



**Ministério da Ciência,
Tecnologia e Inovação**

**Instituto Brasileiro de Informação
em Ciência e Tecnologia**

Temas de pesquisa em **PRESERVAÇÃO DIGITAL**

**Memória e Patrimônio | Museus e Acervos Artísticos | Memória Indígena
Audiovisuais | Periódicos | Repositórios Institucionais | Certificação Digital
Terminologia e Preservação Digital | Dados Pessoais**





MINISTÉRIO DA CIÊNCIA,
TECNOLOGIA E INOVAÇÃO
Instituto Brasileiro de Informação em
Ciência e Tecnologia

Temas de pesquisa em preservação digital

Organizadores

Miguel Ángel Márdero Arellano | Flor de María Silvestre Estela |
Alice da Silva dos Santos

Autores

Alex Pereira de Holanda | Alírio Alcâncer Rungo | Ana Paula Araújo Cabral da Silva | Andréia dos Santos | Cicilia Conceição de Maria | Claudete Fernandes de Queiroz | Diego Abadan Moura Melgarejo | Douglas André Muller | Emanuelle de Oliveira Silva | Fabiano Couto Corrêa da Silva | Flor de María Silvestre Estela | Gildenir Carolino Santos | Igor Dias Ferrer | Ines Aisengart Menezes | João Guilherme Nogueira Machado | José Antônio Euzébio Paiva | José Carlos Abbud Grácio | Julio Santillán-Aldana | Kadidja Valéria Reginaldo de Oliveira | Laura Vilela Rodrigues Rezende | Lenora de Beaurepaire da Silva Schwaitzer | Magda Lúcia Almada Soares | Maria de Fátima Duarte Tavares | Maria de Nazaré Freitas Pereira | Michael dos Santos Leite | Miguel Ángel Márdero Arellano | Pablo Gobira | Pablo Valério Polônia | Priscila Rezende Portugal | Rodrigo da Cunha Méxas | Ronnie Anderson Nascimento de Farias | Telma Campanha de Carvalho Madio | Vania Ferreira da Silva



Brasília, DF

2024

Copyright © 2024 by authors

Licenças de uso

Atribuição-Não Comercial-Compartilha Igual CC BY- NC- SA.

É permitido remixar, adaptar e criar a partir deste trabalho sempre e quando não houver finalidade comercial, houver atribuição do devido crédito autoral desta obra e quando as novas criações forem licenciadas sob termos idênticos.

Ficha catalográfica

Alice da Silva dos Santos
(Bibliotecária)

Edição, diagramação, revisão ortográfica e normativa

Alice da Silva dos Santos

Capa

Douglas Muller

Avaliação textual

Flor de Maria Silvestre Estela

Comissão Editorial desta Publicação

Miguel Ángel Márdero Arellano - Instituto Brasileiro de Informação em Ciência e Tecnologia

Flor de Maria Silvestre Estela - Instituto Brasileiro de Informação em Ciência e Tecnologia

Alice da Silva dos Santos - Instituto Brasileiro de Informação em Ciência e Tecnologia

Catálogo na Publicação (CIP)

T278 Temas de pesquisa em preservação digital [recurso eletrônico / Organizadores: Miguel Ángel Márdero Arellano, Flor de Maria Silvestre Estela, Alice da Silva dos Santos; autores: Alex Pereira de Holanda... [et al.]. — Brasília, DF : Ibict, 2024. 253 p. : il.

Vários autores.

Modo de acesso: World Web Wide (Internet)

ISBN: 978-65-89167-79-2

DOI: 10.22477/9786589167792

1. Preservação digital. 2. Gestão da informação. 3. Repositórios digitais. 4. Dados. 5. Curadoria digital. 6. Povo indígena.

CDU - 025.85(0.0134)

Bibliotecária: Alice da Silva dos Santos (CRB-1ª/3505)

Publicação digital - Brasil

1ª edição - Maio - 2024

ISBN: 978-65-89167-79-2

Análise terminológica de dados de pesquisa científica para preservação digital

Capítulo VIII

*Fabiano Couto Corrêa da Silva
Lenora de Beaurepaire da Silva Schwaitzer
Laura Vilela Rodrigues Rezende*

No mundo acadêmico, a linguagem não é apenas uma ferramenta de comunicação, mas também uma representação da profundidade e precisão do pensamento científico. As palavras escolhidas pelos pesquisadores muitas vezes atuam como portais que abrem novas interpretações, descobertas e debates. Nesse contexto, as nomenclaturas "dados de pesquisa" e "dados científicos" emergem como dois conceitos centrais na produção acadêmica contemporânea. Mas qual é a essência desses termos e como eles se relacionam entre si? E qual é o impacto no emprego dos dois termos quando associados ao contexto de repositórios digitais de dados visando a preservação?

Como hipótese, tem-se o entendimento de que todo dado científico é um dado de pesquisa, mas nem todos os dados de pesquisa são científicos. Esta suposição, embora pareça simplista à primeira vista, é carregada de implicações significativas sobre a natureza da pesquisa e sua prática. Por trás dessa observação, existem nuances de rigor, intenção e metodologia que vão muito além de uma mera diferenciação semântica.

Tomemos, por exemplo, a obra “Dados científicos: perspectivas e desafios” de Dias e Oliveira (2019). Esse livro agrega diversas vozes da academia, com Sayão e Sales (2019) optando pela expressão "dados de pesquisa", enquanto Dias, Anjos e Rodrigues (2019) defendem o termo "dados científicos". Essa divergência não é meramente terminológica, mas reflete as nuances e complexidades inerentes ao mundo da pesquisa.

A distinção se torna ainda mais crítica quando reconhecemos que vivemos em uma era de explosão de informações. Com a facilidade de acesso e disseminação de dados, a academia tem a responsabilidade de garantir que o que é apresentado sob o manto da "ciência" seja, de fato, fruto de investigação rigorosa e confiável.

Esta reflexão é essencial, uma vez que a diferença pode estar na metodologia, no rigor ou na intenção por trás da coleta e análise. Por isso que, parece correto afirmar que a academia, apesar de sua diversidade e amplitude, precisa buscar uma linguagem comum - uma linguagem que não apenas comunique, mas que também valide e confirme a robustez do pensamento científico.

Assim, ao mergulharmos mais profundamente nessa discussão não estamos apenas explorando terminologias, mas também refletindo sobre a natureza da pesquisa, o papel dos dados, as responsabilidades que car-

regamos ao interpretar e comunicar descobertas e, por fim, quais dados iremos coletar, preservar e disseminar em um repositório digital. Em uma era dominada por informações, é imperativo que a ciência se mantenha como um farol de precisão, rigor e confiabilidade. As palavras que escolhermos, assim como os dados que citamos, são fundamentais nesse empreendimento.

No caso do repositório digital, a escolha do termo irá impactar naquilo que iremos reunir em um mesmo ambiente com a intenção de preservar para o futuro e servir de base para novas atividades de investigação científica. Esta decisão consciente irá moldar todas as ações de um determinado organismo, estabelecendo parâmetros para o recolhimento de dados reconhecidos de valor para a guarda permanente.

Apresentado o problema que justifica a pesquisa objeto deste capítulo, passemos aos seus objetivos, tanto geral quanto específicos. O objetivo geral é identificar a essência dos termos "dados científicos" e "dados de pesquisa", buscando apontar caminhos para uma melhor definição de quais objetos poderão ser incorporados em um determinado repositório digital. Para isto, devemos entender o que são dados de uma forma geral e discutir a possível intenção no uso ora do termo "de pesquisa", ora do adjetivo "científicos" na literatura na área de Ciência da Informação.

