

Os diferentes conceitos de dados de pesquisa na abordagem da biblioteconomia de dados

Alexandre Ribas Semeler

Pós-Doutorado pela Universidade Federal de Santa Catarina (UFSC) - Brasil. Doutor em Ciência da Informação pela Universidade Federal de Santa Catarina (UFSC) – SC - Brasil. Bibliotecário de Dados e Gerente de Rede da Universidade Federal do Rio Grande do Sul (UFRGS) - Porto Alegre, RS – Brasil.

<http://lattes.cnpq.br/1564154234554393>

E-mail: alexandre.semeler@ufrgs.br

Adilson Luiz Pinto

Pós-Doutorado pelo Institut de Recherche en Sciences de l'Information et de la Communication (IRSIC) - França. Doutor em Documentación pela Universidad Carlos III de Madrid (UC3M) - Espanha. Professor da Universidade Federal de Santa Catarina (UFSC) - Florianópolis, SC – Brasil.

<http://lattes.cnpq.br/4767432940301118>

E-mail: adilson.pinto@ufsc.br

Data de submissão: 02/10/2018. Data de aprovação: 23/01/2019. Data de publicação: 03/05/2019.

RESUMO

A palavra dados significa uma peça única de informação. Os dados de pesquisa são o resultado de qualquer investigação sistemática que envolva processos de observação, experimentação ou simulação de procedimentos de pesquisa científica. Os estudos sobre dados de pesquisa em biblioteconomia têm origem na metade dos anos 2000. São o resultado do empenho de bibliotecários norte-americanos, britânicos e canadenses em criar serviços e produtos de divulgação, de consultoria, de gerenciamento, preservação e elaboração de esquemas de metadados para efetiva incorporação de dados de pesquisa em coleções de documentos (livros, artigos, relatórios, entre outros). O objetivo deste ensaio é delimitar o conceito de dados de pesquisa e relacioná-lo às diversas disciplinas científicas, ressaltando sua atual relevância para consolidação de uma biblioteconomia orientada ao uso de dados. Está fundamentado na visão do que significam dados de pesquisa para as ciências sociais e para as ciências exatas. Por fim, conclui-se que a ciência com uso intensivo de dados institui nova dimensão para os bibliotecários, a biblioteconomia de dados. Logo, o bibliotecário de dados precisa aprender mais a ciência orientada a dados. A tarefa essencial dos bibliotecários de dados será a de transformar suas atividades quotidianas em bibliotecas, inserindo os dados de pesquisa no contexto da geração de serviços e produtos de informação científica.

Palavras-chave: Dados de pesquisa. Dados de pesquisa disciplinares. Biblioteconomia de dados. Bibliotecário de dados.

The different concepts of research data in the approach to data librarianship

ABSTRACT

The data means a single piece of information. Research data are the result of any systematic investigation involving observation, experimentation, or simulation processes of scientific research procedures. Studies of data in librarianship originate in the mid-2000s. They are the result of the efforts of USA, British and Canadian librarians to create services and products for dissemination, consultancy, management, preservation and elaboration of metadata schemes for effective incorporation of research data into collections of documents (books, articles, reports, among others). The objective of this essay is to delimit the concept of research data and to relate it to the different scientific disciplines, highlighting its current relevance for the consolidation of Librarianship oriented to the use of data. The essay being based on the vision of what research data are for Social Sciences and Hard Sciences. Finally, it is concluded that data-intensive science sets a new dimension for librarians, the Data Librarianship. In this way, the data librarian needs to learn more about this science. The essential task of data librarians will be to transform their day-to-day activities into libraries by inserting research data in the context of the generation of services and products on scientific information.

Palavras-chave: *Research data. Disciplinary research data. Data librarianship. Data librarian.*

Los diferentes conceptos de datos de investigación en el enfoque de la biblioteconomía de datos

RESUMEN

Los datos significan una sola pieza de información. Los datos de investigación son el resultado de cualquier investigación sistemática que involucra procesos de observación, experimentación o simulación de procedimientos de investigación científica. Los estudios sobre datos en Biblioteconomía tienen su origen en la mitad de los años 2000. Son el resultado del compromiso de bibliotecarios norteamericanos, británicos y canadienses en crear servicios y productos de divulgación, de consultoría, de gestión, preservación y elaboración de esquemas de metadatos efectiva incorporación de datos de investigación en colecciones de documentos (libros, artículos, informes, entre otros). El objetivo de este ensayo es delimitar el concepto de datos de investigación y relacionarlo con las diversas disciplinas científicas, resaltando su actual relevancia para la consolidación de Biblioteconomía orientada al uso de datos. Siendo el ensayo fundamentado en la visión de lo que son datos de investigación para las Ciencias y para las Ciencias Exactas. Por último, se concluye que la ciencia con uso intensivo de datos instituye una nueva dimensión para los Bibliotecarios, la Biblioteconomía de datos. De esta forma, el bibliotecario de datos necesita aprender más sobre esta ciencia. La tarea esencial de los bibliotecarios de datos será la de transformar sus actividades cotidianas en bibliotecas, insertando los datos de investigación en el contexto de la generación de servicios y productos de información científica.

Palavras-chave: *Datos de investigación. Datos de investigación disciplinaria. Bibliotecología de datos. Bibliotecario de datos.*

INTRODUÇÃO

A geração de dados está atrelada a todas as coisas usadas no dia a dia. Os dados são colecionados sobre qualquer coisa, a qualquer momento e em qualquer lugar. Por exemplo, um smartphone possui uma série de sensores que capturam, compartilham e recebem dados sobre nossas ações do dia a dia: batimentos cardíacos, contatos pessoais, rotas de trânsito, entre outras informações relevantes para a indústria, a ciência e o setor de negócios.

No entanto, o que são dados? Como definir essa questão com base na aplicação de conceitos, técnicas e metodologias de uma ciência orientada ao seu uso intensivo e principalmente no contexto das bibliotecas? A priori, para este ensaio, dados serão considerados como todo objeto criado em formato digital (*digital-born*¹) ou convertido para o formato digital (*digitalizados*²) que possa ser usado para geração de insights de informação e conhecimento.

Assim, o empoderamento das tecnologias de coleta, manipulação, análise e visualização de dados torna-se uma questão relevante e atual seja para a indústria, seja para a ciência ou seja para os negócios, tornando-se uma variável de transformação comum a quase todas as teorias e práticas contemporâneas. Por exemplo, as tecnologias do Big Data, que estarão relacionadas a conjuntos de dados tão grandes, complexos, diversificados e desorganizados, que não é possível aplicar as antigas técnicas computacionais convencionais para armazenamento, processamento, compartilhamento e preservação dos dados, referindo-se a dados tão imensos que não se encaixam em um só computador. Esses dados são tão desestruturados que não se adequam a linhas e colunas dos bancos de dados tradicionais; possuem fluxo tão contínuo que não podem ser armazenados em um *datawarehouse*³ estático.

No entanto, o volume de dados não é a chave para compreender a tecnologia do Big Data, mas sim a análise e a conversão dos dados em insights, inovações e produtos de dados. Essa tecnologia de dados não se preocupa somente em coletar grandes volumes de dados, mas ela busca transformar os dados em valor real, sendo as dimensões que o consolidam os fenômenos relacionados a propriedades específicas dos dados, como volume, velocidade e variedade (FEDERER, 2016; DAVENPORT, 2014; BURLINGAME; NIELSEN, 2014; VAN DER AALST, 2014).

No que tange à biblioteconomia de dados, os estudos sobre dados têm origem na metade dos anos 2000. São o resultado do empenho de bibliotecários norte-americanos, britânicos e canadenses em criar serviços e produtos de divulgação, de consultoria, de gerenciamento, preservação e elaboração de esquemas de metadados⁴ para efetiva incorporação de dados em coleções de documentos (livros, artigos, relatórios, entre outros). Embora a biblioteconomia de dados tenha suas origens na criação de serviços de qualquer tipo de dados, como os Data Archives Services no Reino Unido e as Data Library Services nos Estados Unidos e no Canadá na década de 1960, o foco da biblioteconomia de dados hoje está mais direcionado à criação de serviços e/ou produtos relacionados ao uso e ao consumo de dados de pesquisa (RICE; SOUTHALL, 2016; KOLTAY, 2017; TENOPIR et al., 2017).

Nesse sentido, dados de pesquisa científica são o resultado de qualquer investigação sistemática que envolva processos de observação, experimentação ou simulação de procedimentos de pesquisa científica. Os dados de pesquisa dependem do assunto ou da disciplina científica e podem diferir segundo suas tipologias.

¹ Produtos binários criados em ambientes computacionais digitais e sem referência ao mundo físico.

² Reprodução digital de um documento ou objeto físico.

