

## O reúso de dados de pesquisa na perspectiva da Ciência da informação

**Crislaine Zurilda Silveira**

Universidade Federal de Santa Catarina - UFSC, Florianópolis-SC, Brasil

**Thiago Magela Rodrigues Días**

Universidade Federal de Santa Catarina - UFSC, Florianópolis-SC, Brasil

ORIGINAL

### Resumo

**Objetivo.** Este artigo busca correlacionar o reúso de dados de pesquisa na agenda das teorias contemporâneas da Ciência da Informação, propostas por Carlos Alberto Ávila Araújo, em 2018.

**Método.** Foram selecionados 32 artigos qualificados e representativos sobre o tema reúso de dados de pesquisa, usando o instrumento ProKnow-C. Após a seleção foi realizada uma análise sistêmica, onde os objetivos dos artigos foram classificados conforme as teorias contemporâneas.

**Resultados.** Os resultados demonstram que na produção e comunicação científica os estudos se voltaram para entendimento dos aspectos que norteiam o conceito de reúso de dados de pesquisa, na identificação dos fatores que limitam o reúso e na identificação dos aspectos relacionados a curadoria de dados de pesquisa. Nos estudos sobre os sujeitos, os artigos buscaram identificar as barreiras para o compartilhamento e reúso de dados, os incentivos que podem estimular os pesquisadores a reusar dados de pesquisa e na identificação das características envolvidas no comportamento busca desses dados. Nos estudos métricos da informação foi possível identificar a métrica de citação como a mais utilizada para identificar o reúso dos dados.

**Conclusões.** Foi verificado que os artigos se concentraram nas teorias voltadas para a Produção e Comunicação Científica, seguida daquelas que envolvem os Estudos sobre os Sujeitos e Estudos Métricos da Informação. Este artigo mostrou que o paradigma da Ciência da Informação vem sendo deslocando para um eixo de estudos sobre dados de pesquisa. Ainda não trata de uma substituição do paradigma vigente, mas mostra que a Ciência da Dados pode trabalhar de maneira conjunta com os paradigmas existentes. O desafio para a Ciência da Informação, neste contexto, consiste relacionar os fatores técnicos tratados no reúso de dados de pesquisa com uma Ciência da Informação mais humana, voltada para a participação dos cidadãos e para a superação de problemas sociais.

Palavras-chave:

*Ciência da Informação; Paradigmas; Proknow-C; Reutilização de dados de pesquisa; Teorias contemporâneas.*

## La reutilización de datos de investigación desde la perspectiva de la Ciencia de la Información

### Resumen

**Objetivo.** Este artículo busca correlacionar la reutilización de datos de investigación en la agenda de las teorías contemporáneas de la Ciencia de la Información, propuestas por Carlos Alberto Ávila Araújo en 2018.

**Método.** Se seleccionaron 32 artículos cualificados y representativos sobre el tema de la reutilización de datos de investigación, utilizando el instrumento ProKnow-C. Después de la selección, se realizó un análisis sistémico, donde se clasificaron los objetivos de los artículos de acuerdo con las teorías contemporáneas.

**Resultados.** Los resultados muestran que en la producción y comunicación científica, los estudios se centraron en la comprensión de los aspectos que orientan el concepto de reutilización de datos de investigación, la identificación de los factores que limitan la reutilización y la identificación de aspectos relacionados con la curación de datos de investigación. En los estudios sobre temas, los artículos buscaban identificar las barreras para compartir y reutilizar datos, los incentivos que pueden animar a los investigadores a reutilizar datos de investigación e identificar las características que intervienen en el

comportamiento de búsqueda de estos datos. En los estudios sobre métricas de información, se pudo identificar la métrica de citas como la más utilizada para identificar la reutilización de datos.

**Conclusiones.** Se constató que los artículos se concentraban en teorías centradas en la Producción y Comunicación Científica, seguidas de las que involucraban Estudios Temáticos y Estudios de Métricas de la Información. Este artículo ha demostrado que el paradigma de la Ciencia de la Información ha ido cambiando hacia un eje de estudios sobre datos de investigación. Esto todavía no sustituye al paradigma actual, pero demuestra que la Ciencia de Datos puede trabajar conjuntamente con los paradigmas existentes. El reto para la Ciencia de la Información en este contexto es relacionar los factores técnicos tratados en la reutilización de datos de investigación con una Ciencia de la Información más humana, orientada a la participación ciudadana y a la superación de problemas sociales.

Palabras clave:

*Ciencia de la información; Paradigmas; Proknow-C; Reutilización de datos de investigación; Teorías contemporáneas.*

## The reuse of research data from the perspective of Information Science

### Abstract

**Objective.** This article seeks to correlate the reuse of research data in the agenda of contemporary theories of Information Science, proposed by Carlos Alberto Ávila Araújo, in 2018.

**Method.** Thirty-two qualified and representative articles on the topic of reuse of research data were selected, using the ProKnow-C instrument. After the selection, a systemic analysis was performed, where the objectives of the articles were classified according to contemporary theories.

**Results.** The results show that in scientific production and communication, studies have focused on understanding the aspects that guide the concept of reuse of research data, identifying factors that limit reuse and identifying aspects related to research data curation. In the studies about the subjects, the articles sought to identify the barriers to sharing and reusing data, the incentives that can encourage researchers to reuse research data and to identify the characteristics involved in the behavior of searching for these data. In the metric studies of information, it was possible to identify the citation metric as the most used to identify the reuse of data.

**Conclusions.** It was found that the articles focused on theories focused on scientific production and communication, followed by those involving studies on subjects and metric studies of information. This article showed that the Information Science paradigm has been shifting towards an axis of studies on research data. It still does not deal with a replacement of the current paradigm, but it shows that Data Science can work together with the existing paradigms. The challenge for Information Science, in this context, is to relate the technical factors dealt with in the reuse of research data with a more humane Information Science, focused on citizen participation and overcoming social problems.

Keywords:

*Information Science; Paradigms; Proknow-C; Research data reuse; Contemporary theories*

---

## 1 Introdução

A Ciência da Informação na perspectiva de Capurro (2003) possui duas raízes, a primeira influenciada pela Biblioteconomia clássica ou, em termos gerais, pelo estudo dos problemas relacionados com a transmissão de mensagens, ou da informação, e a segunda influenciada por um caráter tecnológico, que se refere ao impacto da computação nos processos de produção, coleta, organização, interpretação, armazenagem, recuperação, disseminação, transformação e uso da informação, e em especial, da informação científica registrada em documentos impressos. Sendo que a segunda raiz, é fortemente influenciado pela explosão informacional advinda na Segunda Guerra Mundial.

Atualmente, vem se observando na Ciência da Informação uma mudança de paradigma. Nele a Ciência da Informação vem sofrendo influências da Big Data, E-science e da Ciência Aberta, com isso observa-se um aumento do número de publicações, voltadas para a preocupação com a ciência dos dados. Mais precisamente no meio científico, o paradigma da Ciência da Informação, volta-se para os estudos sobre como a ciência pode fazer a curadoria dos dados que produz de maneira a torná-los mais acessível aos cidadãos, reutilizável para outros pesquisadores e recuperável pelas máquinas.