A metodologia escolhida para a pesquisa foi a bibliográfica a qual, segundo Gil (1991), oferece base teórica que auxilia no exercício crítico sobre o tema em estudo, com abordagens quantitativa e qualitativa. A pesquisa possui natureza básica, pois se propõe a gerar conhecimento novo para o avanço científico (Nascimento, 2016) e, quanto aos objetivos, a pesquisa assume o caráter exploratório, já que a mesma tem por finalidade tornar a questão em debate mais clara (Gil, 1991).

Destaca-se que a metodologia eleita para o desenvolvimento deste capítulo possui abordagem mista quali-quantitativa, tendo como procedimento base a pesquisa bibliográfica no portal de periódicos CAPES. Como a questão gira em torno do adjetivo mais adequado para o termo "dado", tendo como finalidade estabelecer a terminologia mais adequada para o repositório que se propõe a depositar, preservar e disseminar dados obtidos durante um processo de investigação científica, foram eleitas palavras-chave - conforme informação abaixo -

que foram pesquisadas, primeiro de forma individual e, em um segundo momento, com associação do termo “repositório”. O resultado deste segundo momento da pesquisa foi deslocado para a ferramenta **Rayyan**, que auxiliou na eliminação de artigos duplicados, calculou a quantidade de artigos publicados em cada ano e identificou os autores e a quantidade de artigos produzidos.

Em toda a pesquisa bibliográfica, é preciso definir que palavras utilizaremos para delimitação de seu escopo. Em nossa pesquisa, entendemos que a seleção de variantes linguísticas em diferentes idiomas é imprescindível para garantir que nossa pesquisa seja inclusiva e abranja uma ampla gama de literatura global. Cada termo foi escolhido a partir do conhecimento tácito de sua relevância e uso comum em contextos acadêmicos nos respectivos idiomas, tendo como exemplo:

- **Research Data** (Inglês): Este é um termo amplamente utilizado em literatura acadêmica de língua inglesa para se referir a dados coletados ou usados em pesquisas;
- **Datos de Investigación** (Espanhol): Similar ao termo em inglês, é a tradução direta e é comumente usado em publicações acadêmicas em espanhol;
- **Données de Recherche** (Francês): Este termo é a tradução direta do inglês para o francês e é frequentemente usado em contextos acadêmicos de língua francesa;
- **Dados de Pesquisa** (Português): É a tradução direta do termo em inglês e é amplamente utilizado em publicações acadêmicas em português;
- **Dados Acadêmicos** (Português): Este termo foi incluído para abranger uma gama mais ampla de literatura que pode não usar o termo "Dados de Pesquisa", mas ainda se refere a dados em um contexto acadêmico;
- **Dados Científicos** (Português): Este termo foi adicionado para capturar informações que se referem especificamente a dados gerados ou utilizados em contextos científicos, diferenciando-se de dados gerais de pesquisa ou acadêmicos.

Conforme já referido anteriormente, optamos por centralizar nossa busca no portal de periódicos CAPES que, embora diversificado e amplo, e por isto atendendo aos nossos objetivos, não possibilita uma pesquisa de forma exaustiva. Além disso, devemos ressaltar que a inter-

pretação do contexto em que um termo é usado pode ser subjetiva e pode variar entre os pesquisadores. Portanto, os resultados desta pesquisa devem ser interpretados como indicativos, e não conclusivos, das tendências na literatura acadêmica. Feitas estas considerações, destacamos que, da primeira parte da pesquisa realizada, apuramos os seguintes dados:

Quadro 1 - Resultado obtido com o uso das palavras-chave

TERMO	OCORRÊNCIAS	ARTIGOS	CONJUNTO DE DADOS	INGLÊS	PORTUGUÊS	ASSUNTO COM MAIOR OCORRÊNCIA
RESEARCH DATA	12.024.162	11.432.434	20.390	11.997.268	26.241	SCIENCE&TECHNOLOGY - 9.161.895
SCIENTIFIC DATA	24.824	19.942	1.797	23.608	294	SCIENCE&TECHNOLOGY - 13.157
DATOS DE INVESTIGACIÓN	1.543	304	1.208	1.326	79	DATOS DE INVESTIGACIÓN - 50 E RESEARCH DATA - 34
DONNÉES DE RECHERCHE	715	201	451	597	5	CANADA - 78
DADOS DE PESQUISA	1.007	961	27	740	466	EDUCATION&EDUCATION RESEARCH - 36
DADOS ACADÊMICOS	104	102	1	78	49	LITERATURE REVIEW - 4
DADOS CIENTÍFICOS	603	576	7	467	256	SCIENCE&TECHNOLOGY - 20

Fonte: Elaboração própria (2023).

A partir destes dados quantitativos, é possível comprovar que o termo "research data" é de fato prevalente no ambiente acadêmico, que a publicação de artigos se dá predominantemente na língua inglesa e, embora vários assuntos sejam utilizados, o mais frequente é o de "science & technology". Interessante notar, ainda, que o termo "dados de pesquisa" é mais comumente utilizado na área de educação e de pesquisas nesta área, enquanto "dados científicos" tem como assunto mais frequente a ciência e tecnologia e o termo "dados acadêmicos", por fim, está associado ao assunto revisão de literatura.

Ao comparar a incidência e o uso dos termos entre diferentes idiomas e regiões geográficas, observou-se que o termo "Research Data" (Inglês) é mais prevalente em publicações de países de língua inglesa, como EUA e Reino Unido. "Datos de Investigación" foi mais comum em países da América Latina, enquanto "Données de Recherche" foi frequentemente usado em publicações da França e de países africanos

francófonos.

Ao se efetuar a pesquisa utilizando os mesmos termos, mas acrescentando a palavra "repositório", coletamos os seguintes resultados:

Quadro 2 - Resultado obtido com o uso das palavras-chave associadas ao termo "repositório"

TERMO	OCORRÊNCIAS	ARTIGOS	CONJUNTO DE DADOS	INGLÊS	PORTUGUÊS	ASSUNTO COM MAIOR OCORRÊNCIA
RESEARCH DATA	25.206	19704	3244	25.067	74	SCIENCE & TECHNOLOGY - 16717
SCIENTIFIC DATA	825	586	115	809	8	SCIENCE & TECHNOLOGY - 417
DATOS DE INVESTIGACIÓN	49	42	1	27	11	DATOS DE INVESTIGACIÓN - 19
DONNÉES DE RECHERCHE	1					
DADOS DE PESQUISA	79	74	3	74	3	DADOS DE PESQUISA - 13
DADOS ACADÊMICOS	5	5	0	4	1	SEM PREVALÊNCIA
DADOS CIENTÍFICOS	49	45	3	39	21	DADOS CIENTÍFICOS - 6

Fonte: Elaboração própria (2023).

Estes resultados nos possibilitam observar a redução expressiva de ocorrências mantendo-se, no entanto, a constatação de que o termo mais comum é o de "research data" e que a maioria dos artigos são publicados em inglês.