³ Ferramenta de apoio ao processo de extração de conhecimento. Funciona como um banco de dados cuja função é proporcionar aos seus usuários uma única fonte de informação a respeito dos seus negócios. Tem como objetivo básico satisfazer às necessidades dos usuários quanto ao armazenamento dos dados que servirão para realizar consultas e análises necessárias para o apoio a tomada de decisão.

⁴ O termo metadados, segundo Caplan (2003, p. 10), refere-se “[...] a todos os dados utilizados para auxiliar na identificação, descrição e localização de recursos eletrônicos em rede [...]”. A autora (2003, p. 14) explica que “[...] o prefixo meta é comumente usado para significar algo sobre algo, por exemplo, uma metalinguagem usada para descrever outra linguagem [...]”. Eles são informações que se encontram acopladas, embutidas ou diretamente associadas na maioria dos objetos digitais. Servem para descrição, administração e mapeamento da estrutura e do conteúdo desses objetos digitais (CAPLAN, 2003; MOORE; MOREAU, 2001).

Os dados de pesquisa precisam ser identificáveis, citáveis, visíveis, recuperáveis, interpretáveis, contextualizáveis, interoperáveis e reutilizáveis onde quesitos de consistência e procedência são considerados. Dados de pesquisa são objetos digitais e não digitais, como documentos, questionários, avaliações, registros de casos, protocolos de estudo, planilhas, notas de laboratório, notas de campo, diários, filmes, imagens, arquivos digitais de áudio e vídeo, sequências genéticas, coordenadas geográficas, banco de dados, algoritmos, metodologias, protocolos, entre outros tipos de manifestação de pesquisa (ORGANIZATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT, 2007; XIA; WANG, 2014; SAYÃO; SALES, 2015; KOLTAY, 2017; DUDZIAK, 2016; INTERNATIONAL ASSOCIATION FOR SOCIAL SCIENCE INFORMATION SERVICES AND TECHNOLOGY, 2016; HENDERSON, 2017).

Os dados de pesquisa gerados pelas mais diversas disciplinas estão em ascensão no que tange aos estudos sobre a biblioteconomia de dados, principalmente em estudos que se preocupam com o ciclo de vida dos dados⁵ e com o Big Data. Nesse contexto, inclui-se o bibliotecário que, tradicionalmente, desenvolve suas atividades com base no armazenamento, na circulação, na preservação e na gestão de acervos documentais. Deve se preocupar com os estudos sobre os dados no contexto da biblioteconomia; acredita-se que ele possa aplicar suas habilidades tradicionais para oferecer serviços e/ou produtos de dados de pesquisa, como no contexto relacionado às tecnologias como o Big Data ou aos repositórios digitais.

O bibliotecário deve se focar nos dados de pesquisa, pois eles são parte da ciberinfraestrutura⁶ que deve ser investigada para geração de inovação em bibliotecas.

O bibliotecário de dados deve investigar os sistemas e as normas que envolvem a pesquisa com dados de pesquisa em repositórios digitais, como um profissional que deve buscar conhecimentos sobre coleta, manipulação, análise e visualização de dados de pesquisa para oferecer serviços e produtos ligados à gestão e curadoria de dados de pesquisa. Assim, o bibliotecário de dados deve possuir as habilidades necessárias para trabalhar com qualquer tipo de dados, sejam eles observacionais, computacionais ou experimentais. Ele deve compreender como as investigações sobre tecnologia de dados modificam as práticas e as teorias que fundamentam a biblioteconomia tradicional.

As transformações na biblioteconomia tradicional são amplamente influenciadas pelos dados de pesquisa. Cunha (2010) investigou os desafios e a missão das bibliotecas ligadas às inovações tecnológicas e os grandes volumes de dados científicos. Na época o autor apontou a e-Science⁷ como um fator de relevância para o surgimento de novo acervo ligado aos dados de pesquisa (científicos).

⁶ [...] o termo cyberinfrastructure denota a camada de informações, conhecimentos, padrões, políticas, ferramentas e serviços que são compartilhados amplamente pelas comunidades de pesquisa, mas desenvolvidos para fins acadêmicos específicos: a cyberinfrastructure é algo mais específico do que a própria rede, é algo mais geral do que uma ferramenta ou um recurso desenvolvido para um projeto específico [...] Inclui diversas coleções de fontes primárias como bibliotecas, arquivos e museus; bibliografias, auxiliares de pesquisa, sistemas de citações e concordâncias que tornam essa informação recuperável; Os padrões incorporados nos sistemas de catalogação e classificação; as revistas e a imprensa universitária que distribuem a informação; os editores, bibliotecários, arquivistas e curadores que vinculam a operação desta estrutura com os estudiosos que a utilizam. (AMERICAN COUNCIL OF LEARNED SOCIETIES, 2006, p. 1-6, tradução nossa).

⁷ [...] as discussões sobre e-Science no âmbito mundial da ciência da informação encontram-se em estágio inicial, sendo que no Brasil, especificamente, encontra-se numa fase embrionária. Pesquisas precisam ser desenvolvidas para identificar como os dados produzidos em estudos conduzidos nas universidades federais, nos institutos federais, centros de pesquisa ou laboratórios de pesquisa estão sendo tratados. O bibliotecário, principalmente aquele que exerce suas atividades em universidades ou instituições científicas e tecnológicas, precisa se perguntar: como organizar esses dados produzidos nas universidades e centros de pesquisa? Quais serão as habilidades requeridas para o profissional da informação que deseja trabalhar com organização de dados. (COSTA; CUNHA, 2014, p. 203).

⁵ "The research data lifecycle can be depicted in multiple different ways from simple to complex. It provides a useful way to think through the various different stages that data go through during a research project. Key phases of the research data lifecycle created by the Data Documentation Initiative include: Discovery and Planning, Initial Data Collection, Data Preparation and Analysis, Publication and Sharing, and Long-term Management" (TIBBO; JONES 2017, on-line).

Para Cunha (2010), a inclusão desse tipo de acervo em bibliotecas exigirá o treinamento dos bibliotecários que devem assumir novas funções.

Sob o ponto de vista de Cunha (2010); Costa, Cunha (2014), acrescenta-se que as atuais mudanças no modo de conduzir a pesquisa científica oferecem oportunidades para o bibliotecário se envolver com o tratamento de dados. A tarefa do bibliotecário é a gestão de dados. Tal função, já discutida por Costa, Cunha (2014), está diretamente relacionada com a extração de conhecimento-informação de grandes volumes de dados de investigação científica, com o objetivo de identificar relações associativas e correlações existentes em conjuntos de dados que compõem os diversos sistemas onde a informação científica circula.

Assim, os bibliotecários devem focar seus esforços para os dados de pesquisa digitais, pois cada vez mais eles proliferaram devido aos avanços tecnológicos da informação e da computação ligados a fazer ciência, sendo que o uso crescente da tecnologia é uma das principais fontes geradoras de dados de pesquisa.

Nesse contexto, considera-se que os dados de pesquisa são resultados dos mais diversos processos de investigação mediada por dispositivos técnicos, o que transforma todas as etapas da pesquisa, desde a geração de dados, o que pode se dar por meio de um experimento, como um estudo empírico sobre a observação de fenômenos culturais, até a publicação de resultados de pesquisa em um repositório de dados de pesquisa⁸.

Logo, o objetivo deste ensaio será o de apresentar o conceito de dados de pesquisa, sendo o ensaio uma proposta teórica para sistematização dos dados de pesquisa em ciências sociais e ciências exatas.

⁸ Um repositório de dados de pesquisa é um sistema de informação técnico e organizacional que serve para auxiliar pesquisadores na gestão e no armazenamento de dados e também deve facilitar a busca e o acesso a dados de pesquisa em uma ou diversas fontes internas e/ou externas ao repositório. Os repositórios de dados de pesquisa são parte essencial da ciberinfraestrutura de pesquisa e visam à preservação, ao acesso em longo prazo e à reutilização de dados de pesquisa (KINDLING et al., 2017).

Com base em uma pesquisa bibliográfica, a proposta investigativa será discutida à luz da literatura nacional e internacional sobre o tema, ressaltando-se a relevância dos dados de pesquisa para consolidação de uma biblioteconomia orientada ao uso de dados de pesquisa.

A seguir apresenta-se o conceito de dados e dados de pesquisa que fundamentam a proposta deste artigo.

