

DADOS TEXTUAIS GERADOS EM PESQUISAS BIBLIOGRÁFICAS: DEFINIÇÃO E REPRESENTAÇÃO

TEXTUAL DATA GENERATED IN BIBLIOGRAPHIC RESEARCH: DEFINITION AND REPRESENTATION

Laura Rocha Silveira Tavares Da Silva^a
Linair Maria Campos^b

RESUMO

Objetivo: Definir o conceito de dados textuais gerados em pesquisas bibliográficas, visando apoiar o reconhecimento e a valorização deste tipo de dado pela comunidade científica, e, a partir daí, o seu reuso. **Metodologia:** Caracteriza-se como uma pesquisa qualitativa e exploratória. Quanto aos procedimentos técnicos, é uma pesquisa bibliográfica que utiliza, para a exploração e interpretação do material bibliográfico, a Análise de Conteúdo, de Bardin. Como aportes teóricos para a elaboração de definições, recorre-se à Teoria do Conceito, de Dahlberg e às diretrizes fornecidas por Seppälä, Ruttenberg e Smith, autores renomados da área da Ontologia Formal. **Resultados:** Apresenta-se uma proposta de definição do conceito de dados textuais gerados em pesquisas bibliográficas, além de metadados relevantes para caracterizá-lo, como parte de uma proposta inicial de um conjunto de metadados básicos para a representação desse tipo de dado. **Conclusões:** O esforço de definir o conceito de dados textuais gerados em pesquisas bibliográficas consiste em um passo importante para incentivar o reuso e a gestão desse tipo de dado.

Descritores: Dados científicos. Dados textuais. Pesquisa bibliográfica. Definição conceitual. Teoria do conceito.

1 INTRODUÇÃO

Atualmente, se observa um crescente interesse pelos dados de pesquisa no contexto científico, motivado principalmente pela possibilidade de validar os resultados das pesquisas e reutilizar os dados gerados em outros estudos. Assim, eles deixam de ser considerados “meros subprodutos da pesquisa” (Sayão; Sales, 2014) e passam a assumir o papel de protagonistas nesse

^a Doutora pelo Programa de Pós-Graduação em Ciência da Informação da Universidade Federal Fluminense (UFF), Niterói, Brasil. E-mail: laurarochoa@id.uff.br

^b Doutora em Ciência da Informação em Ciência da Informação pelo convênio Universidade Federal Fluminense (UFF) e Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT). Docente do Programa de Pós-Graduação em Ciência da Informação da Universidade Federal Fluminense (UFF), Niterói, Brasil. E-mail: lmcampos@id.uff.br

contexto. Embora não seja novidade a importância dos dados no desenvolvimento de pesquisas, seu tratamento para fins de compartilhamento e reuso tem sido incentivado e institucionalizado pela comunidade científica.

Quando se fala em dados de pesquisa, é comum associá-los ao fenômeno do *Big Data*, em que dados complexos são gerados por instrumentos superpotentes, disseminados em larga escala, demandando um alto custo para serem obtidos e processados etc. (Jin *et al.*, 2015). Contudo, não apenas esses dados são importantes para o desenvolvimento da ciência, também feita por pesquisadores em suas casas, em pequenos laboratórios, sem grandes investimentos. Esses dados integram a chamada cauda longa da ciência, que representa o segmento formado por um grande número de pequenas equipes de pesquisadores e laboratórios independentes, responsáveis por produzir uma ampla e heterogênea variedade de coleções de dados de pesquisa, que, no entanto, costumam ser invisíveis para a comunidade científica (Sales; Sayão, 2018). Heidorn (2008) os denomina como “*dark data*” (dados obscuros), por serem mais difíceis de encontrar e reutilizados ou preservados com pouca frequência, sendo normalmente subutilizados e eventualmente perdidos.

Neste estudo, aborda-se um tipo específico de dado de pesquisa, situado no contexto da cauda longa da ciência: trata-se dos dados gerados em pesquisas baseadas essencialmente em fontes bibliográficas (e não em dados empíricos), que possuem características únicas, devido ao caráter qualitativo inerente às investigações. Eles visam sintetizar e sistematizar conteúdos apreendidos nos textos lidos a partir dos objetivos da pesquisa, sendo normalmente produzidos para apoiar a realização de trabalhos científicos como dissertações, teses, artigos etc. Ressalta-se que não se tratam dados que descrevem de modo genérico as referências bibliográficas de uma pesquisa, comumente armazenados em gerenciadores de referências bibliográficas, ou que descrevem um recurso informacional no âmbito de um catálogo, por exemplo.

Apesar do notório valor desses dados, na prática eles raramente são considerados, tratados e preservados como dados de pesquisa, tanto por seus produtores, quanto pela comunidade científica, dificultando sua descoberta e consequente reuso. Uma vez que não se reconhece a importância deles,

difícilmente surgirão reivindicações, iniciativas e recompensas relativas ao seu compartilhamento. Em vista disso, questiona-se: o que pode ser feito para tornar os dados textuais gerados em pesquisas bibliográficas visíveis para a comunidade científica e passíveis de serem reutilizados em novas pesquisas?

O caminho para tornar visíveis os *dark data* passa pela superação de barreiras pessoais, institucionais, materiais etc. Contudo, para que isso seja possível no contexto dos dados produzidos em pesquisas bibliográficas, é preciso que a comunidade científica não tenha dúvidas sobre o que são e reconheça o valor desses dados. Dessa forma, iniciamos esse esforço caracterizando e definindo o conceito de dados textuais gerados em pesquisas bibliográficas, a partir da Teoria do Conceito, de Ingetraut Dahlberg, e nas recomendações fornecidas pela autora e por Seppälä, Ruttenberg e Smith (2017) para auxiliar na redação, edição e validação de definições.

Este estudo justifica-se por buscar meios para evidenciar e apoiar o reconhecimento dos dados textuais gerados em pesquisas bibliográficas, de modo que eles não fiquem restritos aos seus produtores e sujeitos a serem perdidos com o tempo. Além disso, ao descrever a elaboração da definição do conceito, explicita a necessidade e a importância de se recorrer a princípios teóricos que conferem qualidade e consistência às definições construídas.

O trabalho encontra-se organizado da seguinte maneira: após esta introdução, apresenta-se o referencial teórico utilizado; os princípios para a elaboração de definições; a metodologia adotada; o conceito de dados textuais gerados em pesquisas bibliográficas; os resultados e as considerações finais.

2 REFERENCIAL TEÓRICO

Para se entender o conceito de dados textuais gerados em pesquisas bibliográficas, é preciso primeiramente compreender o que é um dado de pesquisa, um conceito que se mostra fortemente relacionado àquele. Todavia, ao se buscar uma definição sobre dado de pesquisa, observa-se uma falta de consenso sobre sua conceituação. Para Sayão e Sales (2020, p. 33), isso se deve ao fato de que “[...] ele pode ser muitas coisas diferentes para pessoas e circunstâncias diferentes”, podendo variar consideravelmente entre

pesquisadores e áreas do conhecimento. Logo, o contexto se torna crucial para a identificação de algo como sendo dado de pesquisa.

Os dados de pesquisa podem ser definidos como registros factuais (pontuações numéricas, registros textuais, imagens e sons) usados como fontes primárias para a pesquisa científica, que costumam ser aceitos na comunidade científica como necessários para validar os resultados da pesquisa (OECD, 2007). Para Rice e Southall (2016), são dados, registros, arquivos ou outras evidências, independentemente do seu conteúdo ou forma, que compreendem observações de pesquisa, achados ou resultados, incluindo materiais primários e dados já analisados, que são coletados, observados ou criados para fins de análise e produção de resultados de pesquisas originais. Borgman (2007) recorre ao Modelo de Referência *Open Archival Information System*, que considera dado de pesquisa como “uma representação reinterpretável de informações de maneira formalizada adequada para comunicação, interpretação ou processamento” (Consultative Committee for Space Data System, 2012). Já Sayão e Sales (2020), propõem uma definição abrangente:

Dado de pesquisa é todo e qualquer tipo de registro coletado, observado, gerado ou utilizado no âmbito da pesquisa científica, que pode ser interpretado, tratado e aceito como evidência pela comunidade científica e necessário para analisar, validar e produzir resultados de pesquisa (Sayão; Sales, 2020, p. 32).