No recorte do levantamento feito em português, o assunto mais frequente, seja com o termo "dados de pesquisa" ou "dados científicos", foi o próprio termo. Com o uso da ferramenta Rayyan, procedemos à análise dos resultados obtidos com esses termos, obtendo os seguintes resultados:

Quadro 3 - Resultados obtidos com o auxílio da ferramenta Rayyan em português

TERMO	OCORRÊNCIAS	NÃO REPETIDOS	AUTORES MAIS FREQUENTES	ANO DE PUBLICAÇÃO
DADOS DE PESQUISA	79	67	SALES, Luana - 8, SAYÃO, L.F. - 7, VECHIATO, F. L. - 3, SANCHES, F. A. - 3, DIAS, G. A. - 3, MONTEIRO, E. C. - 3	2023 - 6, 2022 - 10, 2021 - 4, 2020 - 12, 2019 - 17, 2016 - 5, 2012 - 1
DADOS ACADÊMICOS	5	0	NÃO RELEVANTES	
DADOS CIENTÍFICOS	49	44	Média de 2 artigos. sem prevalência	2023 - 2, 2022 - 9, 2021 - 5, 2020 - 9, 2019 - 5, 2018 - 1, 2017 - 3, 2016 - 5, 2015 - 2, 2014 - 1, 2012 - 2

Fonte: Elaboração própria (2023).

Por fim, reunimos os resultados obtidos com os termos "dados de pesquisa" e "dados acadêmicos" e percebemos que 13 (treze) artigos aparecem em ambas as pesquisas realizadas no portal de periódicos CAPES.

Em relação à data de publicação desses artigos, identificamos que no ano em que os termos surgem pela primeira vez (2012) o artigo que utilizou o termo "dados de pesquisa" também aparece quando se efetua a consulta com o termo "dados científicos". Já no ano de 2016, quatro artigos se referem a ambos os termos, enquanto nos anos 2019 e 2020, três artigos estão associados a ambos os termos. No ano de 2021, dois resultados aparecem em ambas as pesquisas e um artigo é exibido para a pesquisa com base no ano de 2022. Esses dados nos permitem concluir que, com exceção do ano em que os termos apareceram pela primeira vez, o uso concomitante de ambas as palavras-chave foi reduzindo ao longo do tempo.

Da análise dos dados obtidos, também compreendemos que o termo "dados científicos" foi utilizado anualmente a partir de 2014, enquanto "dados de pesquisa" só passa a ser utilizado de forma sistemática a partir de 2019, coincidentemente, no ano em que a obra de Dias e Oliveira (2019) foi publicada. Ficou claro também que, com exceção dos anos de 2019 e deste ano de 2023, há um equilíbrio na adoção dos termos e que não há uma constância na quantidade de publicações no decorrer dos anos, não revelando nenhuma tendência passível de ser di-

agnosticada. Percebemos, por fim, que não há prevalência de autores que utilizam o termo "dados científicos", enquanto se identifica claramente uma tendência do uso do termo "dados de pesquisa".

Quando nos voltamos para a pesquisa qualitativa dos trabalhos, selecionamos aleatoriamente um subconjunto de publicações que continham cada termo. Dessa forma, foi possível fazer uma análise mais detalhada de cada publicação. Uma parte fundamental dessa análise foi examinar o contexto específico em que o termo foi usado, o que nos ajudou a entender a conotação e o significado do termo no contexto da publicação. Também buscamos entender se o termo foi central para o foco principal da publicação ou se foi mais periférico.

Finalmente, chegamos à interpretação dos resultados. Com base nos dados que coletamos e nas análises que realizamos, discutimos as implicações da prevalência e do uso de cada termo. Esta discussão nos levou a refletir sobre as possíveis razões para as tendências que observamos. Em particular, buscamos entender o que essas tendências puderam revelar sobre a evolução do pensamento acadêmico, especialmente em relação à distinção entre "dados de pesquisa" e "dados científicos", incluindo os repositórios digitais e a futura preservação neste contexto.

Conforme esclarecido na seção anterior, consideramos que é possível identificar a tendência de alguns autores. Sayão e Sales (2012 e 2017), se inclinam para o termo "dados de pesquisa" para identificar o repositório, acompanhado por Sanches e Vechiato (2019a e 2019b). Tanto Dias (2019; 2020) quanto Curty (2019) preferem o termo "dados científicos", enquanto Lucas, Picalho e Caitano (2021) utilizam a expressão "repositórios multidisciplinares de dados científicos abertos". Monteiro (2018a; 2018b e 2023), se preocupa com aspectos administrativos dos repositórios, sendo de pouca relevância estabelecer se são repositórios de dados científicos ou dados de pesquisa. Pavão, Rocha e Gabriel Júnior (2018) focam sua pesquisa no acesso aberto, assim como Cervo, Rockembach e Rocha (2023). Os trabalhos orientados por Caregnato (Medeiros e Caregnato, 2012 e Silveira *et al.*, 2020), além de debaterem o acesso aberto, se preocupam com o uso e o volume de dados.

Entretanto, observa-se que a maioria dos artigos não possui a preocupação central de definir o conceito ou o objetivo dos repositóri-

os, ficando evidente que não houve uma maior reflexão no momento da escolha do termo, seja ele "dados de pesquisa", "dados científicos", "dados acadêmicos", ou qualquer outro termo que seja associado com o objeto "repositório". No artigo publicado em 2016, em que tecem algumas considerações sobre os repositórios digitais de dados de pesquisa, os autores esclarecem que partem do "[...] pressuposto de que dados de pesquisa são, em grande parte, resultados de pesquisas financiadas com dinheiro público [...]" (Sayão; Sales, 2016, p. 92), entretanto sem se aprofundar sobre a amplitude que o termo pode provocar.

Merece destaque também o recente artigo de Sayão e Sales (2022), que apresenta em sua introdução o que os autores entendem como "dados de pesquisa", definindo-os como dados complexos formados por uma variedade de informações fortemente influenciadas pelos contextos específicos de suas áreas disciplinares. Acrescentam, ainda, que a compreensão desses dados está intrinsecamente ligada à riqueza das formas como sua origem e trajetória são representadas. Neste mesmo trabalho, os autores abordam a gestão de dados em repositórios e, neste sentido, reconhecem a existência de

[...] inúmeros repositórios multidisciplinares e de múltiplos propósitos, numa escala que vai de repositórios institucionais, por exemplo, pertencentes a uma única universidade, à repositórios abertos de escopo global tais como *FigShare*, *Dryad*, *Mendeley Data*, *Zenodo*, *DataHub*, *DANS* e *EUDat*, entre outros (Sayão; Sales, 2023, p. 4, grifo dos autores).

Porém, os autores não justificam sua escolha para o termo "dados de pesquisa". Por outro lado, Dias, Anjos e Rodrigues (2019) favorecem a expressão "dados científicos". Esta escolha pode refletir uma ênfase na precisão, replicabilidade e no rigor metodológico, características essenciais para assegurar a validade e confiabilidade das descobertas em um contexto científico mais estruturado.

A coexistência dessas abordagens contrastantes destaca a rica tapeçaria de perspectivas no campo. Essa diversidade, embora possa parecer confusa à primeira vista, também é um testemunho da evolução contínua e da adaptabilidade da academia. No entanto, também sublinha uma questão mais profunda: a ausência de um padrão uniforme ou de um consenso definido em relação a esses termos.

Essa vacilação terminológica não se limita a mera definição conceitual, na verdade, é um reflexo mais amplo do debate que se desenrola no universo acadêmico, um debate que tem implicações tanto pragmáticas quanto epistemológicas. A literatura nos convida a questionar: estamos apenas dividindo uma discussão trivial, ou há nuances metodológicas e teóricas genuínas por trás dessa distinção?

Além disso, essa variação na literatura sugere uma oportunidade: a chance de mergulhar mais fundo, de extrair mais significado e, possivelmente, de chegar a um entendimento mais unificado sobre a natureza e o papel dos dados na pesquisa contemporânea. O caminho à frente também pode oferecer clareza onde, atualmente, encontramos ambiguidade.