Nesse sentido, a epistemologia da Ciência da Informação desloca-se de uma ciência voltada para as práticas informacionais, para uma ciência que busca, a partir dos dados, extrair informações, construir novos conhecimentos para a solução de antigos e novos problemas e assim fortalecer as relações interdisciplinares.

O reúso de dados de pesquisa, que é o foco desse artigo, parte da premissa de que esse tipo de dado é considerado como parte do conjunto dinâmico do fazer científico, sua tipologia, propósito e curadoria dependem do seu contexto. No entanto, as limitações na infraestrutura, tais como metadados não padronizados, conjuntos de dados não curados ou coleções incompletas, determinam quais dados podem ser encontrados e, portanto, reutilizados (Gregory *et al.*, 2020).

Curty *et al.* (2017) identificou que antigamente os dados de pesquisa eram compartilhados apenas entre alguns colegas confiáveis e conhecidos. Mas, com as exigências de agências de fomento, órgãos reguladores e com a necessidade de uma mudança cultural por parte dos cientistas em tornar a ciência mais acessível para todos, a comunidade científica vem adotando práticas informacionais, seja para compartilhamento dos dados ou para seu reúso.

O depósito de dados de pesquisa em repositórios públicos tem várias vantagens, desde o avanço na inovação da pesquisa até a retenção da integridade dos dados por meio da preservação a longo prazo, sendo assim, os pesquisadores podem reutilizar esses dados depositados para reproduzir pesquisas, validar resultados e propor novas pesquisas. (Gregory *et al.*, 2020; He; Nahar, 2016).

Embora as vantagens da reutilização sejam amplamente conhecidas, em algumas áreas do conhecimento, o termo em si, "reutilização de dados de pesquisa", ainda parece não estar definido. O entendimento de reutilização varia entre disciplinas e entre indivíduos (Van de Sandt *et al.*, 2019). Nas Ciências Sociais por exemplo, comumente são utilizados os termos "dados secundários" ou "análise secundária" (Yoon, 2017).

A reutilização de dados de pesquisa ou a reusabilidade é entendida por Thanos (2017) como a facilidade de utilização de dados para pesquisa científica legitimada por uma ou mais comunidades de pesquisa (comunidades de consumidores) e que são produzidas por outras comunidades de pesquisa (comunidades de produtores) com um propósito para estudar um novo problema. Curty *et al.* (2017) também corrobora com essa ideia ao afirmar que a reutilização de dados perpassa pelo uso de um conjunto de dados por alguém que não seja quem coletou para implantar novas pesquisas. Essas definições propostas por Curty *et al.* (2017) e Thanos (2017), no entanto, limitam o reúso para um novo uso dos dados de pesquisa por pessoas que não realizaram a coleta e para realizar novas pesquisas.

Um conjunto de dados produzidos por uma comunidade para ser utilizado por outra comunidade deve ser dotado de propriedades (informações auxiliares) que levem em consideração as características da "relação de usabilidade" entre duas comunidades. Por isso, os dados para serem (re)utilizáveis devem ser complementados com informações apropriadas de metadados (proveniência, contexto, qualidade, incerteza, etc.), seleção dos dados, adição de informações contextuais necessárias para a interpretação (Bhattacharya *et al.*, 2018; Borgman; Scharnhorst; Golshan, 2019; Thanos, 2017).

A disponibilidade de dados para reutilização depende da infraestrutura para tornar esses dados descobertos, recuperáveis, interpretáveis e utilizáveis (Borgman; Scharnhorst; Golshan, 2019). Além da criação de padrões para citar os conjuntos de dados, cuja propriedade é importante para garantir que os pesquisadores que os compartilham recebam crédito adequadamente por sua contribuição (Federer *et al.*, 2015).

Fecher, Friesike e Hebing (2015) e Joo e Kim (2017) elencam que fornecer workshops sobre reutilização de dados, sobre a sua curadoria, estimular cultura de compartilhamento, a documentação de dados e segurança pode ser uma intervenção eficaz para educar os membros da comunidade para a preparação para a sua.

Considerando esse contexto, este artigo busca correlacionar o reúso de dados de pesquisa na agenda das teorias contemporâneas da Ciência da Informação, propostas por Araújo (2018), como uma forma de evidenciar o seu deslocamento para um eixo de estudos sobre dados de pesquisa nessa área.

Após a seção de introdução, na seção dois é descrita uma breve revisão de literatura, na seção três são apresentados os aspectos metodológicos da pesquisa, na seção quatro os dados são organizados e analisados, na seção cinco são realizadas as conclusões, e por fim, são arroladas as referências.

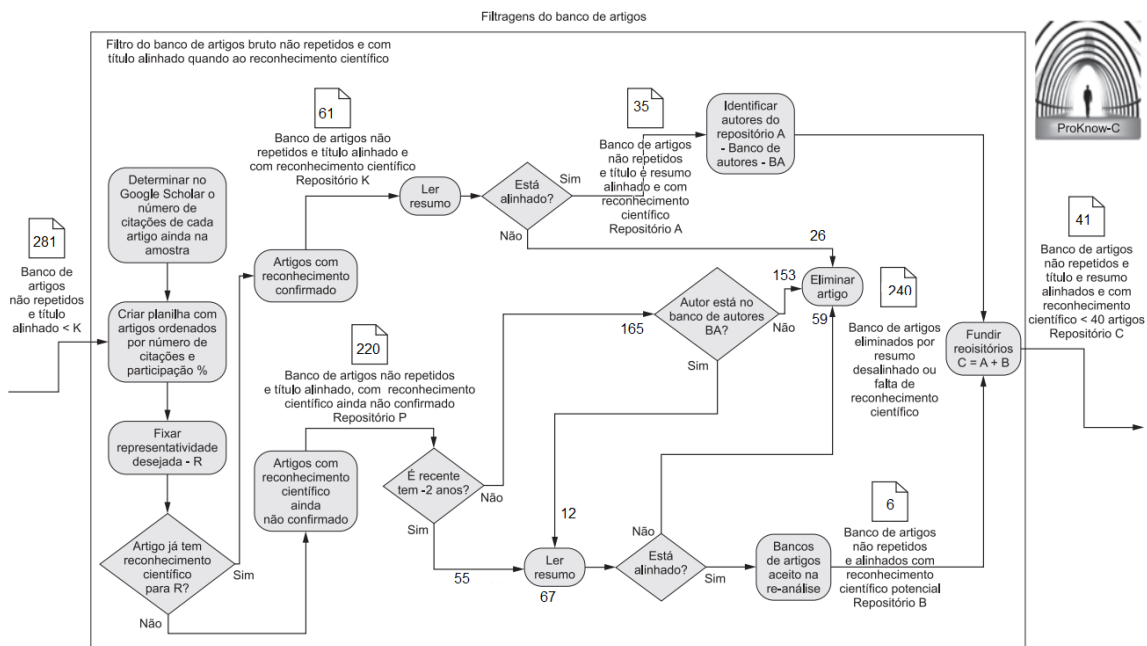
## 2 Metodologia

Trata-se de uma revisão sistematizada da literatura usando o instrumento *ProKnow-C*, cujo objetivo é embasar a construção do conhecimento do pesquisador sobre uma determinada temática a partir da seleção dos artigos relevantes que irão compor um portfólio bibliográfico (Lacerda; Ensslin; Ensslin, 2012). O processo é composto

