

A RELAÇÃO ENTRE OS CONJUNTOS DE DADOS DE PESQUISA DEPOSITADOS EM REPOSITÓRIO DE DADOS E OS PERIÓDICOS CIENTÍFICOS¹

E-mail:
ingridschiessl@gmail.com
fernandam@unb.br

Ingrid Torres Schiessl², Fernanda Passini Moreno³

RESUMO

O estudo tem como objetivo estudar a relação entre os conjuntos de dados depositados em repositórios de dados e os respectivos artigos científicos, durante o período de cinco anos. Para isso foram estudados os conceitos de dados de pesquisa aberto, compartilhamento de dados, princípios e diretrizes para a gestão dos dados de pesquisa, repositório de dados abertos, periódicos científicos, a luz da comunicação científica. A pesquisa se caracteriza como uma pesquisa de descritiva com abordagem quanti-qualitativa, de natureza básica com método indutivo e com o horizonte temporal transversal, pois fará análise de uma quantidade conjuntos de dados em um período específico. Para a coleta de dados será utilizado um *script* para obter a amostra de forma automática, que se dará por meio de *Application Programming Interface* (API), tal recurso é disponibilizado pela DataCite, organização sem fins lucrativos que fornece gratuitamente identificadores persistentes para dados e resultados de pesquisa.

Palavra-chave: Dados de pesquisa abertos. Repositório de dados. Ciência aberta. Identificadores persistentes. Gestão de dados de pesquisa.

ABSTRACT

The study aims to investigate the relationship between data sets deposited in data repositories and their scientific articles over a five-year period. For this purpose we studied the concepts of open research data, data sharing, principles and guidelines for research data management, open data repository, scientific journals, in the light of scientific communication. The research is characterized as a descriptive research with quantitative and qualitative approach, basic in nature with inductive method and with the transverse time horizon, because it will analyze a set of data sets in a specific period. For data collection a script will be used to obtain the sample automatically, which will be done through Application Programming Interface (API), this resource is provided by DataCite, a non-profit organization that provides free persistent identifiers for data and results. of research.

Keywords: Scientific open data. Data repository. Open science. Persistent identifiers. Research data management.

¹ Relatório intermediário de mestrado aprovado em exame de qualificação e desenvolvido no Programa de Pós-graduação em Ciência da Informação da Universidade de Brasília (PPGCINF/UNB).

² Universidade de Brasília. Brasília-DF. Brasil. <https://orcid.org/0000-0001-5815-2574>

³ Universidade de Brasília. Brasília-DF. Brasil. <https://orcid.org/0000-0002-6652-0565>

À medida que o mercado da comunicação acadêmica continua evoluindo, alguns indicadores sugerem que a unidade de moeda da informação está mudando de um foco primário nos artigos de periódicos para uma ênfase mais ampla nos elementos-chave da comunicação acadêmica, isto é, os conjuntos de dados (DAVIS; VICKERY, 2007, tradução nossa). Sendo assim, conforme afirmam Carvalho e Leite (2019) os dados de pesquisa têm introduzido mudanças em relação ao ciclo tradicional da comunicação científica, na medida em que possibilitam aos pesquisadores diferentes maneiras para desenvolver e disseminar suas pesquisas.

A preocupação com os dados de pesquisa levanta questões na comunicação científica, de forma a sobrepor as antigas convenções do que são os produtos da investigação científicas. A comunidade científica, inserida na chamada ciência aberta, vivencia uma nova demanda, que é a disponibilização aberta e de forma inteligível dos dados de pesquisa (CARVALHO; LEITE, 2019).

Disponibilizar dados científicos tratados para pesquisadores não é uma questão nova na CI (SALES; SAYÃO, 2012). Os autores lembram que o Museu Paraense Emílio Goeldi, em fins da década de 1980 e início de 1990, desenvolveu o PRIMATAM, projeto ligado ao Núcleo de Primatologia, cujo tratamento dos dados de pesquisa resultou em um catálogo impresso. Outro exemplo é o projeto Genoma que abriu seus dados à comunicação científica e ao público, desde a década de 1980, International Nucleotide Sequence Database Collaboration (INSDC) mantém o Genbank - um conjunto de bases de dados que trata informações sobre sequenciamento genômico das mais diversas espécies.