O CONCEITO DE DADOS

A palavra dados (*data*)⁹ surge na metade do século XVII; plural do latim *datum*, significa uma única peça de informação. O *Oxford English Dictionary on-line*¹⁰ define dados como fatos e/ou estatísticas reunidas para referência ou análise de algo, as quantidades (informação estatística), os caracteres e os símbolos de operações lógicas executadas por um computador (software). Também é o registro dessas operações armazenado, preservado e transmitido sob forma de sinal elétrico (digital) que possa ser gravado em qualquer tipo de mídia magnética, óptica ou mecânica (hardware). A palavra dados se aplica a uma série de objetos que são originários de ambientes digitais, como textos, números, imagens, vídeos, áudios, softwares, algoritmos, equações, animações, modelos, simulações e etc. (NATIONAL SCIENCE BOARD, 2011; KOLTAY, 2017; OED, 2015).

Nesse sentido, os dados são peças de informação digital distintas, geralmente formatadas de maneira particular. Os dados digitais são considerados um tipo de informação de natureza digital.

⁹ Em latim, *data* é o plural de *datum*. Historicamente e em áreas científicas especializadas, o termo *data*, em inglês, é tratado como plural, tendo um verbo correspondente no plural, como na frase: “dados foram coletados e classificados”. Em se tratando, no entanto, de uso moderno não científico, *data* geralmente não é empregado como um termo plural. Em vez disso, é tratado como um substantivo incontável (*mass noun* ou *uncountable noun*), semelhante a uma palavra como informação, levando um verbo correspondente no singular (SEMIDÃO, 2014, p. 71).

¹⁰ Disponível em: <<https://en.oxforddictionaries.com/definition/data>>. Acesso em: maio de 2016.

Os dados podem ser considerados: [...] representações formais da natureza em sistemas computacionais [...] Os dados podem ser símbolos e representações de informação e conhecimento representáveis [em formato digital]; no entanto, eles não devem ser equivalentes a informações ou a conhecimentos (ZHU; XIONG, 2015, p. 3, tradução nossa)¹¹.

Em decorrência do desenvolvimento das tecnologias digitais, as coisas do mundo natural são cada vez mais armazenadas na forma de dados e compartilhadas em redes como a Internet. Os dados são inseridos, gerados e criados no computador, tornam-se cada vez mais diversos, complexos e fora do controle humano (ZHU; XIONG, 2015).

Os dados são tão complexos e fluídos que virtualmente todos os tipos de informação digital podem ser algum tipo específico de dados. É possível se referir a tudo em meio digital como dados (MANTRA; 2012-13; RICE; SOUTHALL, 2016).

Para os fins deste estudo, os dados podem ser e serão considerados como todo objeto criado em formato digital (*digital-born*) ou convertido para formato digital (digitalizados) que possa ser usado para geração de insights de conhecimento.

Os dados podem ser primários (aqueles dados que não são manipulados ou alterados por pesquisadores), uma vez manipulados por pesquisadores, ou seja, depois de filtrados, analisados ou organizados, os dados primários tornam-se dados secundários ou transformados (NATIONAL RESEARCH COUNCIL, 1999; PIORUN, 2013).

No caso da pesquisa científica, são potencialmente utilizados como recurso primário para geração de resultados de pesquisa. Assim, percebe-se que o conceito atribuído ao termo dados depende do contexto em que ele é produzido e/ou em que fase do ciclo de vida dos dados a informação e/ou conhecimento estarão presentes na pesquisa (MANTRA, 2012-13).

¹¹ “Data is the formal representation of nature in computer systems; information is the phenomena of nature, society, and thinking activities; and knowledge is experience gained through practice. Data can be regarded as symbols and representations of information and knowledge; however, they should not be equivalent to information and knowledge” (ZHU; XIONG, 2015, p. 3).

DADOS DE PESQUISA

O termo dados de pesquisa é formalmente definido, em âmbito internacional, em 2007, pela Organization Economic Co-operation and Development (OECD), por meio da declaração de acesso aos dados científicos com financiamento. Segundo National Science Board (2011), Rodrigues et al. (2010), Pampel et al. (2013), Organization for Economic Co-Operation and Development (2007), Tibbo e Jones (2017) e Costa e Braga (2016), entre outros autores que tratam da temática, dados de pesquisa são “[...] registros factuais usados como fontes primárias na investigação científica, e que são geralmente aceitos pela comunidade científica como necessários para validar os resultados de investigação”¹².

Para o National Science Board (NSB) (2011), essa definição inclui dados analisados e os metadados que descrevem como esses dados foram gerados. Os dados analisados são limitados a informações digitais que descrevem os resultados de pesquisa financiada, incluindo imagens digitais, tabelas publicadas e tabelas usadas para criação de visualizações gráficas (PAMPEL et. al., 2013; NATIONAL SCIENCE BOARD, 2011).

Outra definição de dados de pesquisa inclui a ideia de que esses dados compreendem:

Informação registrada necessária para apoiar ou validar as observações, descobertas ou resultados de um projeto de pesquisa. [...] o que é coletado, observado ou criado de forma digital, para fins de análise e para produzir resultados de pesquisas originais [...] dados de pesquisa são: dados, registros, arquivos ou outras evidências, independentemente do seu conteúdo ou forma (por exemplo, material impresso ou digital), que compreendem observações de pesquisa, achados ou resultados, incluindo materiais primários e dados já analisados (RICE; SOUTHALL, 2016, p. 20, tradução nossa)¹³.

¹² “[...] the recorded factual material commonly accepted in the scientific community as necessary to validate research findings, but not any of the following: preliminary analyses, drafts of scientific papers, plans for future research, peer reviews, or communications with colleagues”

¹³ “The recorded information necessary to support or validate a research project’s observations, findings or outputs’ [...] ‘That which is collected, observed, or created in a digital form, for purposes of analysing to produce original research results’ [...] ‘Research data: The data, records, files or other evidence, irrespective of their content or form (e.g. in print, digital, physical or other forms), that comprise research observations, findings or outcomes, including primary materials and analysed” (RICE; SOUTHALL, 2016, p. 20).

Dados de pesquisa são evidências e/ou insumos de pesquisa que são coletados, observados, registrados e/ou criados para fins de análise e que podem produzir resultados de pesquisa para um estudo científico. Os dados de pesquisa são um produto para pesquisa, bem como um insumo para a pesquisa. Os dados de pesquisa devem poder ser coletados sistematicamente, estruturados e documentados para servir como insumos para futuras pesquisas. Os dados da pesquisa devem ser validados e compartilhados entre a comunidade científica. Dados de pesquisa são situacionais, e da mesma maneira que podem ser considerados válidos para responder a algumas questões de pesquisa, podem não ser. A mesma informação pode ser dados de pesquisa para um pesquisador em um ponto do tempo e em outro ponto do tempo não, dependendo do modo como esse pesquisador usa a informação ou material para análise de pesquisa (KELLAM; THOMPSON, 2016; HENDERSON, 2017; TIBBO; JONES, 2017; PRYOR, 2012; MANTRA, 2012-13).

Como exemplo, Tibbo e Jones (2017) citam definições de dados de pesquisa fornecidos e utilizados por instituições que fomentam pesquisas científicas nos USA, como a National Institutes of Health (NIH), que vê os dados de pesquisa como material fatural registrado e aceito comumente por uma comunidade científica como evidência para documentar e apoiar resultados de pesquisa (NATIONAL INSTITUTES OF HEALTH, 2003). Outra instituição nos USA, a NSF, descreve dados de pesquisa como algo determinado por uma comunidade de interesse por meio do processo de revisão por pares e programas de gerenciamento (NATIONAL SCIENCE BOARD, 2011). Por fim, Tibbo e Jones (2017) apresentam a definição da National Endowment for the Humanities que define dados como materiais gerados ou coletados durante a realização de pesquisas científicas (TIBBO; JONES 2017; NATIONAL INSTITUTES OF HEALTH, 2003; NATIONAL SCIENCE BOARD, 2011).

Os dados de pesquisa podem diferir em suas tipologias. Dependendo do assunto ou da disciplina científica, as definições utilizadas podem incluir uma vasta tipologia de objetos digitais e não digitais, como documentos (texto, palavras), memorandos, notas, avaliações, registros de casos, protocolos de estudo, planilhas, notas de laboratório, notas de campo, diários, questionários, transcrições, filmes, fitas de áudio e vídeo, arquivos digitais de áudio e vídeo, proteínas, sequências genéticas, DNA, cristalografia, coordenadas geográficas, slides, artefatos, espécimes, organismos, coleções de objetos digitais adquiridos e/ou gerados durante o processo de pesquisa, conteúdo em banco de dados (vídeo, áudio, texto, imagens), modelos, algoritmos, scripts de uma aplicação de computador (entrada, saída, arquivos de log para software de análise, software de simulação, esquemas), metodologias, fluxos de trabalho, informações técnicas, procedimentos e protocolos (DUDZIAK, 2016; HENDERSON, 2017).

Dados de pesquisa podem ser caracterizados de várias maneiras, conforme sua natureza, origem ou de acordo com seu status no fluxo de trabalho da investigação científica.

