

Dados e metadados: conceitos e relações

Ana Carolina Simionato Arakaki

Doutora em Ciência da Informação pela Universidade Estadual Paulista (Unesp) - Marília, SP - Brasil.

Professor da Universidade Federal de São Carlos (UFSCar) - São Carlos, SP - Brasil.

<http://lattes.cnpq.br/9896600626524397>

E-mail: acsimionato@ufscar.br

Felipe Augusto Arakaki

Doutor em Ciência da Informação pela Universidade Estadual Paulista (Unesp) - Marília, SP - Brasil.

Professor da Universidade de Brasília (UnB) - Brasília, DF - Brasil.

<http://lattes.cnpq.br/5324289839207169>

E-mail: fe.arakaki@gmail.com

Submetido em: 25/11/2020. Aprovado em: 25/11/2020. Publicado em: 28/07/2021.

RESUMO

A interdisciplinaridade das pesquisas em áreas correlatas pode apresentar muitas contribuições para o desenvolvimento de teorias. Entretanto, também pode proporcionar confusões terminológicas. Dessa forma, destaca-se a importância em discutir esses conceitos e evitar equívocos de nomenclatura. Por essa razão, o objetivo deste trabalho consiste em discutir e relacionar conceitos de dados e metadados na Ciência da Informação. A pesquisa é caracterizada por uma metodologia de análise exploratória, sendo possível identificar elementos conceituais por meio da literatura científica, a analisar as informações a partir do Perspectivismo. Como resultados, são apresentadas as definições e as relações entre 'dados' e 'metadados' identificadas na revisão da literatura. Considera-se que o levantamento e as discussões apontadas no texto demonstram diversas relações entre os conceitos, principalmente, atrelados à *Web Semântica*, *Ciência dos dados*, *Big data*, entre outros, o que leva à necessidade de aprofundamento nas definições e reflexões na área de Ciência da Informação, devido às divergências de perspectivas conceituais.

Palavras-chave: Dados. Metadados. Ciência da Informação. Dado e metadado.

Data and metadata: concepts and relationships

ABSTRACT

The interdisciplinarity of research in related areas can present many contributions to the development of theories. However, it can also provide terminological confusion. Thus, the importance of discussing these concepts and avoiding nomenclature misconceptions is highlighted. For this reason, the objective of this work is to discuss and relate data and metadata concepts in Information Science. The research is characterized by an exploratory analysis methodology, making it possible to identify conceptual elements from the scientific literature, analyzing the information from Perspectivism. As a result, the definitions and relationships between 'data' and 'metadata' identified in the literature review are presented. It is considered that the survey and the discussions presented in the text demonstrate several relationships between the concepts, mainly linked to the Semantic Web, Data Science, Big data, among others, raising the need for further definitions and reflections in the area of Science of the Information, due to divergences in conceptual perspectives.

Keywords: Data. Metadata. Information Science. Data and metadata.

Datos y metadatos: conceptos y relaciones

RESUMEN

La interdisciplinariedad de la investigación en áreas relacionadas puede presentar muchas contribuciones al desarrollo de teorías. Sin embargo, también puede provocar confusiones terminológicas. Por lo tanto, se destaca la importancia de discutir estos conceptos y evitar conceptos erróneos de nomenclatura. Por esta razón, el objetivo de este trabajo es discutir y relacionar los conceptos de datos y metadatos en Ciencias de la Información. La investigación se caracteriza por una metodología de análisis exploratorio, que permite identificar elementos conceptuales de la literatura científica, analizando la información del Perspectivismo. Como resultado, se presentan las definiciones y relaciones entre “datos” y “metadatos” identificados en la revisión de la literatura. Se considera que la encuesta y las discusiones presentadas en el texto demuestran varias relaciones entre los conceptos, principalmente vinculados a la Web Semántica, la Ciencia de Datos, Big Data, entre otros, lo que plantea la necesidad de nuevas definiciones y reflexiones en el área de la Ciencia de la Información, debido a divergencias en perspectivas conceptuales.

Palabras clave: Datos. Metadatos. Ciencia de la información. Datos y metadatos.

INTRODUÇÃO

Diversas áreas apresentam contribuições significativas para a construção e expansão da *Web*. Essa pluralidade de teorias e aplicações para a *Web* gera confusões terminológicas entre áreas correlatas cada vez mais frequentes. Por essa razão, é necessário refletir sobre os conceitos, no intuito de amenizar ou evitar as ambiguidades com relação ao vocabulário conceitual.

O termo dado, por exemplo, foi empregado em diversos contextos, principalmente na perspectiva do atual cenário e das discussões envolvidas nas questões da *Web* semântica, *Linked Data*, *Big Data*, curadoria digital, *e-Science*. Como também, o termo foi relacionado entre outros termos e conceitos nas tendências da Ciência de Dados (*Data Science*), Ciência da Computação e Ciência da Informação.

Nesse sentido, o uso equivocado pode causar confusões em teorias, hipóteses, metodologias e no próprio desenvolvimento científico. Assim, destaca-se a importância de se contextualizar os conceitos e objetos nos estudos, quando é trabalhado com termos que possuem diversas definições e concepções entre duas ou mais áreas.

Na área de Ciência da Informação, isso não é diferente. Com forte relacionamento e interdisciplinaridade com a Ciência da Computação, Ciência de Dados, dentre outras áreas, o termo dado tem tomado inúmeros significados. No caso, foi observado pela literatura que o termo tem sido empregado como sinônimos de metadados, e ainda, na atualização e evolução dos conceitos relacionados a esses dois termos (dados e metadados).

Furner (2019) aponta a necessidade de discutir e comparar o conceito de metadados com outros termos, como exemplo: dados, documento, informação e os dados do próprio registro. Além disso, Baker *et al.* (2011) ressaltam a necessidade de uma discussão conceitual da terminologia empregada em áreas interdisciplinares, como a Ciência da Informação e a Ciência da Computação.