Desta forma, observou-se a criação de diferentes iniciativas de banco de dados com o objetivo de compartilhar os recursos (informação, dados, arquivos, objetos etc) ao longo do tempo. As primeiras iniciativas ocorreram em 1966, com Educational Resources Information Center (ERIC) e o PubMed , este último é um motor de busca de livre acesso a base de dados Medline, esta base de dados é gerida pela *National Library of Medicine* e o *National Institutes of Health* e possui citações bibliográficas de periódicos da área de biomédica. Em 1971, foi criado o Projeto Gutenberg com o objetivo de encorajar a produção e a distribuição de e-books (HART, 2007). Em 1974, o *Deutsches Elektronen Synchrotron* e o *Stanford Linear Accelerator Center* (SLAC) começaram a catalogar literatura eletrônica na área de física.

Foi na década de 1990, com a expansão mundial da Internet, que começaram a proliferar os bancos de dados de artigos científicos de livre acesso. Como destaque tem-se em 1996 o surgimento da *Networked Digital Library of Theses and Dissertations* (NDLTD), que passou a ser o maior banco de teses e dissertações do mundo. Como iniciativa da América Latina há a base de periódicos SciELO (*Scientific Electronic Library On-line*) criada pela Bireme - Centro Latino-Americano e do Caribe de Informação em Ciências da Saúde – com apoio da Fundação de Amparo à Pesquisa do Estado de São Paulo (Fapesp). Os repositórios científicos também desempenharam um papel importante no acesso aberto, permitindo a disponibilização de artigos e documentação de pesquisa produzidos nas universidades e nos centros de investigação.

Ao longo do tempo, estudiosos afirmam que o acesso ao conhecimento e à informação são basais para o desenvolvimento humano, social e econômico. No contexto do político brasileiro, há forte consenso no sentido de que o direito de acesso à informação seja considerado

um direito humano fundamental (MACHADO, 2014). Nesse sentido, a abertura dos dados científicos representa enormes benefícios para a sociedade. Além do aumento de iniciativas open na ciência, que levam ao surgimento de novos paradigmas para produção e distribuição do conhecimento, que de certa maneira gera destaque aos dados de pesquisa na comunicação científica.

Dados de pesquisa têm assumido um maior protagonismo na ciência, pois “passaram a ser valorizados como ativos de pesquisa autônomos, de alto valor intrínseco e publicáveis” (GONÇALVEZ; AVENTURIER, 2017). O estabelecimento de métricas para avaliar a significância e impacto da produção e publicação de dados torna-se cada vez imprescindível, uma vez que, o impacto também vem sendo considerado como a moeda do mundo acadêmico (NASSI-CALÒ, 2015).

O acesso aberto tradicional, como foi debatido nos últimos 10 ou 15 anos, ficou ultrapassado, sendo insuficiente para dar conta dos avanços do uso e reuso de uma base de dados aberta (MACHADO, 2014, p. 202). O autor afirma que, atualmente, se difundem protocolos, formatos, plataformas que permitem maior interoperabilidade, processamento, cruzamento e reusos da informação. Desta forma, conhecer as relações e processos acerca do compartilhamento de dados de pesquisa se faz importante, uma vez que novos modelos, processos, paradigmas são criados, uma nova forma de fazer ciência vem se construindo no mundo digital, de acesso aberto às publicações e aos dados.

No contexto do compartilhamento de dados percebe-se que a transformação ocorre para pesquisas, autores/pesquisadores e disciplinas e/ou área do conhecimento. Esse novo paradigma vivido na ciência impõe desafios e cria oportunidades. A pesquisa realizada pela Springer Nature demonstra que os principais desafios enfrentados pelo pesquisadores para o compartilhamento de dados são: tratamento do dados de uma forma apresentável e útil; falta de informação sobre direitos autorais e licenciamento; não saber qual repositório usar; falta de tempo para depositar dados; custos de compartilhamento de dados (STUART et al., 2018).

No que diz respeito às pesquisas, o ato de compartilhar dados permite a validação da metodologia e a elaboração de estudos relacionados ao estudo. Dessa forma, maiores insumos sobre o tema podem ser produzidos e publicados. Além, disso o compartilhamento de dados reduz o tempo de pesquisa, uma vez que a coleta não precisa ser realizada. Portanto, o compartilhamento de experimentos e resultados desestimula o trabalho de forma isolada e otimiza o tempo da pesquisa.

Em relação as áreas do conhecimento, o compartilhamento de dados exige o comprometimento para a elaboração de políticas e normas para disponibilização, de forma a respeitar os critérios éticos, porque as especificidades de cada área impactam na forma de coleta e análise dos dados e, por conseguinte, no próprio dado. Além disso, a abertura dos dados é mais um aliado na detecção de plágios e fraudes e promove a interdisciplinaridade. Pois, a partir do cruzamento de dados é possível gerar novos estudos. Por fim, a abertura dos dados viabiliza a transparência da pesquisa o que facilita o apoio de agências de fomento à pesquisa.