Segundo o Guia de Gestão de Dados de Pesquisa elaborado pela Comissão de Gestão de Energia Nuclear do governo brasileiro, esses dados são categorizados:

[...] segundo sua natureza: (a) observacionais: gerados por meio de observações diretas que podem ser associadas a lugares e/ou tempo específico, por sua natureza instantânea qualificam-se como registros históricos, pois não podem ser coletados uma segunda vez e, portanto, devem ser submetidos a processos de curadoria que os preservem para sempre; (b) computacionais, são o resultado da execução de modelos computacionais ou de simulações; (c) experimentais, são provenientes de situações controladas em laboratórios. [...] podem ser números, imagens, vídeos, áudio, software, algoritmos, equações, animações, modelos e simulações. Por fim, conforme o fluxo da pesquisa, (a) dados brutos, coletados diretamente de instrumentos científicos de pesquisa; (b) derivados, resultados do processamento ou combinação de dados brutos ou de outros dados; (c) canônicos ou referenciais, coleções de dados consolidados e arquivados geralmente em grandes centros de dados (SAYÃO; SALES, 2015, p. 8-9).

Os dados de pesquisa definem-se conforme sua natureza digital, o que inclui diferentes tipos de dados, níveis de agregação e métodos de investigação científica. O Interagency Working Group on Digital Data do National Science and Technology Council nos USA define dados de pesquisa em formato digital como:

[...] dados para uso e/ou reutilização em pesquisas científicas, técnica e para aplicações educacionais. [...] referem-se a toda a gama de tipos de dados e formatos relevantes para pesquisa científica, de engenharia, de educação e de [humanidades], em contextos locais, regionais, nacionais e globais com a amplitude correspondente às aplicações e a usos científicos em potencial. [...] dados de pesquisa digitais são quaisquer dados digitais, bem como os métodos e as técnicas utilizados na criação e na análise desses dados, [usados para] verificar resultados ou ampliar conclusões científicas, incluindo dados digitais associados a informações não digitais, como metadados associados a amostras físicas (NATIONAL SCIENCE BOARD, 2011, p. 1-33, tradução nossa)¹⁴.

Os dados de pesquisa digitais proliferaram devido aos avanços impulsionados pelas tecnologias da informação e da computação, sendo o uso crescente de tecnologias uma das principais fontes geradoras de dados. Atualmente, os dados de pesquisa são o resultado dos mais diversos processos de investigação mediada por dispositivos técnicos, o que transforma todas as etapas da pesquisa, desde a geração de dados, que pode se dar por meio de um experimento, como um estudo empírico sobre a observação de fenômenos culturais, até a publicação de resultados de pesquisa em um repositório de dados de pesquisa (PAMPEL, 2013). Nesse sentido, o próximo tópico tentará delimitar a definição de dados de pesquisa segundo sua utilização pelas diferentes disciplinas científicas. Principalmente, será questionado o que são dados de pesquisa em ciências sociais e ciências exatas.

DADOS DE PESQUISA E AS DIFERENTES DISCIPLINAS

A definição de dados de pesquisa transforma-se em relação às diferentes disciplinas do conhecimento científico. De disciplina a disciplina, de pesquisador a pesquisador, a expressão dados de pesquisa significa coisas diferentes e depende do contexto científico em que será utilizada.

Os dados de pesquisa são gerados com diferentes propósitos, o que amplia a sua diversidade. Diferentes disciplinas têm e usam linguagem disciplinar específica em torno dos dados de pesquisa (NATIONAL RESEARCH COUNCIL UK, 1999; PIORUN, 2013; SAYÃO; SALES, 2015; KELLAM, THOMPSON, 2016; KOLTAY, 2017; HENDERSON, 2017).

Os dados de pesquisa abrangem diferentes disciplinas acadêmicas, eles representam conjuntos de dados que estão registrados nas mais diversas áreas do conhecimento e que estão distribuídos globalmente.

Assim, percebe-se que os dados de pesquisa irão diferir de acordo com os métodos científicos e que eles dependem de características específicas de cada disciplina científica. Por exemplo, um cientista social pode considerar dados os seus diários, fotografias, resultados de questionários de pesquisas, notas de pesquisa de campo e/ou planilhas. Dados de pesquisa em ciência social são criados conforme os objetivos da investigação. Não são espontâneos, são produzidos e muitas vezes se apresentam como recursos digitais de informação estatística, como indicadores econômicos, políticos e sociais. Por exemplo, pode-se citar um questionário de pesquisa que pode conter informações coletadas a respeito de opiniões, comportamentos e/ou características sociais de um grupo de indivíduos que compõem uma unidade de observação social. As unidades de observação são o backbone dos dados de pesquisa em ciências sociais e são compostas por fatores de agregação, como em representações temporais, unidades de observação, coordenadas geoespaciais, entre outras categorias de conteúdo social (GERACI; HUMPHREY; JACOBS, 2012).

¹⁴ “[...] for use or repurposing for scientific or technical research and educational applications [...] It refers to the full range of data types and formats relevant to all aspects of science and engineering research and education in local, regional, national, and global contexts with the corresponding breadth of potential scientific applications and uses [...]. Digital research data is any digital data, as well as the methods and techniques used in the creation and analysis of that data, that a researcher needs to verify results or extend scientific conclusions, including digital data associated with non-digital information, such as the metadata associated with physical samples” (NATIONAL SCIENCE BOARD, 2011, p. 1-33).

Outro exemplo de dados de pesquisa disciplinar abrangendo a área das ciências exatas está incluso no ponto de vista do cientista da computação, o qual pode utilizar o termo para referir-se ao fluxo de zeros e uns, que fluem em uma rede de computadores.

Para esses cientistas, dados são vistos como valores armazenados em memória artificial, são objetos que podem ser processados e manipulados em um sistema operacional por meio de linguagens de programação:

Tudo em computação pode ser codificado em código binário, o que inclui sistemas operacionais e aplicativos de software, que são arquivos que contêm instruções para direcionar as operações da CPU, bem como as informações a serem manipuladas pela CPU. Os programas de computador são escritos em linguagens de programação de alto nível, ou seja, [...] são convertidos em instruções binárias na CPU por meio de um compilador que converte o código em uma massa de instruções na CPU [...] os cientistas da computação consideram a saída de um compilador, que normalmente é armazenado em um arquivo digital, como dados binários. Da mesma forma, a saída binária do código interpretado também é chamada de dados (GERACI; HUMPHREY; JACOBS, 2012, p. 19, tradução nossa).¹⁵

Scholtz (2016) apresenta os dados geoespaciais como uma subcategoria de dados de pesquisa em ciências exatas, os dados geoespaciais possuem peculiaridades tecnológicas e metodologias específicas no que tange a sua criação, utilização, preservação e compartilhamento.

Esses dados contêm informação a respeito de localizações geográficas e coordenadas geométricas que descrevem e modelam as dimensões e as características do Planeta Terra. Segundo o Guia de Pesquisa da Universidade de York, no Canadá:

¹⁵ “Everything in a computer is encoded in binary, including the operating system and application software, which are files containing instructions to drive the operations of the CPU as well as the information to be manipulated by the CPU. Computer programs written in a higher-level language, that is, a level other than the CPU’s binary instruction set, are converted to binary instructions either through a compiler, which translates code en masse into the CPU’s instruction set, or through an interpreter, which does the translation on the fly. Computer scientists regard the output of a compiler, which is typically stored in a file, as binary data. Similarly, the binary output of interpreted code is also called data” (GERACI; HUMPHREY; JACOBS, 2012, p. 19).

Dados geoespaciais, possuem informações explícitas de posicionamento geográfico, o que inclui dados sobre as redes rodoviárias em um Sistema de Informação Geográfica ou uma imagem de satélite georreferenciada. Os dados geoespaciais podem incluir dados de atributos que descrevem os recursos encontrados em um conjunto de dados. Existem dois tipos de dados geoespaciais – *vetor* e *raster*. Os vetores usam objetos geométricos simples, como pontos, linhas e áreas (polígonos), para representar recursos espaciais. O *raster data* utiliza um *grid* para representar sua informação geográfica. [...] são representados em células individuais, em linhas e em sequências de células vizinhas por áreas de coleção agrupadas. Outros tipos de dados *raster* incluem: imagens de satélite – dados de satélite oriundos de detecção remota; modelos de elevação digital – uma série de dados de elevação uniformemente espaçados; arquivos gráficos – mapas digitalizados, fotografias e imagens em formato TIFF, GIF ou JPEG (YORK UNIVERSITY, 2017, on-line, tradução nossa)¹⁶.