Ao se falar em dado de pesquisa, não se pode deixar de abordar os princípios FAIR, que visam apoiar a descoberta e a inovação do conhecimento através de uma boa gestão de dados, figurando como diretrizes para aqueles que desejam aumentar a reutilização de seus acervos de dados (Wilkinson *et al.*, 2016), tornando-os encontráveis (*Findable*), acessíveis (*Accessible*), interoperáveis (*Interoperable*) e reutilizáveis (*Reusable*) –FAIR – tanto por seres humanos, quanto por máquinas. Essa iniciativa surge da percepção da existência de barreiras no âmbito da *e-Science* relacionadas à publicação e ao uso de dados acadêmicos, as quais impedem que se obtenha o máximo de benefício dos investimentos feitos em pesquisa (Wilkinson *et al.* 2016).

Uma vez apresentadas algumas definições para o dado de pesquisa, pode-se avançar para sua tipologia a partir do trabalho de Sales e Sayão (2019), que propõem uma taxonomia baseada na literatura, a partir de oito

características de divisão que devem ser consideradas na curadoria desses dados: origem; grau de processamento; abordagem da pesquisa; natureza; nível de sensibilidade; materialidade; perenidade; e abertura. Segundo Lyon (2007), apesar da existência de categorias genéricas que podem ser atribuídas aos dados, é crucial reconhecer a complexidade e a heterogeneidade deles.

Quanto à origem, os dados podem ser classificados em dados de pesquisa e dados para a pesquisa. Os primeiros, criados no âmbito de empreendimentos científicos, podem ser divididos em observacionais, computacionais e experimentais (National Science Board, 2005). Já os dados para a pesquisa também fornecem dados úteis para pesquisas científicas (Borgman, 2010), embora não tenham sido gerados com esta finalidade. Quanto ao grau de processamento, podem ser classificados em brutos/crús/primários; intermediários/pré-processados/secundários; e finais/processados/terciários. Quanto à abordagem da pesquisa, podem ser qualitativos ou quantitativos, devendo ser tratados de acordo com a metodologia adotada no desenvolvimento da pesquisa. Quanto à natureza, nota-se uma enorme diversidade em relação aos tipos de dados em termos de formatos, mídias, suportes, expressões, arcabouço tecnológico, etc. (National Science Board, 2005). Quanto ao nível de sensibilidade, podem ser classificados como não sensíveis/ostensivos; confidenciais; e pessoais. Para Sales e Sayão (2019), a importância desta categoria se deve à definição dos níveis de abertura e compartilhamento dos dados durante os processos de gestão e curadoria, bem como à avaliação das intervenções necessárias às coleções, principalmente naquelas que contenham informações pessoais sensíveis e confidenciais. Quanto à materialidade, **podem** ser classificados em convencionais¹ ou digitais, cada um deles demandando tratamentos específicos de acordo com sua forma. Quanto à perenidade, podem ser classificados em canônicos/referenciais ou episódicos: os dados canônicos têm variação mínima, e os episódicos estão em constante mudança, podendo ser únicos no tempo e no espaço (Lyon, 2007). Por último, os dados podem ser classificados segundo o grau de abertura (aberto, fechado, parcialmente aberto,

¹ Originalmente, foi adotado o termo “físicos”, mas preferimos “convencionais”, a fim de evitar o entendimento de que o dado ou documento digital não possui uma materialidade física.

embargado) e o tipo de licença para reuso (Sales; Sayão, 2019).

Diferenças à parte, comum a todos os tipos de dados, em maior ou menor nível, é o fato de que eles, por si só, pouco falam, necessitando de uma descrição mínima capaz de torná-los compreensíveis e úteis para a comunidade científica.

3 PRINCÍPIOS PARA A ELABORAÇÃO DE DEFINIÇÕES

O presente trabalho adota a noção de conceito apresentada pela pesquisadora alemã Ingetraut Dahlberg em sua Teoria do Conceito, a qual fornece elementos para a formação dos conceitos, para a compreensão destes como sínteses de características, para a descoberta de suas características, das relações conceituais, e, por fim, para a construção de sistemas de conceitos de acordo com critérios objetivos (Dahlberg, 1978b).

Na concepção de Dahlberg (1978d), conceitos são abstrações da realidade, já que são produtos e instrumentos da capacidade do homem de pensar e falar sobre a realidade que o circunda. Assim, existe um universo de itens constituído de ideias, objetos, fatos, leis, propriedades, ações etc., em que é possível selecionar um determinado item de referência, sobre o qual podem ser feitas afirmações corretas e verificáveis que o caracterizam (tendo em vista um acordo estabelecido entre os membros de uma comunidade) e que, por sua vez, são sintetizadas na forma verbal por meio de um termo que será aplicado e comunicado no universo dos discursos. Desse modo, emergem três pontos focais para a compreensão do conceito: o item de referência ou referente; as características a ele atribuídas; e a forma verbal escolhida para sintetizar e representar o referente. A construção do conceito envolve, portanto, os passos referencial, predicacionale representacional (Dahlberg, 1978a).

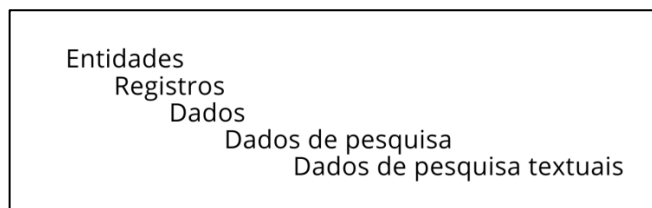
Nessa teoria, as características são elementos indispensáveis para o estabelecimento de um conceito passível de ser representado e recuperado em sistemas de informação. Elas também evidenciam as relações entre os conceitos, pois quando estes possuem características idênticas, opostas ou semelhantes, existem relações entre eles (Gomes; Campos, 2019), as quais compreendem um enfoque quantitativo e um enfoque qualitativo: o primeiro tem por base a observação das características partilhadas do ponto de vista da lógica

formal, observando a quantidade e a similaridade das características (relações de identidade, inclusão, interseção e disjunção); o segundo visa identificar as relações existentes a partir da observação de aspectos formais e materiais (relação formal/categorial; material-paradigmática (relações hierárquicas, partitivas e de oposição); e funcional-sintagmática) (Dahlberg, 1978c, p. 18). Assim, a partir de um relacionamento genérico (com base no gênero próximo e na diferença específica) tem-se uma definição genérica, assim como um relacionamento funcional gera uma definição funcional (ou operacional), em que o referente resulta de uma operação exercida sobre algo/alguém ou de um processo "sofrido por" um referente.

Sobre os conceitos devem ser formuladas afirmações verdadeiras, que representam elementos dos conceitos. Por exemplo, para o objeto geral "Instituição", pode-se afirmar que: é constituída de um grupo de pessoas; que trabalha com determinada finalidade; possui administração comum; localizada em determinado lugar; durante determinado tempo. Os enunciados formados são características do conceito, que podem ser de três tipos: essenciais (necessárias); acidentais; e individualizantes: as primeiras se aplicam a todos os referentes de determinado tipo; as segundas, a alguns referentes de determinado tipo; e as terceiras, a um único referente (Dahlberg, 1978a). Dependendo da espécie de objeto, as características podem variar: em produtos e equipamentos, as características essenciais devem ser determinadas pelas finalidades e pela aplicação, enquanto as acidentais têm relação com a eficiência e valores, em geral, práticos (Dahlberg, 1978d).

Dahlberg também explora a noção de categorias, que são conceitos abrangentes acima dos quais não se pode colocar qualquer conceito (Gomes; Campos, 2019), sendo consideradas essenciais para a sistematização do conhecimento, podendo ser identificadas no momento da determinação do conceito, ao serem inferidas predicções verdadeiras e finais a respeito de um item de referência, conforme mostra a Figura 1.