Por conta disso, consideramos necessário discorrer sobre as palavras-chave utilizadas na pesquisa, buscando identificar o termo que consideramos mais adequado para definir e qualificar repositórios que se propõem à coleta, preservação e disseminação de dados que sejam produto de uma investigação científica.

O conceito de "dados" tem sido amplamente discutido e explorado em diversos campos acadêmicos. A ciência de dados, por exemplo, emergiu como um novo campo acadêmico devido à crescente digitalização da sociedade. Emmert-Streib e Dehmer (2018) utilizaram um modelo de regressão estatística para analisar dados de cientistas e suas citações, buscando, dessa forma, definir a importância dos campos de pesquisa e fornecer uma resposta objetiva à pergunta "O que é ciência de dados?".

Borgman (2015) em uma tentativa de estabelecer a dimensão do que vem a ser dados no contexto científico, prefere indagar: "quando são dados?" ao invés de "o que são dados?". Esta preferência pode ser explicada pelo fato de que o escopo do que vem a ser dados no âmbito científico necessariamente se esbarra nos processos e ações envolvidas na geração destes dados. Como pesquisadores, comunidades, indivíduos criam, selecionam ou usam dados? Quais fatores nestas decisões estão associados aos dados em sua essência? Quais fatores poderiam estar associados aos questionamentos da pesquisa ou métodos científicos?

Se faz necessário pontuar, também, que definir o que são dados sofre relevante influência nas variações expressas pelas áreas do conhe-

cimento e podem evoluir ao longo dos processos investigativos na medida em que são criados e coletados. Assim, é possível afirmar que os dados não são objetos puros ou naturais com uma essência própria. Eles sempre existirão em um contexto particular e com significado específico partindo, assim, da perspectiva de quem os lê (Borgman, 2015).

A importância dos dados não se limita apenas à ciência de dados. Jimerson e Reames (2015) exploraram o conceito de "uso de dados envolvendo estudantes", onde os alunos rastreiam, analisam e interpretam seus próprios dados de maneira formal e estruturada. Além disso, Kang *et al.* (2019) discutiram as dificuldades em definir o que é ou não é ciência de dados urbanos, sugerindo que o termo "ciência de dados" tem valor em aplicações industriais ou populares, mas pode não ser bem compreendido em círculos puramente acadêmicos.

A análise de dados também revelou que diferentes conceitualizações de conhecimento fornecem o cenário para a gestão da pesquisa acadêmica, conforme observado por Sousa e Hendriks (2008). Mayernik (2019) destacou a importância de estudar metadados para fornecer informações sobre a produção de evidências e entender como algo que chamamos de "dados" pode desempenhar um papel evidenciário.

O termo "dados" também tem implicações significativas no jornalismo. Tong e Zuo (2019) sugeriram uma mudança na atenção acadêmica, passando da celebração da objetividade no jornalismo de dados para examinar a epistemologia dos jornalistas de dados. Emmert-Streib, Moutari e Dehmer (2016) argumentaram que, independentemente de quem usou o termo "ciência de dados" primeiro e em que contexto, é um bom termo para indicar que os dados são o foco da pesquisa científica.

Setzer (2001), buscando estabelecer a distinção entre dados, informação, conhecimento e competência, define dados como "[...] uma sequência de símbolos quantificados ou quantificáveis". Abordagem semelhante é a de Valentim (2002), que busca na área da Ciência da Informação apresentar a definição e características de dados, informação e conhecimento. Sobre o dado em si, a autora os conceitua como "Simples observações sobre o estado do mundo [...] estruturado [...] obtido por máquinas [...] transferível" (Davenport; Prusak, 1998, p.

18 *apud* Valentim, 2002, p. 2).

Caminho análogo é o percorrido por Semidão (2010), que entende dado como o estágio inicial de uma articulação de um fenômeno qualquer, que se estrutura para compreensão do fenômeno representado pelo dado e se consolida como conhecimento acumulado a partir da apropriação da informação.

Já para Borgman (2015), os dados podem ser compreendidos como representações de observações, objetos e/ou outras entidades utilizadas como evidências de algum fenômeno para o propósito de alguma pesquisa (científica ou não). A autora adota a definição de entidade como sendo algo que possui uma existência material real, diferentemente de uma mera função, atributo ou relação. Exemplificando, entidades podem ser textos em papel (papiro ou digitais), sinais gerados de sensores e formulários respondidos de uma pesquisa *online*. Todas essas evidências se tornam dados quando utilizadas como evidência (s) de algum fenômeno.

Destaca-se também que o termo “dado”, quando se refere ao conceito em si, em geral é utilizado no singular. Exemplo: dado não é um termo novo, big data é o petróleo dos negócios modernos (Borgman, 2015). Assim, independentemente da abordagem, podemos concluir que dado é a menor unidade cognitiva e que sua compreensão depende de sua estruturação em informação para poder ser apropriada por indivíduos ou máquinas.

O panorama atual da literatura sobre a distinção entre dados de pesquisa e dados científicos, nos mostra o quão dinâmico e, às vezes, ambíguo esse campo de estudo pode ser. Como ilustração dessa complexidade, podemos examinar os livros “Dados científicos: perspectivas e desafios” (Dias; Oliveira, 2019) e “Gestão de dados científicos” (Silva, 2019). Em ambas as obras, os autores adotam uma postura firme e clara ao utilizar o termo “Dados Científicos”. Eles o definem com uma especificidade que destaca sua aplicabilidade direta e exclusiva ao domínio científico. Para esses autores, “Dados Científicos” são aqueles gerados e utilizados com o propósito inequívoco de investigação científica, submetidos a padrões metodológicos rigorosos e destinados a avançar o conhecimento em suas respectivas áreas.

Apesar de não focarmos primariamente em estudos terminológicos, a distinção entre os termos é fundamental para nossa análise subsequente.

Há, de fato, nuances que diferenciam os dois termos. Entendemos que "dados científicos" referem-se estritamente a dados obtidos através de procedimentos rigorosos e metodologias científicas. Já os "dados de pesquisa", em sua definição mais ampla, podem englobar qualquer dado coletado para fins investigativos, independente de sua finalidade ser científica ou não.

A profundidade da linguagem e o rigor na terminologia são dois pilares que sustentam o mundo acadêmico e de pesquisa. O uso impreciso ou mal definido de termos pode levar a conclusões distorcidas ou errôneas, o que conseqüentemente prejudicaria a interpretação e aplicação dos resultados. Portanto, embora não aprofundemos as minúcias da semântica, é imperativo que estabeleçamos um entendimento claro dos termos que empregamos, garantindo uma análise sólida e confiável.

O termo "dados científicos", como já esclarecido, é adotado para descrever informações que são coletadas seguindo métodos e protocolos científicos estritos. Estes dados passam por processos rigorosos de verificação e validação antes de serem considerados aceitáveis para publicação ou uso em estudos posteriores. A integridade e precisão destes dados são de suma importância, pois são esses dados que servem como a base sobre a qual a comunidade científica constrói novas teorias validando hipóteses existentes.

Por outro lado, "dados de pesquisa" é uma terminologia mais inclusiva. Ela não apenas abrange os "dados científicos", mas também outros conjuntos de informações coletadas para propósitos diversos que podem não seguir o rigor científico tradicional. Por exemplo, um jornalista pode coletar "dados de pesquisa" ao conduzir entrevistas para uma reportagem, ou um empresário pode coletar dados de clientes para entender melhor seu mercado. Enquanto esses dados são inegavelmente importantes em seus respectivos contextos, eles não necessariamente passam pelo escrutínio rigoroso que é exigido dos dados científicos (He; Nahar, 2016).