Por fim, os dados de pesquisa no campo das exatas são vistos como dados científicos, sendo utilizados e/ou produzidos por disciplinas como a astronomia e a biologia; caracterizam-se pelo seu grande volume e necessidade de alta infraestrutura tecnológica para armazenamento, preservação e compartilhamento.

O Genoma Humano e as Pesquisas Astronômicas geram alguns dos maiores conjuntos de dados, atingem os petabytes (PB). O Sloan Digital Sky Survey, por exemplo, produz cerca de 5 PB por dia.

¹⁶ “Geospatial data, GIS data or geodata has explicit geographic positioning information included within it, such as a road network from a GIS, or a geo-referenced satellite image. Geospatial data may include attribute data that describes the features found in the dataset. There are two types of geospatial data - vector and raster. Vector data uses the simple geometric objects of points, lines, and areas (polygons) to represent spatial features. Raster data uses a grid to represent its geographic information. Points are represented by single cells, lines by sequences of neighboring cells, and areas by collection of grouping cells. Other types of raster data include: Satellite Imagery - remotely sensed satellite data. Digital Elevation Models- an array of uniformly spaced elevation data. Graphic Files - scanned maps, photographs, and images in TIFF, GIF, or JPEG format” (YORK UNIVERSITY, 2017, on-line).

Já o Large Synoptic Survey Telescope (LSST) grava até 30 trilhões de bytes de dados de imagem diariamente a partir do topo de uma montanha no Chile (GRIGG, 2016, p. 180, tradução nossa)¹⁷.

Os dados científicos são descentralizados e gerados em diversos laboratórios e/ou centros de pesquisa por muitos pesquisadores em diferentes países. Os dados científicos são complexos no que tange à sua análise e manipulação, exigem computadores de alto desempenho e soluções de infraestrutura. Por exemplo, os conjuntos de dados sobre o Genoma Humano e sobre a Pesquisa Astronômica requerem enorme quantidade de espaço para armazenamento e processamento computacional, além de serem dados de custo elevado (GRIGG, 2016).

Esse tipo de dados são aqueles que Cunha (2010); Costa e Cunha (2014) classificam como resultado da e-Science.

Por fim, os dados de pesquisa são produtos de pesquisa heterogêneos e contextualizados conforme as disciplinas do conhecimento. Nesse sentido, os dados de pesquisa devem ser válidos, compartilháveis, heterogêneos e contextualizáveis dentro de uma comunidade científica. No que tange à ciência da informação, os dados de pesquisa são de origem multidisciplinar e dependem da relação que a disciplina possui com a CI.

Um exemplo de fonte para localizar os mais diversos tipos de dados de pesquisa é o Registry of Research Data Repositories (re3data), que é o registro internacional de repositórios de dados de pesquisa. Fundado pelo German Research Foundation (DFG), o re3data congrega iniciativas da Library and Information Services (LIS) of the GFZ German Research Centre for Geosciences, da Library of the Karlsruhe Institute of Technology (KIT), da School of Library and Information Science (BSLIS) na Humboldt-Universität, em Berlin e das Libraries of the Purdue University, na Alemanha.

O re3data.org indexa, internacionalmente, cerca de (2152, set. 2018) repositórios de dados de pesquisa desde o outono de 2012. Ele oferece a pesquisadores, organizações de financiamento, bibliotecas e editores uma visão geral sistemática da paisagem heterogênea dos repositórios de dados de pesquisa (PAMPEL et al., 2013; COSTA; BRAGA, 2016). O re3data abrange diferentes disciplinas acadêmicas, apresenta uma relação de repositórios que servem para o armazenamento permanente e o acesso a conjuntos de dados de pesquisa em que órgãos de financiamento, editores e instituições acadêmicas promovem uma cultura de compartilhamento, acesso e visibilidade aos dados de pesquisa (RE3DATA, 2018).

Atualmente os repositórios de dados de pesquisa possibilitam o acesso a ampla gama de tipos de dados de pesquisa e sua preservação. Usualmente estão hospedados nos países da América do Norte e Europa, são repositórios de dados de pesquisa temáticos, focados no armazenamento de textos e imagens (COSTA, BRAGA, 2016).

Os dados de pesquisa são parte essencial do registro acadêmico científico e a gestão dos dados de pesquisa é cada vez mais vista como uma tarefa elementar para bibliotecas acadêmicas. Assim, constituem uma nova área de pesquisa para os chamados bibliotecários de dados. Nesse sentido, acredita-se que um bibliotecário de dados deve apreender e assimilar novas habilidades ligadas aos jargões disciplinares da gestão dos dados de pesquisa conforme o domínio dos cientistas com quem atua.

Na sequência, apresenta-se a biblioteconomia de dados como modo de delimitar o tema dos dados de pesquisa como uma área de atuação para os bibliotecários de dados.

¹⁷ “Genomics and astronomy generate some of the largest datasets, often reaching one or more petabytes (PB). The Sloan Digital Sky Survey, for example, produces about five PB a day, and the proposed Large Synoptic Survey Telescope (LSST) will record 30 trillion bytes of image data daily from the top of a mountain in Chile, a data volume equal to two daily Sloan Digital Sky Surveys” (GRIGG, 2016, p. 180).

BIBLIOTECONOMIA DE DADOS

A biblioteconomia orientada a dados, em inglês, pode ser chamada de *e-Science Librarianship*¹⁸ (XIA; WANG, 2014), *Data Librarianship* (KOLTAY, 2017; RICE, 2016; RICE; SOUTHALL, 2016)¹⁹ e/ou *Databrarianship* (KELLAM; THOMPSON, 2016)²⁰. Ela corresponde ao interesse de bibliotecários em compreender atividades ligadas ao gerenciamento e à curadoria de todos os tipos de dados, sendo seu foco o tratamento, a gestão e a curadoria de dados de pesquisas em qualquer disciplina científica. Ela busca prover técnicas e serviços para o suporte a coleções de dados de pesquisa, investigando problemas práticos relacionados ao armazenamento, ao descobrimento e ao acesso a dados de pesquisa. Atualmente, é uma tendência de estudos desenvolvida em bibliotecas acadêmicas, principalmente, nos Estados Unidos, no Reino Unido e no Canadá.

As origens da biblioteconomia de dados estão no surgimento das coleções de dados de pesquisa oriundas da utilização de métodos quantitativos de pesquisa em ciências sociais, especialmente dados públicos produzidos por instituições governamentais, como os censos eleitorais e os dados sobre a economia. O surgimento de arquivos e bibliotecas de dados no Reino Unido (Data Archives Services), nos Estados Unidos e no Canadá (Data Library Services), entre as décadas de 1960 e 1970, são fatos fundamentais que dão origem à biblioteconomia de dados.

A forma como serviços de dados surgiram e se desenvolveram difere em relação aos continentes, Europa e a América do Norte. Na Europa, os serviços de dados foram concebidos de maneira centralizada pelos conselhos nacionais de pesquisa, vide o Council of European Social Science Data Archives (CESSDA)²¹. Na América do Norte, os serviços de dados foram oferecidos por pequenos centros de dados e/ou computação de departamentos universitários nos USA e no Canadá (RICE; SOUTHALL, 2016, SEMELER, PINTO, ROZADOS, 2017).²²

As ciências sociais foram um dos primeiros campos disciplinares a adotar e a compartilhar dados de pesquisa para o reuso, isso por razões de economia e eficiência científica. O desenvolvimento de coleções de dados em ciências sociais foi fortemente influenciado pelas ideias computacionais de Alan Turing (1930), pelos métodos de escalonamento e análises multivariadas utilizados pelo World Bank, International Monetary, na década de 1940; pelo surgimento das linguagens de programação Cobol e Fortran; e, respectivamente, no Institute of Social Research em Michigan e no Bureau of Applied Social Research, em Columbia, na década de 1950; pela criação dos *packages* SPSS (*social sciences package*), SAS (*statistical analysis system*), na década de 1960, desenvolvidos pelos serviços de dados locais das universidades de Princeton, Northwestern, British Columbia, North Carolina, Wisconsin e Yale. E, finalmente, a profissão de *data librarian* é apresentada pela IASSIST no Congresso Mundial de Sociologia de Toronto em 1974 (RICE; SOUTHALL, 2016; KOLTAY, 2017; KELLAM; THOMPSON, 2016, SEMELER, PINTO, ROZADOS, 2017).

¹⁸ “Data librarianship in general, which is sometimes also called e-science librarianship, has now become a significant part of the academic library workforce” (XIA WANG, 2014, p. 364).

¹⁹ “The origins of data librarianship and the title data librarian go back to social science data management (particularly publicly available data sets and geospatial data), and they then evolved in the bioinformatics field. The concept and the title now cover library-related work with both purchased and locally produced digital data in any subject domain” (SOEHNER et al., 2010).