Diante desse cenário, o objetivo deste trabalho consiste em discutir e relacionar os conceitos dos termos ‘dado’ e ‘metadado’, no intuito de debater a proximidade conceitual na Ciência da Informação. Portanto, com a apresentação desse texto, busca-se compartilhar as reflexões do Grupo de Pesquisa “Dados e Metadados” sobre a relevância em contextualizar e discutir sobre esses conceitos na literatura.

Apresenta-se o artigo em uma estrutura dividida em: introdução, uma seção para os procedimentos metodológicos adotados, a análise e discussão sobre dados e metadados divididos em duas seções e, por fim, as considerações finais.

PROCEDIMENTOS METODOLÓGICOS

A pesquisa é caracterizada por uma análise exploratória para identificar os elementos conceituais que visam, a partir da literatura científica da área de Ciência da Informação, os conceitos de dados e metadados.

A análise exploratória foi realizada pelo estudo sobre dados e metadados por meio de uma pesquisa bibliográfica. O recorte da pesquisa abrange outras já publicadas internacionalmente e no Brasil, nos idiomas: português, espanhol e inglês. Os procedimentos metodológicos foram divididos em quatro etapas, conforme descrito a seguir:

A primeira etapa está relacionada ao levantamento bibliográfico, que foi realizado em bases de dados como: Base de Dados Referenciais de Artigos de Periódicos em Ciência da Informação (BRAPCI); Repositório Questões em Rede – Coleções (Coleção BENANCIB); *Library and Information Science Abstracts* (LISA); *Library, Information Science and Technology Abstracts* (LISTA); Biblioteca Digital Brasileira de Teses e Dissertações (BDTD); Teses e dissertações da CAPES; OASISBR: portal brasileiro de publicações científicas em acesso aberto do IbiCT; Portal de Periódicos da Capes; *Scientific Electronic Library Online* (SciELO); *Scopus*; e *Web of Science*. Além de outras fontes de informação, como os Anais do *Dublin Core Metadata Initiative International Conference on Dublin Core and Metadata Applications*; *Google Scholar* e as publicações do *World Wide Web Consortium* (W3C).

Para a realização do levantamento, elaborou-se a estratégia de busca com termos principais (dados e metadados) para a localização dos materiais selecionados. Quando possível, fez-se a busca pelo título, por palavras-chave e resumos, na última década, ou seja, 2008-2018.

A tipologia dos trabalhos pesquisados foi: artigos, livros, teses, dissertações, trabalhos em eventos e publicações da W3C.

Após a realização do levantamento e da identificação do *corpus* do trabalho, operou-se a segunda etapa, com uma leitura do resumo e, quando necessário, uma leitura prévia do texto para que pudesse aplicar os seguintes critérios para seleção do material para fundamentação teórica do texto: 1) a partir de uma leitura do resumo verificou-se a relevância da temática do artigo para o escopo da pesquisa; 2) idioma dos documentos (português, inglês e espanhol); e 3) atualidade e pertinência do texto para pesquisa.

A terceira etapa caracterizou-se pela leitura e pelo fichamento dos textos completos. Essa etapa teve como propósito o desenvolvimento da base teórica para a discussão dos diferentes pontos de vista identificados na literatura sobre o tema.

Logo em seguida, durante a quarta etapa, considerou-se o Perspectivismo proposto por Peterson (1996) para analisar os diversos pontos de vista dos conceitos identificados, pois,

[...] a adequação de uma perspectiva é sempre relativa, mas ela não se estabelece de modo arbitrário. Na escolha de uma perspectiva deve-se considerar a sua adequação ao usuário a uma situação determinada, a um processo, a uma ontologia e a uma meta específica. É a análise de cada um destes componentes que se pretende desenvolver no estudo dos processos que atuam nas diretrizes, modelagens e estruturas de sistemas para atendimento de necessidades de sujeitos em ambientes informacionais específicos (SANTOS; VIDOTTI, 2009, *não paginado*).

Com essa perspectiva, na quarta etapa, caracterizou-se a análise e o estabelecimento das características fundamentais extraídas da literatura sobre dados e metadados, para elucidação do problema de pesquisa. Criou-se, assim, a base teórica para a elaboração (redação) da pesquisa.

DISCUSSÃO CONCEITUAL SOBRE OS METADADOS

Ao longo dos anos, autores como Alves (2010), Alves e Santos (2013), Joudrey, Taylor e Wisser (2018), Méndez Rodríguez (2002), Pomerantz (2015) e Zeng e Qin (2008, 2016) têm discutido e estruturado definições sobre o termo metadados na perspectiva da Ciência da Informação. Dentre esses autores, o que é possível identificar como ponto em comum é que o conceito de metadados está atribuído a uma informação estruturada para as ações de identificação, descoberta, seleção, uso, acesso e gerenciamento.

O termo metadado é conceituado por meio das “[...] informações de valor agregado que criam para organizar, descrever, rastrear e melhorar o acesso a objetos de informação e itens físicos e coleções, relacionados a esses objetos” (GILLILAND, 2016, p. 2, *tradução nossa*). Zeng e Qin (2016) indicam as variações de definição do conceito de metadados por meio de diferentes comunidades de prática, ou ainda, que metadados são informações sobre coisas específicas.

Os metadados, inicialmente, foram identificados pela expressão ‘dados sobre dados’, cunhada na década de 1960 para se referir a um conjunto de declarações sobre os dados (POMERANTZ, 2015). No entanto, é perceptível que os metadados fazem parte da rotina de diversas classes profissionais que projetam, criam, descrevem, preservam e usam sistemas e recursos informacionais (GILLILAND, 2016).