Além disso, a distinção se torna ainda mais importante quando consideramos a maneira como os dados são interpretados e aplicados. Enquanto os "dados científicos" são frequentemente usados para formular, testar e refinar teorias em um ambiente acadêmico, os "dados de pesquisa" podem ser utilizados em uma variedade de contextos e pa-

ra uma gama mais ampla de propósitos. Nesse sentido, é vital reconhecer e respeitar essa distinção. Ao fazer isso, asseguramos que nossa análise e discussão subsequentes sejam baseadas em uma compreensão sólida e inequívoca dos dados em mãos, evitando, assim, mal-entendidos e garantindo que nossas conclusões sejam tão robustas quanto precisas. A terminologia, em última análise, serve não apenas como uma ferramenta de comunicação, mas também como um guia, ajudando a navegar e interpretar o complexo terreno da pesquisa e do conhecimento.

A compreensão é simples: todos os "dados científicos" se enquadram como "dados de pesquisa", mas a recíproca não é necessariamente verdadeira. Esta distinção é vital, pois nos permite navegar pela literatura com uma compreensão clara de que termo está sendo empregado e em que contexto.

Ao adentrarmos o domínio acadêmico e de pesquisa, a sutileza com que as terminologias são utilizadas pode ter implicações profundas em relação à interpretação e aplicação do conhecimento gerado. Assim, a clareza terminológica não é apenas uma questão de semântica, mas uma necessidade para preservar a integridade da investigação e das conclusões derivadas dela.

Os "dados científicos" representam o ápice da precisão em coleta, análise e interpretação, servindo como a base para hipóteses, teorias e descobertas que moldam e expandem nosso conhecimento. Quando nos referimos a eles na literatura, estamos evocando uma certa garantia de qualidade e autenticidade (Hug; Mcneill, 2008).

Em contrapartida, os "dados de pesquisa" constituem um espectro mais amplo. Eles podem ser oriundos de estudos científicos, mas também de outras formas de investigação que podem não se submeter ao mesmo rigor metodológico dos estudos científicos. A diversidade de fontes e métodos associados aos "dados de pesquisa" torna imperativo entender a origem e a natureza dos dados antes de fazer inferências ou aplicações baseadas neles.

Aqui reside a essência da distinção: enquanto todos os "dados científicos" são um subconjunto dos "dados de pesquisa", a amplitude dos "dados de pesquisa" engloba uma variedade de informações, algumas das quais podem não ter passado pelos padrões de validação que caracterizam o rigor científico. Ignorar essa distinção pode levar a

equivocos, onde dados coletados em contextos menos rigorosos são tratados com a mesma confiança que os dados científicos.

Além disso, ao ter uma clara demarcação entre esses termos, os pesquisadores e leitores são capazes de avaliar, com mais precisão, a robustez e a aplicabilidade das informações apresentadas. Isso é fundamental para garantir que decisões, sejam elas em pesquisa, práticas profissionais ou políticas públicas, sejam baseadas em evidências de alta qualidade.

Adicionalmente, outra nomenclatura surge embora mais raramente: "dados acadêmicos". Este termo, conforme entendido, pode ser alinhado com "dados científicos", e conseqüentemente, encontra-se sob o guarda-chuva de "dados de pesquisa".

A menção ao termo "dados acadêmicos" nos conduz a mais um nível de complexidade na já intrincada teia de terminologias associadas à pesquisa. Sua presença, embora menos frequente do que os outros termos já discutidos, sugere nuances e contextos específicos que merecem nossa atenção.

Os "dados acadêmicos" são frequentemente associados ao ambiente de instituições de ensino e pesquisa, como universidades, faculdades e institutos. Eles não se limitam apenas aos resultados de investigações científicas, mas também podem englobar uma série de outras informações geradas no âmbito acadêmico. Por exemplo, estes podem incluir dados relacionados a avaliações de estudantes, pesquisas de opinião sobre currículos, informações administrativas, entre outros. Estes dados são todos gerados dentro do domínio acadêmico, mas não necessariamente fruto de investigação científica rigorosa (Walford, 2017).

Por outro lado, quando os "dados acadêmicos" são oriundos de pesquisas realizadas em instituições acadêmicas, eles se alinham mais estreitamente com o que chamamos de "dados científicos". Isso porque a pesquisa acadêmica, em sua maioria, é conduzida seguindo padrões rigorosos e metodologias aprovadas. No entanto, é fundamental entender que nem todos os "dados acadêmicos" são "dados científicos", dada a diversidade de informações produzidas em um ambiente acadêmico.

A incorporação deste termo à discussão amplia a necessidade de clareza e especificidade. Sem uma compreensão adequada dos termos,

corre-se o risco de formar suposições errôneas. Por exemplo, alguém poderia presumir, equivocadamente, que todos os "dados acadêmicos" têm a mesma validade e confiabilidade dos "dados científicos", o que pode não ser o caso.

Por fim, ao considerar "dados acadêmicos" sob o termo mais amplo de "dados de pesquisa", reconhecemos que todas as informações coletadas no contexto acadêmico têm valor e potencial de contribuição. No entanto, é importante discernir sua origem, método de coleta e propósito para garantir que sejam utilizados de maneira adequada e que as inferências derivadas deles sejam fundamentadas e válidas.

No decorrer deste capítulo, buscamos estabelecer a distinção entre os termos "dados de pesquisa" e "dados científicos", por entendermos que a terminologia adotada irá impactar não somente em diversos aspectos da absorção de conhecimento, mas também na identificação e no conteúdo que será incorporado a um determinado tipo de repositório digital. Neste contexto, julgamos conveniente destacar que os dados científicos são evidências obtidas durante o percurso investigativo visando comprovar seus resultados que, desde o século XVI, são disseminados por meio das denominadas revistas científicas.

Vê-se, portanto, que a "revista científica" surge da prática de se reunir diversos artigos que resumem investigações científicas em curso ou já concluídas em uma linguagem impessoal e está associada eminentemente à forma de se disseminar o conhecimento, antigamente somente em papel, e hoje em formato digital. A credibilidade dessas publicações, principalmente na atualidade, decorre do processo de revisão cega por pares, que atribui valor ao que está sendo publicado e robustece o caráter eminentemente científico da investigação descrita nos artigos. Essas revistas ficavam à disposição dos consulentes em bibliotecas, para que novos conhecimentos fossem produzidos a partir das mesmas.

Em uma realidade digital, o conhecimento científico tem gerado dados digitais cada vez mais concentrados em repositórios que assumem denominações múltiplas. Eles podem ser considerados repositórios institucionais, repositórios digitais, repositórios digitais confiáveis, repositórios de acesso aberto, repositórios de dados de pesquisa, ou até mesmo repositórios de dados científicos. Da mesma forma, como a carência de precisão terminológica entre “dados de pesquisa” e “dados

científicos" evidencia a evolução contínua da academia, ela também denuncia a ausência de uma padronização ou consenso em relação ao objetivo e escopo desses repositórios.

Um repositório digital é um tipo de ferramenta tecnológica que objetiva agregar e disseminar informações para os consulentes ou usuários, sendo

[...] um serviço de armazenamento de objetos digitais que tem a capacidade de manter e gerenciar materiais por longos períodos de tempo e prover o seu acesso apropriado. Os repositórios digitais dividem-se em temáticos, institucionais e centrais (Márdero Arellano, 2008, p. 124).