Nesse contexto, o problema de pesquisa pode ser sintetizado pela seguinte pergunta: “Como se caracteriza o cenário sobre o compartilhamento de dados de pesquisa, que tenha relação com artigos científicos publicados?”.

Para responder à questão dessa pesquisa tem-se como objetivo geral: Analisar a relação dos conjuntos de dados depositados em repositórios de dados com os periódicos científicos. Os objetivos específicos são: 1) Compreender como as diferentes áreas do conhecimento

compartilham dados; 2) Analisar o nível de descrição dos dados realizados pelas diferentes áreas do conhecimento; 3) Avaliar as políticas de depósito de dados de pesquisa de periódicos com intuito de verificar a relação entre as políticas e a adesão dos pesquisadores.

2 REVISÃO DE LITERATURA

Comunicar os resultados de pesquisa aos pares e aos leigos, de forma eficiente, incide diretamente no progresso da ciência (MUELLER, 2000). Nesse sentido, a Ciência da Informação tem estudado, desde a sua concepção, questões relacionadas com a comunicação científica, a qual surgiu em virtude da preocupação dos cientistas, dos tecnólogos e dos documentalistas com os fluxos da informação em ciência e tecnologia (COSTA; LEITE, 2018).

De acordo com Leite (2006, p. 66), a comunicação científica “é um processo fundamental e inseparável da atividade científica”. Por isso, as diferentes áreas do conhecimento formalizam as práticas para circulação e trocas de informação entre os pesquisadores e o público geral (PINTO; COSTA, 2018, p. 146). De forma que, para compreender os processos e as práticas construiu-se diversos modelos sobre a comunicação da ciência. Contudo, tais modelos podem variar consoante os interlocutores, as características e os meios que estão ao dispor de quem produz conhecimento (PINTO; COSTA, 2018, p. 147).

As áreas do conhecimento concebem padrões e normas, que em alguns aspectos são semelhantes e em outros diferentes, isto é, existem formas diferentes de “fazer ciência” e, provavelmente, de comunicar a pesquisa. Nesse sentido, em 1979, Kuhn (1979, p. 312) afirma que no processo de pesquisa científica a comunidade científica não funciona de maneira homogênea e seu comportamento está relacionado com uma divisão em grupos.

O modelo de comunicação científica de Garvey e Griffith (Figura 1) é um dos pioneiros, foi desenvolvido para uma área do conhecimento específica, a Psicologia, e considerava os periódicos impressos (COSTA, 1999). A partir de meados da década de 1980, com os avanços tecnológicos e o surgimento dos periódicos eletrônicos, novos modelos foram apresentados. Hurd (2000) acrescentou ao modelo de Garvey e Griffith o componente eletrônico. Assim, os canais informais tradicionais, como o telefone e o contato pessoal, são substituídos pelo correio eletrônico ou pelas listas de discussão (LEITE, 2006).

Costa (1999) propõe um modelo híbrido, pois segundo a autora era necessário fazer a “ponte” entre o impresso e o digital. O estudo desenvolvido pela autora investiga comunicação na área das Ciências Sociais e elucida, que apesar da preponderância dos meios eletrônicos em determinadas fases, a comunicação formal ainda era realizada com base no modelo tradicional, em que livros e periódicos impressos se impõem aos novos formatos digitais (COSTA, 1999).

Outro modelo de comunicação científica, é o modelo UNISIST criado em 1971, que oferece a ótica sócio técnica sobre as atividades de comunicação científica (SØNDERGAARD; ANDERSEN; HJØRLAND, 2003). Segundo os autores, o modelo define um sistema formado por várias unidades organizacionais e documentárias, em que todas colaboram para o compartilhamento de atividades na comunicação científica. Esse modelo dá enfoque para a comunicação da informação entre produtores e usuários de conhecimento.

De acordo com Leite (2011), os canais informais de comunicação “são utilizados quando produtores e usuários conhecem um ao outro e trocam informação por meio de correspondências pessoais, distribuição de pré-prints, indicações de leituras, etc”. Por outro lado, os canais formais

de comunicação são divididos em documentos publicados e documentos não publicados. Por fim, os canais tabulares comunicam dados técnicos e científicos. É por meio deles que os dados são apresentados em forma de tabelas, ao contrário do fluxo linear do texto falado ou escrito, como afirma o autor.