Ressalta-se que, por meio dessa rotina, os metadados são ora entendidos como produtos, ora como processo. Mayernik (2020, *não paginado, tradução nossa*) apresenta em seu trabalho um panorama sobre os metadados. Destaca que “[...] o uso de padrões para criar metadados estruturados resulta no que pode ser caracterizado como ‘produtos de metadados’”. Isto é, o registro informacional gerado é o produto de metadados.

Edwards *et al.* (2011, p. 684, *tradução nossa*) esclarecem ainda que “[...] os produtos de metadados bem codificados aumentam a precisão com a qual um conjunto de dados que pode ser ajustado para finalidades para as quais não foi originalmente planejado ou pode ser reutilizado por pessoas que não participaram da criação”. Por outro lado, o processo, por exemplo, pode ser relacionado como a interpretação de regras de instrumentos de representação, como o *Anglo-American Cataloguing Rules 2nd ed.* (AACR2r) no domínio bibliográfico, para apoiar no processo de decisão do catalogador durante a definição de informações que serão inseridas para compor o registro informacional.

Nesse sentido, Haynes (2004) afirma que a complexidade de compreensão dos metadados e as suas funções são essenciais para as atividades dos setores que envolvem conhecimento, informação, cultura e aprendizagem.

As funções dos metadados estão direcionadas à sua tipologia. Segundo Michael Buckland (2017), o primeiro e original uso dos metadados é a descrição de documentos, enquanto o uso adicional dos metadados é permitir a pesquisa. “Os metadados podem ser usados para fornecer estruturas que suportam a pesquisa e a descoberta consistentes de informações em uma ampla variedade de documentos. Os metadados também podem, potencialmente, permitir distinções entre tipos ou documentos semelhantes” (BUCKLAND, 2017, p. 118, *tradução nossa*).

A partir da análise terminológica realizada por Arakaki (2019), observou-se que, durante os últimos anos, houve uma disparidade na definição das tipologias dos metadados. Mayernik (2020) discute que essa variedade de categorias encontrada na literatura indica uma compreensão ampla do que o conceito de metadados pode abranger, além de representar tarefas ou ações específicas.

Para Mayernik (2020), a única semelhança entre essas várias definições e categorizações de metadados é que estes são criados para serem usados, por alguns propósitos e, ainda, por pessoas ou pelos aplicativos de computador. De acordo com Arakaki (2019, p. 80-81), as tipologias são identificadas como:

- *Metadados administrativos*: usados para gerenciar e administrar coleções e recursos informacionais, para auxiliar na tomada de decisão e na manutenção dos registros e recursos informacionais. Fornecem informações sobre a origem e a manutenção de um objeto;
- *Metadados de autenticação*: são informações que possibilitam a identificação, integridade, legitimidade de um recurso informacional;
- *Metadados de preservação*: estão relacionados com informações de preservação e conservação dos recursos informacionais;
- *Metadados de proveniência*: estão relacionados às informações de procedência, fornecem dados sobre entidades, criação e modificações e seus relacionamentos;
- *Metadados técnicos*: estão relacionados a como um sistema funciona, fornecem informações do sistema ou do recurso;
- *Meta-metadados (Metametadata)*: correspondem às informações sobre o registro criado, ou informações da criação de um conjunto de dados;
- *Metadados descritivos*: identificam características identificadoras e os contextos intelectuais dos recursos de informação para fins de descoberta, identificação, seleção, aquisição, contexto e compreensão;
- *Metadados de direitos*: estão relacionados às informações sobre propriedade e direitos autorais;
- *Metadados de acesso e uso*: são informações de como um recurso informacional foi acessado e utilizado, como restrições de circulação e acesso, registros de exposições, entre outros;

- *Metadados estruturais*: estão relacionados à composição e à organização do recurso informacional;
- *Linguagens de marcação (Markup languages)*: integram metadados e sinalizações para outros recursos estruturais ou semânticos.

As tipologias dos metadados estão presentes na seleção do padrão de metadados a ser utilizado no sistema informacional. Conforme apontado por Zeng e Qin (2008), há uma intrínseca ligação no estabelecimento de metadados e formatos de metadados. A construção de um padrão de metadados exige a adoção de procedimentos metodológicos para a definição dos metadados, assim como eles precisam estar em uma estrutura de descrição padronizada. Santos, Simionato e Arakaki (2014) apontam que a definição dos metadados deve ser uma ação consensual para que o sistema contenha interoperabilidade de seus dados e, ainda, quando há dados ambíguos e que necessitam de um elevado detalhamento do recurso informacional.

Ademais, os metadados comprovam a autenticidade e o grau de completude do recurso, estabelecem o seu contexto, identificam suas relações estruturais com outros recursos, provêm diversos pontos de acesso para diferentes tipos de usuários e podem fornecer informações que são geralmente obtidas por meios tradicionais (GILLILAND, 2016).

RELAÇÕES TERMINOLÓGICAS DO TERMO DADO

Os dados são destacados, inicialmente, pelo conceito atribuído por Santos e Sant’Ana (2013, p. 205), os quais definem dado como

“[...] uma unidade de conteúdo necessariamente relacionada a determinado contexto e composta pela tríade entidade, atributo e valor, de tal forma que, mesmo que não esteja explícito o detalhamento sobre contexto do conteúdo, ele deverá estar disponível de modo implícito no utilizador, permitindo, portanto, sua plena interpretação”.

A partir da definição de Santos e Sant’Ana (2014), observa-se que o dado é composto pela tríade entidade, atributo e valor. Apesar de implícito, o dado sempre está atrelado a um contexto.