Dessa forma, os repositórios digitais diferem quanto à própria ferramenta utilizada, o tipo de material coletado e quanto à garantia na preservação ao longo do tempo de forma confiável.

Na área arquivística, há um consenso quanto ao repositório digital confiável ser uma solução de gerenciamento de materiais digitais, constituído por uma associação entre: *hardware*; *software*; metadados de uma infraestrutura organizacional; e procedimentos normativos e técnicos capazes de manter a organicidade e autenticidade dos documentos arquivísticos digitais, preservá-los e assegurar seu acesso ao longo do tempo (CONARQ, 2015).

Em relação às denominadas fontes de informação tradicionalmente oferecidas em uma biblioteca, costuma-se denominar essa ferramenta tecnológica como repositórios institucionais que, para Márdero Arellano (2008, p. 124), "[...] são sistemas de informação que armazenam, preservam, divulgam e dão acesso à produção intelectual de instituições e comunidades científicas, em formato digital [...]".

Estabelecidas estas diferenças em relação aos tipos de repositórios existentes, destaca-se que o enfoque aqui será àqueles cuja finalidade é o acesso a conjuntos de dados, conferindo a estabilidade, a preservação, o uso, a coleta de estatística e a implementação de política de gestão de grande volume de dados obtidos a partir da pesquisa científica. O que viabiliza, assim, a Ciência Aberta nos moldes propostos pela UNESCO.

Rezende *et al.* (2023), enfatizam a importância dos repositórios ao oferecerem de forma aberta o conhecimento científico em diferentes idiomas. Esses repositórios devem facilitar o acesso e a reutilização desse conhecimento por todos, fomentar colaborações e o compartilha-

mento de informações em prol da ciência e da sociedade. Além disso, é crucial que esses processos sejam abertos não apenas para a comunidade científica, mas também para outros participantes da sociedade.

Existindo finalidade e demandando ferramenta específica para sua operacionalização, parece-nos evidente que é preciso que este tipo de repositório receba um tratamento exclusivo que o diferencie daquele que se propõe a reunir materiais bibliográficos ou que organize documentos arquivísticos. Tais repositórios trazem ainda embutida a ideia de um procedimento específico, o de planejamento e implantação de uma gestão de dados voltados para o desenvolvimento da Ciência Aberta, o que o distingue dos demais.

Ao pensarmos em atribuir qualidade aos repositórios de dados digitais, destaca-se a certificação *Core Trust Seal*, que visa conferir os repositórios de dados interessados, um atestado, ou selo de qualificação com base em 16 (dezesesseis) requisitos a serem autoavaliados e subsequentemente conferidos pela equipe avaliadora da entidade responsável pela emissão do selo (Core Trust Seal, 2023). O primeiro requisito a ser verificado e providenciado pelo repositório é a definição de sua missão e contexto que o qualifica. É preciso definir quais objetos digitais são aceitos para serem arquivados, tipo de repositório (se é um repositório multidisciplinar, institucional, temático, etc.). qual a área do conhecimento contemplada, se for o caso, além do público alvo ao qual os conjuntos de dados se destinam.

Assim, parece correto afirmar que a terminologia adequada para se definir os objetos digitais, que neste caso são os dados científicos, de pesquisa ou até mesmo acadêmicos, poderá conferir maior precisão no processo de qualificação do que poderá ser contemplado pelo repositório situando tanto produtores quanto consumidores de dados.

Bode e Sousa (2015) trazem um outro ponto a ser considerado no contexto dos repositórios que é sobre a linguagem, a qual pode ser entendida segundo os autores como uma capacidade humana que permite uma série de desdobramentos em relação ao pensamento e conhecimento humanos, em sua interação com os demais indivíduos e a cultura social. Os autores destacam a importância de se definir como os sistemas atuais que recuperam informações, nos referindo aqui aos repositórios especificamente, são estruturados para levar em considera-

ção o aspecto linguagem, se é que o fazem, e quais aspectos são considerados ou não. Isso se faz necessário se considerarmos por exemplo que os indivíduos assumem papéis na governança dos repositórios que podem resultar em usos específicos da língua.

Os autores exemplificam, ainda, que é possível considerar que uma pessoa assuma pelo menos o papel de produtor de documentos e informações, pesquisador de documentos e informações e gestor (indexador, classificador e/ou catalogador) de documentos e informações (Bode e Sousa, 2015). Tendo a preservação digital dos dados em repositórios como escopo reflexivo, isso pode implicar em tentar equacionar problemas como a língua escrita utilizada por alguém. Esse caso se aplica quando pensamos no papel de indexador no século passado como também na atualidade, quando uma pessoa assume um papel de pesquisador.

Os resultados da análise indicam uma tendência crescente na utilização do termo "dados de pesquisa" em publicações acadêmicas. No entanto, é fundamental destacar a importância e a relevância do termo "dados científicos". Este termo, além de evocar um conceito mais robusto e preciso, é amplamente utilizado em contextos que enfatizam rigor, metodologia e validação, características essenciais para a integridade e confiabilidade da pesquisa.

A revisão da literatura evidenciou que "dados científicos" é frequentemente associado a trabalhos que estão categorizados nos campos como ciências naturais, medicina e engenharia, onde a precisão e a replicabilidade são fundamentais. Além disso, o termo foi predominantemente usado em publicações que discutiam rigor metodológico, validação e replicabilidade, reforçando sua importância no cenário acadêmico.

Embora o termo "dados acadêmicos" esteja emergindo em alguns contextos, especialmente relacionados à educação e avaliações institucionais, ele não possui a mesma profundidade e abrangência que "dados científicos". A variação no uso dos termos entre diferentes idiomas e regiões geográficas destaca a diversidade da pesquisa global. No entanto, para garantir a precisão e a relevância da comunicação acadêmica, é essencial que a comunidade científica adote e promova o uso correto da terminologia.

Com base nas evidências apresentadas, recomenda-se fortemente a

utilização do termo "dados científicos" como padrão na literatura acadêmica quando se tratar de repositórios que armazenam, preservam e disseminam dados resultantes de investigações científicas, dada sua clareza conceitual e sua ampla aceitação em campos de pesquisa rigorosos.

Em se tratando de linguagem no contexto dos repositórios de dados e preservação digital, deve-se considerar questões que envolvem a individualidade no momento de se realizar o processo descritivo dos dados a serem depositados. Assim considera-se, finalmente, que definições previamente estabelecidas sobre terminologias e padrões quanto ao preenchimento dos metadados nos repositórios, favorecem a inteligibilidade futura destes insumos que foram, então, previamente preservados.

REFERÊNCIAS

ANJOS, R. L. D.; DIAS, G. A.; RODRIGUES, A. A. Dados científicos: as práticas de gestão dos pesquisadores brasileiros na ciência da informação. ENCONTRO NACIONAL DE PESQUISA E PÓS-GRADUAÇÃO EM CIÊNCIA DA INFORMAÇÃO, n. XVIII ENANCIB, 2017. **Anais** [...]. Disponível em: <http://hdl.handle.net/20.500.11959/brapci/104605>. Acesso em: 11 ago. 2023.

BODE, E. C., SOUSA, R. T. B. de. Preservação digital, recuperação da informação e linguagem. **Revista Ibero-Americana De Ciência Da Informação**, 8(2), 122–141. 2015. Disponível em: <https://doi.org/10.26512/rici.v8.n2.2015.2061>. Acesso em: 19 fev. 2024.