²⁰ “[...] we use the term databrarianship to describe a field that is characterized by a diversity of interests but united by our commitment to ensuring access to data, be they primary research data created by our institutions’ researchers or secondary data used for analysis” (KELLAM; THOMPSON, 2016, p. 1).

²¹ CESSDA. Disponível em: <http://www.unidata.unimib.it/?page_id=2136>. Acesso em: maio 2017.

²² Cronograma de surgimento de Bibliotecas de dados disponível em: <https://docs.google.com/spreadsheets/d/1qmC_z50UDHh3Jwdu6wGrz1xjdfumBaad6P6cUpBdY/edit?pli=1#gid=0>. Acesso em: set. 2017.

A biblioteconomia orientada a dados não é um novo ramo da biblioteconomia, fundamenta-se em uma diversidade de habilidades já incorporadas e conhecidas por bibliotecários. Práticas tradicionais são retomadas com o objetivo de serem aplicadas a conjuntos de dados. Por exemplo, a catalogação e a organização de materiais; a preservação e a curadoria; serviços de usuários e de referência; consultoria e treinamentos são repensados para serem aplicados junto às novas práticas, que envolvem o gerenciamento, a curadoria e o compartilhamento de dados. No entanto, a biblioteconomia de dados não serve apenas como apoio à descoberta de novas habilidades em biblioteconomia, ela envolve-se com a geração e a preservação de dados de pesquisa, preocupa-se com praticamente todas as funções tradicionais do trabalho bibliotecário, como aquisição e desenvolvimento de coleções, organização, catalogação de acervos e implementação de serviços de referência aos usuários (RICE; SOUTHALL, 2016; KELLAM; THOMPSON, 2016; SEMELER, PINTO, ROZADOS, 2017).

Uma das principais tarefas da biblioteconomia de dados é ser *upstream*²³ (GOLD, 2007a), o que significa que ela não deve apenas se preocupar com informação publicada, mas também com fontes potenciais de dados, sendo o seu objetivo compreender como os diferentes tipos de dados podem ser capazes de gerar informação útil para pesquisa (XIA; WANG, 2014; KOLTAY, 2017).

Os bibliotecários de dados, entre outras habilidades, devem se preocupar com avaliação e retenção de dados, promoção de dados, marketing de dados, conscientização de dados, coordenação de práticas com dados em toda unidade e instituição, habilidades de negociação e habilidades de gerenciamento de expectativas de dados. O bibliotecário de dados compartilhará as mesmas habilidades de um gestor de dados, no que tange à preservação e à avaliação dos valores econômicos atribuídos aos dados.

Em relação ao uso de tecnologias, o bibliotecário de dados deve ter em mente a habilidade *hacker* de um cientista de dados (XIA; WANG, 2014).

O bibliotecário de dados deve possuir as habilidades necessárias para trabalhar com qualquer tipo de dados, sejam eles observacionais, computacionais ou experimentais. Ele deve compreender como as investigações sobre dados modificam as práticas e as teorias que fundamentam a biblioteconomia. Além de deter o papel de ajudar pesquisadores a coletar, localizar e recomendar recursos de dados apropriados, como o uso de provedores de serviços de dados de pesquisa, os bibliotecários de dados podem oferecer serviços de consultoria sobre o preparo e a aquisição de dados para análise de questões de pesquisa em disciplinas específicas (DEKKER; LACKIE, 2016).

Assim, o trabalho de um bibliotecário de dados pode estar ligado à manipulação de dados, mas também se conecta com o trabalho humano de interpretação, tomada de decisão e comunicação de dados. McCaffrey e Giesbrecht (2016) listam algumas das habilidades e/ou competências necessárias a um bibliotecário de dados. Os autores sugerem três áreas distintas para a atuação desse profissional: (a) gestão e curadoria de dados; (b) visualização de dados e representação de dados geoespaciais; e (c) serviços avançados de referência a dados de pesquisa. Kennan (2016, p. 1-10) apresenta um estudo sobre as habilidades requeridas por aqueles que contratam bibliotecários de dados. A autora lista as habilidades exigidas por organizações de pesquisa e universidades nos Estados Unidos, Canadá, Reino Unido e Austrália, conforme quadro 1.

²³ “In a seminal article on cyberinfrastructure, data and libraries in D-Lib Magazine Anna Gold has characterized these areas that fall outside the usual comfort zone of academic librarians as ‘working upstream’ in the research process, before the point of publication” (GOLD, 2007a, p. 12).

Quadro 1 – Habilidades requeridas aos bibliotecários de dados

Habilidades	Descrição
Interpessoais e características comportamentais	Capacidade relacionada à comunicação formal, como a escrita de documentação técnica e estudos de caso. Outro ponto ligado a este perfil é a capacidade de se adaptar a novas experiências, de estar constantemente em busca de atualização.
Conhecimento contextual sobre ambientes institucionais	Essa habilidade envolve conhecimentos a respeito de políticas de financiamento oferecidas por agências que fomentam a investigação científica. Ressalta-se que essa habilidade exige a compreensão de processos éticos do processo científico, métodos de pesquisa disciplinar, formas de comunicação científica, propriedade intelectual, formas de acesso, marcos legais e regulatórios (copyright e Creative Commons) e políticas de direito autoral.
Uso de dados	Inclui a compreensão de tipos de dados (quantitativos, qualitativos), padrões e esquemas de metadados, (Dublin Core, RDF), assim como questões relacionadas a identificadores únicos (Digital Object Identifiers) e preservação de dados digitais.
Conhecimentos sobre tecnologias de manipulação de dados	Os conhecimentos em tecnologia mais relevantes aos bibliotecários de dados abrangem linguagens de programação (Python, SQL, Java, XML entre outras), design e estrutura de bases de dados, APIs de recuperação de dados, o design centrado no usuário, ferramentas de processamento de linguagem natural, internet of things e Big Data.

Fonte: Kennan (2016, p. 1-10).

Por fim, Ferder (2018) realizou um estudo sobre as habilidades e conhecimentos necessários em biblioteconomia de dados. A autora selecionou bibliotecários que trabalham com dados, indagou a esses profissionais sobre sua formação educacional e solicitou que eles a avaliassem a relevância de um conjunto de habilidades e conhecimentos relacionados a dados para seu trabalho. Os entrevistados consideraram ampla gama de habilidades e conhecimentos importantes para o seu trabalho, especialmente habilidades pessoais e características pessoais, como habilidades de comunicação e a capacidade de desenvolver relacionamentos com pesquisadores. As habilidades tradicionais de biblioteca, como catalogação e desenvolvimento de coleções, eram consideradas menos relevantes. A autora revelou também que existem dois tipos de bibliotecários de dados: os generalistas de dados, que tendem a fornecer serviços de dados em vários campos, e especialistas de assunto, que tendem a fornecer serviços mais especializados para uma disciplina distinta.

Assim, é preciso um esforço multidisciplinar de aprendizagem, por parte dos bibliotecários de dados, que devem combinar as habilidades e os conhecimentos necessários à manipulação e à gestão de dados. O papel fundamental dos bibliotecários de dados é buscar adquirir experiência técnica para aquisição, análise e interpretação de dados.

Por fim, o bibliotecário de dados deve transformar os dados em algo reutilizável, compartilhável e preservável em longo prazo. Ele deve conhecer os esquemas de metadados, os repositórios de dados, os conjuntos de dados que são utilizados por seus usuários pesquisadores, além de ter uma boa noção sobre as mais diversas metodologias e disciplinas científicas. Mas, principalmente, ele deve possuir habilidades ligadas a atividades de coleta, de descrição, de ensino (data literacy), de construção de sistemas, de elaboração de serviços de referência e de criação de políticas de gerenciamento de dados institucionais (MCCAFFREY; GIESBRECHT, 2016; KENNAN, 2016; DEKKER, LACKIE, 2016; SEMELER, PINTO, ROZADOS, 2017).

O interesse de bibliotecários em adquirir habilidades técnicas para gerenciamento de dados de pesquisa demonstra que o uso de dados em biblioteconomia de dados pode estar ligado à elaboração de produtos e serviços de dados vinculados ao gerenciamento de dados de pesquisa em bibliotecas. Assim, a próxima questão a ser discutida revela o papel dos bibliotecários de dados em relação ao uso, à gestão e à curadoria de dados de pesquisa.

Revisto o conceito, a origem e algumas das características dos estudos sobre dados, dados de pesquisa e biblioteconomia de dados, percebe-se que a complexidade do conceito de dados revela a necessidade de uma ciência que estude habilidades, métodos, técnicas e tecnologias relacionadas ao uso intensivo de dados em biblioteconomia.

CONSIDERAÇÕES FINAIS

A ciência com uso intensivo de dados institui uma nova dimensão para os cientistas, o universo dos dados, causando uma revolução no processo de pensamento científico, a qual consolida teorias e práticas científicas que se utilizam para as diversas dimensões do conceito de dados.