No intuito de traçar uma evolução do conceito do termo dado, Furner (2016) faz um levantamento da definição de dados ao longo dos séculos. Por uma perspectiva histórica, discute a mudança do sentido. De acordo com o autor, o conceito dos dados pode possuir diversas abordagens, tais como:

- *Abordagem extensional*: busca caracterizar coisas ou tipos de coisas que se enquadram no conceito de “dados”;
- *Abordagem intencional*: identifica as propriedades que algo deve ter, caso seja tratado como dados;
- *Abordagem classificatória*: reconhece um conceito individual como “dados” pode ter, ou possuir múltiplos sentidos, e que esses sentidos podem ser categorizados de acordo com similaridades em função e contexto;
- *Abordagem histórica*: conduz análises lógicas e/ou computacionais das propriedades necessárias dos conceitos, permitir que os autores considerem o desenvolvimento culturalmente específico dos significados de termos, como dados ao longo do tempo.

A partir dessas abordagens (extensional, intencional, classificatória e histórica), Furner (2016) destaca as diversas interpretações que os dados podem assumir:

- *A interpretação clássica*: origem do termo do latim *dātŭm*, como o verbo ‘dar’. Furner (2016) discute a origem do termo em latim *dātŭm*, por volta do ano 100a.C., utilizado muitas vezes como verbo (dar);
- *A interpretação documental*: dados como metadados. De acordo com Furner (2016), ainda por volta do ano de 100a.C., o termo ‘dado’ (*dātŭm*) começou a ser usado como substantivo, como informação sobre algo, ou seja, como metadado;

- *A interpretação eclesiástica*: dados como dons de Deus. Furner (2016) relata que por volta de 1614, o termo ‘dados’ começou a ser utilizado em sermões, com o significado de ‘com a graça de Deus’;
- *A interpretação geométrica*: dados como premissas geométricas. No contexto da geometria, o termo ‘dado’ passou a ser empregado por volta de 1645, para representar os valores dos lados e ângulos (FURNER, 2016);
- *A interpretação matemática*: dados como premissas matemáticas. Após a interpretação geométrica, por volta de 1704, o conceito de dados foi ampliado para qualquer aplicação matemática, independente da área de aplicação (FURNER, 2016);
- *A interpretação epistêmica*: dados como evidência. Por volta de 1648, o termo ‘dado’ foi incorporado em alguns dicionários atribuindo o conceito de dados como fatos;
- *A interpretação informacional*: dados como valores de atributo. Na segunda metade do século XIX, há uma mudança na interpretação dominante do conceito de dados, sendo uma das principais alterações que o termo não apenas se refere a informações numéricas;
- *A interpretação computacional*: dados como *bits*. O termo ‘dado’ na computação foi aplicado pela primeira vez em 1953, com a publicação do IBM701 *Electronic Data Processing Machine*. O termo, porém, só foi constar nos dicionários terminológicos da computação em 1980. A princípio, utilizou-se o termo ‘dado’ para definir o valor do atributo em um banco de dados. Posteriormente, encontrou-se também como sinônimo de *bits*;
- *A interpretação diafórica*: por volta dos anos 2000, o termo relaciona-se à sua pluralidade. Os dados são atribuídos como realidade objetiva, aparências subjetivas, observações, ideias, significados, ou mesmo expressões linguísticas de observações individuais.

Ao analisar do ponto de vista da interpretação informacional dos dados destacada por Furner (2016), observa-se forte relação com a definição e a aproximação do conceito de metadados.

Diante disso, Mayernik (2020) discute a possibilidade de categorizar os metadados como uma tipologia dos dados e conclui que “[...] se os dados são entidades usadas como evidência, os metadados são os processos e produtos que permitem que essas entidades sejam responsabilizadas como evidência” (MAYERNIK, 2019, p. 735, tradução nossa).

Corroborando com essa discussão focada em ambientes de bibliotecas, Baker *et al.* (2011) atribuem aos dados que são produzidos por meio de uma informação ou curadoria. São nomeados como ‘dados de biblioteca’. Ainda, descrevem recursos ou ajudam a sua descoberta. Os dados de bibliotecas dividem-se em: *datasets* (conjuntos de dados), *metadata element set* (conjuntos de elementos) e *value vocabularies* (vocabulários de valor) (BAKER *et al.*, 2011).

Os *datasets* são coleções estruturadas de metadados para descrição de recursos, como livro. Isto é, um conjunto de registros de metadados. Os *datasets* equivalem aos registros de bibliotecas que consistem em declarações, elementos da entidade e seus valores. Os elementos são definidos a partir de padrões, como *Machine Readable Cataloging* (MARC21) ou *Dublin Core* (DC) e os valores por vocabulários de valores, como a *Library of Congress Subject Headings* (LCSH) (ISAAC *et al.*, 2011). Pode-se considerar exemplos, como a *British National Bibliography*, o catálogo da *Hungarian National Library*, o *CrossRef* e a *Europeana* (BAKER *et al.*, 2011).

Furner (2016, p. 303) argumenta que todos os conjuntos de dados são documentos, afirmando que ‘o conjunto de dados é uma espécie de documento’. Portanto, se os metadados são um tipo especial de dados, conforme observado acima, os metadados também existem como documentos, não como conceitos abstratos ou informações que existem sem forma material. Como tal, os metadados podem ser analisados através do mesmo aparato conceitual dos documentos (MAYERNIK, 2020, não paginado, tradução nossa).

Os *value vocabularies* definem os valores dos elementos para descrição de um recurso. Eles não estabelecem informações de um recurso, mas sim conceitos relacionados a um recurso como pessoas, assuntos, idiomas, países etc. (ISAAC *et al.*, 2011). Como exemplos: o *Library of Congress Subject Headings*, o *Virtual International Authority File* (VIAF), a Classificação Decimal de Dewey (CDD) e o *GeoNames*.

Metadata element set é determinado como “[...] um conjunto de elementos de metadados que definem as classes e atributos utilizados para descrever entidades de interesse” (ISAAC *et al.*, 2011, não paginado, tradução nossa). O conjunto de elementos também designa um conjunto completo de elementos de metadados, assim como a codificação dos elementos e estrutura em uma linguagem de marcação (ZENG; QIN, 2008). Podem ser exemplificados: o *Dublin Core Metadata Terms*, os elementos do *Resource Description and Access* (RDA), do *Simple Knowledge Organization System* (SKOS) e o *Friend of a Friend vocabulary* (FOAF) (BAKER *et al.*, 2011).