BORGMAN, C. L. Research data: who will share what, with whom, when, and why? **Ratswd working paper**, v. 161, n.10, 2010. Disponível em: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1714427. Acesso em: 11 ago 2023.

CERVO, M.; ROCKEMBACH, M.; ROCHA, A. L. C. Dados abertos antropológicos: o que são e quais suas problemáticas? **Encontros Bibli: Revista Eletrônica de Biblioteconomia e Ciência da Informação**. Florianópolis: UFSC, 2022. Vol. 28,(2023), e90091, 2023. Disponível em: <https://www.lume.ufrgs.br/bitstream/handle/10183/257670/001166852.pdf?sequence=1&isAllowed=y>. Acesso em: 18 fev. 2024.

CONARQ. **Diretrizes para a implementação de repositórios arquivísticos**

digitais confiáveis - RDC-Arq. Rio de Janeiro: Arquivo Nacional, 2015. PDF.

COSTA, M. M.; CUNHA, M. B. A necessidade de uma política nacional para a gestão de dados de pesquisa no Brasil. **Liinc em Revista**, [S. l.], v. 15, n. 2, 2019. DOI: 10.18617/liinc.v15i2.4763. Disponível em: <https://revista.ibict.br/liinc/article/view/4763>. Acesso em: 31 jan. 2024.

CURTY, R. Abordagens de reuso e a questão da reusabilidade dos dados científicos. **Liinc em Revista**, [S. l.], v. 15, n. 2, 2019. DOI: 10.18617/liinc.v15i2.4777. Disponível em: <https://revista.ibict.br/liinc/article/view/4777>. Acesso em: 18 fev. 2024.

CURTY, R. G.; AVENTURIER, P. O paradigma da publicação de dados e suas diferentes abordagens. In: XVIII ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO (XVIII ENANCIB), n. XVIII ENANCIB, 2017. **Anais [...]**. Disponível em: <http://hdl.handle.net/20.500.11959/brapci/105144>. Acesso em: 11 ago. 2023.

DIAS, G. A.; OLIVEIRA, B. M. J. F. **Dados científicos: perspectivas e desafios.** João Pessoa: Ed. UFPB, 2019.

DIAS, G. A.; SIEBRA, S. A.; SOUSA, R. P. M.; SOUSA, M. R. F. Publicando dados de pesquisa: contextualizando as principais etapas e elementos envolvidos no processo. **Ciência da Informação**, [S. l.], v. 49, n. 3, 2020. DOI: 10.18225/ci.inf.v49i3.5501. Disponível em: <https://revista.ibict.br/ciinf/article/view/5501>. Acesso em: 18 fev. 2024.

EMMERT-STREIB, F.; MOUTARI, S.; DEHMER, M. The process of analyzing data is the emergent feature of data science. **Frontiers in Genetics**, v. 7, p. 12, 2016. Disponível em: <https://www.frontiersin.org/articles/10.3389/fgene.2016.00012/full>. Acesso em: 31 jan. 2024.

EMMERT-STREIB, F.; DEHMER, M. Defining data science by a data-driven quantification of the community. **Mach. Learn. Knowl. Extr.**, v. 1, n. 1, p. 29-59, 2018. Disponível em: <https://www.mdpi.com/2504-4990/1/1/15>. Acesso em: 31 jan. 2024.

GIL, A. C. **Métodos e técnicas de pesquisa social.** São Paulo: Atlas, 1991.

HE, L.; NAHAR, V. Reuse of scientific data in academic publications: a Dryad digital repository investigation. **Aslib J. Inf. Manag.**, [S. l.], v. 68, n. 2, p. 207-223, 2016. Disponível em: <https://doi.1108/AJIM-01-2016-0008/full/html2016-0008/full/html>. Acesso em: 31 jan. 2024.

HUG, B.; MCNEILL, K. L. Use of first-hand and second-hand data in science: does data type influence classroom conversations? **International Journal of Science Education**, [S. l.], v. 30, n. 13, p. 1725-1751, 2008. Disponível em: [10.1080/09500690701506945](https://doi.10.1080/09500690701506945). Acesso em: 30 jan. 2024.

JIMERSON, J.; REAMES, E. H. Student-involved data use: establishing the evidence base. **Journal of Educational Change**, v. 16, n. 2, p. 125-142, 2015. Disponível em: <https://link.springer.com/article/10.1007/s10833-015-9246-4>. Acesso em: 30 jan. 2024.

KANG, W. *et al.* A roundtable discussion: defining urban data science. **Environment and Planning B: Urban Analytics and City Science**, v. 46, n. 9, p. 1596-1608, 2019. Disponível em: <https://journals.sagepub.com/doi/10.1177/2399808319882826>. Acesso em: 3 jan. 2024.

LUCAS, E. R. O.; PICALHO, A. C.; CAITANO, V. M. H. Mapeamento e descrição de características de repositórios multidisciplinares de dados científicos abertos. **BiblioCanto**, [S. l.], v. 7, n. 1, p. 37-61, 2021a. DOI: 10.21680/2447-7842.2021v7n1ID25703. Disponível em: <https://periodicos.ufrn.br/bibliocanto/article/view/25703>. Acesso em: 18 fev. 2024.

MANNHEIMER, S.; PIENTA, A.; KIRILOVA, D.; ELMAN, Colin; WUTICH, A. Qualitative data sharing: data repositories and academic libraries as key partners in addressing challenges. **American Behavioral Scientist**, [S. l.], v. 62, n. 5, p. 688-707, 2018. Disponível em: <https://journals.sagepub.com/doi/full/10.1177/0002764218784991?journalCode=absb>. Acesso em: 30 jan. 2024.

MÁRDERO ARELLANO, M. À. **Critérios para a preservação digital da informação científica**. 2008, 356 f. Tese (Doutorado em Ciência da Informação) - Universidade de Brasília, Brasília, 2008. Disponível em: http://eprints.rclis.org/12649/1/Tese_Miguel_%C3%81ngel_M%C3%A1rdero_Arellano.pdf. Acesso em: 30 jan. 2024.

MAYERNIK, M. Metadata accounts: achieving data and evidence in scientific research. **Social Studies of Science**, v. 49, n. 3, p. 400-423, 2019. Disponível em: <https://journals.sagepub.com/doi/10.1177/0306312719863494>. Acesso em: 30 jan. 2024.

MEDEIROS, J. S.; CAREGNATO, S. E. Compartilhamento de dados e e-Science: explorando um novo conceito para a comunicação científica. **LIINC em revista**. Laboratório Interdisciplinar sobre Informação e Conhecimento. Rio de Janeiro, vol. 8, n. 2 (set. 2012), p. 311-322, 2012. Disponível em: <https://www.lume.ufrgs.br/handle/10183/254028>. Acesso em: 18 fev. 2024.

MONTEIRO, E. C. S. A.; AFFONSO, E. P.; BORBA, V. U.; SANT'ANA, R. C. G. A privacidade e os planos de gerenciamento de dados de repositórios de dados científicos. **Informação & Tecnologia**, [S. l.], v. 4, n. 1, p. 35-53, 2018. DOI: 10.22478/ufpb.2358-3908.2017v4n1.37586. Disponível em: <https://periodicos.ufpb.br/index.php/itec/article/view/37586>. Acesso em: 19 fev. 2024.

MONTEIRO, E. C. S. A.; SANT'ANA, R. C. G. Plano de gerenciamento de dados em repositórios de dados de universidades. **Encontros Bibli**: revista eletrônica de Biblioteconomia e Ciência da informação, v. 23, n. 53, p. 160-173, 2018. Disponível em: <https://periodicos.ufsc.br/index.php/eb/article/view/1518-2924.2018v23n53p160>. Acesso em: 18 fev. 2024.