Dentre os modelos de comunicação, há o proposto por Björk (2007), que também divide o processo de comunicação científica em uma parte informal e uma formal. No modelo, é adotada uma visão mais funcional, onde os estágios antes da publicação formal (documentos de trabalho, disponibilização de preprints etc.) são agrupados com a publicação formal tradicional. O diagrama elaborado pelo autor leva em consideração o fato de que os cientistas não apenas publicam documentos textuais, mas também dados e modelos.

Dentre os modelos citados, observa-se a presença do artigo de periódico científico como um canal de comunicação dos resultados de pesquisa. Na literatura acerca do tema, discute-se sobre a importância dos artigos de periódico como um dos principais meios para a disseminação de informações científicas (MEADOWS, 1999; MUELLER, 2000; MUELLER; PASSOS, 2000; PINTO; COSTA, 2018; TENOPIR; KING, 2001).

Os primeiros periódicos científicos de que se tem notícia são o *Journal de Sçavans*, fundado pelo francês Denis de Sallo, primeiro fascículo foi publicado em 5 de janeiro de 1665; e *Philosophical Transactions of the Royal Society*, que surgiu em Londres, este dedicado exclusivamente ao registro das experiências científicas, publicado em menos de três meses após o *Journal de Sçavans* (MUELLER, 2000, p. 74).

O periódico científico se consolidou como um meio de comunicação científica. Por isso, tem como funções: garantir a memória da ciência, apontar o grau de evolução da ciência, estabelecer a propriedade intelectual, legitimar novos campos de estudos e disciplinas, constituir-se como fonte para o início de novas pesquisas e entre outras (BIOJONE, 2003; MARCHIORI; ADAMI, 2005; MEADOWS, 1999; STUMPF, 1996).

Como a tecnologia modificou os meios de comunicação científica, conseqüentemente os periódicos também foram impactados. O desenvolvimento muito rápido da Internet e, em particular, dos serviços disponíveis na rede desde 1994, modificaram profundamente o acesso à informação, inclusive os periódicos. Esse meio de comunicação formal, passou pela migração para o suporte eletrônico, o que modificou tanto o mercado editorial da comunicação científica como as estratégias de visibilidade do conhecimento (WEITZEL, 2015). Em conseqüência, os artigos de periódicos também se modificaram, de acordo com Meadows (1999), tais mudanças ocorreram, pois, era necessário melhorar a eficiência na comunicação entre pesquisadores de comunidades que cresceram e se tornaram mais complexas.

Para Morrison (2012), em sua tese de doutorado, as editoras comerciais possuem uma receita lucrativa. No estudo, o autor apresenta dois fatos, primeiro é rentabilidade da comercialização de periódicos científicos e; o segundo, é que o mercado tem se mostrado em crescimento, apesar das alternativas propostas pelo Movimento de Acesso Aberto. Ainda segundo ao autor, as editoras comerciais alcançaram altas taxas de lucro, entre 32% e 42%, que crescem ao longo dos anos, entre 3% e 13%. Para Costa (2017, p 28), a lucratividade da comercialização da comunicação científica foi um ponto fortemente questionado no início dos anos 2000, nas articulações do Movimento de Acesso Aberto, que propõe uma alternativa a esse modelo.

Os princípios que sustentam a ideia da ciência aberta de acordo com Boulton (2013) são os mesmos que serviram de base para as revoluções científicas dos séculos XVIII e XIX, porque,

em ambos os contextos, a motivação seria tornar a ciência pública. Para o autor, a prática de tornar públicas as evidências que sustentam uma teoria favorecem o princípio científico da refutação e do que ele chamou de *scientific self-correction*.

Nesse sentido, a Ciência Aberta pode significar um conjunto de conceitos, “que vão desde o democrático direito de acesso ao conhecimento (por exemplo, acesso aberto às publicações), até a demanda por incluir o público na pesquisa (por exemplo, ciência cidadã) para o uso de ferramentas de colaboração e de compartilhamento”, como defendem Fecher e Friesike (2014).

No contexto da Ciência Aberta, os dados de pesquisa estão no centro de uma nova discussão, visto que, por muito tempo, os cientistas se preocupavam apenas em disseminar os resultados de pesquisa, como nos modelos de comunicação científica propostos por Garvey e Griffith (1979), Hurd (2000) e Bjork (2005). Mas conforme advoga Molloy (2011), comunicar os dados de pesquisa melhora a ciência, em termos de transparência, reprodutibilidade, eficiência e, em última instância, traz maior benefício para a sociedade.

