identificados na pesquisa, nenhum deles é do escopo específico da Enfermagem, o que aponta para um viés investigativo sobre o não escoamento desse tipo de publicação em periódicos especializados da área.

O critério “revisão por pares” presente nas revistas científicas e considerado pela ciência como um critério “validador” da qualidade científica das publicações periódicas foi observado de forma pouco criteriosa na pesquisa. Contudo, consideramos relevante apontá-lo por tratar-se de um aspecto legitimador de credibilidade científica.

Dos 12 títulos representados no Gráfico 3, apenas um não apresentou revisão por pares (conforme Gráfico 4), nem fez menção à proposta de uma revisão do tipo *open peer review*, modelo que vem sendo propagado pelo movimento da Ciência Aberta. De modo bastante simplificado, a revisão do tipo: *open peer review* propõe um processo de validação científica mais inclusivo e ampliado.

Gráfico 4 - Revisão por pares em *data paper* via periódico na área de Enfermagem



Fonte: A autora, 2021.

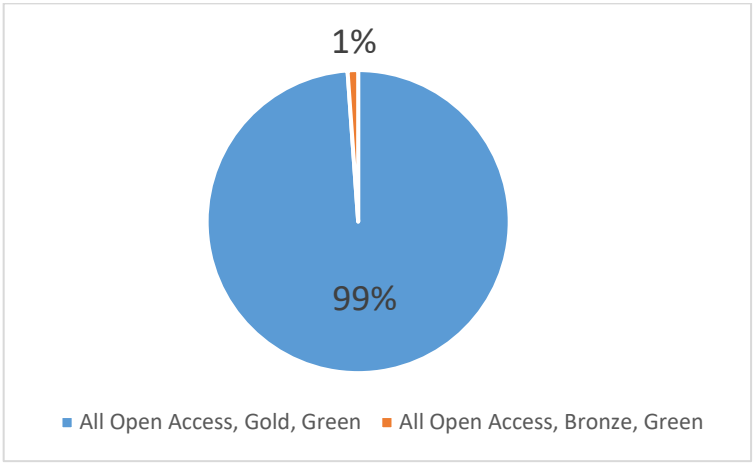
O Gráfico 4 refere-se à revisão por pares em *data papers* publicados em periódicos, e não os publicados automaticamente nos repositórios. Como pôde ser observado, a revisão por pares em *data papers*, presente em 92% dos títulos identificados na pesquisa, apresenta-se como uma nova prática científica, e como tal, vem sendo acolhida pela comunidade científica. Apenas 8% da amostra, representado em números absolutos por 1 título não adota a revisão por pares em *data papers*.

A forma de acesso aos artigos de dados também foi observada na pesquisa. Na maioria dos casos, são de acesso aberto via verde[55], variando o tipo de licença *Creative Commons (CC)*. Apenas duas publicações apresentaram acesso aberto via bronze[56], o que revelou uma majoritária aceitação ao movimento da Ciência Aberta pela área da Enfermagem, no que se refere à disponibilização e acesso à informação científica, recomendado pelo movimento, conforme veremos no gráfico a seguir:

[55] O acesso aberto via verde ou *Green* consiste no depósito em repositórios de artigos de pesquisa que já foram publicados ou espera-se que serão publicados em revistas científicas, revisados pelos pares, frequentemente envolvendo o auto-arquivamento do artigo num repositório pelos autores. (Pinfield et al., 2014) Disponível em: https://asistdl-onlinelibrary-wiley.ez24.periodicos.capes.gov.br/doi/full/10.1002/asi.23131

[56] No acesso aberto via bronze o conteúdo é gratuito para leitura e/ou download no site da editora, mas não é publicado sob uma licença aberta que permite o compartilhamento ou reutilização. O editor pode retirar o acesso a qualquer momento. Essa forma de acesso 'aberto' é frequentemente usada para tornar o conteúdo de leitura livre por apenas um breve período, talvez imediatamente após a publicação ou em resposta a um evento catastrófico como a pandemia COVID-19. No entanto, uma vez que não existe uma licença aberta, ele não é, de fato, um Acesso Aberto. (Barnes, 2018). Disponível em: https://doi.org/10.11647/OBP.0173.0089 Acesso em: 12/11/2021

Gráfico 5 - Acesso aos dados de pesquisa de Enfermagem via *data paper*



Fonte: A autora, 2021.

O tipo de acesso, apontado no Gráfico 5 como "*All Open Access, Gold, Green*", representa por 99% dos títulos identificados, e refere-se as publicações de acesso aberto, publicado em periódico e repositório de acesso aberto. A característica *Gold* e *Green* refere-se à via de publicação, conforme descrito anteriormente. Esse tipo de publicação não gera custos de acesso, contudo, a escolha do periódico onde será publicado pode demandar custos elevados para os autores da publicação.

Já o tipo de acesso "*All Open Access, Bronze, Green"* representado na pesquisa por apenas 1% dos títulos recuperados, caracterizam as publicações que podem estar temporariamente com acesso aberto. A característica bronze refere-se à ausência de licença aberta, que permite o compartilhamento ou reutilização, contudo, o editor pode retirar o acesso a qualquer momento.

De maneira geral, os dados aqui apresentados revelaram uma aceitabilidade pela Enfermagem à prática da publicação de dados de pesquisa via periódico por meio de *data papers*. Além disso, os dados sugerem que a Enfermagem está em acordo com a premissa do movimento da Ciência Aberta, tendo em vista a juventude do campo, no que se refere à pesquisa científica datada por Avelino et al. (2010) em meados do século XX.

Apresentaremos a seguir, sob a lente dos repositórios de dados, o comportamento da Enfermagem no que se refere à prática da publicação de *datasets* em repositório de dados de pesquisa.

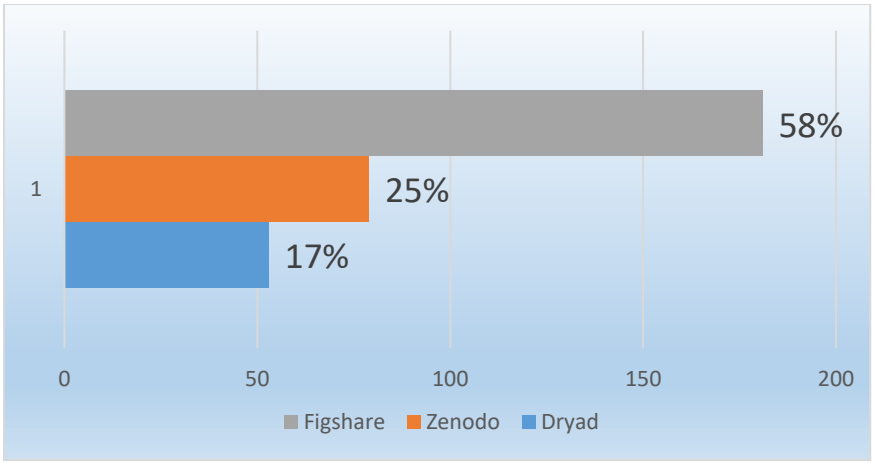
7.2 PRÁTICA DE PUBLICAÇÃO DE DADOS VIA REPOSITÓRIO

A publicação de dados de pesquisa via repositório na área de Enfermagem foi evidenciada a partir da exploração dos seguintes repositórios: *Figshare*, *Zenodo* e *Dryad*. Tal prática envolve, de maneira simplificada, o depósito ou *upload* de um *dataset* em um repositório seguro, onde esses dados poderão ser localizados, preservados e reutilizados. Além disso, pressupõe também uma cultura de gestão, abertura, transparência e, até mesmo, de desapego aos dados de pesquisa, que é uma das bandeiras do movimento da Ciência Aberta.

A implementação dessa cultura revela um amadurecimento e desprendimento do campo, pois amplia as possibilidades de pesquisa e colaboração científicas de um modo geral.

O Gráfico 6 apresenta a prática da publicação de dados de pesquisa através de *datasets* via repositório.

Gráfico 6 - Publicação de dados de pesquisa de Enfermagem em repositórios via *dataset*



Fonte: A autora, 2021.

Observou-se por meio do Gráfico 6, uma evidente “preferência” por um repositório em detrimento aos demais. O *Figshare* desponta como preferido pela área, com 58% dos resultados. Em seguida, vem o *Zenodo,* com a fatia de 25% das publicações de *datasets* e por último o *Dryad* com 17% de publicações de dados em

repositórios. Em números absolutos, temos um quantitativo de 181 *datasets* publicados no *Figshare*, seguido de 79 publicados no *Zenodo* e 53 publicados no Dryad.

As razões pelas quais um pesquisador opta por um determinado repositório em detrimento a outros podem ser diversas, e, na pesquisa em questão, não foi possível, nem era o objetivo identificar tais motivos. Contudo, podemos buscar pistas que nos levem a entender tal comportamento. A primeira delas é o ano de lançamento, que respectivamente apresenta-se da seguinte maneira: *Dryad* (2009), *Figshare* (2011) e *Zenodo* (2013).

Inversamente proporcional ao quantitativo de *datasets* publicados, o *Dryad* (2009), que foi o primeiro a ser lançado*,* é o que menos contém depósitos de *datasets* do campo da Enfermagem. A primazia no lançamento poderia ser uma das justificativas para o povoamento do repositório. Porém, não foi isso que encontramos na prática. Apesar de ter sido o primeiro a ser lançado e não apresentar dificuldades para o *upload* dos conjuntos de dados, o *Dryad* apresenta um *layout* “pobre” em termos de apresentação dos resultados de busca, filtros e serviços.