Acredita-se que elas representem um novo desafio de aprendizagem e que se caracterizem como uma nova oportunidade aos profissionais bibliotecários e cientistas da informação. O interesse pelos dados de pesquisa proporciona ir gerar novos métodos e práticas que podem contribuir e modificar as investigações desses profissionais. Principalmente, com a inclusão das tecnologias de dados nas práxis bibliotecárias, percebe-se uma transformação no modo como o bibliotecário vê suas próprias atividades. As atuais mudanças na forma de se conduzir a pesquisa científica oferecem oportunidades para os bibliotecários de dados se envolverem com a ciência de uso intensivo de dados; mas em específico, eles precisam adquirir maiores conhecimentos sobre os dados de pesquisa e as tecnologias utilizadas para sua manipulação e gestão.

Conclui-se que uma ciência fundamentada na exploração dos dados de pesquisa pode ser considerada como uma nova área de estudos para bibliotecários que procuram se envolver com as questões ligadas à coleta, manipulação, análise e visualização de dados de pesquisa. Essa corrente investigativa levantará questões a respeito de práticas, modelos e metodologias utilizadas para se obter informação útil a partir de dados, *insights* de conhecimento a partir de ambientes heterogêneos e complexos de dados.

Assim, o bibliotecário de dados precisa aprender mais sobre essa ciência. A tarefa essencial dos bibliotecários de dados será a de transformar suas tarefas e atividades quotidianas nas bibliotecas, inserindo os dados de pesquisa no contexto da geração de produtos e/ou serviços em bibliotecas.

Para concluir, é relevante destacar a dificuldade de localizar literatura sobre o tema dados de pesquisa e biblioteconomia de dados; estas temáticas carecem de atenção por parte dos profissionais da área da informação. Nesse sentido, para adquirir os conhecimentos teóricos mínimos em sobre biblioteconomia de dados e dados de pesquisa, recomendam-se algumas obras essenciais.

A temática de dados de pesquisa é revista exhaustivamente por Margaret E. Henderson no livro *Data Management A Practical Guide for Librarians*, de 2017. A autora apresenta os princípios básicos da gestão e das melhores práticas com dados de pesquisa. Outra obra, organizada por Lynda Kellam e Kristi Thompson em 2016, é o livro de artigos *Databrarianship: the academic data librarian in theory and practice*, que apresenta diversas questões relacionadas à biblioteconomia de dados e ao uso de dados de pesquisa pelas distintas disciplinas científicas. Também sobre biblioteconomia de dados, mas com o foco no campo de atuação do bibliotecário de dados, está a obra *The Data Librarian's Handbook*, escrita em 2016 por Robin Rice e John Southall.

No que tange à realização de trabalhos futuros, sugere-se que os bibliotecários investiguem temas relacionados ao uso de dados de pesquisa em contextos diversos como Big Data, Internet das coisas, inteligência artificial, métricas informacionais, comunicação científica, entre outras temáticas. Seria interessante a realização de um estudo sobre as possíveis análises que podem ser aplicadas aos metadados e ao conteúdo depositado em repositórios de dados de pesquisa. Outra maneira de abordar os dados de pesquisa em biblioteconomia de dados poderia ser a aproximação dos bibliotecários de dados do processo de elaboração de planos de gestão de dados de pesquisa.

Por fim, o conceito de dados de pesquisa aponta forte tendência de estudos e práticas profissionais que está se desenvolvendo no cerne das investigações realizadas por bibliotecários de dados e outros cientistas focados no uso e consumo de dados. Os dados de pesquisa são uma oportunidade para os bibliotecários de dados oferecerem serviços e produtos de dados a seus usuários. Os dados de pesquisa são um produto para pesquisa, bem como um insumo para a pesquisa. Esse tipo de dados assume papel relevante no que tange aos novos processos de comunicação científica. No futuro, os bibliotecários de dados devem pensar em serviços que liguem pesquisadores a produtos de dados de pesquisa. Os dados de pesquisa devem poder ser coletados sistematicamente, estruturados, documentados e armazenados para servir como insumos para futuras pesquisas.

Assim, conclui-se que as atuais tendências de pesquisa em biblioteconomia de dados estão voltadas para os estudos sobre a proveniência, as melhores práticas, o planejamento, a curadoria, entre outras atividades científicas relacionadas à teoria e à práxis com dados, principalmente quando esses dados se relacionam com os processos de pesquisa científica em meio digital. Basicamente, o campo da biblioteconomia de dados está relacionado com a prática, com o uso e com o consumo de dados de pesquisa.

REFERÊNCIAS

- AMERICAN COUNCIL OF LEARNED SOCIETIES. *Our cultural commonwealth*: The report of the American Council of Learned Societies Commission on Cyberinfrastructure for the Humanities and Social Sciences, 2006. Disponível em: http://www.acls.org/uploadedFiles/Publications/Programs/Our_Cultural_Commonwealth.pdf. Acesso em: maio 2015.
- BURLINGAME, N; NIELSEN, L. *A Simple Introduction to Data Science*. Wickford: New Street Communications, 2012.
- CAPLAN, P. *Metadata fundamentals for all librarians*. La Vergne: Lightning Source, 2003.
- CESSDA. Disponível em: http://www.unidata.unimib.it/?page_id=2136. Acesso em: maio 2017.
- COSTA, M.; CUNHA, M. B. da. O bibliotecário no tratamento de dados oriundos da e-Science: considerações iniciais. *Perspectivas em Ciência da Informação*, v. 19, n. 3, p. 189-206, 2014.
- COSTA, M; BRAGA, T. Repositórios de dados de pesquisa no mundo. *Cadernos BAD*, n. 2, p. 80-95, 2016. Disponível em: <https://www.bad.pt/publicacoes/index.php/cadernos/article/view/1585>. Acesso em: 15 set. 2016.
- CUNHA, M. B. da. A biblioteca universitária na encruzilhada. *DataGramaZero: Revista de Ciência da Informação*, v. 11, n. 6, 2010.
- DAVENPORT, T. H. *Big Data at Work: Dispelling the Myths, Uncovering the Opportunities*. Boston: Harvard Business Review Press, 2014.
- DEKKER, H.; LACKIE, P. Technical Data Skills for Reproducible Research. In: KELLAM, L.; THOMPSON, K. *Introduction to Data Librarianship: The Academic Data Librarian in Theory and Practice*. Chicago: Association of College and Research Library, 2016.
- DUDZIAK, E. *Dados de Pesquisa agora devem ser armazenados e citados*. 2016. Disponível em: <http://www.sibi.usp.br/?p=6189>. Acesso em: 10 ago. 2016.
- FEDERER, L. Research data management in the age of big data: Roles and opportunities for librarians. *Information Services & Use*, v. 36, n. 1-2, p. 35-43, 2016. Disponível em: <http://content.iospress.com/articles/information-services-and-use/isu797> . Acesso em: 10 ago. 2017.
- FEDERER, L. Defining data librarianship: a survey of competencies, skills, and training. *Journal of the Medical Library Association*, v. 106, n. 3, p. 294-303, 2018.
- GERACI, D.; HUMPHREY, C.; JACOBS, J. *Data Basics: An Introductory Text*. Michigan: ICPSR, 2008.
- GOLD, A. Cyberinfrastructure, data, and libraries, part 1: A cyberinfrastructure primer for librarians. *D-Lib Magazine*, v. 13, n. 9-10, 2007a. Disponível em: <http://www.dlib.org/dlib/september07/gold/09gold-pt1.html>. Acesso em: 15 maio 2016.

GOLD, A. Cyberinfrastructure, data, and libraries, part 2: Libraries and the data challenge: Roles and actions for libraries. *D-Lib Magazine*, v. 13, n. 9-10, 2007b. Disponível em: <http://www.dlib.org/dlib/september07/gold/09gold-pt2.html>. Acesso em: 15 maio 2016.

GRIGG, K. Data in the Sciences. In: KELLAM, L.; THOMPSON, K. *Introduction to Databrarianship: The Academic Data Librarian in Theory and Practice*. Chicago: Association of College and Research Library, 2016.

HENDERSON, M. *Data Management: a practical guide for librarians*. Lanham: Rowman & Littlefield, 2017.

INTERNATIONAL ASSOCIATION FOR SOCIAL SCIENCE INFORMATION SERVICES AND TECHNOLOGY (IASSIST). *Defining data librarian - call for comments*, 2016. Disponível em: <http://www.iassistdata.org/blog/defining-data-librarian-call-comments>. Acesso em: jun. 2016.