O surgimento de iniciativas que passam a destacar a importância dos dados de pesquisa, entendidos como processos intermediários, define um novo momento na comunicação científica. Pois, se antes apenas importava os resultados da pesquisa, hoje discute-se que além do produto final da pesquisa os processos intermediários também precisam de visibilidade.

3 PROCEDIMENTOS METODOLÓGICOS

O presente trabalho se caracteriza como uma pesquisa de descritiva. Pois, visa analisar as relações dos conjuntos de dados depositados em repositório de dados com os seus respectivos artigos de periódicos correspondentes. Em relação à abordagem do problema considera-se como uma pesquisa quanti-qualitativa. Porque, a fonte são os conjuntos de dados depositados no repositório e os artigos científicos correspondentes, desta forma apresenta-se a relação entre o mundo objetivo e subjetivo. A pesquisa possui natureza básica, pois, não tem o objetivo gerar conhecimentos para aplicação. A pesquisa possui o horizonte temporal transversal, pois fará análise de uma quantidade conjuntos de dados em um período específico, visto que a coleta na base de dados – e essa cresce diariamente - deverá ser realizada em um dia e um horário estabelecidos. O método será o indutivo, pois a partir da identificação dos conjuntos de dados; dos metadados e elementos utilizados para descrição e; dos artigos de periódicos correspondentes, será possível inferir conclusões.

Para a coleta de dados foram elaboradas folhas de estilo (*scripts*), escritas em linguagem *Python* na ferramenta *Jupyter Notebook*. Essa ferramenta é considerada uma ferramenta de narrativa computacional. De acordo com, Perez e Granger (2014, tradução nossa) “A missão do *Project Jupyter* é criar ferramentas de código aberto para computação científica interativa e a ciência de dados no que tange a pesquisa, a educação e a indústria, com ênfase em usabilidade, colaboração e reprodutibilidade”, portanto, pode-se dizer que a ferramenta está alinhada a filosofia open. Além do mais, a ferramenta é bastante elogiada pelos cientistas de dados pois oferta funcionalidades simples e efetivas (PERKEL, 2018).

Para a coleta de dados foi selecionado a base de dados do DataCite. O DataCite é uma organização global sem fins lucrativos que fornece identificadores persistentes (DOIs especificamente) para dados de pesquisa e outros resultados de pesquisa. As comunidades

científicas juntam-se ao DataCite como membros para poder atribuir DOIs a todos os resultados de pesquisa. Dessa forma, todos os resultados de pesquisa tornam-se detectáveis e os metadados associados são disponibilizados à comunidade. Atualmente, o DataCite possui 170 membros, distribuídos em mais de 20 países, entre repositório de dados, bibliotecas, agências governamentais, universidades etc. O site do DataCite fornece documentação e informações sobre o esquema de metadados e como realizar a coleta via API. Fator que corrobora para elaboração do *script*.

Além do DataCite, outra fonte consultada foi a organização CrossRef. Fundada por editores científicos, no ano de 2000, a Crossref surgiu da necessidade de fornecer um serviço colaborativo central de referência – usando DOIs – que seria acessível e gerenciado pela comunidade global de comunicação científica (WOOD, 2018). Atualmente, a CrossRef é uma organização sem fins lucrativos autossustentável que visa tornar todos os resultados de pesquisa não apenas fáceis de encontrar, citar e vincular, mas também fáceis de avaliar e reutilizar (WOOD, 2018).

O *Directory of Open Access Journals* (DOAJ) ou Diretório de Revistas de Acesso Aberto foi lançado em 2003 na Universidade de Lund, na Suécia, com 300 revistas de acesso aberto. Atualmente, contém cerca. 12000 revistas de acesso aberto cobrindo todas as áreas conhecimento ciência, tecnologia, medicina, ciências sociais e humanidades (DIRECTORY OF OPEN ACCESS JOURNALS, [s.d.]). Sua missão do é aumentar a visibilidade, acessibilidade, reputação, uso e impacto de periódicos de pesquisa acadêmica de qualidade, revisados por pares e de acesso aberto em todo o mundo, independentemente da disciplina, geografia ou idioma (DIRECTORY OF OPEN ACCESS JOURNALS, [s.d.]). É uma organização sem fins lucrativos gerenciada pelo *Infrastructure Services for Open Access C.I.C.* com sede no Reino Unido, e depende inteiramente de doações voluntárias de seus membros e de verbas de patrocínio recebidas.