por quatro etapas: (a) seleção de um portfólio de artigos sobre o tema da pesquisa; (b) análise bibliométrica do portfólio; (c) análise sistêmica; e a (d) definição da pergunta e do objetivo de pesquisa (Ensslin; Ensslin; Pinto, 2013). Neste artigo é relatado apenas a etapa de análise sistêmica, realizada partir de um portfólio de artigos representativos sobre o tema, cuja seleção está descrita no processo descrito na figura 1.

Na figura 1, é possível observar que ao final do processo 41 artigos passaram para a fase de leitura integral dos textos, a fim de concluir a filtragem dos artigos conforme alinhamento da pesquisa. Ao fim da leitura, 32 artigos, arrolados no quadro 1, foram selecionados para compor o portfólio de pesquisa, a ser verificado através da análise sistêmica.

**Figura 1 – Processo seleção dos artigos conforme alinhamento da pesquisa**



Fonte: adaptado de Lacerda, Ensslin e Ensslin (2012)

## Quadro 1 – Artigos que fazem parte do Portfólio bibliográfico

Referências dos artigos selecionados	Citações recebidas
FECHER, B.; FRIESIKE, S.; HEBING, M. What Drives Academic Data Sharing? <i>PLoS ONE</i> , [s.l.], v. 10, n. 2, 2015.	328
BHATTACHARYA, S. <i>et al.</i> ImmPort, toward repurposing of open access immunological assay data for translational and clinical research. <i>Scientific Data</i> , [s.l.], v.5, p. 1-9, 2018.	273
DANCIU, I. <i>et al.</i> Secondary use of clinical data: The Vanderbilt approach. <i>Journal of Biomedical Informatics</i> , [s.l.], v.52, p. 28-35, 2014.	195
FEDERER, L. M. <i>et al.</i> Biomedical data sharing and reuse: attitudes and practices of clinical and scientific research staff. <i>PLoS ONE</i> , [s.l.], v.10, n. 6, 2015.	107
PĂNESCU, A.; MANTA, V. Smart contracts for research data rights management over the Ethereum blockchain network. <i>Science &amp; Technology Libraries</i> , [s.l.], p. 1-12, 2018	86
CARROLL, S.R. <i>et al.</i> The CARE Principles for Indigenous Data Governance. <i>Data Science Journal</i> , [s.l.], v. 19, n. 43, p. 1-12, 2020.	86
CURTY, R.G. <i>et al.</i> Attitudes and norms affecting scientists' data reuse. <i>PLoS ONE</i> , [s.l.], v. 12, n. 12, 2017.	80
YOON, A. Data reusers' trust development. <i>Journal of the Association for Information Science and Technology</i> , [s.l.], v. 68, n. 4, p. 946-956, 2017.	69
CHAUVETTE, A.; SCHICK-MAKAROFF, K.; MOLZAHN, A. E. Open data in qualitative research. <i>International Journal of Qualitative Methods</i> , [s.l.], v. 18, p. 1-6, 2019.	69
CHILDS, S. <i>et al.</i> Opening research data: issues and opportunities. <i>Records Management Journal</i> , [s.l.], v. 24, n. 2, p. 142-162, 2014.	61
IRWIN, S.; WINTERTON, M. Qualitative secondary analysis and social explanation. <i>Sociological research online</i> , [s.l.], v. 17, n. 2, p. 1-12, 2012.	57
POOLE, A. H. The conceptual landscape of digital curation. <i>Journal of Documentation</i> , [s.l.], v. 72, n. 5, p. 961-986, 2016.	48
PARK, H.; WOLFRAM, D. An examination of research data sharing and re-use: implications for data citation practice. <i>Scientometrics</i> , [s.l.], v. 111, p. 443-461, 2017.	46
GREGORY, K. M. <i>et al.</i> Understanding data search as a socio-technical practice. <i>Journal of Information Science</i> , [s.l.], v. 46, n. 4, p. 459-475, 2020	44
KIM, Y.; NAH, S. Internet Researchers' data sharing behaviors: an integration of data reuse experience, attitudinal beliefs, social norms, and resource factors. <i>Online Information Review</i> , [s.l.], v. 42, n. 1, p. 1-31, 2018.	42
BORGMAN, C L.; SCHARNHORST, A.; GOLSHAN, M. S. Digital data archives as knowledge infrastructures: Mediating data sharing and reuse. <i>Journal of the Association for Information Science and Technology</i> , [s.l.], v. 70, n. 8, p. 888-904, 2019.	42
WOMACK, R. P. Research Data in Core Journals in Biology, Chemistry, Mathematics, and Physics. <i>PLoS ONE</i> , [s.l.], v. 10, n. 12, 2015.	39
VAN DE SANDT, S. <i>et al.</i> The definition of reuse. <i>Data Science Journal</i> , [s.l.], v. 18, n. 22, p. 1-19, 2019	36
ZUIDERWIJKA, A.; SPIERS, H. Sharing and re-using open data: A case study of motivations in astrophysics. <i>International Journal of Information Management</i> , [s.l.], v. 49, p. 228-241, 2019.	35
TENOPIR, C. <i>et al.</i> Research data sharing: practices and attitudes of geophysicists. <i>Earth and Space Science</i> , [s.l.], v. 5, p. 891-902, 2018.	31
JOO, S.; KIM, S.; KIM, Y. An exploratory study of health scientists' data reuse behaviors: examining attitudinal, social, and resource factors. <i>Aslib Journal of Information Management</i> , [s.l.], v. 69, n. 4, p. 1-29, 2017.	29
PERRIER, L.; BLONDAL, E.; MACDONALD, H. The views, perspectives, and experiences of academic researchers with data sharing and reuse: a meta-synthesis. <i>PLoS ONE</i> , [s.l.], v. 15, n. 2, 2020.	29
HE, L.; NAHAR, V. Reuse of scientific data in academic publications: an investigation of Dryad Digital Repository. <i>Aslib Journal of Information Management</i> , [s.l.], v. 68, n. 4 p1-23, 2016.	28
THANOS, C. Research data reusability: conceptual foundations, barriers and enabling Technologies. <i>Publications</i> , [s.l.], v. 5, n. 2, p. 1-19, 2017.	25
SINACI, A. A. <i>et al.</i> From raw data to FAIR Data: the FAIRification workflow for health research. <i>Methods of Information in Medicine</i> , [s.l.], v. 59, p. e21–e32. 2020.	24
JOO, Y. K.; KIM, Y. Engineering researchers' data reuse behaviours: a structural equation modelling approach. <i>The Electronic Library</i> , [s.l.], v. 35, n. 6, 2017.	13
YOON, A.; LEE, Y. Y. Factors of trust in data reuse. <i>Online Information Review</i> , [s.l.], v. 43, n. 7, p. 1245-1262, 2019.	11
CHATFIELD, S. L. Recommendations for Secondary Analysis of Qualitative Data. <i>The Qualitative Report</i> , [s.l.], v. 25, n. 3, p. 833-842, 2020	9
PARK, M. S.; PARK, H. An examination of metadata practices for research data reuse: Characteristics and predictive probability of metadata elements. <i>Malaysian Journal of Library &amp; Information Science</i> , [s.l.], v. 24, n. 3, p. 61-75, 2019.	5
IMKER, H. J. <i>et al.</i> An examination of data reuse practices within highly cited articles of faculty at a research university. <i>The Journal of Academic Librarianship</i> , [s.l.], v. 47, p.1-11, 2021.	3
GREGORY, K. M. A dataset describing data discovery and reuse practices in research. <i>Scientific Data</i> , [s.l.], v. 7, n. 232, 2020.	3
KRÄMER, T. <i>et al.</i> Data-seeking behaviour in the Social Sciences. <i>International Journal on Digital Libraries</i> . <i>International Journal on Digital Libraries</i> , [s.l.], v. 22, p. 175-195, 2021.	2