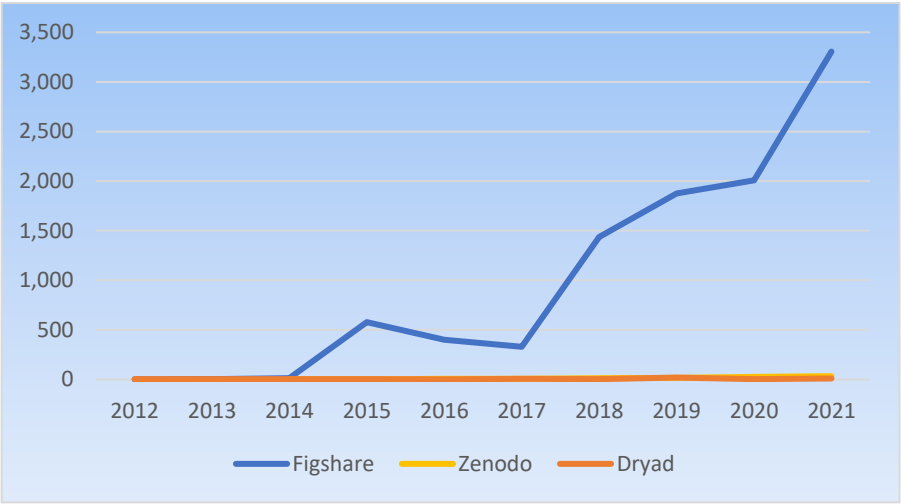
Além das questões de interface de busca e recuperação dos resultados, acreditamos que o fator mais determinante que possa justificar a baixa opção por esta ferramenta é a cobrança de taxa para publicação. O *Dryad* cobra uma taxa pela publicação de dados que subsidia a curadoria e preservação dos conjuntos de dados publicados no valor de U$ 120. Tendo em vista a possibilidade de outros repositórios que oferecem o mesmo serviço e não cobram pelo armazenamento dos dados, talvez este seja um dos fatores que justifiquem a baixa demanda por este repositório pela área investigada.

Dando prosseguimento à análise a partir da data de lançamento do repositório, aparece o *Figshare* (2011), que é o campeão em depósitos de *datasets* no campo da Enfermagem. Acreditamos que a preferência por esta ferramenta possa estar relacionada a dois fatores: o primeiro, a simplicidade no *upload* e descrição dos dados. O segundo, refere-se à gratuidade do serviço. Não é cobrado taxa de curadoria e preservação dos dados ali depositados. Além disso, o *Figshare* é o único repositório, dos três explorados que oferece serviço de altimetria para os *datasets*.

Por último, o *Zenodo,* que foi lançado em 2013 e que, em termos de número de depósitos de *datasets,* aparece entre os dois anteriormente citados. Sua posição intermediária talvez possa ser justificada por ser o mais jovem dos três repositórios explorados. A interface de navegação e serviços são equiparados aos demais, além de não cobrar taxa de depósito. Apresenta simplicidade no *upload* de *datasets*, contudo, exige o preenchimento mais detalhado de metadados.

Outro aspecto observado na pesquisa refere-se ao período em que os *datasets* começam a ser depositados nos repositórios pela Enfermagem. De acordo com o Gráfico 7, podemos verificar que a prática da publicação de dados em repositório tem início em 2012. O *Figshare* e o *Dryad* dão o pontapé inicial e mais tarde o *Zenodo* entra no circuito. Destacamos a liderança do *Figshare* nesse processo e o crescimento exponencial do número de depósitos de *datasets* pela área da Enfermagem. Cabe destacar que alguns destes *datasets* são de pesquisas brasileiras e que poderão ser identificadas no *dataset*[57] desta pesquisa.

Gráfico 7 – Publicação de dados de pesquisa de Enfermagem em repositório por ano



Fonte: A autora, 2021.

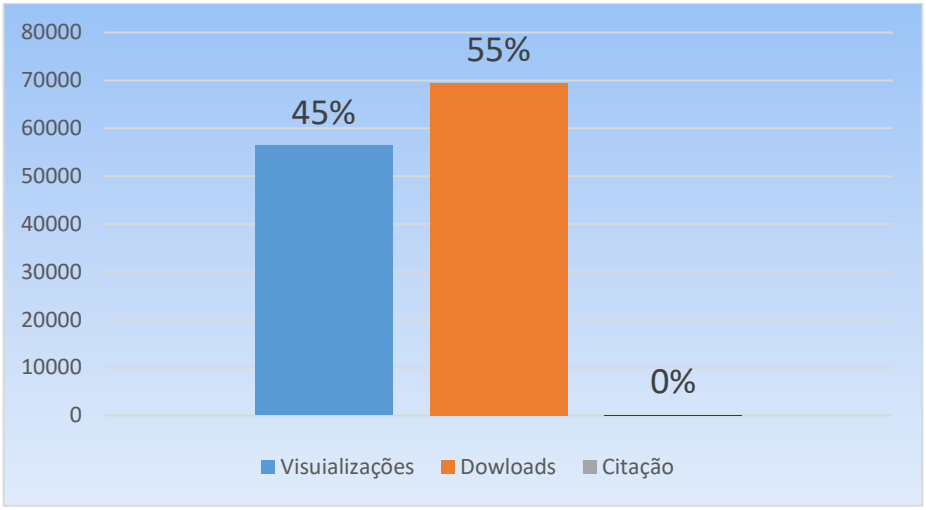
Estes dados revelaram um traço de comportamento da área, uma vez que, ao analisarmos em comparação ao início do movimento da Ciência Aberta nos anos 2000, temos uma área que primeiro publica dados em repositórios (2012), ou seja,

[57] Cf. https://doi.org/10.6084/m9.figshare.19790425.v1

disponibiliza seus *datasets* antes de descrevê-los por meio de *data paper* em periódico científico (2015). Tal reflexão deve considerar a existência de repositórios de dados que geram seus *data papers* de maneira automática. Embora esse não tenha sito um critério observado, cabe o esclarecimento para que não comprometa a reflexão proposta.

O interesse em *datasets* pela comunidade científica foi outro aspecto observado. Por meio do Gráfico 8, pudemos observar a procura/pesquisa por conjuntos de dados através dos índices de visualização, *downloads* e citação.

Gráfico 8 – Visualização, downloads e citação de dados de pesquisa de Enfermagem



Fonte: A autora, 2021.

A representação gráfica demonstra que o número de *downloads*, representado pela amostra por 55% dos dados da pesquisa, é superior ao número de visualizações (45%), o que à primeira vista parece-nos um pouco contraditório. Contudo, essa discrepância se dá graças à possibilidade de acesso ao conjunto de dados via link externo ao repositório investigado. Desse modo, nem sempre o *download* é feito via repositório.

Outro ponto que nos chamou atenção é o baixíssimo número de citações para os *datasets.* Representado por 0% dos dados recuperados na pesquisa, o equivalente a 122 citações em números absolutos, esse retrato demostra uma fragilidade e abre lacunas a serem investigadas sobre a citação de dados de pesquisa. Nossa hipótese é a de que tal discrepância deve-se principalmente a três fatores: o primeiro está

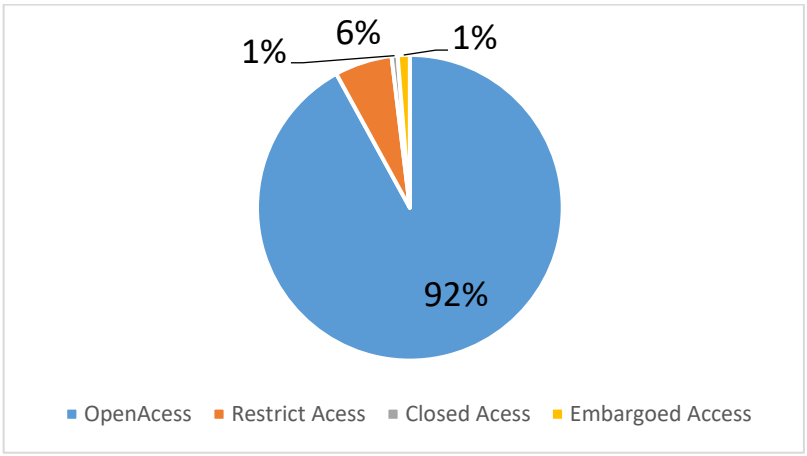
relacionado à falta de conhecimento sobre a possibilidade de citação dos dados; o segundo, a questões de indexação dos repositórios onde tais *datasets* estão depositados e o terceiro deve-se à insuficiência de interoperabilidade semântica e/ou tecnológica, que permitiria a comunicação entre *datasets* e sua compreensão para reutilização futura.

A verticalização de estudos que abordem o índice de citação dos *datasets* possibilita a abertura de um viés de discussão que promova visibilidade das pesquisas conduzidas no âmbito das instituições participantes, de seus pesquisadores e áreas de pesquisa.

Percebemos a existência de uma lacuna temática que aborde a reutilização e citação dos *datasets* de pesquisa. Mesmo com uma tímida prática de citação de *datasets,* aqui representada, vimos uma mudança em curso no que se refere às práticas de publicação científica no campo da Enfermagem.

Através dos dados produzidos pela pesquisa, foi possível identificar o nível de abertura dos *datasets* depositados nos repositórios. Como pode ser visto no Gráfico 9, em sua maioria, são dados abertos, sem nenhuma restrição para acesso e *downloads*, devendo apenas respeitar os critérios de uso e citação dos dados previstos nas licenças adotadas.

Gráfico 9 – Nível de abertura dos *datasets* de Enfermagem



Fonte: A autora, 2021.