Figura 1 – Cadeia de características/conceitos



Fonte: Elaboração própria com base em Dahlberg (1978a, p. 145).

Neste exemplo, as características são organizadas de forma hierárquica – das gerais às específicas –, formando cadeias de características/conceitos, e devem ser lidas de baixo para cima. Assim, temos dados de pesquisa textuais, que são dados de pesquisa, que são dados, que por sua vez consistem em registros que, por fim, são entidades. Portanto, observa-se que do conceito de dados de pesquisa textuais chega-se à categoria entidades – localizada no topo da cadeia de características/conceitos –, a qual constitui a característica mais abrangente em que se encontra o referido conceito.

Segundo Campos (2001), Dahlberg utiliza a noção de categoria como um recurso para o entendimento da natureza do conceito e para a formação de estruturas conceituais. Desse modo, a Teoria do Conceito fornece um método de raciocínio analítico-sintético para organizar os conceitos de um domínio em grandes categorias, a partir da análise da sua definição (pensamento analítico) e da identificação de características comuns que revelam categorias para agrupar esses conceitos (pensamento sintético).

As características podem ser explicitadas por meio de uma definição, que consiste na “delimitação ou fixação do conteúdo de um conceito (conteúdo do conceito = intensão, ou conjunto de características ou atributos)” (Dahlberg, 1978c, p. 106). A partir das chamadas definições conceituais, é possível delimitar a intensão de determinado conceito, distinguindo-o de outros com características idênticas; ou seja, apresenta-se o conhecimento contido sobre o conceito em um dado contexto de aplicação; nesse caso, o *definiens* (equivalência textual do *definiendum*) contém as características necessárias de um referente nomeado pelo *definiendum* (expressão verbal) (Dahlberg, 1978c, 1983).

Para orientar a elaboração de definições, Campos (2017) recorre à Terminologia, relacionando a abordagem onomasiológica à teoria de Dahlberg,

na medida em que ela se orienta por uma perspectiva extralinguística, em que o entendimento do referente, num domínio, leva à formulação do termo e à determinação do conceito. A abordagem contrária a esta é a semasiológica, baseada em uma perspectiva centrada no termo. Assim, na abordagem onomasiológica, os termos adotados refletem o acordo estabelecido por um grupo de especialistas sobre os objetos do mundo real, que são reunidos em classes que partilham características comuns (Gomes; Campos, 2019).

Apresentados os princípios teóricos para a elaboração de definições, torna-se possível definir o conceito de dados textuais gerados em pesquisas bibliográficas, a partir da compreensão de suas características. Entretanto, as definições encontradas na literatura nem sempre atendem a padrões e critérios, levando Dahlberg (1981) a recomendar regras para sua elaboração e avaliação. Por exemplo, quanto à forma, devem ser usadas somente características necessárias de um referente; e não devem ser usados termos equivalentes ao *definiendum*. Quanto ao conteúdo, o *definiendum* e o *definiens* devem ter um único referente (um não pode ser um processo e, o outro, um objeto).

No contexto das ontologias, Seppälä, Ruttenberg e Smith (2017) destacam a importância da definição para que ontologistas e curadores de dados compreendam o significado pretendido dos termos da ontologia e os utilizem de forma consistente com o contexto. Eles fornecem diretrizes para auxiliar na redação, edição e validação de definições, as quais são suficientemente amplas para serem estendidas e aplicadas para além das ontologias. Por exemplo, recomenda-se evitar plurais, expressões generalizantes, informação enciclopédica, termos negativos, exemplos etc., assim como categorizadores, pois o próprio gênero da definição tem a função de categorizar a coisa denotada sob um tipo mais geral. A circularidade é outro aspecto negativo, pois a definição não deve ser definida em termos de si mesma, usando rótulos anexados à classe, sinônimos ou alguma forma gramaticalmente derivada. Indica-se, ainda, que deve-se definir apenas um item por vez, com uma única definição textual. Para serem informativas, as definições devem destacar a intensão do *definiens*, especificando as características e dizendo o que é a coisa definida (definição por intensão), e não listar instâncias (definição por extensão). Se possível, deve-se

adotar a forma gênero-diferença, em que o gênero está alinhado com uma entidade de um nível de generalidade superior e a diferença especifica os casos abrangidos pelo termo superior (Campos; Rondinelli; Campos, 2022).

Seppälä, Ruttenberg e Smith (2017) indicam que a definição não deve incluir o termo que está sendo definido. Recomendam ainda que a cópula “é” não seja inserida (em seu lugar, sugerem o uso dos dois-pontos “:”) e que sinais de pontuação que não sejam vírgulas – por exemplo, parênteses, dois-pontos, barras (como em 'e/ou'), ou ponto-e-vírgula – sejam evitados.

Com base no que foi apresentado, o Quadro 1 reúne recomendações gerais para elaboração de definições, mantendo-se a divisão entre forma e conteúdo de Dahlberg, com os aspectos e explicações complementados pelas contribuições de Seppälä, Ruttenberg e Smith (2017).

Quadro 1 – Recomendações gerais para elaboração de definições

	ASPECTO	EXPLICAÇÃO
FORMA	Simplicidade	Usar somente as características necessárias e suficientes de um referente. Evitar informações não definidoras, enciclopédicas, termos negativos e definições ostensivas.
	Clareza	Usar palavras e termos claros que já tenham sido definidos no contexto/ambiente em que a definição se insere.
	Nível	Usar expressões de acordo com a linguagem e área de assunto dos destinatários, como termos técnicos para cientistas.
	Justaposição	Não usar termos equivalentes, sinônimos ou quase-sinônimos do <i>definiendum</i> : isso não produz uma definição orientada ao referente.
	Conformidade com as convenções	Obedecer às convenções linguísticas e lexicográficas da língua, assim como convenções tipográficas e diretrizes editoriais.
CONTEÚDO	Correspondência com o referente	Definir a coisa denotada pelo <i>definiendum</i> (referente), não o termo em si ou a classe que o representa. <i>O definiendum e o definiens</i> devem ter um único referente.
	Completude	Incluir todas as características necessárias do referente de forma estruturada.
	Adequação da extensão e do escopo	Usar as características necessárias do referente, sem ampliar/restringir inadequadamente a extensão e o escopo da definição. Definir apenas uma coisa, com uma única definição.
	Inclinação	Não incluir pontos de vista, declarações subjetivas e avaliativas.
	Mistura de conceitos	Não fornecer a um conceito uma interpretação especial por falta de um termo adequado.
	Circularidade	Não definir usando um dos rótulos anexados à classe ou um sinônimo dela, ou alguma forma gramaticalmente derivada.
	Forma gênero-diferença	Usar a forma gênero-diferença, onde o gênero ancora a entidade definida a uma entidade conhecida em um nível mais alto de

	generalidade, e a(s) diferença(s) escolhe(m) os casos dentro desse nível mais alto que se enquadram no termo definido. Deve-se incluir um e apenas um gênero, sendo ele o mais próximo (o termo pai).
Grau apropriado de generalidade	Usar o grau de generalidade apropriado, mais geral, não sendo recomendado o uso de expressões generalizantes (p. ex. “geralmente”), exemplos, listas, termos indexicais e dêiticos.

Fonte: Elaboração própria com base em Dahlberg (1981) e Seppälä, Ruttenberg e Smith (2017).

4 METODOLOGIA

Este trabalho adota uma abordagem qualitativa, do ponto de vista do problema, e, quanto aos objetivos, caracteriza-se como uma pesquisa exploratória. Quanto aos procedimentos técnicos, caracteriza-se como uma pesquisa bibliográfica, com emprego da técnica de cultivo de pérolas (Hadfield, 2020); para a exploração e interpretação do material, utiliza a Análise de Conteúdo (Bardin, 2011). Para o levantamento bibliográfico, as seguintes fontes de informação foram utilizadas: Base de Dados em Ciência da Informação (BRAPCI), Portal Capes, Biblioteca Digital Brasileira de Teses e Dissertações (BDTD) e Google Acadêmico. A busca priorizou a seleção de textos completos; publicados em periódicos revisados por pares; de acesso aberto; nos idiomas inglês, espanhol ou português; em geral, o recorte temporal não foi aplicado.