Fonte: elaborado pelos autores.

Antes de proceder para a análise sistêmica, as referências foram classificadas com base nas teorias contemporâneas da Ciência da Informação, proposta por Araújo (2018), descritas no quadro 2.

**Quadro 2 – Teorias contemporâneas da Ciência da Informação e os estudos envolvidos**

Teorias contemporâneas	Estudos envolvidos
Economia política da informação	Nessa subárea se caracterizam os estudos sobre arqueologia da ideia de sociedade da informação, ética intercultural da informação, regimes de informação,
Estudos métricos da informação	Nessa subárea se caracterizam os estudos sobre análise de citação, altmetria,
Estudos sobre os sujeitos	Nessa subárea se caracterizam os estudos sobre práticas informacionais, mediação da informação, apropriação e competência crítica em informação.
Gestão da Informação	Nessa subárea se caracterizam os estudos sobre cultura organizacional, cultura informacional, orientação informacional
Memória, patrimônio e documento	Nessa subárea se caracterizam os estudos sobre memória, neodocumentação, bibliografia e humanidades digitais
Produção e comunicação científica	Nessa subárea se caracterizam os estudos sobre acesso livre a informação e curadoria digital,
Representação e organização da informação	Nessa subárea se caracterizam os estudos sobre ontologias, websemântica, classificação facetada, folksonomias, análise de domínio, informação e semiótica e recuperação da informação.

Fonte: elaborado pelos autores, a partir de Araújo, 2018.

A definição das teorias e a classificação dos estudos, realizada a partir dos objetivos, culminou para a definição de lentes que foram utilizadas na análise sistêmica, sendo que essas estão descritas no quadro 3.

**Quadro 3 – Lentes de pesquisa**

Teorias contemporâneas	Lentes
Economia política da informação	- Quais estudos tratam de economia política da informação?
Estudos métricos da informação	- Quais estudos tratam de estudos métricos? - Quais métricas foram relatadas? - De que maneira essas métricas podem ajudar na construção coletiva da ciência?
Estudos sobre os sujeitos	- Quais os estudos tratam sobre os sujeitos? - Quais as barreiras para o reúso de dados? - Quais as barreiras para o compartilhamento de dados? - Quais os incentivos para o compartilhamento e reúso de dados? - Quais as características envolvidas na busca de dados de pesquisa?
Gestão da Informação	Não houve estudos vinculados a esta teoria.
Memória, patrimônio e document	Não houve estudos vinculados a esta teoria.
Produção e comunicação científica	- Quais estudos tratam sobre a produção e comunicação científica? - Quais aspectos norteiam o conceito de reúso de dados de pesquisa? - Quais fatores limitam o reúso de dados de pesquisa qualitativa? - Quais os aspectos relacionados com a curadoria de dados de pesquisa?
Representação e organização da informação	Não houve estudos vinculados a esta teoria.

Fonte: elaborado pelos autores, a partir de Araújo, 2018.

A definição das lentes de pesquisa em formato de perguntas, contribuiu para a identificação das oportunidades de pesquisa na Ciência da Informação, a fim de preencher as lacunas do reúso de dados na área.

### 3 Resultados

A partir da análise das teorias contemporâneas propostas por Araújo (2018) e dos objetivos dos artigos, foi realizada a classificação dos estudos, conforme quadro 4.