Os procedimentos metodológicos adotados para a coleta dos dados podem ser divididos em duas etapas:

Etapa 1 - Elaboração do script

Passo 01 - Recorte temporal da base de dados

Passo 02 – Definição dos metadados para coleta no DataCite

Passo 03 – Consulta na agência CrossRef e no DOAJ

Etapa 2 - Exportação dos dados para planilhas

Passo 1 - Tratamento dos campos Data

Passo 2 - Aplicação do recurso filtro do editor de planilhas

A partir dos dados tratados, isto é, a conclusão da Etapa 2, espera-se que seja possível realizar as análises, de forma a averiguar:

1. A quantidade de conjunto de dados do tipo dataset;
2. A quantidade de conjuntos de dados com documentos relacionados;
3. A quantidade de conjuntos de dados compartilhados por ano;
4. A quantidade de identificadores persistentes (DOI) válidos e inválidos;
5. O histórico do compartilhamento de dados por período;
6. Qual momento há mais compartilhamento.
7. Consultar os DOIs dos documentos relacionados em agências DOI, o que permitirá a

8. identificação do tipo de documento, ou seja, se são artigos de periódicos;
9. Identificar a quais revistas os artigos pertencem;
10. Identificando as revistas é possível coletar os metadados sobre a revista, principalmente, a respeito da área do conhecimento.
11. Identificar quais delas estão presentes do Directory of Open Access Journals (DOAJ), de forma a verificar se o compartilhamento de conjunto de dados ocorre em revistas de acesso aberto;
12. Identificar a área do conhecimento com maior número de conjunto de dados compartilhados.

O script foi elaborado para realizar o recorte na base de dados do DataCite. O primeiro critério estabelecido foi delimitar o período de 10 anos (2008-2019). Devido a problemas com a comunicação via *Application Programming Interface* (API). Observou-se que a requisição máxima permitida era de 300 registros por ano. Entende-se registro como qualquer arquivo depositado nos repositórios de dados membros do consórcio DataCite

4 CONSIDERAÇÕES PRELIMINARES

A breve discussão apresentada acerca dos temas pertinentes ao estudo divide-se em dois elementos. O primeiro relaciona-se com o desenvolvimento da comunicação científica, perpassando os modelos de comunicação, os meios de comunicação com ênfase no periódico científico, até as discussões sobre ciência aberta, com enfoque nos dados abertos. O segundo elemento diz respeito aos fatores envolvidos no processo de comunicação dos dados de pesquisa, definindo o conceito de dado de pesquisa e apresentando o contexto do compartilhamento de dados e repositórios de dados.

O primeiro tópico da fundamentação teórica trata da comunicação científica. Aborda a essência dos processos da comunicação científica para o próprio “fazer ciência”. Em seguida, aponta os modelos de comunicação sistematizados ao longo do tempo, com enfoque nos três processos basilares, uso, produção e compartilhamento de informação científica, enfatizando-se os dados de pesquisa. Na discussão sobre a comunicação científica foi ainda abordado o tema da ciência aberta, como fenômenos atuais e relevantes para a discussão da comunicação científica na contemporaneidade.

O segundo aspecto da fundamentação teórica desse estudo tratou especificamente da comunicação dos dados de pesquisa. Primeiramente, definiu-se elementos essenciais para a discussão, tal como o conceito de dado de pesquisa e suas formas de classificação. Na discussão observou-se que a definição de conceitos e classificações é formulada dentro das áreas do conhecimento, respeitando as diferenças de cada uma. Em seguida, abordou o tema de compartilhamento de dados, levantando as questões sobre políticas de compartilhamento, quais existem e quem as desenvolve; também aborda sobre curadoria dos dados de pesquisa, visto que o compartilhamento de dados está inserido nesse tema. Assim, como há diferentes definições e classificação para cada área do conhecimento, o mesmo ocorre para o compartilhamento de dados de pesquisa. As áreas do conhecimento se desenvolveram de maneiras diferente em relação a esse processo. Finalmente, o último tópico é sobre repositório de dados abertos, o qual traz a caracterização e as iniciativas a respeito.

A literatura acerca dos temas pertinentes ao estudo, possibilitaram analisar, conforme afirmam Sayão e Sales (2012) que a ciência orientada por dados cria um ponto de inflexão no ciclo tradicional da comunicação científica. Observa-se que o conceito de publicação científica pode ser entendido como um conglomerado de produtos científicos, isto é, tornam visíveis tanto os resultados quanto o processo para atingi-los (PAMPEL et al., 2013). No domínio da comunicação dos dados de pesquisa são inúmeras as reflexões que se podem fazer face aos impactos do compartilhamento de dados de pesquisa, reuso de dados de pesquisa, da publicação e da citação de coleções de dados. De uma forma geral, a comunicação de dados de pesquisa promove uma certa velocidade ao ciclo da comunicação científica, na medida em que, oferta aos pesquisadores diversos dados, ou seja, dados tratados, que asseguram a fidedignidade de seu significado e a reconstrução correta.