MONTEIRO, E. C. S. A.; SANT'ANA, R. C. G. Anomia de pesquisadores no compartilhamento de dados. **Em Questão**, v. 29, p. e-122627, 2023. Disponível em: <https://www.academia.edu/download/66356073/21006.pdf>. Acesso em: 18 fev. 2024.

OLIVER, J.; KOLLEN, C.; HICKSON, B.; RIOS, F. Data science support in university library. **Journal of Library Administration**, [S. l.], v. 59, n. 4, p. 418-431, 2019. Disponível em: <https://www.tandfonline.com/doi/full/10.1080/01930826.2019.1583015>. Acesso em: 30 jan. 2024.

PAVÃO, C. G.; ROCHA, R. P.; GABRIEL JUNIOR, R. F. Proposta de criação de uma rede de dados abertos da pesquisa brasileira. **RDBCI**: Revista Digital de Biblioteconomia e Ciência da Informação, v. 16, n. 2, p. 329-343, 2018. Disponível em: <https://periodicos.sbu.unicamp.br/ojs/index.php/rdbci/article/view/8651180>. Acesso em: 18 fev. 2024.

REZENDE, L. V. R. *et al.* Concepção de uma ferramenta brasileira de elaboração de planos de gestão de dados de pesquisa: desafios rumo ao modelo de planos acionáveis por máquina, MaDMP. **BiD**: textos universitaris de biblioteconomia i documentació. Barcelona: Universitat de Barcelona. n. 50, jun. 2023. DOI: 10.1344/BiD2023.50.07. Disponível em: <https://bid.ub.edu/en/50/vilela.htm>. Acesso em: 3 out. 2023.

SALES, L. F.; SAYÃO, L. F. O impacto da curadoria digital dos dados de pesquisa na comunicação científica. **Encontros Bibli**: revista eletrônica de biblioteconomia e ciência da informação, [S. l.], p. 118–135, 2012. DOI: 10.5007/1518-2924.2012v17nesp2p118. Disponível em: <https://periodicos.ufsc.br/index.php/eb/article/view/1518-2924.2012v17nesp2p118>. Acesso em: 18 fev. 2024.

SANCHEZ, F. A.; VECHIATO, F. L. Avaliação de repositórios de dados de pesquisa segundo critérios da encontrabilidade da informação. **Liinc em Revista**, [S. l.], v. 15, n. 2, 2019a. DOI: 10.18617/liinc.v15i2.4774. Disponível em: <https://revista.ibict.br/liinc/article/view/4774>. Acesso em: 18 fev. 2024.

SANCHEZ, F. A.; DA SILVA, N. B. P.; VECHIATO, F. L. Padrões de metadados para representação e organização da informação em repositórios de dados de pesquisa. **Informação & Tecnologia**, [S. l.], v. 5, n. 1, p. 37–51, 2019b. Disponível em: periodicos.ufpb.br/. Acesso em: 18 fev. 2024.

SAYÃO, L. F.; SALES, L. F. **Guia de gestão de dados para bibliotecários e pesquisadores**. Rio de Janeiro: CNEN/IEN, 2015. 90 p. Disponível em: [https:// Guia de gestao de dados de pesquisa para bibliotecarios e pesquisadores](https://Guia de gestao de dados de pesquisa para bibliotecarios e pesquisadores). Acesso em: 11 ago. 2022.

SAYÃO, L. F.; SALES, L. F. **Algumas considerações sobre os repositórios digitais de dados de pesquisa**. 2016, p. 90-115. DOI: 10.5433/1981-8920.2016v21n2p90. Disponível em: <https://ojs.uel.br/revistas/uel/index.php/informacao/article/view/27939>. Acesso em: 25 set. 2023.

SAYÃO, L. F.; SALES, L. F. Plataformas de gestão de dados de pesquisa: expandindo o conceito de repositórios de dados. **Palavra Clave (La Plata)**, [S. l.], v. 12, n. 1, p. e171, 2022. DOI: 10.24215/18539912e171. Disponível em: <https://www.palavraclave.fahce.unlp.edu.ar/article/view/pce171>. Acesso em: 25 set. 2023.

SEMELER, A. R.; PINTO, A. L. Os diferentes conceitos de dados de pesquisa na abordagem da biblioteconomia de dados. **Ciência da Informação**, [S. l.], v. 48, n. 1, 2019. DOI: 10.18225/ci.inf.v48i1.4461. Disponível em: <https://revista.ibict.br/ciinf/article/view/4461>. Acesso em: 30 jan. 2024.

SEMIDÃO, R. A. M. Atributos de dados, informação e conhecimento enquanto elementos de fundamentação conceitual para abordagens abrangentes em ciência da informação. **Revista de Iniciação Científica da FFC (Cessada)**, v. 10, n. 3, 2010. Disponível em: <https://revistas.marilia.unesp.br/index.php/ric/article/view/333>. Acesso em: 30 jan. 2024.

SETZER, V. W. **Os meios eletrônicos e a educação: uma visão alternativa**. São Paulo: Editora Escrituras. (Coleção Ensaios Transversais). v. 10, 2011. Disponível em: <https://www.ime.usp.br/~vwsetzer/livro-meios.htm>. Acesso em: 30 jan. 2024.

SILVA, F. C. C. **Gestão de dados científicos**. Rio de Janeiro: Interciência, 2019. 146 p. ISBN 9788571934351.

SILVEIRA, L.; BARBOSA, A. D.; FERREIRA, M. K.; CAREGNATO, S. E. Citação de dados científicos: scoping review. **Encontros Bibli: Revista Eletrônica de Biblioteconomia e Ciência da Informação**, [S. l.], v. 25, p. 01–31, 2020. DOI: 10.5007/1518-2924.2020.e72153. Disponível em: <https://periodicos.ufsc.br/index.php/eb/article/view/1518-2924.2020.e72153>. Acesso em: 19 fev. 2024.

SOUSA, C. A.; HENDRIKS, P. Connecting knowledge to management: the case of academic research. **Organization**, v. 15, n. 6, p. 869-888, 2008. Disponível em: <doi/10.1177/1350508408091004>. Acesso em: 30 jan. 2024.

STUMPF, I. R. C. Passado e futuro das revistas científicas. **Ciência da Informação**, [S. l.], v. 25, n. 3, 1996. DOI: 10.18225/ci.inf.v25i3.637. Disponível em: <https://revista.ibict.br/ciinf/article/view/637>. Acesso em: 24 set. 2023.

TONG, J.; ZUO, L. The inapplicability of objectivity: understanding the work of data journalism. **Journalism Practice**, v. 13, n. 6, p. 686-701, 2019. Disponível em: <https://www.tandfonline.com/doi/full/10.1080/17512786.2019.1698974>. Acesso em: 30 jan. 2024.

VALENTIM, M. L. P. Inteligência competitiva em organizações: dado, informação e conhecimento. **DataGamaZero, Revista de Ciência da Informação**, v. 3, n. 4, ago, 2002. Disponível em: <https://pesquisa.bvsalud.org/portal/resource/pt/lis-10722>. Acesso em: 30 jan. 2024.

WALFORD, A. Raw data: making relations matter. **Social Analysis**, [S. l.], v. 61, n. 2, p. 74-91, 2017. Disponível em: <https://discovery.ucl.ac.uk/id/eprint/10051061/>. Acesso em: 30 jan. 2024.