Para a sistematização dos princípios teóricos para a elaboração de definições, foram recuperados 52 textos, de autores da Ciência da Informação e da ontologia aplicada, dos quais 14 foram selecionados, a partir do emprego, de forma combinada e incluindo suas variações na língua inglesa, das seguintes expressões de busca: “teoria do conceito”, “definição conceitual” e “organização do conhecimento”. Para a descrição dos conceitos de dado de pesquisa e dados textuais gerados em pesquisas bibliográficas, foram recuperados 103 textos, dos quais 31 foram selecionados, a partir do emprego, de forma combinada e incluindo suas variações na língua inglesa, das seguintes expressões de busca: “dado de pesquisa”, “gestão de dados de pesquisa” e “pesquisa bibliográfica”.

As etapas da Análise de Conteúdo são descritas na sequência. Inicialmente, na etapa de pré-análise, aplicou-se a técnica da leitura flutuante para nortear a seleção dos textos resultantes do levantamento bibliográfico, considerando-se os títulos, resumos e palavras-chave desses. Após a seleção

dos textos, procedeu-se à determinação dos objetivos do estudo.

Na etapa de exploração do material, foi feita a leitura dos textos selecionados, dos quais foram extraídos trechos (unidades de contexto) relacionados aos temas (unidades de registro) considerados relevantes aos propósitos estabelecidos, o que orientou a identificação de três categorias de análise consideradas úteis aos propósitos da pesquisa, a saber: fundamentação teórica para a elaboração de definições; orientação objetiva para a elaboração de definições; e definição de dado de pesquisa. A partir da identificação da importância de fundamentação teórica e objetiva para a elaboração de definições, adotou-se a Teoria do Conceito de Dahlberg para este fim.

Já na etapa de tratamento e interpretação dos resultados, a partir das categorias de análise identificadas, foi possível interpretar a literatura explorada no marco teórico e derivar os resultados obtidos, favorecendo uma melhor compreensão sobre os aspectos que são importantes para caracterizar os dados textuais gerados em pesquisas bibliográficas.

5 DADOS TEXTUAIS GERADOS EM PESQUISAS BIBLIOGRÁFICAS

Os dados textuais gerados em pesquisas bibliográficas são um tipo de dado de pesquisa e, portanto, herdam as características gerais aplicáveis a qualquer dado de pesquisa. Eles são criados a partir da leitura e análise de textos científicos no âmbito de uma pesquisa bibliográfica, como fruto de um movimento de pesquisa do pesquisador para assimilar e interpretar as ideias contidas no texto tendo em vista os problemas e objetivos de sua pesquisa. Dessa forma, resultam de uma análise interpretativa ou crítica dos textos; e apresentam um componente contextual, na medida em que a análise considera o contexto teórico mais amplo em que os textos estão inseridos, bem como sua ligação teórica com outros textos e com a pesquisa em desenvolvimento. Assim, eles abrangem três focos: o conteúdo temático do texto; o contexto teórico em que este se insere; e sua aplicação na pesquisa, conforme mostra a Figura 2.

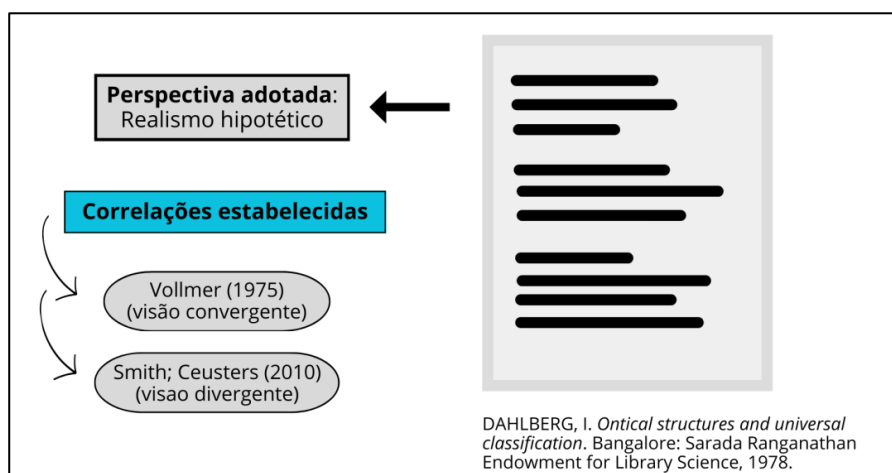
Figura 2 – Dados textuais gerados em pesquisas bibliográficas



Fonte: Elaboração própria.

Num primeiro momento, o pesquisador busca a compreensão do texto, o que envolve a temática abordada, objetivos, conceitos centrais, ou seja, seu **conteúdo temático**. Concomitante a esse exercício de compreensão, é importante situar o texto no **contexto teórico** em que ele foi produzido e a partir do qual podem ser estabelecidas relações diversas que contribuem para uma visão ampla e crítica sobre o tema abordado; por exemplo, a partir da identificação da perspectiva adotada no texto, podem ser estabelecidas correlações que forneçam alguma contribuição relacionada, como visões teóricas semelhantes e/ou contrárias (Figura 3). Por fim, segue-se o esforço de relacionar os conteúdos apreendidos com sua **aplicação na pesquisa**.

Figura 3 – Correlações estabelecidas no contexto teórico



Fonte: Elaboração própria.

Com base no exposto, ressalta-se que os dados textuais gerados em pesquisas bibliográficas não devem ser confundidos com os dados bibliográficos tipicamente abordados no domínio da Biblioteconomia para identificar e recuperar um item informacional, como título, assunto, notas, dados de publicação etc. Embora se refiram a uma fonte bibliográfica, os aspectos por eles enfatizados são descritivos, diferentemente do foco dos dados textuais gerados em pesquisas bibliográficas, que enfatizam o conteúdo do texto, de forma estruturada, viabilizando uma análise teórica deste.

A fim de fornecer exemplos de dados textuais gerados em pesquisas bibliográficas, realizou-se um estudo preliminar de análise de um texto, tendo em vista a identificação de metadados voltados à representação desses dados. A fundamentação teórica que norteou a obtenção dos metadados, visando uma representação ecológica capaz de possibilitar não apenas o uso dos dados por terceiros, mas também a interação e a construção coletiva do conhecimento, se baseia na Organização do Conhecimento, com ênfase na representação de dados de pesquisa (Caplan, 2003; Sales; Sayão, 2012); na influência de perspectivas ontológicas e epistemológicas na representação do conhecimento (Hjørland, 2009); e na decolonialidade do saber (Garcez; Sales, 2021) e na teoria dos objetos de fronteira (Star; Griesemer, 1989). Além disso, diferentes esquemas de metadados voltados para a representação de dados de pesquisa foram analisados, bem como métodos e técnicas como a leitura informativa, a leitura analítica, a análise de conteúdo e a revisão sistemática da literatura, com vistas à identificação de metadados que possam refletir os tipos de dados resultantes da aplicação dos métodos. Entretanto, está fora do escopo deste trabalho detalhar a derivação dos metadados a partir desses aportes.

Cabe ressaltar que, embora a referida representação deva ser pensada considerando-se o contexto digital e a possibilidade de processamento automático, nesse primeiro momento não nos debruçamos sobre esse contexto, o qual envolve questões como o uso de padrões e vocabulários e o depósito dos dados em repositórios, dentre outras.