Em números significativamente inferiores, temos alguns *datasets* de acesso restrito, acesso fechado ou com algum tipo de embargo. Para estes casos, há de se

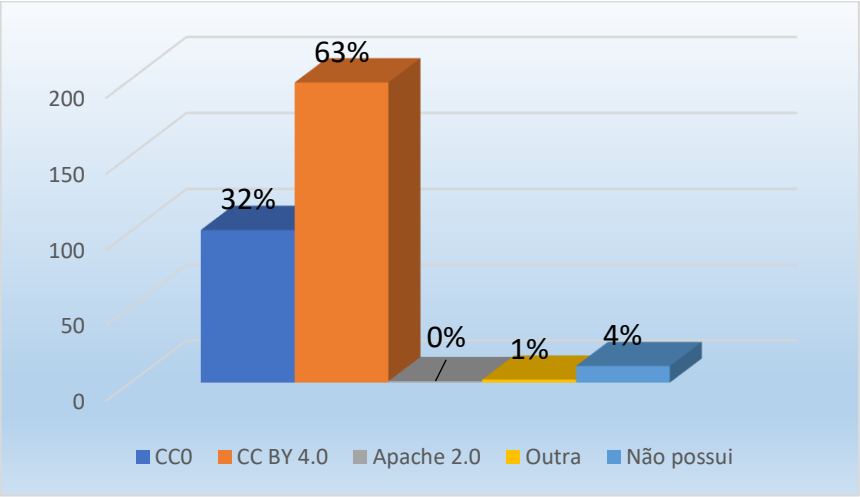
buscar outra alternativa para acessar os *datasets* de maneira responsável e ética. Ainda assim, mesmo com algum tipo de restrição, o ato de dar publicidade à pesquisa e aos dados produzidos pela Enfermagem por meio dos repositórios configura-se como uma prática positiva e favorável à Ciência Aberta, e que vai ao encontro do Programa *Horizon2020,* que ressalta a importância de os dados estarem “tão abertos quanto possível, tão fechados quanto necessário”. De acordo com o documento: *The “A” of FAIR – As Open as Possible, as Closed as Necessary:*

> De acordo com as Diretrizes do *Programa H2020* sobre Dados FAIR, os dados devem ser "tão abertos quanto possível e tão fechados quanto necessário", "abertos" a fim de promover a reutilização e acelerar a pesquisa, mas ao mesmo tempo devem ser "fechados" para salvaguardar a privacidade dos sujeitos.[58] (LANDI et al., 2020, p.50, tradução nossa).

Em consonância com o movimento em curso, observamos a adoção de licenças para utilização dos *datasets*. De acordo com o Gráfico 10, a maioria deles, cerca de 63% dos *datasets* possuem licença do tipo CCBY 4.0, de atribuição internacional, onde o usuário tem permissão para compartilhar os dados, copiar e redistribuir o material em qualquer suporte ou formato. Pode também adaptar, transformar e criar a partir do material para qualquer fim, mesmo que comercial. Essa licença também prevê a atribuição de crédito, identificação do link para a licença e indicar as mudanças que foram feitas.

[58] The H2020 Program Guidelines on FAIR Data (H2020 Program) introduced the concept that “data should be as open as possible and as closed as necessary”.

Gráfico 10 – Licença de uso dos *datasets*



Fonte: A autora, 2021.

Na sequência, aparece a licença do tipo CC0, representando 32% dos datasets identificados na pesquisa. Nela, o proprietário ou criador do *dataset* atribui característica de domínio público aos dados. Apenas um d*ataset* informou a licença do tipo Apache 2.0, que consiste em ser uma licença de software livre. Os demais não possuíam licença ou informavam na descrição que era “outra” licença.

A determinação da tipologia de licença atribuída aos *datasets* são extremamente relevantes para o tipo de uso, que pode ser feito dos dados publicados. Questões legais e de direitos autorais se estabelecem neste circuito e compõem o cenário da prática da Ciência Aberta.

A análise dos dados produzidos pela pesquisa permitiu-nos observar um outro viés de discussão proposto pelo movimento: a escolha da extensão do arquivo de dados. Por meio da identificação da tipologia de arquivo utilizado para armazenamento dos dados, podemos ampliar a discussão sobre a necessidade em atender aos Princípios FAIR (*Findable, Accessible, Interoperable, Reusable*) preconizados pelo movimento para os dados de pesquisa. Tais princípios são atribuídos ao conjunto de dados de pesquisa e têm o objetivo de promover o compartilhamento e reuso dos dados.

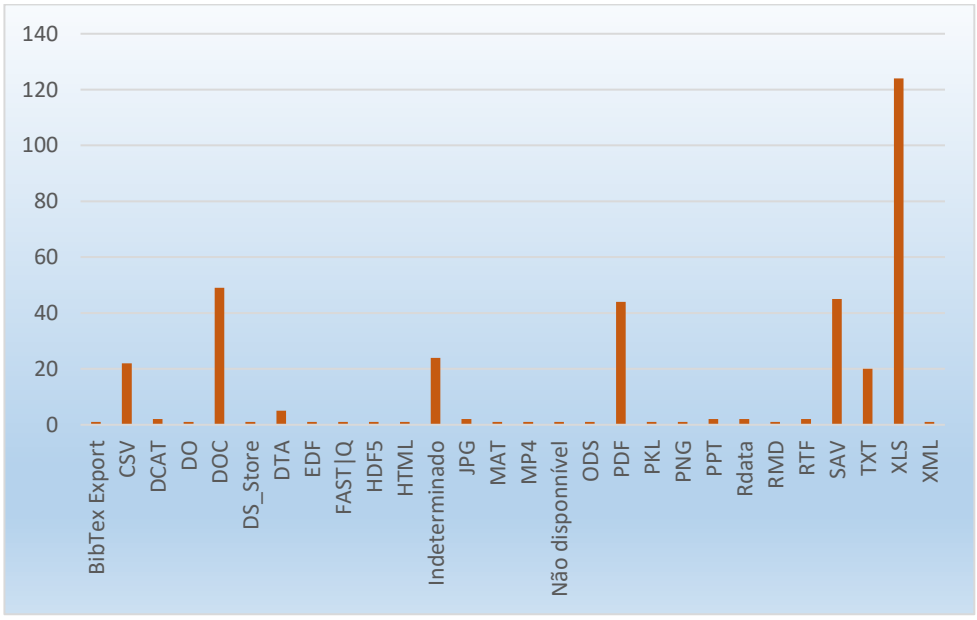
Consideramos relevante trazer para a reflexão a implicação dos dados estarem de acordo com tais princípios. De acordo com Veiga et al. (2019),

> Os princípios FAIR surgem como resultado da conferência internacional "*Jointly designing a data FAIRPOR"* realizada em 2014, com o intuito de contribuir para o aumento do compartilhamento e reuso dos dados de pesquisa [...]. A partir de então, esses princípios passaram a ser reconhecidos internacionalmente, sendo considerados os elementos norteadores para a descoberta, acesso, interoperabilidade e reuso dos dados, por oferecerem orientações basilares no tratamento, reuso e publicação de conjunto de dados, sob condições claramente entendidas por humanos e máquina (VEIGA et al., 2019, p. 277).

No conjunto de dados produzido por esta pesquisa, foi criado um metadado denominado "tipo de arquivo". Este, descreve a extensão escolhida pelos pesquisadores para armazenamento e gestão dos dados produzidos pela pesquisa e que serviram de termômetro para compreendermos o quão FAIR estão os dados produzidos pela Enfermagem.

O Gráfico a seguir revela o formato de dados mais publicado pelo campo da Enfermagem.

Gráfico 11 - Formato de dados publicado pela Enfermagem



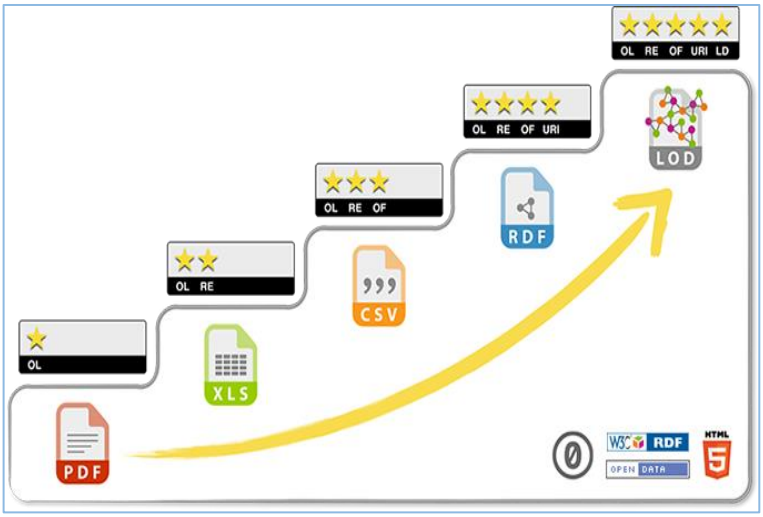
Fonte: A autora, 2021

Como pôde ser observado no Gráfico 11, o formato de arquivo .XLS é o mais utilizado pelo campo da Enfermagem, seguido respectivamente pelos arquivos: .DOC, .Sav, .PDF, Indeterminado, CSV, TXT e demais arquivos.

Cabe traduzir o metadado “Indeterminado” para o conjunto de dados produzido pela pesquisa. O termo para definir arquivos do tipo “Indeterminado” está relacionado à condição de acesso ao *dataset*. Ou seja, nos *datasets* que são do tipo “*Closed acess*”, “*Restrict acess*” ou possuem algum tipo de embargo, não foi possível identificar a tipologia de formato escolhido para armazenamento dos dados de pesquisa. Portanto, para esses *datasets,* o tipo de arquivo foi estabelecido como indeterminado.