Nesse sentido, a literatura aponta que, em muitos casos, dados e metadados são tratados como sinônimos. Segundo Wickett *et al.* (2013, não paginado, tradução nossa), “[...] os componentes de dados e metadados estão entrelaçados: nenhuma distinção estrutural permite uma discriminação imediata entre dados e metadados”.

Exemplificando uma situação no contexto das bibliotecas, Jeffery *et al.* (2014, não paginado, tradução nossa) distinguem que “[...] para o pesquisador, o registro da biblioteca são metadados para descobrir um livro ou artigo de interesse. Para o bibliotecário, o registro pode ser utilizado como dados para analisar a completude relativa das coleções por assunto, por editora, por ano etc.”. Ou seja, o mesmo objeto (dado/metadado) pode ser considerado metadado, como forma de descoberta e busca de um recurso informacional no catálogo, por exemplo.

Como também, poderá ser avaliado como dado, no momento em que o bibliotecário começar a analisar os registros bibliográficos contidos na biblioteca como um todo, para realizar análises do acervo, de empréstimos, entre outras possibilidades.

De acordo com Hyvönen (2012, p. 10, *tradução nossa*), “[...] em torno de 2005, as ideias sobre *Linked Data* e *Web* de dados começaram a ganhar impulso como uma abordagem simples para *Web* Semântica, focada na publicação de grandes conjuntos de dados existentes e usando apenas ontologias RDF simples e leves”. Esse processo foi um dos movimentos para ressaltar a importância dos dados e a ampliação das pesquisas na área.

Além disso, os termos são essenciais nas discussões sobre a *Web* Semântica e para os princípios *Linked Data*. Os quatro princípios, definidos por Tim Berners-Lee em 2006, são sugestões de boas práticas para a estruturação de dados. No entanto, não são tidos como requerimentos obrigatórios. Esses princípios são:

1. Use URIs como nomes para as coisas;
2. Use HTTP URIs para que as pessoas possam procurar esses nomes;
3. Quando alguém procurar um URI, fornecer informações úteis, usando os padrões (RDF, SPARQL);
4. Incluir links para outros URIs, para que eles possam descobrir mais coisas.

No contexto dos metadados, Pomerantz (2015, p. 158, *tradução nossa*) faz uma releitura dos princípios do *Linked Data*. Dessa forma, Pomerantz (2015) enfatiza a questão da identificação dos recursos ao invés de coisas e ressalta a persistência dos metadados, conforme destacados nos princípios 3 e 4. Sendo que os princípios são redefinidos para:

1. Use URIs como identificadores para recursos;
2. Formate as URIs de acordo com o HTTP, para que os recursos possam ser encontrados facilmente, usando a tecnologia estabelecida;

3. Use padrões como o RDF para fornecer tanto o recurso e metadados sobre o recurso;
4. Fornecer *links* juntamente com esses metadados para outros URIs, para que mais recursos possam ser encontrados.

Paralelo a esse movimento, às tecnologias e aos novos paradigmas, tanto da Ciência como a *e-Science*, quanto da *Web* de documentos para *Web* de dados, impulsionou-se a importância dos dados para os processos do dia-a-dia. Nesse contexto, novos conceitos relacionados aos dados têm surgido, como a definição de *Big Data*, e mesmo as discussões de um campo específico para tratar de dados, como a Ciência de Dados.

Diante dos problemas e dos desafios de estudos sobre metadados na Ciência de Dados, Greenberg (2017) posiciona a importância dos metadados. De acordo com a autora, a Ciência de Dados é contextualizada como “[...] um campo interdisciplinar que possui como objetivo o estudo e avanço dos dados para obter *insights*. Uma empresa de Ciência de Dados pode permitir prever um fenômeno ou automatizar a tomada de decisões” (GREENBERG, 2017, p. 22, *tradução nossa*). A gama de dados encontrada pela Ciência de Dados concentra-se em diferentes formas. Greenberg (2017) destaca que são dados pequenos, grandes, estáticos, estruturados, não estruturados ou em fluxo contínuo e, ainda, são aplicadas metodologias científicas e estatísticas para o aprendizado dos dados.

No entanto, com relação aos metadados, Greenberg (2017) afirma que os esforços da Ciência de Dados dependem de uma descrição precisa dos dados, um subsídio necessário para os processos e para as aplicações. A autora também destaca que as estruturas de metadados garantem a confiabilidade dos dados e identificam uma infinidade de tópicos de pesquisa relacionados a metadados, muitos dos quais encontrados em relatórios governamentais e disciplinares similares em todo o mundo.

O uso e aplicação dos metadados e a importância dos cientistas de dados realizarem o tratamento da informação ou dos dados é destacada por Stanton *et al.* (2012) e por Curty e Serafim (2016). De acordo com Curty e Serafim (2016, p. 312), “[...] os cientistas de dados são responsáveis pela identificação, coleta, tratamento, transformação, análise, visualização e curadoria de grandes conjuntos de dados heterogêneos.”. Esses aspectos foram descritos por Mishra e Chang (2015) e por Rautenberg e Carmo (2019, p. 62), que em seus trabalhos ressaltam a importância do gerenciamento, da proveniência, da curadoria e do arquivamento de dados e dos metadados, principalmente no âmbito do *Big Data*. Segundo Rautenberg e Carmo (2019, p. 62):

[...] a Ciência da Informação contribui no fomento das competências de organização e representação de dados e informação, privilegiando os serviços de coleta, registro, filtragem, classificação e entrega de dados e seus metadados às atividades reservadas à camada da Ciência de Dados.