**Quadro 4 – Classificação dos estudos conforme as teorias contemporâneas da Ciência da Informação**

Autor(es)	Ano	Objetivo	Teoria contemporânea na CI
He; Nahar	2016	Identificar como e por que pesquisadores reutilizaram dados compartilhados por outros em diferentes campos de pesquisa.	Estudos métricos da informação
Park; Wolfram	2017	Examinar as características de compartilhamento e reutilização de dados em Genética e Hereditariedade, onde a citação de dados é mais comum.	Estudos métricos da informação
Womack	2015	Fornecer evidências empíricas para a presença ou ausência das práticas de citação, compartilhamento, reprodutibilidade, publicação de dados em revistas científicas de quatro disciplinas: Biologia, Química, Matemática e Física.	Estudos métricos da informação
Curty <i>et al.</i>	2017	Desenvolver um modelo teórico dos fatores que precedem a reutilização de dados e testar sua importância com dados empíricos.	Estudos sobre os sujeitos
Fecher; Friesike; Hebing	2015	Oferecer uma análise interdisciplinar das barreiras e facilitadores predominantes para o compartilhamento de dados na academia, a partir de uma revisão sistemática.	Estudos sobre os sujeitos
Federer <i>et al.</i>	2015	Investigar as diferenças nas experiências e percepções sobre o compartilhamento de dados, bem como as barreiras ao compartilhamento entre pesquisadores clínicos e de ciências básicas, através de entrevistas.	Estudos sobre os sujeitos
Gregory	2020	Analisar as necessidades de dados, comportamentos de descoberta, critérios e estratégias na avaliação de dados para reutilização.	Estudos sobre os sujeitos
Gregory <i>et al.</i>	2020	Integrar estudos cientométricos e métodos qualitativos, combinando bibliometria e entrevistas para obter evidências empíricas sobre as práticas de pesquisa de dados dos pesquisadores.	Estudos sobre os sujeitos
Imker <i>et al.</i>	2021	Investigar se os indivíduos que produziram os dados compartilhavam perspectivas semelhantes sobre a reutilização de dados com os indivíduos que os citaram.	Estudos sobre os sujeitos
Joo; Kim	2017	Investigar os fatores que influenciam os comportamentos de reutilização de dados de pesquisadores de engenharia.	Estudos sobre os sujeitos
Joo; Kim; Kim	2017	Examinar como os fatores atitudinais, sociais e de recursos dos cientistas da saúde afetam seu comportamento de reutilização de dados.	Estudos sobre os sujeitos
Kim, Nah	2018	Examinar como a experiência na reutilização de dados, crenças de atitude, normas sociais e fatores de recursos influenciam pesquisadores da Internet a compartilhar dados com outros pesquisadores fora de suas equipes.	Estudos sobre os sujeitos
Krämer <i>et al.</i>	2021	Observar o comportamento de busca de informações de cientistas sociais em busca de dados de pesquisa para revelar impedimentos e identificar oportunidades para infraestrutura de pesquisa de dados.	Estudos sobre os sujeitos
Perrier; Blondal; MacDonald	2020	Relatar as opiniões e experiências dos pesquisadores sobre compartilhamento e reutilização de dados.	Estudos sobre os sujeitos
Tenopir <i>et al.</i>	2018	Investigar motivos, atitudes e práticas de dados da comunidade da Terra e geofísicos planetários.	Estudos sobre os sujeitos
Yoon	2017	Explorar os processos dos reutilizadores de dados para fazer julgamentos de confiança sobre os dados.	Estudos sobre os sujeitos
Yoon; Lee	2019	Examinar quantitativamente os fatores de confiança na reutilização de dados na perspectiva dos reutilizadores	Estudos sobre os sujeitos
Zuiderwijk; Spiers	2019	Fornecer uma visão aprofundada sobre a complexa interação de fatores que influenciam as motivações para compartilhar e reutilizar dados de pesquisa dentro da disciplina de astrofísica.	Estudos sobre os sujeitos
Bhattacharya <i>et al.</i>	2018	Descrever as práticas de curadoria adotadas no Immport - repositório de dados sobre imunologia	Produção e comunicação científica
Borgman; Scharnhorst; Golshan	2019	Relatar as práticas de curadoria de dados do Dutch Data Archiving and Networked Services Institute (DANS) para a área de Ciências Sociais.	Produção e comunicação científica

Autor(es)	Ano	Objetivo	Teoria contemporânea na CI
Carroll <i>et al.</i>	2020	Relatar a criação de um padrão de governança de dados de pesquisa sobre povos indígenas.	Produção e comunicação científica
Chatfield	2020	Encorajar os pesquisadores a usar os dados de pesquisa qualitativa existentes de maneira ponderada e criativa.	Produção e comunicação científica
Chauvette; Molzahn; Schick-Makaroff	2019	Argumentar nem todos os dados de pesquisa qualitativa são apropriados para acesso aberto, pois questões epistemológicas, metodológicas, legais e éticas podem se tornar preocupações.	Produção e comunicação científica
Childs <i>et al.</i>	2014	Apresentar o tema de dados de pesquisa abertos, descreve brevemente os dois projetos de pesquisa de RDM conduzidos pelos autores e os desafios de gerenciamento de dados vivenciados pelos pesquisadores participantes.	Produção e comunicação científica
Danciu <i>et al.</i>	2014	Fornecer um resumo do uso da abordagem Vanderbilt University Medical Center no uso secundário de dados clínicos para domínio de pesquisa.	Produção e comunicação científica
Irwin; Winterton	2012	Debater sobre o reúso de dados nas pesquisas qualitativas.	Produção e comunicação científica
Park; Park	2019	Explorar as práticas de metadados na relação com a reutilização de dados em Biologia.	Produção e comunicação científica
Pănescu, Manta	2018	Propor um novo método de compartilhamento de dados de pesquisa de forma que permita aos autores supervisionar como eles estão sendo acessados e reutilizados.	Produção e comunicação científica
Poole	2016	Definir e descrever curadoria digital, um campo emergente de teoria e prática nas profissões da informação que abrange preservação digital, curadoria de dados e gerenciamento de ativos de informação ao longo de seu ciclo de vida.	Produção e comunicação científica
Sinaci <i>et al.</i>	2020	Abordar o projeto de arquitetura de uma solução tecnológica aberta construída sobre o processo FAIRification proposto pelo "GO FAIR".	Produção e comunicação científica
Thanos	2017	Abordar a dimensão tecnológica da reusabilidade de dados. Os fundamentos conceituais da reutilização de dados, assim como as barreiras que impedem a reutilização de dados.	Produção e comunicação científica
Van de Sandt <i>et al.</i>	2019	Avaliar critérios que distinguem a reutilização de todas as outras ações relacionadas e encontrar uma definição que torne a reutilização mensurável e compreensível para um público mais amplo.	Produção e comunicação científica

Fonte: elaborado pelos autores.

Com base no quadro 4, é possível observar que os estudos se concentram nas teorias voltadas para a Produção e Comunicação Científica, seguida daquelas que envolvem os Estudos sobre os Sujeitos e Estudos Métricos da Informação. Pode-se observar também, que são ausentes os estudos nas subáreas de Representação e Organização da Informação, Gestão da Informação, Economia Política da Informação e estudos sobre Memória, Patrimônio e Documento.

### 3.1 Produção e comunicação científica

Os estudos nessa área, de maneira geral, buscam identificar e analisar a matriz coletiva de produção do conhecimento científico (Araújo, 2018). Uma das iniciativas mais significativas nessa área, recentemente, é o movimento em prol do acesso livre ao conhecimento científico e tecnológico, por meio das discussões em torno da Ciência Aberta, cujo objetivo é contribuir para uma mudança real e duradoura no comportamento dos pesquisadores para garantir que o fazer científico se torne a norma, a partir de uma prática coletiva de fazer ciência, permitindo que outros pesquisadores possam colaborar e contribuir entre si (FOSTER, [202-], tradução nossa).

As discussões em torno do reúso de dados, como uma prática da Ciência aberta e como promotora do acesso livre ao conhecimento, são direcionadas por Thanos (2017) e Van de Sandt *et al.* (2019), além disso, são destacadas algumas características inerentes ao reúso de dados de pesquisa qualitativa norteadas por Chatfield (2020), Chauvette, Molzahn e Schick-Makaroff (2019), Childs *et al.* (2014) e Irwin e Winterton (2012).



A reutilização de dados de pesquisa ou a reusabilidade é entendida por Thanos (2017) como a facilidade de utilização de dados para pesquisa científica legitimada por uma ou mais comunidades de pesquisa (comunidades de consumidores) e que são produzidas por outras comunidades de pesquisa (comunidades de produtores), com o propósito para estudar um novo problema. No entanto, essa proposta de Thanos (2017), limita o reúso para um novo uso dos dados de pesquisa por pessoas que não realizaram a coleta e para realizar novas pesquisas.