A partir da identificação dos formatos de dados utilizados pela Enfermagem, consideramos relevante trazer para o debate a proposta de *Tim Berners-Lee*[59], o mentor do *Web and Linked Data*[60], que recomenda um esquema de pontuação baseado em cinco estrelas para dados abertos, conforme Figura 13:

Figura 14- Esquema de pontuação baseado em cinco estrelas para dados abertos



Fonte: As cinco estrelas dos dados abertos, 2019. [61]

O esquema de pontuação baseado em cinco estrelas identifica o potencial FAIR dos dados publicados enquanto objetos digitais, enriquecendo a discussão proposta

[59] Tim Berners-Lee é um físico britânico, cientista da computação e professor do Massachusetts Institute of Technology - MIT. É o criador da *World Wide Web* e diretor do World Wide Web Consortium (W3C), que supervisiona o desenvolvimento continuado da web. Fundador da World Wide Web Foundation e pesquisador sênior e titular e fundador da cadeira de 3Com no Laboratório de Inteligência Artificial e Ciência da Computação do MIT (CSAIL). É diretor da *The Web Science Research Initiative* (WSRI) e membro do conselho consultivo do Centro de Inteligência Coletiva do MIT.

[60] A Web and Linked Data prevê a publicação de dados estruturados vinculados.

[61] Disponível em: https://h5p.org/node/520377

no âmbito da Ciência Aberta sobre a publicação de dados e, consequentemente a abertura dos dados de pesquisa.

Para complementar a análise, reproduzimos no Quadro 1 a descrição e características da pontuação baseada em cinco estrelas proposta por *Tim Berners-Lee*, onde:

Quadro 1- Pontuação proposta por Tim Berners-Lee para dados abertos

Pontuação	Características	Extensão do arquivo
★	Dados publicados com uma licença aberta, não importando o formato, segundo a classificação original de Berners-Lee. A informação fica acessível, mas a manipulação e o processamento dos dados abertos são dificultados.	PDF DOC
★★	Os dados precisam estar em um formato legível por máquina e ter uma licença aberta. Refere-se a disponibilização do conjunto de dados de forma estruturada.	XLS SAV
★★★	Os dados devem estar em formato não proprietário (por exemplo, csv em vez de excel). Além da licença aberta, o **formato** do arquivo também precisa ser aberto. Isto é, o formato deve ser especificado como um **padrão aberto**, ao invés de ser propriedade exclusiva de uma única empresa.	ODS CSV
★★★★	Além da licença aberta e o formato de arquivo aberto, precisa ter um identificador único para este recurso na Internet. Um **endereço URL único**.	RDF
★★★★★	Agrupa as características das 4 estrelas e acrescenta a condição de os dados estarem linkados entre si. .	LOD

Fonte: As cinco estrelas dos dados abertos, 2019.[62]

Como pôde ser constatado no Gráfico 11, o formato de arquivo XLS caracterizou-se como o formato mais utilizado pela Enfermagem, e, de acordo com o esquema de pontuação de dados abertos representado pela Figura 13 associado ao *Quadro* 1, é classificado como um formato duas estrelas, ou seja, um formato que, embora apresente os dados de forma estruturada, ainda não atende satisfatoriamente

[62] Disponível em: https://h5p.org/node/520377

aos princípios FAIR para dados abertos de pesquisa. Supomos que tal escolha deve-se a uma gestão de dados minimamente preocupada com a possibilidade de manipulação, processamento e possível reutilização do *dataset* publicado. Por ser um formato proprietário, o .XLS pode apresentar barreiras de processamento e manipulação dos dados.

As próximas três tipologias de arquivos que aparecem com mais frequência na sequência de formatos escolhidos para publicação de dados pela Enfermagem são: .DOC, .Sav, e .PDF., respectivamente. Enquadram-se nas classes 1, 2 de estrelas proposto por *Berners-Lee* e apresentam formato proprietário, característica que pode dificultar a acessibilidade do conteúdo, principalmente para os arquivos do tipo .SAV. Este último, embora seja arquivo do tipo estruturado, necessita de software específico para acessar o conteúdo do arquivo. Os arquivos do tipo .DOC e .PDF são classificados como não estruturados e, como tal, podem apresentar entraves para manipulação e processamento dos dados.

Em números absolutos, dos 313 *datasets* analisados, apenas um apresentou arquivo de dados do tipo .ODS. Os formatos do tipo .CSV representaram um quantitativo de 22 arquivos, ambos classificados por Berners-*Lee* como três estrelas. Os formatos citados são estruturados e não-proprietários, aproximando-se da proposta do movimento para dados abertos.

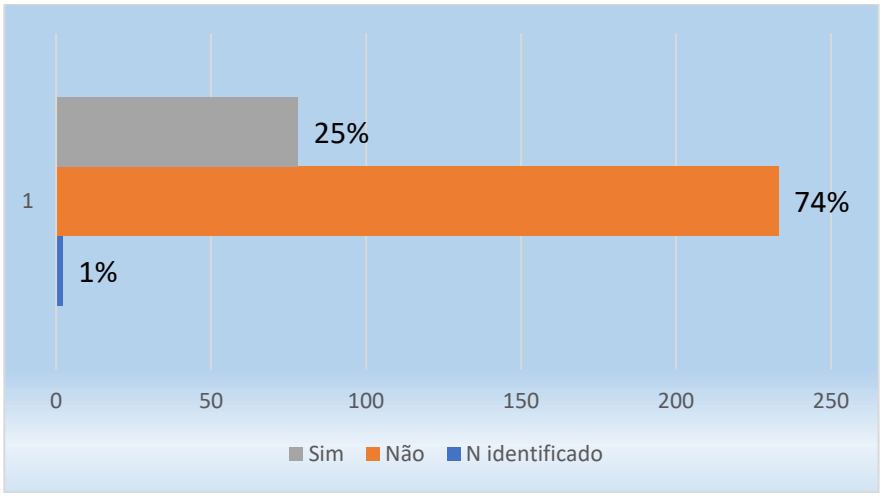
Nenhum dos *datasets* mapeados pelo estudo foram classificados como quatro ou cinco estrelas, o que demonstra uma prática a ser aprimorada rumo às boas ações em Ciência Aberta no quesito publicação de dados de pesquisa.

Contudo, a pontuação proposta por *Berners-Lee* nos auxilia a buscar alcançar as métricas FAIR dos objetos digitais aqui observados. De acordo com HANNING et. al. (2019),

> As métricas FAIR devem definir, com certa precisão, um conjunto mensurável de propriedades e comportamentos que avaliam o nível de FAIRness dos objetos digitais. O termo FAIRness, significa o grau em que um recurso digital adere aos princípios FAIR. [...] tais princípios são considerados aspiracionais, na medida em que não definem estritamente como alcançar um estado de FAIRness, mas descrevem um conjunto de características, atributos e comportamentos que moverão um recurso digital para mais perto desse objetivo (HANNING et. al., 2019, p. 180).

O último aspecto explorado pela análise dos *datasets* foi quanto à existência de arquivos complementares ou descrição dos dados de pesquisa. O Gráfico 12 retrata a existência ou não de arquivos que contenham os metadados da pesquisa, bem como sua descrição para melhor compreensão e reutilização dos dados de pesquisa.

Gráfico 12– Arquivos complementares ou descrição dos metadados em *datasets* de Enfermagem



Fonte: A autora, 2021.

Conforme representação gráfica, a maioria dos *datasets,* representado por 74% dos datasets observados na pesquisa, não apresentaram a descrição dos metadados, o que aponta uma lacuna na gestão dos dados. Tal comportamento pode vir a impactar a reutilização dos dados, tendo em vista à ausência de conteúdo informativo sobre a origem e o contexto de produção dos mesmos. Os *datasets* nos quais não foi possível identificar a existência de arquivos complementares representam 1% do total e foram os classificados como: *closed acess*, *restrict acess* ou com algum tipo de embargo. Somente 25% dos conjuntos de dados apresentaram arquivos complementares ou metadados que contextualizassem os dados publicados nos repositórios.

Este ponto da análise resgata o exemplo utilizado no capítulo dois sobre a escrita cuneiforme e a necessidade de descrição para sua interpretação nos dias

atuais. A publicação do conjunto de dados sem uma prévia gestão nem sempre garante seu entendimento e possibilidade de reuso em novas pesquisas. A ausência de uma descrição cuidadosa dos dados, por meio de sua gestão, é um fator determinante para o reuso dos dados em novas pesquisas.

De acordo com os princípios FAIR, este viés observado na pesquisa interfere na interoperabilidade dos dados, uma vez que na maioria dos *datasets* observados os metadados não apresentam linguagem formal, acessível e de ampla aplicabilidade para a representação do conhecimento.

Observamos, ainda, que em alguns dos resultados recuperados em repositórios foram depositados mais de uma versão de *datasets*. Algumas destas versões estão disponíveis para *download* e outras não. Isso implica em expectativas quanto ao acesso à informação que nem sempre está disponível ou completa.

Quanto à recuperação dos *datasets* em repositórios, vimos que em alguns casos o resultado não correspondia ao comando de busca, ou seja, retornaram somente o documento com a descrição dos dados. Não havia a disponibilização do *dataset*. Tal constatação levou em consideração o filtro “tipo de documento: *dataset*” utilizado na busca em repositório. O que nos leva a um questionamento sobre os critérios de curadoria dos dados.

Sobre o viés da gestão dos dados, observamos que em alguns casos os títulos dos *datasets* nem sempre representavam o conteúdo dos dados, tais como: Dataset, *Nursing dataset*, *Table* 1, *File*, entre outros. A ausência de gestão dos dados reduz a possibilidade de localização e reutilização dos mesmos.

Outro fator referente à curadoria dos *datasets* identificado como entrave na localização dos dados foi a ausência do preenchimento do metadado: resumo, em alguns dos conjuntos de dados recuperados pela pesquisa. Acredita-se que esta falha na gestão se configura como um complicador para a compreensão do conteúdo dos dados produzidos.