Nesse esforço, uma questão que se coloca é se a descrição será igualmente útil e relevante tanto para o produtor dos dados (pesquisador), quanto

para possíveis interessados em reutilizá-los. Em outras palavras, questiona-se se aos aspectos atribuídos pelo pesquisador em sua análise devem ser acrescentados outros que possam ser necessários ou importantes para otimizar o reuso dos dados, o qual consiste no objetivo final dos princípios FAIR. A reutilização depende de metadados e documentação ricos que atendam aos padrões de relevância da comunidade e forneçam informações sobre a procedência, incluindo os processos de redução ou transformação de dados para torná-los mais utilizáveis, compreensíveis ou "prontos para a ciência" (Hodson *et al.*, 2018). Sobre essa questão, considera-se importante prezar pelo interesse do pesquisador quando da definição dos metadados a serem empregados na descrição proposta, evitando-se adicionar muitos elementos que, por ventura, possam ser vistos como barreiras ao preenchimento e, por conseguinte, ao compartilhamento dos dados.

Todavia, entende-se como necessária a atribuição de metadados descritivos, relacionados aos propósitos de descoberta, identificação e seleção de um recurso, contemplando elementos como título, autor, resumo, palavras-chave e identificador persistente (Caplan, 2003), que fornecem uma visão básica do recurso descrito. Em relação a isso, pode ser útil recorrer a tecnologias e iniciativas voltadas à representação e interligação de dados no ambiente web, cuja discussão está fora do escopo do presente artigo. Sendo assim, embora nosso foco recaia sobre os metadados atribuídos pelo pesquisador, destinados ao conteúdo do texto e sua relação com a pesquisa em desenvolvimento, optou-se por indicar, além desses, alguns aspectos básicos relativos à identificação do texto. Ressalta-se que a questão do reuso é essencial para a descrição dos dados textuais gerados em pesquisas bibliográficas, devendo ser aprofundada com vistas a uma descrição completa deste tipo de dado, o que não se enquadra no propósito deste artigo, que tem como foco a definição do referido conceito.

Dessa forma, no Quadro 2 buscamos explicitar, ainda que de maneira preliminar e não exaustiva, aspectos pelos quais esses dados podem ser representados, de forma estruturada, para que possam ser compartilhados e reutilizados. Ressalta-se que cada artigo considerado central para uma pesquisa gera um conjunto de metadados com seus respectivos dados associados, sendo

este conjunto concebido como um bloco unitário de dados. Para o exemplo, analisou-se o texto “Teoria do Conceito”, de Dahlberg, traduzido por Astério Tavares Campos, cuja escolha justifica-se por sua importância para a elaboração do presente artigo (mais especificamente, para a elaboração do conceito de dados textuais gerados em pesquisas bibliográficas). Destaca-se que a análise do referido texto não se limita à sua leitura, sendo por vezes necessário recorrer a leituras complementares para sua contextualização teórica.

Quadro 2 – Exemplo de descrição básica para dados textuais gerados em pesquisas bibliográficas

METADADO	VALOR
DOI	10.18225/ci.inf..v7i2.115
Título	Teoria do Conceito
Autor	Dahlberg, Ingetraut, 1927-2017 VIAF: http://viaf.org/viaf/12834306 WIKIDATA: https://www.wikidata.org/wiki/Q1097111
Minibiografia do autor	Filósofa e cientista da informação alemã. Exerceu grande influência nos estudos de Organização do Conhecimento e Classificação no âmbito da Ciência da Informação. Atuou em importantes instituições na área de Documentação e em atividades relacionadas ao desenvolvimento de linguagens documentárias. Foi uma das fundadoras da Knowledge Organization e da ISKO .
Idioma	Português
Ano de publicação	1978
País de publicação	Brasil
Tipo de fonte	Artigo científico
Tipo de artigo	Artigo de análise
Formato do artigo	Artigo completo
Resumo	Com a ajuda das linguagens naturais é possível formular enunciados a respeito de conceitos individuais e conceitos gerais. Todo enunciado sobre objetos contém um elemento do respectivo conceito, que se identifica como sua característica. A categorização formal dos conceitos tem importância na formação de sistemas e na combinação dos mesmos. As definições corretas dos conceitos são muito importantes, pois o contínuo desenvolvimento do conhecimento e da linguagem leva à utilização de novos termos e conceitos cujo domínio nem sempre é fácil manter.
Palavras-chave	Conceito; Método analítico-sintético; Categorização; Definição; Sistema de classificação.
Área de conhecimento	Ciência da Informação; Organização do Conhecimento.
Problema	A imprecisão na definição dos conceitos no âmbito das linguagens especializadas pode acarretar consequências negativas para a representação, que deve acompanhar o contínuo desenvolvimento do conhecimento e da linguagem.
Objetivo geral	Descrever o conceito e seus elementos, bem como as formas de analisá-lo, relacioná-lo com outros conceitos e defini-lo.
Resultados	Descrição detalhada do método analítico-conceitual, que possibilita que os conceitos sejam conhecidos e definidos de maneira precisa, o que é fundamental para a construção de sistemas de classificação.

Teorias abordadas	Teoria do Conceito
Domínio de aplicação	Construção de sistemas de classificação; Definição de conceitos.
Conceitos centrais definidos ²	Linguagem natural; Linguagem artificial; Definição de conceitos; Conceito; Conceito individual; Conceito geral; Característica essencial; Característica accidental; Hierarquia de características; Espécies de características; Intensão do conceito; Extensão do conceito; Relação entre conceitos; Categoria de conceitos.
Objetivo da pesquisa atrelado	Identificar princípios teóricos e metodológicos para a definição de conceitos.
Posicionamentos convergentes	VOLLMER, Gerhard. Evolutionäre Erkenntnistheorie . Stuttgart: S. Hirzel Verlag, 1975. 209 p.
Posicionamentos divergentes	SMITH, Barry; CEUSTERS, Werner. Ontological realism: a methodology for coordinated evolution of scientific ontologies. Applied Ontology , v. 5, n. 3-4, p. 139-188, nov. 2010.

Fonte: Elaboração própria.

Apesar de considerarmos que os dados textuais gerados em pesquisas bibliográficas já são produzidos pelos pesquisadores no desenvolvimento de suas pesquisas, entendemos que a ênfase no seu uso como dado de pesquisa que pode ser representado de forma estruturada, privilegiando diferentes aspectos, consiste em uma nova contribuição, estando sujeita a aperfeiçoamentos, como toda pesquisa. Assim, aspectos como a perspectiva epistemológica, a natureza da pesquisa ligada ao texto, bem como os relativos à preservação digital do dado de pesquisa não foram contemplados aqui, pois não fazem parte do foco de nossa proposta.

6 RESULTADOS: A CONSTRUÇÃO DA DEFINIÇÃO DO CONCEITO

A definição do conceito de dados textuais gerados em pesquisas bibliográficas foi feita a partir de uma abordagem onomasiológica, apoiada no método analítico-conceitual proposto na Teoria do Conceito, bem como nas recomendações estabelecidas por Dahlberg e Seppälä, Ruttenberg e Smith (2017) para a elaboração de definições. De forma mais detalhada, identificou-se, com base na Teoria do Conceito, o item de referência relativo ao conceito em questão, suas características (em forma de enunciados) e a forma verbal mais

² O conteúdo de alguns metadados, como é o caso deste, pode ser alvo de padronização por meio de um vocabulário controlado ou de uma ontologia de domínio, o que aumentaria as chances de recuperação do texto alvo da descrição.

adequada para sua representação, para, por fim, elaborar sua definição.

Em nosso caso, o item de referência compreende qualquer registro que possua as características e funções de um dado textual produzido no âmbito de uma pesquisa bibliográfica, podendo ser um resumo, fichamento, categoria de análise, conceito etc. Passando para o passo predicacional, buscam-se as informações e definições já apresentadas sobre o conceito de dados textuais gerados em pesquisas bibliográficas. Em um primeiro momento, a teoria de Dahlberg foi aplicada para entender a relação deste conceito com o de dado de pesquisa, que é o gênero próximo. Para tanto, a definição de dado de pesquisa foi analisada de modo a explicitar e qualificar seus elementos constituintes, conforme ilustrado abaixo:

Definição:

Dado de pesquisa: qualquer registro coletado, observado, gerado ou utilizado no âmbito da pesquisa científica, que pode ser interpretado, tratado e aceito como evidência pela comunidade científica e necessário para analisar, validar e produzir resultados de pesquisa.