Van de Sandt et al. (2019) amplia a abrangência do reúso de dados de pesquisa como o uso de qualquer recurso de pesquisa independente de quando é usado, a finalidade, as características dos dados e seu usuário, ou seja, nessa perspectiva o reúso pode ser caracterizado pelo uso dos dados pela pessoa o coletou para a mesma finalidade ou para outra.

Em relação ao reúso aos dados qualitativos, Irwin e Winterton (2012) pontuam que a pesquisa primária, ou seja, aquela para a qual a finalidade dos dados foi coletada de maneira inicial e a pesquisa secundária que seria aquela derivada do reúso de dados, têm posicionamentos diferentes em relação ao pesquisador. Para os autores, o pesquisador primário e o pesquisador secundário estão situados de forma diferente em relação ao dado disponível, sendo que os primeiros têm uma relação privilegiada com os dados que geraram, mas não necessariamente têm uma reivindicação privilegiada sobre os argumentos que podem ser feitos a partir desses dados, que é o posicionamento dos pesquisadores secundários. Portanto, a maneira como os dados serão usados como evidência no desenvolvimento, teste e construção de reivindicações de conhecimento é o que mais importa neste contexto.

Ainda no caso de dados de pesquisa qualitativos, alguns fatores podem limitar o reúso, são eles: a possibilidade de que os dados existentes não sejam “receptivos” à análise para um novo propósito; a desidentificação, ou seja, remoção de referências a pessoas, lugares e entidades, para proteger a identidade dos participantes pode deturpar os resultados, o risco de infringir as condições de consentimento da pesquisa, na qual os participantes se voluntariam para compartilhar suas experiências para uma questão de pesquisa específica, logo, a reutilização desses dados para uma questão de pesquisa diferente pode acarretar numa infração ética (Chatfield, 2020; Chauvette; Molzahn; Schick-Makaroff, 2019, Childs et al., 2014).

Embora esses desafios pareçam intransponíveis, a reutilização de dados de pesquisa tem muitas vantagens, o que inclui economia de tempo na amostragem, processamento e coleta de dados, possibilidade de revisão mais rápida pelos comitês de ética em pesquisa institucional, uma vez que a pesquisa já passou por uma primeira avaliação, no caso de dados qualitativos, o potencial acesso a dados fornecidos por participantes de difícil acesso, ou dados relacionados a tópicos controversos ou desconfortáveis podem reduzir a carga emocional sobre os participantes, a maximização do valor das contribuições dos participantes para a pesquisa, possibilita dar voz àqueles, cujos pensamentos e ideias eram menos valorizados ou ignorados devido a circunstâncias ou prioridades culturais, sociais ou políticas (Chatfield, 2020).

A curadoria de digital, de acordo com Araújo (2018), pode ser caracterizada como a prática e o estudo dos processos de seleção, preservação, manutenção, coleção e arquivamento de dados digitais, com a consequente criação de repositórios e/ou plataformas digitais participativas, tendo como preocupação a fragilidade dos dados, a possibilidade de um melhor uso e reúso. Nesta perspectiva os trabalhos de Bhattacharya et al. (2018); Borgman, Scharnhorst e Golshan (2019), Carroll et al. (2020), Danciu et al. (2014), Park e Park (2019), Pănescu e Manta (2018), Poole (2016), Sinaci et al. (2020) discutem temas correlacionados com a curadoria de dados.

A curadoria de dados de pesquisa exige uma visão a longo prazo, nem sempre os resultados são vistos de maneira imediata. Borgman, Scharnhorst e Golshan (2019) identificaram que muitas partes interessadas estão envolvidas, são os estudiosos e as equipes que produziram os dados, agências de financiamento que fornecem os recursos para conduzir pesquisas, universidades e outras instituições de pesquisa onde as investigações foram baseadas ou conduzidas, formuladores de políticas de pesquisa em organizações públicas e privadas, usuários atuais e potenciais desses dados, e as bibliotecas e arquivos que podem adquirir e gerenciar.

Além das partes interessadas, principalmente no que condiz ao compartilhamento e reúso de dados, Danciu et al. (2014) pontua que a construção e manutenção de dados em grande escala e orientados para a pesquisa, requer métodos e especialistas de vários domínios. Especialistas em arquitetura de dados e infraestrutura técnica, profissionais de informática com conhecimento conceitual e aplicado de processamento de linguagem natural, cientistas da computação com conhecimento e a compreensão dos riscos de desidentificação e

reidentificação, especialistas em privacidade e ética, especialistas em modelagem preditiva e aprendizado de máquinas são necessários para converter dados em informações acionáveis para alimentar protocolos de pesquisa ou melhorias no fluxo de trabalho.

Ainda em relação às partes envolvidas na curadoria, Carrol et al. (2020), identificou que a governança de dados indígenas depende tanto das comunidades, isto é, partes sobre os quais os dados foram coletados, quanto das comunidades de pesquisa que coletaram os dados. A relação entre essas partes é necessária para entender os conceitos operacionais e aplicá-los preventivamente em ecossistemas de dados e ciclos de vida.

Na pesquisa qualitativa, seu objetivo é desenvolver a compreensão dos fenômenos por meio da obtenção de opiniões e experiências de participantes humanos, logo os dados qualitativos são descritivos, não numéricos e geralmente não estruturados, por exemplo, uma transcrição de entrevista (Childs et al., 2014) Logo, o processo de curadoria que vise o reuso dos dados de pesquisa, além de envolver os sujeitos, sobre os quais os dados foram coletados, é preciso se atentar os processos de anonimização, definição de licenças de uso e reuso, inclusão da documentação, tais como protocolos de processamento e análise, ferramentas de coletas de dados, digitalização de registros em papel, além de assegurar, que todos esses arquivos estejam vinculados ao conjunto de dados (Carrol et al., 2020; Childs et al., 2014).

Na pesquisa quantitativa, os dados são mais homogêneos e a maior parte do processamento pode ser automatizada, consumindo menos recursos no processamento, no entanto, fornecer informações contextuais suficientes para permitir a reutilização, também é um desafio para este tipo de pesquisa (Childs et al., 2014). Portanto, em ambas as pesquisas, a descrição precisa dos metadados e a inclusão e vinculação de arquivos relacionados aos dados, são imprescindíveis para um potencial reuso.

Bhattacharya et al. (2018), ao relatar as práticas de curadoria adotadas no ImmPort, um repositório de dados sobre Imunologia, perceberam que a documentação de metadados de proveniência de dados pode desempenhar um papel crucial na superação dos desafios que envolvem a descrição e a precisão dos dados. Os autores também destacam o trabalho com vários consórcios para facilitar a curadoria e o compartilhamento de dados entre os laboratórios participantes (Bhattacharya et al., 2018).

Em relação aos metadados. Park e Park (2019) destacam que cada contexto exige uma descrição de dados suficiente para permitir o desenvolvimento de novas pesquisas científicas por reutilizadores de dados e que é importante estar ciente que as diferentes disciplinas precisam incluir uma descrição detalhada com informações precisas e relevantes de proveniência que atendem a sua comunidade de maneira específica.