Mabe (2010) corrobora esse entendimento ao propor uma reflexão sobre dimensões da comunicação científica. Consideramos a dimensão da prática da publicação de dados enquanto “ato de comunicação”, e, como tal, deve estar preocupada com a transmissão da informação ali contida. Segundo o autor, a tecnologia, representada nesta pesquisa pelos repositórios e periódicos digitais,

configura-se como “novas ferramentas para propósitos antigos.” (MABE, 2010, p. 134). E, sendo assim, necessita de metadados que garantam os antigos propósitos de tornar público o conteúdo científico produzido pela área.

O panorama apresentado a partir da exploração das bases e repositórios de dados permitiu-nos identificar um alinhamento da área para a proposta do movimento da Ciência Aberta. Contudo, este mesmo panorama aponta para obstáculos a serem superados e para a necessidade de aprimoramento das práticas de dados abertos de pesquisa. Embora algum dos requisitos proposto pelos princípios FAIR tenham sido atingidos, verificou-se que existe um longo percurso a ser trilhado quanto à gestão dos dados produzidos em pesquisas na área.

No próximo capítulo, apresentaremos as considerações finais do estudo, considerando o objetivo principal de identificar a prática de publicação de dados de pesquisa pela área da Enfermagem, enquanto novas práticas em Ciência Aberta.

8 CONSIDERAÇÕES FINAIS

Cabe-nos aqui apresentar as considerações finais produzidas no estudo, à luz dos objetivos propostos, da cientificidade da área observada e da promoção de novas práticas científicas impulsionadas pelo Movimento da Ciência Aberta.

Inicialmente nos propomos a responder, com base nos dados produzidos, as seguintes questões que nortearam o estudo: 1) Se os pesquisadores da área da Enfermagem estão publicando os dados produzidos em pesquisas e, em caso positivo, 2) Quais canais de comunicação científica estão escolhendo para publicá-los?

Sim, os pesquisadores da área da Enfermagem estão publicando os dados produzidos em pesquisas. Os dados consolidados apresentados e discutidos no capítulo 7 demonstraram que a prática da publicação de dados de pesquisa já é uma realidade adotada pela área da Enfermagem, mesmo que de forma tímida.

Tal constatação pode ser confirmada graças ao número de publicações recuperadas nos periódicos e repositórios explorados. Este resultado reflete uma transformação científico-cultural no comportamento da comunidade científica observada.

A publicação de dados de pesquisa é uma das bandeiras levantada pelo movimento da Ciência Aberta e, como tal, pôde ser considerada como uma boa prática que vem crescendo e sendo disseminada pela área da Enfermagem.

A segunda questão norteadora pretendeu identificar quais os canais de comunicação científica a área da Enfermagem escolheu para dar publicidade aos dados produzidos por suas pesquisas. Vimos que tais dados vêm sendo publicados em periódicos e repositório de dados, como canais formais de comunicação científica.

Abre-se uma ressalva para o reconhecimento da limitação da análise quanto aos canais identificados para publicar os dados de pesquisa, tendo em vista as diversas possibilidades de recursos e tecnologias disponíveis para o depósito e descrição dos dados produzidos pela comunidade científica.

Percebemos que essa constatação, friamente descrita, pode induzir a uma avaliação equivocada sobre a prática investigada. A exploração dos repositórios

permitiu-nos identificar artigos e outros documentos depositados nos repositórios como *datasets*, o que possa, talvez, justificar um maior número de ocorrência para esta ferramenta.

O mesmo não aconteceu com os periódicos que publicam dados de pesquisa via *datapaper*. Embora exista uma variedade de *templates* para a publicação de dados em periódicos, a tipologia artigo não se confunde com conjunto de dados, imagens ou outro tipo de veículo de comunicação científica. Artigo é artigo. Contém a descrição dos dados produzidos pela pesquisa, diferentemente dos repositórios explorados, onde, para cada resultado de pesquisa, tínhamos uma variedade de tipologia de publicações (imagem, texto, planilhas, artigos, dentre outros tipos de dados) que interferiram na identificação dos *datasets*.

Percebemos que, com ou sem curadoria, os repositórios de dados analisados permitem a inclusão de *datasets*, independente do conceito e entendimento estabelecido pela literatura e pela ciência de modo geral.

Desse modo, identificamos um maior rigor para a prática da publicação de dados via periódico em comparação à publicação de dados via repositório, tendo em vista os critérios de revisão por pares e a curadoria dos dados dos repositórios explorados pelo estudo, respectivamente.

Os objetivos propostos pelo estudo conduziram o caminho trilhado pela pesquisa e nos ajudaram a responder as questões norteadoras do estudo. Por meio do mapeamento das publicações de dados via periódico e via repositório, associado a identificação dos canais de publicação de dados, foi possível identificar o comportamento da área no contexto da Ciência Aberta. Tal comportamento aponta para uma transformação de *habitus científico* da Enfermagem, ao atender um chamamento global de abertura de dados de pesquisa, considerando a proposta do programa *Horizon2020*, que reforça a importância de os dados estarem “tão abertos quanto possível, tão fechados quanto necessário”.

Quanto ao potencial de mobilização de mundo identificado pela pesquisa através da publicação de dados e projetado pela TAR de Bruno Latour, foi possível observar que, ao trabalhar sob a ótica do composicionismo, Latour propõe um cenário científico onde as ciências trabalham de forma integrada e não mais dispersas e de forma isolada. Necessitam de reflexões interdisciplinares para o enriquecimento das

análises e discussões. Neste caso específico, o cenário científico foi representado pela *e-science* e orientado pelo movimento da Ciência Aberta.

A publicação de dados confere transparência e mobilidade a uma realidade já mapeada, já testada e descrita. Além disso, possibilita novas utilizações, arranjos, interpretações e experimentos a partir do reuso do conjunto de dados adequadamente descritos. A interoperabilidade semântica e tecnológica pretendida pelos princípios FAIR facilitam essa mobilização tanto na esfera humana quanto na esfera das máquinas.

A pesquisa reforçou uma aproximação entre a prática da publicação de dados e a Ciência da Informação. Esta última ampara a discussão por propor estudos que discutem a descrição adequada dos dados FAIR no contexto da pesquisa e da comunicação científica, além de buscar instrumentos para representação temática (metadados) que mais se adequem ao contexto e proveniência dos dados produzidos em pesquisas.

O estudo acompanha a ciência em ação, conforme pode ser observado a partir da composição multidisciplinar entre Enfermagem, Ciência da Informação e a Comunicação científica. A investigação sobre a prática da publicação de dados pela área da Enfermagem pressupõe a quebra de paradigma no sentido de transformação de mentalidades quanto ao *modus operandi* científico da área. Além disso, entende que a proposta da Ciência Aberta proporciona uma vasta rede de relacionamentos e representações que podem ser oportunizados pela publicação de dados de pesquisa.

Um outro viés observado, ao estudarmos a prática da publicação de dados, foi a característica de "objeto de fronteira"[63]. Tal proposição considerou o seguinte entendimento: ao serem produzidos com uma finalidade específica de pesquisa, em um dado domínio do conhecimento, podem estes mesmos dados serem reutilizados em outras pesquisas, cujos propósitos ou domínios inicialmente não se comunicavam. Embora este não seja um aspecto explorado pela pesquisa, consideramos relevante destacá-lo para sinalizar a possibilidade de verticalização em outros estudos.

O arcabouço teórico da Ciência da Informação proporciona o entrelaçamento de áreas e a composição com outras disciplinas para proposição de novos estudos.

[63] Conceito proposto por Star e Griesemer, 1989.

Desse modo, consideramos oportuno apontar a possibilidade de lançar novos olhares para outras comunidades e práticas de publicação de dados e assim identificar novos *habitus científicus* na contemporaneidade.

Sabe-se que esta recente prática de comunicar a ciência encontra-se em processo de apropriação e experimentação nacional e internacional entre pares, e apostamos que não tardará a se tornar uma realidade nos moldes de publicação científica no Brasil.

Atualmente, o Brasil encontra-se como um participante inicial do processo de publicação de dados de pesquisa em periódicos científicos nacionais. Em algumas áreas do conhecimento, pesquisadores buscam alternativas para solucionar este problema. A publicação em periódicos estrangeiros ou internacionais vêm sendo uma saída para esta lacuna no campo da comunicação científica de autores brasileiros.

O mesmo já não acontece para o depósito do conjunto de dados em repositórios especializados. Este tem se tornado uma prática exigida por periódicos e agências de fomento, e, no contexto brasileiro, já se apresenta como uma realidade em algumas áreas do conhecimento.

Considerando todos os desafios e obstáculos que a prática investigada impõe a área observada, ainda assim podemos afirmar, com base nos dados produzidos pela pesquisa, que a Enfermagem está respondendo positivamente ao apelo do movimento da Ciência Aberta no que se refere à publicação de dados de pesquisa. Tal constatação pode ser considerada como um caminhar rumo às boas práticas em Ciência Aberta e demostra que a área está no centro da discussão que se coloca.

Cabe ressaltar que o olhar informacional direcionado pelo estudo compreende a prática observada como uma oportunidade de comunicar e divulgar a ciência para além dos seus pares. Além disso, a descrição e representação temática da informação que compõe a prática reforça a riqueza da parceria entre o pesquisador e o profissional da informação. Os aspectos críticos disciplinares dos metadados e a importância da parceria entre pesquisadores, profissionais da informação e tecnologia são pilares significativos para garantir o sucesso da prática observada.

Contudo, sabemos que tal prática demanda uma mudança cultural científica para que essa realidade esteja cada vez mais próxima do modo real de comunicar e compartilhar a ciência. A infraestrutura tecnológica para publicação de dados via

periódico especializado atualmente não é suficiente. Os repositórios dão conta dessa demanda, contudo, acredita-se que por meio do amadurecimento das práticas é que serão oportunizados o escoamento da produção científica de dados via periódico e/ou via repositório.