Análise:

Gênero: registro (objeto)

Diferença específica 1: coletado, observado, gerado ou utilizado no âmbito da pesquisa científica (contexto)

Diferença específica 2: que pode ser interpretado, tratado e aceito como evidência pela comunidade científica (propósito 1)

Diferença específica 3: e necessário para analisar, validar e produzir resultados de pesquisa (propósito 2)

Com base nas informações coletadas, confirma-se que os dados textuais gerados em pesquisas bibliográficas são um dado de pesquisa, configurando, assim, uma relação hierárquica, de gênero/espécie, em que os dois conceitos possuem as mesmas características, sendo que um deles possui uma característica a mais que o outro. Nesse caso, dado de pesquisa é o conceito mais genérico, enquanto dados textuais gerados em pesquisas bibliográficas é o conceito mais específico, que possui todas as características daquele, mas se distingue como uma espécie. Dessa forma, tendo em vista a elaboração da definição, interessa-nos agora identificar as características essenciais e acidentais desse tipo de dado, aquelas relativas somente a ele – e não a todo dado de pesquisa. Assim, a determinação dessas características é fundamental

para que a definição construída seja suficientemente completa e informativa, a fim de evitar dificuldades no entendimento e consequente mau uso do conceito.

Posto isso, foram atribuídas predicções sobre o item de referência, a fim de se identificar suas características essenciais, que seguem listadas abaixo:

- é um dado de pesquisa textual;
- é produzido no desenvolvimento de uma pesquisa bibliográfica;
- é originado de uma fonte bibliográfica (por exemplo, um artigo ou livro);
- é fruto de uma análise do pesquisador sobre a fonte bibliográfica em questão, no contexto de sua pesquisa, através de uma abordagem majoritariamente qualitativa;
- expressa o conteúdo temático e o contexto teórico pertinente ao texto, este último contemplando uma perspectiva ecológica;
- descreve como se dá a aplicação do texto na pesquisa em desenvolvimento;
- visa apoiar a análise do conteúdo de um texto considerado central para a pesquisa.

Ainda no que tange à análise das características, para melhor compreensão do conceito, faz-se importante identificar a categoria a qual ele pertence. Assim, baseando-se em Dahlberg, constata-se que o conceito de dados textuais gerados em pesquisas bibliográficas trata de uma entidade, termo que pode ser usado para se referir a objetos concretos ou abstratos (Gomes; Campos, 2019), nesse caso, um objeto concreto.

No passo representacional, é feita a escolha da forma verbal do conceito em questão: “dados textuais gerados em pesquisas bibliográficas”, em detrimento de “dado de pesquisa bibliográfico”, opção inicialmente cogitada para refletir o contexto de produção dos dados – a pesquisa bibliográfica. Tal escolha foi feita tendo em vista a possibilidade de má interpretação da opção anterior como dos dados bibliográficos tipicamente abordados no contexto da catalogação no domínio da Biblioteconomia. Por sua vez, considera-se que a opção escolhida detalha e reflete os principais aspectos dos dados em questão.

Por fim, uma vez identificados os elementos da tríade conceitual de Dahlberg e a partir das recomendações do Quadro 1, procedeu-se à elaboração da definição do conceito de dados textuais gerados em pesquisas bibliográficas, do tipo conceitual. Nela, destaca-se a explicitação da característica essencial de ser um dado de pesquisa, ou seja, a explicitação do gênero próximo. Também buscou-se apresentar o conhecimento contido sobre este conceito, distinguindo-o de outros com características idênticas; ou seja, do conceito de dado de

pesquisa. Além disso, seu contexto de uso e finalidade também foram inseridos. Abaixo, tem-se a definição proposta: à esquerda, o *definiendum*; à direita, o *definiens*; e entre eles, uma cópula que expressa a equivalência. Logo após, conforme feito por Dahlberg (1983), tem-se a análise da definição.

Definição:

Dados textuais gerados em pesquisas bibliográficas: dado de pesquisa textual produzido no desenvolvimento de uma pesquisa bibliográfica; originado diretamente de uma fonte bibliográfica; como fruto de uma análise do pesquisador sobre a fonte bibliográfica em questão, no contexto de sua pesquisa; visando expressar o conteúdo temático e o contexto teórico pertinente ao texto, em uma perspectiva ecológica, descrever como se dá a aplicação do texto na pesquisa em desenvolvimento, e apoiar a análise do conteúdo de um texto considerado central para a pesquisa.

Análise:

Gênero: dado de pesquisa textual (objeto)

Diferença específica 1: produzido no desenvolvimento de uma pesquisa bibliográfica (contexto)

Diferença específica 2: diretamente de uma fonte bibliográfica (origem)

Diferença específica 3: como fruto de uma análise do pesquisador sobre a fonte bibliográfica em questão, no contexto de sua pesquisa (meio de geração)

Diferença específica 4: visando expressar o conteúdo temático e o contexto teórico pertinente ao texto, em uma perspectiva ecológica (propósito)

Diferença específica 5: descrever como se dá a aplicação do texto na pesquisa em desenvolvimento (propósito)

Diferença específica 6: apoiar a análise do conteúdo de um texto considerado central para a pesquisa (propósito).

Uma vez finalizado o processo de elaboração da definição, considera-se que ela está em consonância com as recomendações fornecidas no Quadro 2.

Quanto à forma, atende aos aspectos de simplicidade (constitui-se de características essenciais do referente); clareza (os termos foram previamente explicados e definidos neste artigo); nível (as expressões coincidem com a linguagem e área de assunto dos destinatários, aqui considerados estudantes e profissionais da informação); justaposição (não adota termos equivalentes ou sobrepostos); e conformidade com as convenções (obedece às convenções linguísticas e lexicográficas da língua portuguesa, assim como ao padrão adotado por Seppälä, Ruttenberg e Smith (2017) para a redação da definição).

Quanto ao conteúdo, apresenta correspondência com o referente (tanto o *definiendum* quanto o *definiens* referem-se ao dados textuais gerados em pesquisas bibliográficas); completude, (reúne todas as características essenciais

do conceito, além de características acidentais; adequação da extensão e do escopo (não incorpora características relativas ao dado de pesquisa, em geral, para não ampliar de forma inadequada a extensão e o escopo do conceito definido); forma gênero-diferença (apresenta um único – e mais próximo – gênero (dado de pesquisa) e características que refletem as diferenças da espécie para o gênero); e o grau apropriado de generalidade (não inclui expressões generalizantes, exemplos, listas, termos indexicais e dêiticos). Por sua vez, a definição não apresenta os aspectos a serem evitados: inclinação, (não apresenta subjetividade); mistura de conceitos (o próprio artigo caracteriza o referido conceito, evidenciando sua especificidade em relação ao conceito de dado de pesquisa); e circularidade (não há uso de um *definiens* como o *definiendum* de outro, ou vice-versa).

Portanto, a consulta e a verificação dos aspectos e explicações reunidos no Quadro 1 foram fundamentais para a elaboração da definição proposta, antes e após a escrita, o que evidencia sua eficácia para a construção e para a avaliação e revisão de definições.

7 CONSIDERAÇÕES FINAIS

O aporte teórico da Teoria do Conceito mostra-se fortemente aplicável ao objetivo de elaborar uma definição para o conceito de dados textuais gerados em pesquisas bibliográficas, de forma consciente e justificada. A teoria, no contexto da pesquisa, elucida pontos importantes a serem considerados no entendimento do conceito proposto, partindo da identificação de um item de referência com existência no mundo real, passando por sua caracterização, descrição e nomeação. Além disso, as recomendações de Dahlberg (1981) e Seppälä, Ruttenberg e Smith (2017) são úteis principalmente devido ao seu caráter objetivo, juntamente com exemplos que facilitam o entendimento.