Em relação a segurança dos dados nos os repositórios, Pănescu e Manta (2018), sugerem o uso do uso do blockchain, como forma de facilitar o registro dos dados, possibilitando que todas as entidades envolvidas rastreiem todo e qualquer acesso e reutilização de um conjunto de dados compartilhado.

Sinaci et al. (2020) desenvolveram um fluxo de trabalho, norteado pelos princípios FAIR, acrônimo para *Findable* (Localizáveis), *Accessible* (Acessíveis), *Interoperable* (Interoperáveis) e *Reusable* (Reutilizáveis), para ajudar a otimizar a reutilização dos conjuntos de dados de pesquisa em saúde, levando em consideração funcionalidades específicas para curadoria de dados, validação de dados, desidentificação/pseudonimização de dados e controle de versão de dados.

Por fim, em relação ao planejamento de um projeto de curadoria, Poole (2016) destaca seis questões para serem consideradas, são elas: em que ponto(s) do ciclo de vida a contribuição e o apoio dos curadores digitais podem ser mais valiosos? O que um exame mais aprofundado das práticas dos pesquisadores pode sugerir sobre as necessidades comuns de curadoria digital em comunidades de prática, domínios, subdisciplinas, disciplinas, instituições e nações? Como os arquivistas podem expandir melhor seu papel incipiente no trabalho de curadoria digital? Quais estratégias podem ser usadas para garantir o financiamento? Que estratégias de divulgação e conscientização podem ser bem-sucedidas? Por fim, quais são as métricas mais adequadas para medir o sucesso no trabalho de curadoria digital?

Ao se verificar os trabalhos sobre reuso de dados pode-se elencar algumas oportunidades de pesquisa voltadas a Produção e Comunicação científica na Ciência da Informação, por exemplo: como minimizar os fatores que limitam o reuso de dados qualitativos? Quais os papéis práticos dos bibliotecários na curadoria de dados de pesquisa? Na ótica dos pesquisadores que reusam dados, tem-se alcançados novos conhecimentos a partir dos dados existentes? Quais as ferramentas da Ciência da Informação têm contribuído para a curadoria dos dados?

### 3.2 Estudos sobre os sujeitos

Os estudos sobre os sujeitos são considerados por Araújo (2018) como uma evolução do campo dos chamados estudos de usuários, eles passaram a privilegiar não mais as questões cognitivas (tipos de lacuna de informação, tipos de informação a preencher essas lacunas), mas as compreensões dessas questões, voltando-se para enfoques mais interpretativos das práticas informacionais dos usuários num contexto coletivo e influenciados pelas mudanças tecnológicas. Sendo as práticas o “[...] movimento por meio do qual os indivíduos agem no mundo, conformados pela cultura, e ao mesmo tempo constituem essa cultura que os influencia e a realidade em que atuam.” (Araújo, 2018, p. 55).

Em relação às práticas informacionais dos sujeitos, nos estudos analisados elas se centram nas práticas relacionados ao compartilhamento e reúso de dados conforme Curty *et al.* (2017), Fecher, Friesike e Hebing (2015), Federer *et al.* (2015), Imker *et al.* (2021), Joo e Kim (2017), Joo, Kim e Kim (2017), Kim e Nah (2018), Perrier, Blondal e MacDonald (2020), Tenopir *et al.* (2018), Yoon (2017), Yoon e Lee (2019) e Zuiderwijk e Spiers (2019), em geral práticas estas voltadas para disciplinas específicas.

Curty *et al.* (2017) identificaram que a reutilização e o compartilhamento de dados não estão vinculados, sendo apenas moderadamente correlacionado, isto é, os cientistas possuem comportamentos diferentes para compartilhamento e reutilização de dados. Nas questões que envolvem o reúso, Curty *et al.* (2017) verificaram que embora os pesquisadores estejam cientes das armadilhas potenciais da reutilização de dados, eles aparentemente sentem que podem superá-las em sua própria prática. Joo e Kim (2017) também, identificaram que, no caso das Engenharias, o esforço percebido envolvido na reutilização de dados não foi um preditor das atitudes dos pesquisadores em relação à reutilização de dados ou na intenção de reutilização de dados. Isso pode estar relacionado às características dos dados de engenharia que tendem a ser mais sistemáticos e estruturados, comparados aos dados das Ciências Sociais (Joo; Kim, 2017).

No entanto, Imker *et al.* (2021) verificaram que o conceito de compartilhamento e reutilização de dados parece abstrato e distante dos fluxos e práticas de trabalho diários dos pesquisadores das áreas que pertencem ao grupo de disciplinas da STEM (Science, Technology, Engineering, and Mathematics). Sendo assim, pode-se inferir que os campos diferem no tipo de dados usados e na facilidade com que esses dados podem ser reutilizados (Curty *et al.*, 2017).

Para Yoon (2017) identificou que os reutilizadores de dados não dependem apenas de instituições culturais e/ou acadêmicas para adquirir materiais como dados, eles usam dos relacionamentos interpessoais para obter as formas brutas de informações e dados que não são processados, gerenciados ou curados por profissionais.

Em relação a confiança na reutilização, Yoon (2017) identificou, nos pesquisadores das áreas de Saúde Pública e Serviço Social, fatores que influenciam na confiança no reúso de dados de pesquisa. São eles: as avaliações dos dados ajudaram os participantes a desenvolverem confiança ao buscar confirmação de fontes já confiáveis; o fato de alguém ter publicado um artigo usando os dados também reforçou a confiança; o número de vezes que os dados foram usados foi outro indicador; o fato dos estudos originais terem sido financiados por organizações governamentais ou sem fins lucrativos pode ser uma indicação de avaliação por pares e reconhecimento das comunidades, pois sugere o reconhecimento de pesquisas de destaque; embora a reputação geral dos investigadores originais possa ser outra maneira de verificar diretamente sua competência, os participantes também buscaram evidências de competência por conta própria verificando a participação dos investigadores originais em uma Comunidade de Prática (CoP) e a documentação foi comumente discutida pelos participantes como um fator que aumentou (ou diminuiu) a confiança (Yoon, 2017).

Joo, Kim e Kim (2017) identificaram que apesar dos benefícios amplamente percebidos da reutilização de dados, existem preocupações significativas sobre os dados, dentre as quais destacam-se: questões legais relacionadas à privacidade do assunto, barreiras culturais, desafios técnicos, competição científica, falta de incentivos na academia para compartilhar dados e a qualidade do uso de dados secundários.

Em relação ao compartilhamento de dados pesquisa Zuiderwijk e Spiers (2019) identificaram fatores positivos que favorecem o compartilhamento e o reúso em astrofísica: os dados brutos raramente são coletados diretamente pelos pesquisadores, os recursos limitados incentivam a colaboração, os dados públicos externos são compartilhados automaticamente, a existência de uma cultura de compartilhamento de dados de apoio, a falta de preocupações com a ética e o potencial comercial dos dados e uma experiência positiva com a reutilização de dados abertos.