Ambos os processos (publicação via periódico e via repositório) não são antagônicos e deseja-se que ocorram de forma concomitante. Para tal reforçamos a necessidade da curadoria dos dados, sua adequada representação, além da criação de repositórios disciplinares para absorver esse rico universo de dados especializados.

Ainda assim, vislumbra-se na promoção da prática de publicação de dados de pesquisa uma estratégia científica possível para ampliarmos a maneira de comunicar, colaborar e compartilhar a ciência, o que supomos promover maior representatividade científica da Enfermagem no Brasil e no mundo.

REFERÊNCIAS

ALMEIDA, M. de C. Talhas numéricas e o antigo testamento. **Revista Brasileira de História da Matemática**, v. 2, n. 4, p. 119-139, 2002. DOI: 10.47976/RBHM2002v2n4119-139. Disponível em: https://rbhm.org.br/index.php/RBHM/article/view/252. Acesso em: 10 jan. 2022.

ALBAGLI, S.; CLINIO, A.; RAYCHTOCK, S. Ciência aberta: correntes interpretativas e tipos de ação. **Liinc em Revista**, Rio de Janeiro. v. 10, n.2, p.434-450, nov. 2014. Disponível em: https://doi.org/10.18617/liinc.v10i2.749. Acesso em: 02 fev. 2020.

ASSANTE, M.; CANDELA, L.; CASTELLI, D.; TANI, A. Are scientific data repositories coping with research data publishing? **Data Science Journal**, v. 15, n. 6, p. 1–24, 2016. Disponível em: http://dx.doi.org/10.5334/dsj-2016-006. Acesso em: 5 mar. 2020.

ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. **NBR 6022**: informação e documentação -artigo em publicação periódica técnica e/ou científica. ABNT: Rio de Janeiro, 2018.

AUSTIN, C. et al. Key components of data publishing: using current best practices to develop a reference model for data publishing**. International Journal on Digital Libraries**, [*s. l.*], v. 18, n. 2, p. 77–92, 2017. DOI 10.1007/s00799-016-0178-2. Disponível em: http://search-ebscohost-com.ez24.periodicos.capes.gov.br/login.aspx?direct=true&db=aph&AN=122919195&lang=pt-br&site=ehost-live. Acesso em: 5 mar. 2020.

AUSTRALIAN National Data Service. **Data and journals**. Disponível em: https://www.ands.org.au/working-with-data/publishing-and-reusing-data/data-journals. Acesso em: 02 fev. 2019.

AVELINO, L. N. T.; OLIVEIRA, B. M. de; SILVA, E. B. da; SANTOS, S. S. S. dos; FERREIRA, C. M. de Arruda. A pesquisa científica da Enfermagem brasileira: uma abordagem histórica. In: **Encontro Latino Americano de Iniciação Científica**, 14. e Encontro Latino Americano de Pós-Graduação, 10. Universidade do Vale do Paraíba, 2010. Disponível em: http://www.inicepg.univap.br/cd/INIC_2010/anais/arquivos/0563_0804_01.pdf. Acesso em: 11 nov. 2021.

BARNES, L. Verde, Ouro, Diamante, Preto o que tudo isso significa? **Blog Open Book Publishers**, 22 out 2018. Disponível em: https://doi.org/10.11647/OBP.0173.0089 Acesso em: 12 nov. 2021.

BARRETO, A. de A. Uma história da ciência da informação. In: TOUTAIN, L. M. B. B. (Org.). **Para entender a ciência da informação.** Salvador: EDUFBA, 2007. Disponível: https://repositorio.ufba.br/ri/bitstream/ufba/145/1/Para%20entender%20a%20ciencia%20da%20informacao.pdf. Acesso em: 01 jul. 2020.

BLOOM, T.; DALLMEIER-TIESSEN, S.; MURPHY, F.; AUSTIN, C.; WHYTE, A.; TEDDS, J.; NURNBERGER, A.; RAYMOND, L.; STOCKHAUSE, M.; VARDIGAN, M.. **Workflows for Research Data Publishing: Models and Key Components**, 2015. Disponível em: https://zenodo.org/record/20308#.Xs2UMjdKjGg. Acesso em: 01 fev. 2019

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

BORKO, H. Information Science: What is it? **American Documentation**, v.19, n.1, p.3-5, Jan. 1968. (Tradução Livre). Disponível em: https://edisciplinas.usp.br/pluginfile.php/2532327/mod_resource/content/1/Oque%C3%A9CI.pdf. Acesso em: 01 jul. 2022.

BRASIL. Ministério da Transparência e Controladoria-Geral da União. **Open Government Partnership – OGP**: 4º Plano de Ação Nacional em Governo Aberto. Brasília, 2018. Disponível em: http://governoaberto.cgu.gov.br/a-ogp/planos-de-acao/4o-plano-de-acao-brasileiro/4o-plano-de-acao-nacional_portugues.pdf. Acesso em: 01 jun 2020.

BRASIL. Controladoria-Geral da União. **CNPq e IBICT lançam Lattes Data**, 05 fev. 2020. Disponível em: https://www.gov.br/cgu/pt-br/governo-aberto/noticias/2020/2/cnpq-e-ibict-lancam-lattes-data. Acesso em: 01 jul. 2022.

CALLAGHAN, S., MURPHY, F.,TEDDS, J., ALLAN, R.,KUNZE, J.,LAWRENCE, R.,MAYERNIK, M.S., WHYTE , A. Processes and procedures for data publication: a case study in the geosciences. **Int. J. Digital Curation,** v.8, n.1, 2013. Disponível em: doi:10.2218/ijdc.v8i1.253. Acesso em: 19 fev. 2019.

CAMPOS, M. L. de A. **A organização de unidades do conhecimento em hiperdocumentos**: o modelo conceitual como um espaço comunicacional para a realização da autoria. Rio de Janeiro: IBICT/UFRJ, 2001. Tese. Disponível em: http://www.conexaorio.com/biti/tesemlcampos/TeseMLCampos.pdf. Acesso em: 5 mar. 2020.

CAMPOS, M.; FUHR, I. As origens da Matemática e os variados modos de operação com seus conceitos. **Projeção e docência**, v. 8, n.1, p.79-90, 2017.

CANDELA, L.; CASTELLI, D.; MANGHI, P; TANI, A. Data Journals: A Survey. **Journal of the Association for Information Science and Technology**, v. 66, n.

9,1747–1762, 2015. Disponível em: https://onlinelibrary.wiley.com/doi/full/10.1002/asi.23358 Acesso em 02 fev. 2019.

CHAVAN, V; PENEV, L. The data paper: a mechanism to incentivize data publishing in biodiversity Science. **BMC Bioinformatics**, v.12, supl.15, n.S2, 2011. Disponível em: https://doi.org/10.1186/1471-2105-12-S15-S2. Acesso em: 5 mar. 2020.

CHRISTOVÃO, H. T. Da comunicação informal a comunicação formal: identificação da frente de pesquisa através de filtros de qualidade. **Ciência da Informação**, Rio de Janeiro, v. 8, n. 1, 3-36, 1979. Disponível em: https://ridi.ibict.br/bitstream/123456789/373/1/TARDINCI1979.pdf. Acesso em: 02 jul. 2020.

COSTAS, R.; MEIJER, I.; ZAHEDI, Z.; WOUTERS, P. **The Value of Research Data Metrics for datasets from a cultural and technical point of view**. Copenhagen, Denmark: Knowledge Exchange Report, 2013. Disponível em: https://www.researchgate.net/publication/307547241_The_Value_of_Research_Data__Metrics_for_datasets_from_a_cultural_and_technical_point_of_view_A_Knowledge_Exchange_Report_April_2013. Acesso em: 01 fev. 2019.

CRONIN, B. Thinking About Data. **JASIST**, v. 64, n. 3, p. 435–436, 2013. Disponível em: https://doi.org/10.1002/asi.22928. Acesso em: 22 maio 2020.

CURTY, R. G.; AVENTURIER, P. O paradigma da publicação de dados e suas diferentes abordagens. IN: **Encontro nacional de pesquisa e pós-graduação em ciência da informação** - ENANCIB, 18., UNESP – Marília, 2017. Disponível em: http://hdl.handle.net/20.500.11959/brapci/105144. Acesso em: 01 jun. 2020.

DAHLBERG, I. Teoria do conceito. **Ciência da Informação**, Rio de Janeiro, v.7, n.2, p.101-107, 1978.

DECLARAÇÃO da Sorbonne sobre os Direitos de Dados de Pesquisa [Publicado originalmente no site da LERU em janeiro/2020] [online]. **SciELO em Perspectiva**, 2020. Disponível em: https://blog.scielo.org/blog/2020/02/13/declaracao-da-sorbonne-sobre-os-direitos-de-dados-de-pesquisa-publicado-originalmente-no-site-da-leru-em-janeiro-2020/. Acesso em: 02 jun 2020.

DIAS, J.; SZTUTMAN, R.; MARRAS, S. Múltiplos e animados modos de existência: entrevista com Bruno Latour. **Revista de Antropologia**, v. 57, n. 1, p. 499-519, 11 nov. 2014. Disponível em: http://www.revistas.usp.br/ra/article/view/87772/90699. Acesso em: 22 maio 2020.

DILLO, I; HODSON, S; SIERMAN, B et al. **Publishing Data Interest Group** - Cost Models and Cost Recovery Working Group - Draft Case Statement. [2014]. Disponível em: https://www.rd-alliance.org/sites/default/files/RDA-WDS_Publishing_Data_Cost_Models_Recovery_WG_DraftCaseStatement.pdf. Acesso em: 16 maio 2020.