Em relação aos procedimentos metodológicos, entende-se que a amostra é aleatória, portanto, não pode se dizer que partir dela poderá ser possível inferir conclusões acerca da população de artigos e/ou conjunto de dados. Entretanto, ela representa uma porção específica, útil ao objeto deste estudo.

Foram executados pré-testes com o script, deste modo, observou-se que o DataCite não permite a extração de todos os conjuntos de dados, no período definido, isto é, não é possível extrair a base completa. O recurso API disponibilizado possui restrições quanto a quantidade de dados a serem coletados. Entretanto, é possível extrair a quantidade de aproximadamente 47 mil conjuntos de dados, cabe ressaltar que os itens recuperados são fornecidos aleatoriamente.

Outra forma de abordagem foi pesquisar em revista intituladas Open Journal, pois na literatura afirma-se que o compartilhamento de dados de pesquisa ocorre também por causa da ideologia do autor, isto é, o pesquisador é a favor do movimento de acesso aberto e; entende e apoia o compartilhamento dos dados. Então partindo do pressuposto que pesquisadores com essa ideologia publicam em revistas de acesso aberto, realizou-se testes para recuperar os conjuntos de dados a partir dos artigos científicos publicados em periódicos com tais características. Mas como as políticas de depósito de dados ainda estão sendo construídas e a cultura de compartilhamento ainda está se estabelecendo, a coleta automática não foi possível e, portanto, a análise inviável.

REFERÊNCIAS

BIOJONE, M. R. **Os periódicos científicos na comunicação da ciência**. São Paulo: Educ/Fapesp, 2003.

BJÖRK, B. C. A lifecycle model of the scientific communication process. **Learned Publishing**, v. 18, n. 3, p. 165–176, 1 jul. 2005.

BJÖRK, B.-C. A model of scientific communication as a global distributed information system. **Information Research**, v. 12, n. 2, 2007

BOULTON, G. Reinventing Open Science for the 21st Century. In: RODRIGUES, E.; SWAN, A.; BAPTISTA, A. A. (Eds.). **Uma década de acesso aberto na UMinho e no Mundo**. Lisboa: Universidade do Minho, 2013. p. 239–250.

CARVALHO, É. R. S. DE; LEITE, F. C. L. Diferenças na produção, no compartilhamento e no (re)uso de dados de pesquisa: a percepção de pesquisadores de Química, Antropologia e Educação. **Em Questão**, v. 25, n. 3, p. 321–347, 2019.

COSTA, S. M. DE S.; LEITE, F. C. L. Imbricações teóricas entre comunicação e gestão da informação e do conhecimento na ciência da informação. In: COSTA, S. M. DE S.; LEITE, F. C. L.; TAVARES, R. B. (Ed.). **Comunicação da informação, gestão da informação e gestão do conhecimento**. Brasília: Ibict, 2018. p. 17–41.

COSTA, S. M. DE S. **The impact of computer usage on scholarly communication amongst academic social scientists**. [s.l.] Loughborough University, 9 set. 1999.

DAVIS, H. M.; VICKERY, J. N. Datasets, a Shift in the Currency of Scholarly Communication: Implications for Library Collections and Acquisitions. **Serials Review**, v. 33, n. 1, p. 26–32, 6 mar. 2007.

GARVEY, W. D.; GRIFFITH, B. C. Communication and information processing within scientific disciplines: empirical findings for psychology. In: GARVEY, W. D. (Ed.). **Communications, the essence of science: facilitating information exchange students**. Oxford: Pergamon Press, 1979. p. 127–174.

GONÇALVES, R. C.; AVENTURIER, P. **O paradigma da publicação de dados e suas diferentes abordagens**. XVIII ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO. Anais...Marília: ENANCIB, 2017Disponível em: <http://enancib.marilia.unesp.br/index.php/xviiienancib/ENANCIB/paper/viewFile/468/820>. Acesso em: 13 ago. 2019.

HART, M. **Project Gutenberg mission statement**. Disponível em: http://www.gutenberg.org/wiki/Gutenberg:Project_Gutenberg_Mission_Statement_by_Michael_Hart. Acesso em: 13 ago. 2019.

HURD, J. M. The transformation of scientific communication: A model for 2020. **Journal of the American Society for Information Science**, v. 51, n. 14, p. 1279–1283, 1 jan. 2000.