Os autores destacam as habilidades e a interdisciplinaridade da Ciência da Informação e Ciência da Computação. Tal relacionamento auxilia e possibilita aos gestores uma melhor visão para o processo de tomada de decisão guiada por dados, mesmo que seja uma grande proporção de dados, como no *Big Data*. Rautenberg e Carmo (2019, p. 62) salientam que os profissionais da Ciência da Informação e da Computação podem atuar juntos em diversas frentes da Ciência de Dados e do *Big Data*.

A relação entre Ciência da Informação e *Big Data* já foi relatada por diversos autores como: Dias e Vieira (2013), Souza, Ribeiro e Porto (2013), Ribeiro (2014), Araújo Júnior e Souza (2016), Dutra e Macedo (2016), Rodrigues, Dutra e Dias (2017), Coneglian, Gonzalez e Santarém Segundo (2017), Souza, Martins e Ramalho (2018), Andrade, Gonzalez, Berti Júnior, Baptista e Coneglian (2020), Reis e Sá (2020), dentre outros. Nesta relação, os autores realçam o termo dado ao tratamento e processamento de grande volume de dados.

Araújo Júnior e Souza (2016) relacionam as etapas do ciclo documentário (seleção, aquisição, registro ou tombamento, descrição, análise ou condensação, indexação, armazenamento dos documentos, armazenamento da representação, processamento da informação, produtos do processamento, interrogação e busca e recuperação da informação) proposto por Robredo (2005) como subsídio para a definição do tratamento dos dados no *Big Data*.

Corroborando com essa linha, outra questão que reflete é sobre as atividades da curadoria digital propostas por Higgins (2008), que incluem a descrição e representação, planejamento da preservação, acompanhamento e participação da comunidade, curadoria e preservação, conceitualização, criação e/ou recepção, avaliação e seleção, admissão, ações de preservação, armazenamento, acesso, uso e reuso, transformação, descarte, reavaliação e migração. Dutra e Macedo (2016) e Rautenberg e Carmo (2019) propõem que essas atividades sejam empregadas no contexto do *Big Data*.

Neste contexto, os metadados são fundamentais para possibilitar o gerenciamento de grande volume de dados. Triques, Arakaki e Castro (2020) discutiram e apresentaram elementos que relacionam a curadoria digital com a representação da informação, em especial, utilizando os metadados para auxiliar nas atividades curatoriais. Os metadados são evidenciados no contexto do *Big Data* por Reis e Sá (2020). De acordo com os autores,

A proliferação de formatos de arquivos digitais e a explosão informacional vivida atualmente, com a produção de conteúdo digital por smartphones e outros dispositivos conectados à internet, geram uma grande quantidade de metadados. A quantidade disponível e sua relação indissociável do cotidiano do indivíduo tornam os metadados a mais rica fonte para a análise e extração de informações sobre consumidores e seus hábitos, despertando o interesse de empresas que buscam atingir vantagem competitiva através de ferramentas de análise de dados. Os metadados são, portanto, o principal insumo do *Big Data* (REIS; SÁ, 2020, p. 237).

Conforme observado, os metadados são a fonte primária para o *Big Data*, seguindo na perspectiva da interpretação documental (dados como metadados), conforme destacado por Furner (2016).

Nesta perspectiva interpretativa documental de Furner (2016), do que são os dados tidos como metadados, Magalhães *et al.* (2014) esclarecem que o *Big Data* é uma grande massa de dados com características estruturadas e não-estruturadas que busca a autenticidade dos objetos (dado/metadado) para dar sentido às informações que podem agregar alguma forma de valor a empresas e governos.

A partir dessa relação, os textos de Coneglian, Gonzalez e Santarém Segundo (2017, p. 138), e Reis e Sá (2020) apontam o papel do bibliotecário e do profissional da informação no *Big Data*, que pode facilitar e permitir a descoberta e recuperação dos dados, inferir na manutenção e qualidade dos dados, arquivamento, representação e preservação de dados, além de possibilitar a agregação de valor dos dados por meio dos metadados.

O termo também foi utilizado para fundamentar o *Big Data* e sua relação com os padrões de representação *Resource Description Framework* (RDF); *Simple Knowledge Organization System* (SKOS) e *Ontology Web Language* (OWL) pelos autores Souza, Martins e Ramalho (2018). Além disso, os autores relacionam esses padrões com os conceitos da *Web Semântica* e concluem que “[...] foi verificado que os modelos de representação analisados contribuem para interligar grandes volumes de dados sem perder o contexto no qual são originados, favorecendo um melhor entendimento do *Big Data* e os novos paradigmas de representação em ambientes digitais.” (SOUZA; MARTINS; RAMALHO, 2018, p. 18).

Conforme visto, diversos autores relacionam a Ciência de Dados e o *Big Data* com as questões do tratamento e gerenciamento de metadados, utilizando-se de aportes teóricos e metodológicos, além de técnicas da Representação e Organização da Informação e do Conhecimento para estruturação dos dados.

CONSIDERAÇÕES FINAIS

Diante do levantamento e das discussões apresentadas no texto, observou-se que na literatura da área de Ciência da Informação há diversas relações entre os conceitos de dados e metadados. Como visto, os metadados podem ter diversas interpretações, como clássica, documental, eclesiástica, geométrica, matemática, epistêmica, informacional, computacional e diafórica. Em muitos casos, a depender da interpretação e do contexto que são tratados, os dados podem ser entendidos como metadados ou documento.

Na tentativa de esclarecer essas relações, em muitos casos, pode depender da perspectiva em que analisa os dados e de seu contexto. Ao analisar da ótica do usuário da biblioteca, o profissional da informação, bibliotecário, constrói registros informacionais, extrai e descreve os metadados referentes a algum recurso informacional para posterior busca pelo usuário. Assim, esta busca por uma informação a partir de metadados. Quando esse conjunto de registros criados são usados para a criação de relatórios ou pesquisa, pode-se considerar essas informações como dados, referindo-se ainda a esses registros como uma forma de documento e uma evidência dos recursos que constam na biblioteca, por exemplo.