EDMUNDS, S. C., LI, P., HUNTER, C.I. et al. Experiences in integrated data and research object publishing using GigaDB. **International Journal on Digital Libraries**, v.18, 99-111, 2017. Disponível em: https://doi.org/10.1007/s00799-016-0174-6 Acesso em: 13 março 2020.

EYNDEN, V. D.; CORTI, V. L. Advancing research data publishing practices for the social sciences: from archive activity to empowering researchers. . **International Journal on Digital Libraries**, v.18, 113–121, 2017. Disponivel em: https://doi.org/10.1007/s00799-016-0177-3. Acesso em: 13 março 2020.

FARACE, D.; FRANTZEN, J.; SMITH, P.L. Data Papers are Witness to Trusted Resources in Grey Literature: a project use case. IN: Nineteenth International Conference on Grey Literature - **Public Awareness and Access to Grey Literature,** Rome, Italy 23-24 October, v. 19, p. 27-32, 2017. Disponível em: https://doi.org/10.26069/GREYNET-2018-000.002-GG. Acesso em: 22 maio 2020.

FERREIRA, V.B. **E-science e políticas públicas para ciência, tecnologia e inovação no Brasil** [*on-line*]. Salvador: EDUFBA, 2018, 256 p. ISBN: 978-85-232-1865-2. https://doi.org/10.7476/9788523218652.

FOOTE, K. D. **A Brief history of big data.** 14 de dezembro de 2017.Disponível em: https://www.dataversity.net/brief-history-big-data/# Acesso em: 23 set 2021.

FOSTER Open Science. **Open Data Journals**, 2019. Disponível em: https://www.fosteropenscience.eu/taxonomy/term/114. Acesso em: 22 maio 2020.

GALVÃO, T. F.; PEREIRA, M. G. Revisões sistemáticas da literatura: passos para sua elaboração. **Epidemiol. Serv. Saúde**, v.23, n.1, Jan-Mar 2014. Disponível em: https://doi.org/10.5123/S1679-49742014000100018. Acesso em: 02 jun 2020.

GARCÍA-GARCÍA, A.; LÓPEZ-BORRULL, A.; PESET, F. Data journals: eclosión de nuevas revistas especializadas en datos**. El profesional de la información,** v. 24, n. 6, pp. 845-854, 2015. http://dx.doi.org/10.3145/epi.2015.nov.17 Acesso em: 22 maio 2020.

GARCÍA-GARCÍA, A.; LÓPEZ-BORRULL, A.; PESET, F. **Análisis de Data Journals: la eclosión de nuevas revistas especializadas en datos**. (dataset). Disponivel em: http://dx.doi.org/10.6084/m9.figshare. Acesso em: 22 maio 2020.

GONZÁLEZ DE GÓMEZ, M. N. Regime de informação: construção de um conceito. **Informação & Sociedade:** Estudos, João Pessoa, v. 22, n. 3, p. 43-60, set./dez. 2012.

GRAUNT, J. **Natural and political observations mentioned in a following index, and made upon the Bills of mortality / By John Graunt.** With reference to the government, religion, trade, growth, ayre, diseases, and the several changes of the said city. Londres, 1662. Disponível em: https://wellcomecollection.org/works/ear22f7e. Acesso em 08 out. 2021.

HENNING, P. C. et al. Desmistificando os princípios FAIR: conceitos, métricas, tecnologias e aplicações inseridas no ecossistema dos dados FAIR. **Pesq. Bras. em Ci. da Inf. e Bib**., João Pessoa, v. 14, n. 3, p. 175-192, 2019. Disponível em: https://periodicos.ufpb.br/index.php/pbcib/article/view/46969/27455. Acesso em: 02 jun 2020.

HRYNASZKIEWICZ, I; SHINTANI, Y. Scientific Data: an open access and open data publication to facilitate reproducible research.:Johokanri. **Journal of Information Processing and Management**. v. 57, n.9, p. 629-640, 2014. Disponível em: http://dx.doi.org/10.1241/johokanri.57.629. Acesso em: 22 maio 2020.

INSTITUTO BRASILEIRO DE INFORMAÇÃO EM CIÊNCIA E TECNOLOGIA. **Manifesto de Acesso Livre à Informação Científica do IBICT**. Brasília: IBICT, 2005.

Instituto Information Management (IIMA). **Uma breve história do big data desde 18.000 A.C**., 2015. Disponível em: https://docmanagement.com.br/04/22/2015/uma-breve-historia-do-big-data-desde-18-000-a-c/. Acesso em: 23 set 2021.

Instituto Information Management (IIMA). **Manifesto de Acesso Aberto a Dados da Pesquisa Brasileira para Ciência Cidadã**. Brasília: IBICT, 2016. Disponível em: <http://migre.me/v8BZ0>. Acesso em: 01 fev 2019.

Instituto Information Management (IIMA). **Plano de dados abertos**, 2017-2018. Brasília: IBICT, 2017. Disponível em: <http://wiki.dados.gov.br/GetFile.aspx?File=%2fPlanos%20de%20Dados%20Abertos%20Publicados%2fPLANO%20DE%20DADOS%20ABERTOS%20(IBICT).pdf>. Acesso em: 02 fev. 2019.

Kitāb Dīsqūrīdis fī mawādd al-'ilāj كتاب ديسقوريدس في مواد العلاج Dīsqūrīdis (Dioscorides) ديسقوريدس**, Biblioteca Britânica**: Oriental Manuscritos, ou 3366, na Biblioteca Digital Qatar < https://www.qdl.qa/archive/81055/vdc_100022531380.0x000001 > Acesso em 6 de outubro de 2021.

KRATZ, J; STRASSER, C. Data publication: consensus and controversies [version 3; peer review: 3 approved]. **F1000Research,** v.3, n. 94, 2014. Last updated: 16 may 2019. Disponível em:< https://doi.org/10.12688/f1000research.3979.3>. Acesso em: 25 maio 2020.

LANDI, A. et al. The “A” of FAIR – As Open as Possible, as Closed as Necessary. **Data Intelligence** 2020, v. 2, n.1-2, p.47–55, 2020. Disponível em: https://doi.org/10.1162/dint_a_00027. Acesso em: 21 dez 2020.

LATOUR, B. **Reagregando o social**. Salvador: EDUFBA; Bauru: EDUSC, 2012.

LATOUR, B**. Ciência em ação**: como seguir cientistas e engenheiros sociedade afora. São Paulo: Ed. UNESP, 2000.

LAWRENCE, B., JONES, C., MATTHEWS, B., PEPLER, S., CALLAGHAN, S.: Citation and peer review of data: moving toward formal data publication. **Int. J. Digital Curation**, v.6, n.2, 2011. Disponível em: doi:10.2218/ijdc.v6i2.20r Acesso em: 19 fev. 2019.

LÉVY, P. **As tecnologias da inteligência**. São Paulo: Editora 34, 1993.

LI, Kai; GREENBERG, J.; DUNIC, J. Data Objects and Documenting Scientific Processes: an analysis of data events in biodiversity data papers **JASIST - Journal of the Association for Information Science and Technology**, v. 71, n. 2, p.172–182, 2020. Disponível em: https://doi.org/10.1002/asi.24226. Acesso em: 25 maio 2020.

LYNCH, C. Jim Gray's Fourth Paradigm and the Construction of the Scientific Record. In: HEY, Tony; TANSLEY, Stewart; TOLLE, Krist in (Eds.). **The fourth paradigm: data-intensive scientific discovery**. Redmond, Washington: Microsoft Research, 2009. Disponível em: https://www.microsoft.com/en-us/research/wp-content/uploads/2009/10/Fourth_Paradigm.pdf. Acesso em: 01 fev. 2019.

MABE, M. A. Scholarly Communication: a long view. **New Review of Academic Librarianship**. v. 16, n.S1, p.132-144, 2010. Disponível em: https://doi.org/10.1080/13614533.2010.512242. Acesso em: 25 maio 2020.

MCEWEN, M.; WILLS, E. M. Filosofia, ciência e Enfermagem. In: **Bases teóricas para Enfermagem**. Porto Alegre: Artmed, 2009. Disponível em: https://statics-americanas.b2w.io/sherlock/books/firstChapter/124979871.pdf. Acesso em 22 dez 2021.

PR_ISHANGO_01_NL. In: Royal Belgian Institute of Natural Science. Disponível em: https://www.naturalsciences.be/en/content/prishango01nl. Acesso em 22 dez 2021.

NEWMAN, Paul; CORKE, Peter. Data Papers – peer reviewed publication of high quality data sets. **The International Journal of Robotics Research**, v. 28, n. 5, p. 587, may 2009. Disponível em: https://doi.org/10.1177/0278364909104283. Acesso em: 01 fev. 2019.

NIGHTINGALE, F. **Army sanitary administration and its reform under the late Lord Herbert**. London: Printed by McCorquodale & Co., Cardington Street, 1862. Disponível em: https://archive.org/details/armysanitaryadmi00nigh/page/n3/mode/2up. Acesso em: 07 jan. 2022.

PARSONS, M.A.; FOX, P.A. Is data publication the right metaphor? **Data Science Journal**, 12, 2013. Disponível em: https://www.jstage.jst.go.jp/article/dsj/12/0/12_WDS-042/_pdf/-char/en. Acesso em: 25 maio 2020.

PINFIELD, S.; SALTER, J.; BATH, P.A, HUBBARD, B.; MILLINGTON, P., ANDERS, JH e HUSSAIN, A. (2014), Open-Access Repositories Worldwide, 2005–2012: Past Growth, Current Características e possibilidades futuras. **Journal of the Association for Information Science and Technology**, v. 65, 2404-2421. Disponível em: <https://doi-org.ez24.periodicos.capes.gov.br/10.1002/asi.23131>. Acesso em: 12 nov. 2021.