Assim, considerando-se o interesse em propor um novo conceito, faz-se essencial recorrer a princípios teóricos sólidos que confirmam consistência e qualidade à sua definição. De outro modo, poder-se-ia elaborar uma definição de maneira intuitiva, sem buscar apoio em qualquer teoria, o que é uma prática comum, que, no entanto, gera dificuldades no entendimento e na utilidade da

definição. Uma vez observados os aspectos estabelecidos pela teoria, percebe-se como as falhas e equívocos são frequentes quando se procede sem um embasamento teórico: a forma verbal nem sempre evoca o sentido do conceito, as definições muitas vezes contemplam informações irrelevantes ou deixam de incluir outras que fornecem a identidade do conceito, entre outros.

Portanto, ao se aplicar o aporte teórico da Teoria do Conceito e as recomendações para a elaboração de definições no âmbito das ontologias na definição do conceito de dados textuais gerados em pesquisas bibliográficas, buscou-se tornar seu entendimento claro e preciso no âmbito da comunidade científica, de modo que sua adoção seja difundida e incentive a aplicação de ações de curadoria nesse tipo de dado, que deve estar visível e ser passível de reutilização por outros pesquisadores.

Por fim, considerando-se a necessidade de dar continuidade aos estudos, incentiva-se o desenvolvimento de trabalhos futuros para aprofundar os metadados que descrevem o texto alvo, bem como que explorem a criação de vocabulários controlados ou ontologias para a descrição do conteúdo de alguns desses metadados.

REFERÊNCIAS

BARDIN, Laurence. **Análise de Conteúdo**. ed. rev. e ampl. Lisboa: Edições 70, 2011.

BORGMAN, Christine L. **Scholarship in the digital age: information, infrastructure and the internet**. London: The MIT Press, 2007.

BORGMAN, Christine L., **Research Data: Who Will Share What, with Whom, When, and Why?** RatSWD Working Paper No. 161, 2010. Disponível em: <https://ssrn.com/abstract=1714427>. Acesso em: 25 dez. 2026.

CAMPOS, Linair Maria; RONDINELLI, Rosely Curi; CAMPOS, Maria Luzia de Almeida. O suporte do documento arquivístico digital: uma proposta de definição conceitual apoiada nos princípios da teoria do conceito e da ontologia formal. **Acervo**, Rio de Janeiro, v. 35, n. 2, p. 1-22, mai./ago. 2022. Disponível em: <https://revista.an.gov.br/index.php/revistaacervo/article/view/1803>. Acesso em: 8 dez. 2024.

CAMPOS, Maria Luiza de Almeida. A elaboração de modelos de domínio em ontologias: a abordagem onomasiológica e a função da definição. **Ciência da**

Informação, Brasília, v. 46, n. 1, p. 23-33, jan./abr. 2017. Disponível em: <https://revista.ibict.br/ciinf/article/view/4016>. Acesso em: 8 dez. 2024.

CAMPOS, Marua Lucia de Almeida. **Linguagem documentária**: teorias que fundamentam sua elaboração. Niterói: EdUFF, 2001.

CAPLAN, Priscilla. **Metadata fundamentals for all librarians**. Chicago: American Library Association, 2003.

CONSULTATIVE COMMITTEE FOR SPACE DATA SYSTEMS (CCSDS). Reference Architecture for Space Data Systems. Recomendação para Sistemas de Dados Espaciais, 2012. Disponível em: <https://public.ccsds.org/Pubs/351x0m1.pdf>. Acesso em: 26 dez. 2024.

DAHLBERG, Ingetraut. A Referent-oriented, analytical concept theory of interconcept. **International Classification**, [S. l.], v. 5, n. 3, p. 122-151, 1978a. Disponível em: https://www.nomos-elibrary.de/10.5771/0943-7444-1978-3-142.pdf?download_full_pdf=1. Acesso em: 9 dez. 2024.

DAHLBERG, Ingetraut. Conceptual definitions for Interconcept. **Knowledge Organization**, [S. l.], v. 8, n. 1, p. 16-22, 1981.

DAHLBERG, Ingetraut. Fundamentos teórico-conceituais da classificação. **Revista de Biblioteconomia de Brasília**, Brasília, v. 6, n. 1, p. 12, jan./jun. 1978b. Disponível em: <https://periodicos.unb.br/index.php/rbbsb/article/view/29057>. Acesso em: 9 dez. 2024.

DAHLBERG, Ingetraut. **Ontical Structures and Universal Classification**. Bangalore: Sarada Ranganathan Endowment, 1978c. 64 p.

DAHLBERG, Ingetraut. Teoria do Conceito. **Ciência da Informação**, Rio de Janeiro, v. 7, n. 2, p. 101-107, 1978d. Disponível em: <https://revista.ibict.br/ciinf/article/view/115>. Acesso em: 9 dez. 2024.

DAHLBERG, Ingetraut. Terminological definitions: characteristics and demands. *In: PROBLÈMES de la définition et de la synonymie en terminologie*. Québec: GIRSTERM, 1983.

GARCEZ, Dirnele Carneiro; SALES, Rodrigo de. Olhares decoloniais em organização do conhecimento: uma análise das publicações do periódico knowledge organization (2000-2020). *In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO*, 21., 2021, Rio de Janeiro. **Anais eletrônicos** [...]. Rio de Janeiro: UFRJ, 2021. Disponível em: <https://enancib.ancib.org/index.php/enancib/xxienancib/paper/view/277>. Acesso em: 9 dez. 2024.

GOMES, Hagar Espanha; CAMPOS, Maria Luiza de Almeida. **A organização do conhecimento na Web**: contribuições de Shiyali Ramamrita Ranganathan e Ingetraut Dahlberg. Niterói: IACS/UFF, 2019. Disponível em:

https://eocci.uff.br/wp-content/uploads/sites/319/2020/09/eocci_ca-1.pdf.
Acesso em: 25 jun. 2022.

HADFIELD, Ruth. Pearl growing as a strategy in systematic literature searches. **Mediwrite**, [S. l.], 2020. Disponível em: <https://www.mediwrite.com.au/medical-writing/pearl-growing/>. Acesso em: 20 dez. 2024.

HEIDORN, P. Bryan. Shedding light on the dark data in the long tail of science. **Library Trends**, [S. l.], v. 57, n. 2, p. 280-299, dec. 2008.

HJØRLAND, Birger. Concept Theory. **Journal of the American Society for Information Science and Technology**, [S. l.], v. 60, n. 8, p. 1519-1536, apr. 2009.

HODSON, Jones *et al.* **Turning FAIR data into reality**: interim report of the European Commission Expert Group on FAIR data. 2018. Disponível em: <https://www.rd-alliance.org/wp-content/uploads/2024/05/FAIR-Data-EG-Interim-Report.pdf>. Acesso em: 8 dez. 2024.

JIN, Xiaolong; WAH, Benjamin W.; CHENG, Xueqi, WANG, Yuanzhuo. Significance and challenges of big data research. **Big Data Research**, Paris, v. 2, n. 2, p. 59-64, jun. 2015.

LYON, Liz. Dealing with data: roles, rights, responsibilities and relationships. **Consultancy Report**, UKOLN, University of Bath, p. 1-65, jun. 2007. Disponível em: <https://researchportal.bath.ac.uk/en/publications/dealing-with-data-roles-rights-responsibilities-and-relationships>. Acesso em: 9 dez. 2024.

NATIONAL SCIENCE BOARD. **Long-lived digital data collections**: enabling research and education in the 21st Century. Arlington: NSF, 2005. Disponível em: <https://www.nsf.gov/geo/geo-data-policies/nsb-0540-1.pdf>. Acesso em: 9 dez. 2024.

OECD. **OECD Principles and guidelines for access to research data from public funding**. USA: OECD, 2007. Disponível em: <https://www.oecd-ilibrary.org/docserver/9789264034020-en-fr.pdf?expires=1733713880&id=id&accname=guest&checksum=1A7919365FDEEE63A42E532A8D631065>. Acesso em: 8 dez. 2024.