KELLAM, L.; THOMPSON, K. *Introduction to Databrarianship: The Academic Data Librarian in Theory and Practice*. Chicago: Association of College and Research Library, 2016.

KENNAN, M. Data management: Knowledge and skills required in research, scientific and technical organizations. In: IFLA WLIC, 82., 2016, Columbus. *Proceedings[...]*. Columbus, 2016. Disponível em: <http://library.ifla.org/1466/1/221-kennan-en.pdf>. Acesso em: jun. 2016.

KINDLING, M.; PAMPEL, H. Informationsinfrastrukturangebote für digitale Forschungsdaten. *E(bren)Journal*, p. 15-33, 2017. DOI: 10.18452/2341.

KOLTAY, T. Data literacy for researchers and data librarians. *Journal of Librarianship and Information Science*, v. 49, n. 1, 2017. Disponível em: <http://journals.sagepub.com/doi/abs/10.1177/0961000615616450>. Acesso em: 4 ago. 2017.

MANTRA. *Do-It-Yourself Research Data Management Training Kit for Librarians*. 2012-13. Disponível em: <http://mantra.edina.ac.uk/libtraining.html>. Acesso em: ago. 2017.

MCCAFFREY, M.; GIESBRECHT, W. Teaching Data Librarianship to LIS Students. In: KELLAM, L.; THOMPSON, K. *Introduction to Databrarianship: The Academic Data Librarian in Theory and Practice*. Chicago: Association of College and Research Library, 2016.

MOORE, G.; MOREAU, L. *From metadata to links*. In: WORKSHOP ON OPEN HYPERMEDIA SYSTEMS- OHS, 6, 2001, Berlin. *Proceedings[...]*. Berlin: Springer-Verlag, 2001. Disponível em: <http://eprints.soton.ac.uk/id/eprint/254475>. Acesso em: 15 maio 2012.

NATIONAL INSTITUTES OF HEALTH. *NIH Data Sharing Policy and Implementation Guidance*. 2003. Disponível em: https://grants.nih.gov/grants/policy/data_sharing/data_sharing_guidance.htm. Acesso em: ago. 2017.

NATIONAL RESEARCH COUNCIL. *A question of balance: Private rights and the public interest in scientific and technical databases*. Washington, DC: The National Academies Press, 1999. Disponível em: <http://www.nap.edu/read/9692/chapter/1>. Acesso em: maio 2016.

NATIONAL RESEARCH COUNCIL (NRC). *A question of balance: Private rights and the public interest in scientific and technical databases*, 1999. Disponível em: <http://www.nap.edu/read/9692/chapter/1>. Acesso em: jun. 2016.

NATIONAL SCIENCE BOARD (NSB). *Cyberinfrastructure vision for 21st century discovery*, 2005. Disponível em: <http://www.nsf.gov/pubs/2007/nsf0728>. Acesso em: jun. 2016.

NATIONAL SCIENCE BOARD (NSB). *Digital Research Data Sharing and Management*, 2011. Disponível em: <https://www.nsf.gov/nsb/publications/2011/nsb1124.pdf>. Acesso em: ago. 2017.

OED Online. *Datum, n*. Oxford University Press, 2015. Disponível em: <http://www.oed.com/view/Entry/47434>. Acesso em: jun. 2016.

ORGANIZATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT (OECD). *Declaration on Access to Research Data from Public Funding Organization for Economic Co-operation and Development* (OECD), 2004. Disponível em: http://acts.oecd.org/Instruments/ShowInstrumentView.aspx?InstrumentID=157#_ftn0. Acesso em: 5 out. 2016.

PAMPEL, H. *et al*. Making research data repositories visible: the re3data.org Registry. *PLoS One*, v. 8, n. 11, 2013. Disponível em: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3817176/>. Acesso em: 15 maio 2016.

PIORUN, M. E. *E-Science as a Catalyst for Transformational Change in University Research Libraries*. 2013. 369f. Tese (Doutorado em Filosofia) – Faculty of the Simmons College Graduate School of Library and Information Science. University of Massachusetts Medical School, 2013. Disponível em: http://escholarship.umassmed.edu/cgi/viewcontent.cgi?article=1151&context=lib_articles. Acesso em: 14 ago. 2016.

PRYOR, G. *Managing Research Data*. London: Facet Publishing, 2012.

RE3DATA 2018. Disponível em: <http://www.re3data.org/about>. Acesso em: nov. 2018.

RICE, R.; SOUTHALL, S. *The data librarian's handbook*. London: Facet Publishing, 2016.

RODRIGUES, E. *et al*. *Os repositórios de dados científicos: estado da arte*. 2010.

Disponível em: http://projeto.rcaap.pt/index.php?option=com_remository&Itemid=2&func=startdown. Acesso em: 5 jun. 2016.

- SAUNDERS, T. Data Science and Data Scientists: What's in a Name? *Information Management*, nov., 2013. Disponível em: <http://www.cbiconsulting.com/wp-content/uploads/2014/03/data-scientist.pdf>. Acesso em: 14 ago. 2016.
- SAYÃO, L. F.; SALES, L. F. *Guia de gestão de dados de pesquisa para bibliotecários de pesquisadores*. Rio de Janeiro: CNEN, 2015. Disponível em: http://carpedien.ien.gov.br:8080/bitstream/ien/1624/1/GUIA_DE_DADOS_DE_PES. Acesso em: 5 out. 2016.
- SCHOLTZ, N. Supporting Geospatial Data. In: KELLAM, L.; THOMPSON, K. *Introduction to Databrarianship: The Academic Data Librarian in Theory and Practice*. Chicago: Association of College and Research Library, 2016.
- SEMLER, A. R.; PINTO, A. L. ; ROZADOS, H.B. Data science in data librarianship: Core competencies of a data librarian. *Journal of Librarianship and Information Science*, v. 50, p. 096100061774246-10, 2017.
- SEMIDÃO, R. A. M. *Dados, Informação e Conhecimento enquanto elementos de compreensão do universo conceitual da Ciência da Informação*: contribuições teóricas. 2014. 198 f. Dissertação (Mestrado) – Programa de Pós-Graduação em Ciência da Informação, Faculdade de Filosofia e Ciências, Universidade Estadual Paulista – UNESP, Marília, 2014. Disponível em: <https://repositorio.unesp.br/bitstream/handle/11449/110783/000799485.pdf?sequence=1>. Acesso em: ago. 2017.
- SOEHNER, C.; STEEVES, C.; WARD, J. *E-science and data support services: A study of ARL member institutions*. Washington, DC: Association of Research Libraries, 2010.
- TENOPIR, C. et al. Research Data Services in European Academic Research Libraries. *LIBER Quarterly*, v. 27, n.1, p. 23–44, 2017. Disponível em: <https://www.liberquarterly.eu/articles/10.18352/lq.10180/> . Acesso em: jun. 2017.
- TENOPIR, C. et al. Research data management services in academic research libraries and perceptions of librarians. *Library & Information Science Research*, v. 36, p. 84–90, 2014. Disponível em: <http://www.sciencedirect.com/science/article/pii/S0740818814000255> . Acesso em: jun. 2016.
- TIBBO, H.; JONES, S. *Research data management and sharing*, 2017. Disponível em: <https://www.coursera.org/learn/data-management>. Acesso em: jun. 2017.
- VAN DER AALST, W. M. P. Data Scientist: The Engineer of the Future. In: MERTINS, K. et al. (Ed.). *Enterprise Interoperability VI: Interoperability for Agility, Resilience and Plasticity of Collaborations*. Berlin: Springer, 2014.
- INTERAGENCY WORKING GROUP ON DIGITAL DATA. *Harnessing the Power of Digital Data for Science and Society*, 2009. Disponível em: https://www.nitrd.gov/about/harnessing_power_web.pdf . Acesso em: 4 ago. 2015.
- WORLD DATA CENTRE FOR GLACIOLOGY. Disponível em: <https://www.spri.cam.ac.uk/about/>. Acesso em: set. 2017.
- XIA, J.; WANG, M. Competencies and Responsibilities of Social Science Data Librarians: An Analysis of Job Descriptions. *College & Research Libraries*, v. 75, n. 3. p. 362-388, 2014. Disponível em: <http://crl.acrl.org/content/75/3/362.abstract> . Acesso em: 4 ago. 2015.
- YORK UNIVERSITY. *Geospatial Data: What is Geospatial* 2017. Disponível em: <http://researchguides.library.yorku.ca/c.php?g=679467&p=4793119>. Acesso em: ago. 2017.
- ZHU, Y.; XIONG, Y. Towards Data Science. *Data Science Journal*, v. 14, p. 8, 2015. Disponível em: <http://doi.org/10.5334/dsj-2015-008>. Acesso em: jun. 2016.

AGRADECIMENTOS

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (Capes) - Código de Financiamento 001.