PINTO, F. M. A. G. **Publicação de dados de pesquisa_Enfermagem**. Figshare. [Dataset]. Disponível em: https://doi.org/10.6084/m9.figshare.19790425.v1. Acesso em: 08 jun. 2022.

PINTO, F. M. A. G., PINHEIRO, L. V. R. O periódico de dados e as transformações na comunicação científica. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 20., 2019, Florianópolis. **Anais**. GT-7 – Produção e Comunicação da Informação em Ciência, Tecnologia & Inovação. Disponível em: https://conferencias.ufsc.br/index.php/enancib/2019/paper/view/664/968. Acesso em: 01 fev. 2019.

PINHEIRO, L. V. R. Campo interdisciplinar da Ciência da Informação: fronteiras remotas ne recentes. In: **Ciência da Informação, ciências sociais e interdisciplinaridade**. Brasília: IBICT, 1999.

REIS, C. K. **História da escrita**: uma contextualização necessária para o processo de alfabetização. Uberlândia, MG, 2019. Trabalho de conclusão de curso. Universidade Federal de Uberlândia. Disponível em: https://repositorio.ufu.br/bitstream/123456789/28854/1/Hist%C3%B3riaEscritaUma.pdf. Acesso em: 25 maio 2020.

ROA-MARTÍNEZ ,Sandra M.; VIDOTTI, Silvana A. B.; SANTANA, Ricardo C. Estructura propuesta del artículo de datos como publicación científica. **Revista Española de Documentación Científica**, v.40, n.1, enero-marzo 2017. Disponível em: http://dx.doi.org/10.3989/redc.2017.1.1375 Acesso em: 19 fev. 2019.

SALES, L. F.; SAYÃO, L. F. A ciência invisível: revelando os dados da cauda longa da pesquisa. In: Encontro Nacional de Pesquisa e Pós-graduação em Ciência da Informação - ENANCIB, 19., 2018. **Anais**. Disponível em: http://hdl.handle.net/20.500.11959/brapci/103678. Acesso em: 17 maio 2022.

SALES, L. F.; SAYÃO, L. F. Infraestruturas de gestão de dados de pesquisa: proposta de modelo de avaliação, 2019 . IN: **Edicic 2019, Barcelona**. Disponível em: http://eprints.rclis.org/34472/. Acesso em: 02 jun 2020.

SALLES, E. B.; BARREIRA, I de A. Formação da comunidade científica de Enfermagem no Brasil. **Texto Contexto Enferm**, Florianópolis, jan-mar 2010; v. 19, n. 1, 137-46. Disponível em: https://www.scielo.br/j/tce/a/kgbCHWhdxkJZjw3bcFSRtHQ/abstract/?lang=pt#. Acesso em: 21 dez. 2021.

SAYÃO, L.; SALES, L. A ciência invisível: revelando os dados da cauda longa da pesquisa. In: **ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO**, 20., 2019, Florianópolis. Anais. Disponível em: https://www.researchgate.net/publication/330449075. Acesso em: 01 fev. 2019.

SAMPAIO, F. A. de. **História dos reinos vagetal, animal e mineral do Brasil pertencente a Medicina Tomo II**. Cachoeira, BA: [*s.n*.], 1789. 255 p., 20 estampas. Disponível em: http://objdigital.bn.br/objdigital2/acervo_digital/div_manuscritos/mss22950/mss22950.pdf. Acesso em: 11 out. 2021.

SANTOS, C. Os números primos de *Ishango*. **Revista Brasileira Multidisciplinar**, v. 22, n.2, 2019. Disponível em: DOI: 10.25061/2527-2675/ReBraM/2019.v22i2.638. Acesso em 07 Jan. 2022.

SANTOS, T. C. F.; GOMES, M. L. B. Nexos entre pós-graduação e pesquisa em Enfermagem no Brasil. **Revista Brasileira de Enfermagem** [online]., v. 60, n. 1, p. 91-95, 2007. Disponível em: <https://doi.org/10.1590/S0034-71672007000100017>. Epub 31 Mar 2008. ISSN 1984-0446. https://doi.org/10.1590/S0034-71672007000100017. Acesso em 18 Maio 2022.

SANTOS, S. M. dos; MUGNAINI, R. Comunicação científica em periódicos e a internacionalização das ciências brasileiras. In: CARNEIRO, F. F. B.; FERREIRA NETO, A.; SANTOS, W. dos (Orgs). **Comunicação científica em periódicos**. Curitiba: Appris, 2019. p. 73-93

SCHÖPFEL, J.; FARACE, D.; PROST, H.; ZANE, A. Data papers as a new form of knowledge organization in the field of research data. In: COLLOQUE INTERNATIONAL ISKO-FRANCE 2019: **Données et mégadonnées ouvertes en SHS** : de nouveaux enjeux pour l'état et l'organisation des connaissances?, 12.ISKO France, Oct 2019, Montpellier, France. (Dataset). Disponível em: https://doi.org/10.17026/dans-zk3-jkyb. Acesso em: 04 jun 2020.

SIMMS, S. **Data Publishing–the First 500 Years**. University of California Curation Center (UC3), March 26, 2012. Disponível em: https://uc3.cdlib.org/2012/03/26/data-publishing-the-first-500-years/. Acesso em: 26 maio 2020.

STAR, S. L.; GRIESEMER, J. Institutional ecology, 'Translations', and Bound ary objects: Amateurs and professionals on Berkeley's museum of vertebrate zoology. **Social Studies of Science**, v. 19, n. p. 387-420, 1989.

STOKER, R.; MELZACK, G.; MCLEAN, J. Developing Culturally Competent Data Publication Resources. **International Journal of Digital Curation,** v. 14, n. 1, p. 168–179, 2019. Disponível em: DOI: 10.2218/ijdc.v14i1.648. Acesso em: 26 maio 2020.

SCHWARTZMAN, S. **Um espaço para a ciência**: a formação da comunidade científica no Brasil. Brasília: Ministério de Ciência e Tecnologia, 2001. (Prefácio à nova edição). Disponível em: http://www.schwartzman.org.br/simon/spacept/pdf/Prefacio%20a%20nova%20edicao.pdf. Acesso em 01 jul. 2020.

STUART, D. 2017. Data bibliometrics: metrics before norms. **Online Information Review** , v. 41, 428–435. Disponível em: https://doi.org/10.1108/OIR-01-2017-0008. Acesso em: 01 jul. 2020.

SWAN, A.; BROWN, S. **To share or not to share**: Publication and quality assurance of research data outputs. Report commissioned by the research information network, 2008. Disponível em: https://eprints.soton.ac.uk/266742/1/Published_report_-_main_-_final.pdf. Acesso em: 07 jan. 2022.

SWEET, V. Far more than a lady with a lamp. ***The New Yourk Time***, March 3, 2014. Disponível em: https://www.nytimes.com/2014/03/04/health/florence-nightingales-wisdom.html. Acesso em 07 jan. 2022.

TASCH, B. Marie Curie's belongings will be radioactive for another 1,500 years. **Sciencealert**, 31 de agosto de 2015. Disponível em: https://www.sciencealert.com/these-personal-effects-of-marie-curie-will-be-radioactive-for-another-1-500-years. Acesso em 08 out. 2021.

TORINO, E.; ROA-MARTÍNEZ, S. M.; VIDOTTI, S. A. B. G. Dados de pesquisa: disponibilização ou publicação?. In: SHINTAKU, M.; SALES, L. F; COSTA, M. (org). **Tópicos sobre dados abertos para editores científicos**. Botucatu, SP: ABEC, 2020. p. 183-201. Disponível em: 10.21452/ 978-85-93910-04-3.cap15. Acesso em: 02 fev. 2019.

TOMASELLI, C. The Vienna Dioscurides. **Smarthistory**, 5 abril de 2021. Disponível em: https://smarthistory.org/the-vienna-dioscurides/. Acesso em 07 outubro 2021.

UNIVERSITY OF SIDNEY. **Data publication:** home. Disponível em: https://libguides.library.usyd.edu.au/datapublication. Acesso em: 26 maio 2020.

VALENTIM, M. L. P. Construção do conhecimento científico. In: **Métodos qualitativos de pesquisa em Ciência da Informação**. São Paulo: Polis, 2005.

VEIGA, V. S. de O., HENNING, P., Dib, S., PENEDO, E., LIMA, J. da C., SILVA, L. O. B. da, & Pires, L. F. Plano de gestão de dados fair: uma proposta para a Fiocruz. **Liinc Em Revista**, v. 15, n. 2, 2019. Disponível em: https://doi.org/10.18617/liinc.v15i2.5030. Acesso em 06 jan. 2022.

VIANA, V.; BUCO, C.; SANTOS, T. dos; SOUSA, L. D. Arte rupestre. In: GRIECO, B; TEIXEIRA, L.; THOMPSON, A. (Orgs.). **Dicionário IPHAN de Patrimônio Cultural.**

2. ed. rev. ampl. Rio de Janeiro, Brasília: IPHAN/DAF/Copedoc, 2016. (verbete). ISBN 978-85-7334-299-4.

WARD, P. **Sources of dataset peer review**, 2016. Disponível em: https://www.wiki.ed.ac.uk/display/datashare/Sources+of+dataset+peer+review Acesso em: 26 maio 2020.

WYBORN, L.; LEHNERT, K. Exploiting the long tail of scientific data: Making small data BIG. In: ERESEARCH AUSTRAULASIA CONFERENCE, Melbourne, Australia, 10-14 Oct. 2016. **Anais**. Melbourne, Australia, 2016. Disponível em: https://eresearchau.files.wordpress.com/2016/03/eresau2016_paper_88.pdf. Acesso em: 26 maio 2020.

ZIMAN, J. **Conhecimento público.** São Paulo: Ed. USP, 1979.