O desenvolvimento tecnológico, atrelado às discussões e à formalização da *Web Semântica*, Ciência dos dados, *Big data*, entre outros conceitos, evidenciou a importância dos dados. Entretanto, verificou-se que há poucos estudos que contextualizam os metadados na Ciência dos Dados, e uma tendência de estudos sobre *Big Data*.

Ao final, não foi possível delimitar uma definição única dos dois conceitos, mas pôde-se verificar a diversidade epistemológica encontrada sobre a temática. Apesar disso, muitos autores discordam com esses conceitos apresentados e se destaca a necessidade de aprofundamento em trabalhos futuros, com a finalidade de fomentar discussões sobre a temática e contextualizá-la. Além disso, ainda é necessário delinear posicionamentos e perspectivas divergentes entre pesquisas.

REFERÊNCIAS

- ANDRADE, M. C. *et al.* Ciência responsável dos dados: imparcialidade, precisão, confidencialidade, e transparência dos dados. *Informação & Informação*, v. 25, n. 2, p. 26-48, 2020. Disponível em: <http://www.uel.br/revistas/uel/index.php/informacao/article/view/38467>. Acesso em: 25 set. 2020.
- ALVES, R. C. V. *Metadados como elementos do processo de catalogação*. 2010. 132p. Tese (doutorado em Ciência da Informação) – Faculdade de Filosofia e Ciências, Universidade Estadual Paulista Júlio Mesquita Filho, Marília, 2010. Disponível em: <http://repositorio.unesp.br/handle/11449/103361>. Acesso em: 8 set. 2019.
- ALVES, R. C. V.; SANTOS, P. L. V. A. C. *Metadados no domínio bibliográfico*. Rio de Janeiro: Intertexto, 2013.
- ARAKAKI, F. A. *Metadados administrativos e a proveniência dos dados: modelo baseado na família PROV*. 2019. 139p. Tese (doutorado em Ciência da Informação) - Faculdade de Filosofia e Ciências, Universidade Estadual Paulista Júlio Mesquita Filho, Marília, 2019. Disponível em: <https://repositorio.unesp.br/handle/11449/180490>. Acesso em: 8 set. 2019.
- ARAÚJO JÚNIOR, R. H.; SOUSA, R. T. B. Estudo do ecossistema de big data para conciliação das demandas de acesso, por meio da representação e organização da informação. *Ciência da Informação*, v. 45, n. 3, 2016. Disponível em: <http://revista.ibict.br/ciinf/article/view/4057>. Acesso em: 25 set. 2020.
- BAKER, T. *et al.* *Library Linked Data Incubator Group Final Report*. W3C Incubator Group Report, 2011. Disponível em: <http://www.w3.org/2005/Incubator/lld/XGR-ld-20111025/>. Acesso em: 8 set. 2019.
- BERNERS-LEE, T. *Linked Data: Design Issues*. [s.l.]: W3C, 2006. Disponível em: <http://www.w3.org/DesignIssues/LinkedData.html>. Acesso em: 25 jun. 2015.
- BUCKLAND, M. *Information and Society*. Cambridge: MIT Press, 2017.
- CONEGLIAN, C. S.; GONÇALEZ, P. R. V. A.; SANTARÉM SEGUNDO, J. E. O profissional da informação na era do big data. *Encontros Bibli: Revista Eletrônica de Biblioteconomia e Ciência da Informação*, v. 22, n. 50, p. 128-143, 2017. Disponível em: <https://periodicos.ufsc.br/index.php/eb/article/view/1518-2924.2017v22n50p128>. Acesso em: 12 mar. 2021.
- CURTY, R. G.; SERAFIM, J. da S. A formação em ciência de dados: uma análise preliminar do panorama estadunidense. *Informação & Informação*, v. 21, n. 2, p. 307-331, dez. 2016. Disponível em: <http://www.uel.br/revistas/uel/index.php/informacao/article/view/27945>. Acesso em: 24 set. 2020.
- DIAS, G. A.; VIEIRA, A. A. N. Big data: questões éticas e legais emergentes. *Ciência da Informação*, v. 42, n. 2, 2013. Disponível em: <http://revista.ibict.br/ciinf/article/view/1380>. Acesso em: 12 mar. 2021.
- DUTRA, M. L.; MACEDO, D. D. J. Curadoria digital: proposta de um modelo para curadoria digital em ambientes big data baseado numa abordagem semi-automática para a seleção de objetos digitais. *Informação & Informação*, v. 21, n. 2, p. 143-169, 2016. Disponível em: <http://www.uel.br/revistas/uel/index.php/informacao/article/view/27176>. Acesso em: 12 mar. 2021.
- EDWARDS, P. *et al.* Science Friction: Data, Metadata, and Collaboration. *Social Studies of Science*, v. 41, n. 5, p. 677-690, 2011.
- FURNER J. “Data”: The data. In: KELLY, M.; BIELBY J. (Eds.) *Information Cultures in the Digital Age*. Wiesbaden: Springer VS, 2016.
- FURNER, J. Definitions of “Metadata”: A Brief Survey of International Standards. *Journal of the Association for Information Science and Technology*, 2019.
- FURNER, J. Philosophy of data: why?. *Education For Information*, v. 33, n. 1, p. 55-70, 2017.
- GILLILAND, A. J. Setting the Stage. In: BACA, M. (Org.) *Introduction to Metadata*. 3. ed. Los Angeles: Getty Research Institute, 2016. Disponível em: <http://www.getty.edu/publications/intrometadata/>. Acesso em: 8 set. 2019.
- GREENBERG, J. Big Metadata, Smart Metadata and Metadata Capital: toward greater synergy between data science and metadata. *Journal of Data and Information Science*, v. 2, n. 3, p. 19-36, 2017.
- HAYNES, D. *Metadata for information management and retrieval*. [s.l.]: Facet Publishing, 2004.
- HIGGINS, S. The DCC Curation Lifecycle Model. *The International Journal of Digital Curation*, n. 1, v. 3, 2008.
- HYVÖNEN, E. *Publishing and Using Cultural Heritage Linked Data on the Semantic Web*. EUA: Morgan & Claypool Publishers, 2012.
- ISAAC, A. *et al.* *Library Linked Data Incubator Group: Datasets, Value Vocabularies, and Metadata Element Sets: W3C Incubator Group Report 25 October 2011*. W3C, 2011. Disponível em: <http://www.w3.org/2005/Incubator/lld/XGR-ld-vocabdataset20111025/>.
- JEFFERY, K. *et al.* A 3-Layer Model for Metadata.
- INTERNATIONAL CONFERENCE ON DUBLIN CORE AND METADATA APPLICATION, 13., Portugal, *Anais...* DCMI, EUA. 2014. Disponível em: <http://dcevents.dublincore.org/IntConf/dc-2013/paper/view/199/199>. Acesso em: 8 set. 2019.
- JOUDREY, D. N.; TAYLOR, A. G.; WISSER, K. M. *The organization of information*. 4. ed. California: Libraries Unlimited, 2018.
- MAGALHAES, V. R. V. *et al.* O uso do Big Data na violação da privacidade dos usuários para estratégias de negócios. In: ENCONTRO REGIONAL DE COMPUTAÇÃO E SISTEMAS DE INFORMAÇÃO, 3.; ENCONTRO REGIONAL DE COMPUTAÇÃO E SISTEMAS DE INFORMAÇÃO, *Anais...*, Manaus, 2014.