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

SALES, Luana Farias; SAYÃO, Luís Fernando. A ciência invisível: revelando os dados da cauda longa da pesquisa. *In*: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO (ENANCIB), 19., 2018, Londrina. **Anais** [...]. Londrina: ANCIB, 2018. Disponível em: http://enancib.marilia.unesp.br/index.php/XIX_ENANCIB/xixenancib/paper/view/Paper/1538. Acesso em: 10 dez. 2024.

SALES, Luana Farias; SAYÃO, Luís Fernando. O impacto da curadoria digital dos dados de pesquisa na Comunicação Científica. **Encontros Bibli: Revista Eletrônica de Biblioteconomia e Ciência da Informação**, [S. l.], v. 17, n. esp. 2, p. 118-135, 2012. Disponível em: <https://periodicos.ufsc.br/index.php/eb/article/view/1518-2924.2012v17nesp2p118>. Acesso em: 10 dez. 2024.

SALES, Luana Farias; SAYÃO, Luís Fernando. Uma proposta de taxonomia para dados de pesquisa. **Conhecimento em Ação**, Rio de Janeiro, v. 4, n. 1, p. 31-48, jan./jun. 2019. Disponível em: <https://revistas.ufrj.br/index.php/rca/article/view/26337>. Acesso em: 9 dez. 2024.

SAYÃO, Luís Fernando; SALES, Luana Farias. Afinal, o que é dado de pesquisa? **Biblos: Revista do Instituto de Ciências Humanas e da Informação**, Rio Grande, v. 34, n. 02, p. 32-51, jul./dez. 2020. Disponível em: <https://periodicos.furg.br/biblos/article/view/11875>. Acesso em: 10 dez. 2024.

SAYÃO, Luís Fernando; SALES, Luana Farias. Dados abertos de pesquisa: ampliando o conceito de acesso livre. **RECIIS: Revista Eletrônica de Comunicação, Informação e Inovação em Saúde**, [S. l.], v. 8, n. 2, p. 76-92, jun. 2014. Disponível em: <https://www.reciis.icict.fiocruz.br/index.php/reciis/article/view/611>. Acesso em: 10 dez. 2024.

SEPPÄLÄ, Selja; RUTTENBERG, Aalan; SMITH, Barry. Guidelines for writing definitions in ontologies. **Ciência da Informação**, Brasília, DF, v. 46, n. 1, p. 73-88, jan./abr. 2017. Disponível em: <https://revista.ibict.br/ciinf/article/view/4015>. Acesso em: 10 dez. 2024.

SMITH, Barry; CEUSTERS, Werner. Ontological realism: a methodology for coordinated evolution of scientific ontologies. **Applied Ontology**, v. 5, n. 3-4, p. 139-188, nov. 2010. Disponível em: <https://pmc.ncbi.nlm.nih.gov/articles/PMC3104413/>. Acesso em: 10 dez. 2024.

STAR, Susan Leigh; GRIESEMER, James R. Institutional ecology, "translations" and boundary objects: amateurs and professionals in Berkeley's Museum of Vertebrate Zoology, 1907-39. **Social Studies of Science**, [S. l.], v. 19, n. 3, p. 387-420, aug. 1989. Disponível em: https://lchc.ucsd.edu/mca/Mail/xmcamail.2012_09.dir/pdfuaCxVBhVe5.pdf. Acesso em: 10 dez. 2024.

VOLLMER, Gerhard. **Evolutionäre Erkenntnistheorie**. Stuttgart: S. Hirzel Verlag, 1975. 209 p.

WILKINSON, Mark D.; DUMONTIER, Michel; AALBERSBERG, Ijsbrand Jan; APPLETON, Gabrielle; AXTON, Myles; BAAK, Arie; BLOMBERG, Niklas ; BOITEN, Jan-Willem; SANTOS, Luiz Bonino da Silva; BOURNE, Philip E.; BOUWMAN, Jildau; BROOKES, Anthony J.; CLARK, Tim; CROSAS, Mercè; DILLO, Ingrid; DUMON, Olivier; EDMUNDS, Scott; EVELO, Chris T.; FINKERS,

Richard; GINZALEZ-BELTRAN, Alejandra; GRAY, Alasdair J. G.; GROTH, Paul; GOBLE, Carole; GRETHE, Jeffrey S.; HERINGA, Jaap; HOEN, Peter A. C't; HOOFT, Rob; KUHN, Tobias; KOK, Ruben; KOK, Joost; LUSHER, Scott J.; MARTONE, Maryann E.; MONS, Albert; PACKER, Abel L.; PERSSON, Bengt; ROCCA-SERRA, Philippe; ROOS, Marco; SCHAIK, Rene van; SANSONE, Susanna-Assunta; SCHULTES, Erik; SENGSTAG, Thierry; SLATER, Ted; STRAWN, George; SWERTZ, Morris A.; THOMPSON, Mark; LEI, Johan van dar; MULLIGEN, Erik van; VELTEROP, Jan; WAAGMEESTER, Andra; WITTENBURG, Peter; WOLSTENCROFT, Katherine; ZHAO, Jun; MONS, Barend. The FAIR Guiding Principles for scientific data management and stewardship. **Scientific data**, [S. l.], v. 3, n. 1, p. 1-9, mar. 2016. Disponível em: <https://www.nature.com/articles/sdata201618>. Acesso em: 10 dez. 2024.

TEXTUAL DATA GENERATED IN BIBLIOGRAPHIC RESEARCH: DEFINITION AND REPRESENTATION

ABSTRACT

Objective: Define the concept of textual data generated in bibliographic research, aiming to support the recognition and appreciation of this type of data by the scientific community, and, from there, its reuse. **Methodology:** It is characterized as qualitative and exploratory research. As for the technical procedures, it is bibliographic research that uses Bardin's Content Analysis for the exploration and interpretation of the bibliographic material. As theoretical contributions for the elaboration of definitions, Dahlberg's Concept Theory and the guidelines provided by Seppälä, Ruttenberg and Smith, renowned authors in Formal Ontology, are used. **Results:** A proposal is presented to define the concept of textual data generated in bibliographic research, in addition to relevant metadata to characterize it, as part of an initial proposal for a set of basic metadata to represent this type of data. **Conclusions:** The effort to define the concept of textual data generated in bibliographic research is an important step to encourage the reuse and management of this type of data.

Descriptors: Scientific data. Textual data. Bibliographic research. Conceptual definition. Concept theory.

DATOS TEXTUALES GENERADOS EN LA INVESTIGACIÓN BIBLIOGRÁFICA: DEFINICIÓN Y REPRESENTACIÓN

RESUMEN

Objetivo: Definir el concepto de datos textuales generados en la investigación bibliográfica, con el objetivo de apoyar el reconocimiento y valoración de este tipo de datos por parte de la comunidad científica, y, a partir de ahí, su reutilización. **Metodología:** Se caracteriza por ser una investigación cualitativa y exploratoria. En cuanto a procedimientos técnicos, se trata de una investigación bibliográfica que utiliza, para la exploración e interpretación del material bibliográfico, el Análisis de Contenido de Bardin. Como aportes teóricos para la elaboración de definiciones se utiliza la Teoría

de Conceptos de Dahlberg y los lineamientos brindados por Seppälä, Ruttenberg y Smith, autores reconocidos en el campo de la Ontología Formal. **Resultados:** Se presenta una propuesta para definir el concepto de dato textual generado en la investigación bibliográfica, así como metadatos relevantes para caracterizarlo, como parte de una propuesta inicial de un conjunto de metadatos básicos para representar este tipo de datos. **Conclusiones:** El esfuerzo por definir el concepto de datos textuales generados en la investigación bibliográfica es un paso importante para incentivar la reutilización y gestión de este tipo de datos.

Descriptores: Datos científicos. Datos textuales. Investigación bibliográfica. Definición conceptual. Teoría del concepto.

Recebido em: 25.07.2024

Aceito em: 03.12.2024