KAUARK, F.; MANHÃES, F. C.; MEDEIROS, C. H. **Metodologia da pesquisa: guia prático**. Itabuna: Via Litterarum, 2010.

KUHN, T. S. Reflexões sobre os meus críticos. In: LAKATOS, I.; MUSGRAVE, A. (Eds.). **A crítica e o desenvolvimento do conhecimento**. São Paulo: Cultrix, 1979. p. 285–343.

LEITE, F. C. L. **Gestão do conhecimento científico no contexto acadêmico: proposta de um modelo conceitual**. [s.l.] Universidade de Brasília, 2006.

LEITE, F. C. L. **Modelo genérico de gestão da informação científica para instituições de pesquisa na perspectiva da comunicação científica e do acesso aberto**. [s.l.] Universidade de Brasília, 2011.

MACHADO, J. Dados abertos e ciência aberta. In: ALBAGLI, S.; MACIEL, M. L.; ABDO, A. H. (Eds.). . **Ciência aberta, questões abertas**. Brasília: IBICT, 2014. p. 201–228.

MARCHIORI, P. Z.; ADAMI, A. Autoria e leitura de artigos por docentes pesquisadores: motivações e barreiras. In: FERREIRA, S. M. S. P.; TARGINO, M. DAS G. (Eds.). **Preparação de revistas científicas: teoria e prática**. São Paulo: Reichmann & Autores, 2005. p. 73–100.

MEADOWS, A. J. **A comunicação científica**. Brasília: Briquet de Lemos, 1999.

MOLLOY, J. C. The Open Knowledge Foundation: Open Data Means Better Science. **PLoS Biology**, v. 9, n. 12, p. e1001195, 6 dez. 2011.

MORRISON, H. **Freedom for scholarship in the internet age**. Tese de doutorado—Canadá: Simon Fraser University, 2012.

MUELLER, S. P. M.; PASSOS, E. J. L. As questões da comunicação científica e a ciência da informação. In: MUELLER, S. P. M.; PASSOS, E. J. L. (Eds.). . **Comunicação científica**. [s.l.] Ciência da Informação, 2000. p. 13–22.

MUELLER, S. P. M. O periódico científico. In: SANTOS, B.; CAMPELLO, B. V.; CENDÓN, J. M. K. (Eds.). **Fontes de informação para pesquisadores e profissionais**. Belo Horizonte: Ed. UFMG, 2000. p. 73–96.

NASSI-CALÒ, L. Projeto Making Data Count incentiva compartilhamento de dados de pesquisa [online]. **SciELO em Perspectiva**, 2015.

PAMPEL, H. et al. Making research data repositories visible: The re3data.org registry. **PLoS ONE**, v. 8, n. 11, 4 nov. 2013.

PEREZ, F.; GRANGER, B. E. Project Jupyter: computational narratives as the engine of collaborative data science. **Jupyter Blog**, v. 11, n. 207, 2014

PERKEL, J. M. Why Jupyter is data scientists' computational notebook of choice. **Nature**, v. 563, n. 7729, p. 145–146, 30 nov. 2018.

PINTO, C. S.; COSTA, J. L. Padrões de comunicação em diferentes comunidades científicas. In: COSTA, S. M. DE S.; LEITE, F. C. L.; TAVARES, R. B. (Eds.). . **Comunicação da informação, gestão da informação e gestão do conhecimento**. Brasília: Ibict, 2018. p. 145–159.

SALES, L. F.; SAYÃO, L. F. O impacto da curadoria digital dos dados de pesquisa na Comunicação Científica. **Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação**, v. 0, n. 0, p. 118–135, 18 dez. 2012.

STUART, D. et al. **Whitepaper: practical challenges for researcher in data sharing**. [s.l.: s.n.].

STUMPF, I. R. C. Passado e futuro das revistas científicas. **Ciência da Informação**, v. 25, n. 3, 1996.

SØNDERGAARD, T. F.; ANDERSEN, J.; HJØRLAND, B. Documents and the communication of scientific and scholarly information revising and updating the UNISIST model. **Journal of Documentation**, v. 59, n. 3, p. 278–320, 12 jun. 2003.

TENOPIR, C.; KING, D. W. A importância dos periódicos para o trabalho científico. **Revista de Biblioteconomia de Brasília**, v. 25, n. 1, p. 15–26, 2001.

WEITZEL, S. DA R. E-prints: modelo da comunicação científica em transição. In: FERREIRA, S. M. S. P; TARGINO, M. DAS G. (Ed.). **Preparação de revistas científicas: teoria e prática**. São Paulo: Reichman & Autores Editores, 2015. p. 161–193.