- MAYERNIK, M. Metadata Accounts: Achieving Data and Evidence in Scientific Research. *Social Studies of Science*, v. 49, n. 5, p. 732-757, 2019.
- MAYERNIK, M. Metadata. In: HJØRLAND, B.; GNOLI, C. *Encyclopedia of Knowledge Organization*. ISKO, 2020. Disponível em: <https://www.isko.org/cyclo/metadata>. Acesso em: 30 abr. 2020.
- MÉNDEZ RODRÍGUEZ, E. *Metadatos y recuperación de información*. Gijón: Ediciones Trea, 2002.
- PETERSON, D. *Forms of representation: an interdisciplinary theme for cognitive science*. Wiltshire: Cromwell Press, 1996.
- POMERANTZ, J. *Metadata*. Cambridge: The MIT Press, 2015.
- RAUTENBERG, S.; CARMO, P. R. Big data e ciência de dados. *Brazilian Journal of Information Science*, v. 13, n. 1, p. 56-67, 2019.
- REIS, L. C. R.; Sá, M. I. F. E. Big data: um novo campo de atuação para bibliotecários. *Prisma.com*, Portugal, n. 41, p. 231-250, 2020.
- RIBEIRO, C. J. S. Big data: os novos desafios para o profissional da informação. *Informação & Tecnologia*, v. 1, n. 1, p. 96-105, 2014.
- ROBREDO, J. *Documentação de hoje e de amanhã*. Brasília: Edição de autor, 2005.
- RODRIGUES, A. A.; DUARTE, E. N.; DIAS, G. A. Desafios da gestão de dados na era do big data: perspectivas profissionais. *Informação & Tecnologia*, v. 4, n. 2, p. 63-79, 2017.
- SANTOS, P. L. V. A. da C.; SANTANA, R. C. G. Dado e Granularidade na perspectiva da Informação e Tecnologia: uma interpretação pela Ciência da Informação. *Ciência da Informação*, v. 42, n. 2, jan. 2013. Disponível em: <http://revista.ibict.br/index.php/ciinf/article/view/228>. Acesso em: 8 set. 2019.
- SANTOS, P. L. V. A. C.; SIMIONATO, A. C.; ARAKAKI, F. A. Definição de metadados para recursos informacionais: apresentação da metodologia BEAM. *Informação & Informação*, Londrina, v. 19, n. 1, p. 146-163, fev. 2014. Disponível em: <http://repositorio.unesp.br/handle/11449/114736>. Acesso em: 8 set. 2019.
- SANTOS, P. L. A. C.; VIDOTTI, S. A. B. G. Perspectivismo e tecnologias de informação e comunicação: acréscimos à Ciência da Informação. *DataGramaZero: revista de Ciência da Informação*, Rio de Janeiro, v. 10, n. 3, 2009.
- SOUZA, J. L.; MARTINS, P. G. M.; RAMALHO, R. A. S. Modelos de representação semântica na era do big data. *Brazilian Journal of Information Science*, v. 12, n. 3, p. 34-40, 2018.
- STANTON, J. *et al.* Interdisciplinary data science education. In: XIAO, N.; MCEWEN, L. R. *Special Issues in Data Management*. Washington: American Chemical Society, 2012. p. 97-113.
- TRIQUEZ, M. L.; ARAKAKI, A. C. S.; CASTRO, F. F. Aspectos da representação da informação na curadoria digital. *Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação*, Florianópolis, v. 25, p. 1-21, maio 2020. Disponível em: <https://periodicos.ufsc.br/index.php/eb/article/view/1518-2924.2020.e69898>. Acesso em: 25 set. 2020.
- WICKETT, K. M. *et al.* Identifying content and levels of representation in scientific data. *Proceedings of The American Society For Information Science And Technology*, v. 49, n. 1, p. 1-10, 2013.
- ZENG, M. L. QIN, J. *Metadata*. New York: Neal-Schuman Publishers, 2008.
- ZENG, M. L. QIN, J. *Metadata*. 2. ed. London: facet publishing, 2016.

AGRADECIMENTOS

Agradecemos ao Grupo de Pesquisa “Dados e Metadados” e o apoio do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), processo 431612/2016-1.