Para Tenopir *et al.* (2018) os geofísicos, em geral, reconhecem que o compartilhamento de dados de pesquisa pode ter um impacto positivo no progresso científico em relação à economia de tempo e eficiência da pesquisa, mas quando se trata de compartilhar seus próprios dados de pesquisa, os cientistas têm preocupações, referentes a má utilização e má interpretação, preocupação também identificada por Perrier, Blondal e MacDonald (2020).

Além disso, os pesquisadores têm preocupação com questões relacionadas à qualidade dos dados, uso indevido de dados, proteção de dados, falta de tempo e habilidades e deficiências de infraestrutura e suporte, falta de tempo para preparar a documentação para disponibilizar os dados (PERRIER; BLONDAL; MACDONALD, 2020). Em nível institucional, a falta de procedimentos, políticas, orientações, incentivos e benefícios contribuem para os desafios no compartilhamento de dados. (Perrier; Blondal; Macdonald, 2020).

Especificamente, no caso da astrofísica, dois fatores que desmotivam os pesquisadores a compartilharem e reutilizarem abertamente dados de pesquisa aberta, primeiro, é o enorme volume de alguns conjuntos de dados, alguns conjuntos de dados não podem ser facilmente armazenados e baixados e a segunda é a falta de condições facilitadoras para a descrição dos dados, isto é, compartilhar dados de maneira útil para outros pesquisadores, pode ser difícil, pois é necessário descrever metadados e ferramentas usadas (Zuiderwijk; Spiers, 2019).

Embora o incentivo ao compartilhamento seja importante, mudanças regulatórias e políticas podem ser necessárias para remover barreiras, mitigar consequências negativas não intencionais e possibilitar que o maior número possível de dados seja compartilhado (Fecher; Friesike; Hebing, 2015; Federer *et al.*, 2015). As medidas de política estratégica podem, portanto, ir em duas direções: primeiro, podem fornecer incentivos para o compartilhamento de dados e, segundo impedir que os pesquisadores não compartilhem. Os possíveis incentivos podem incluir o reconhecimento formal adequado na forma de citação de dados, autoria e perspectivas de carreira. (Fecher; Friesike; Hebing, 2015; Kim; Nah, 2018).

Mecanismos como revisão por pares, acordos de compartilhamento de dados que especificam claramente como um conjunto de dados podem ser usados e a aprovação ou isenção de projetos por conselhos de revisão institucionais podem ajudar a garantir que os dados sejam reutilizados com respeito aos sujeitos e aos pesquisadores originais que coletaram os coletaram (Federer *et al.*, 2015).

Em relação aos incentivos que envolvem a carreira, sugestões de incentivos incluem a oferta de bolsas de pesquisa que se concentram especificamente na reutilização de dados gerados a partir de bolsas anteriores, a criação de sistemas que garantem que o crédito seja concedido aos geradores de dados (Perrier; Blondal; Macdonald, 2020). Além disso, as formas de financiamento reembolso, ajudam a aumentar o esforço individual para disponibilizar os dados (Fecher; Friesike; Hebing, 2015).

Uma outra maneira de incentivar o compartilhamento e reuso de dados é educar os pesquisadores, através de workshops em grandes conferências ou palestras (Joo; Kim, 2017). Um número crescente de bibliotecas e repositórios oferece sessões educacionais para ensinar pesquisadores práticas e planejamento de gerenciamento de dados, mas muitos deles se concentram no planejamento de gerenciamento de dados como parte dos requisitos dos financiadores – uma interação única em vez de esforços contínuos e compartilhados de gerenciamento (Yoon; Lee, 2019).

Por fim, pode-se verificar que o compartilhamento de dados e o reuso na academia é um esforço multidimensional que inclui um conjunto diversificado de *stakeholders*, entidades e interesses individuais (Fecher; Friesike; Hebing, 2015).

Ainda nas teorias contemporâneas dos sujeitos, destaca-se a **mediação e a apropriação da informação**, como uma ação dialógica entre os serviços e sistemas de informação e os usuários, de modo que esses deixam de ser apenas artifícios de transferência de conteúdos informacionais para se constituírem em verdadeiros dispositivos produtores de sentidos, tendo os usuários como sujeitos ativos do processo (Araújo, 2018). Essas teorias se voltaram para estudos sobre o comportamento de busca dos usuários por dados de pesquisa, sobre as quais Gregory *et al.* (2020) e Krämer *et al.* (2021) identificaram algumas características.

Gregory *et al.* (2020) identificou que a busca por dados de pesquisa são práticas sociais que perpassam pelo contato entre os utilizadores de dados e os autores (produtores dos dados), onde os as interações sociais são realizadas para localizar, avaliar e desenvolver a confiança nos dados.

Pensando nessa interlocução entre usuários, autores e mediadores, Gregory *et al.* (2020) ressalta que não basta pensar nos usuários de dados como pesquisadores em uma disciplina com práticas fixas, ou seja, comunidades, interesses e práticas de pesquisa são dinâmicos, por vezes influenciados pelo desenvolvimento de novas técnicas de pesquisa e análise (por exemplo, ciência dos dados). Nem todos os que procuram dados são pesquisadores, ou seja, bibliotecários, gestores de literatura e estudantes também procuram dados, usando diferentes estratégias e métodos de avaliação. Indivíduos fora do meio acadêmico, incluindo pessoas que trabalham na indústria e cidadãos preocupados, estão interessados em encontrar e usar dados também. Além desses, outro possível "usuário" também está surgindo: a máquina (Gregory *et al.*, 2020).

Em relação aos dados de pesquisa, eles nem sempre são dados de pesquisa, isto é, metadados, textos, logs de servidor, especificações de dispositivos e mensagens de mídia social, todos são usados para fins de primeiro plano e de fundo na pesquisa, mas não se enquadram no que tradicionalmente é considerado como "dados de pesquisa" (Gregory *et al.*, 2020). Por fim, dentre as características dos dados de pesquisa, Gregory *et al.* (2020) afirmam que os dados não são estáticos, de interesse apenas para a comunidade onde são produzidos. Eles se movem entre situações e comunidades, existindo em diferentes contextos e sendo adaptados a diferentes propósitos. Os caminhos que os dados percorrem também dependem muito do contexto - o contexto da sua criação e os contextos de descoberta e reutilização.

Krämer *et al.* (2021), ao investigar o comportamento de busca de dados de cientistas sociais, verificaram que a pesquisa de literatura é uma parte importante da pesquisa de conjuntos de dados, as ferramentas para a busca são desconhecidas, pois a maioria usou ferramentas populares como por exemplo, Google Scholar. A acessibilidade aos conjuntos de dados foi considerada limitada, procedimentos de acesso, como registro ou autorização, causaram interrupções no fluxo de trabalho de pesquisa do conjunto de dados. Em relação a avaliação dos resultados, ela se tornou muito complexa, pois para encontrar todas as informações relevantes, os participantes tiveram que inspecionar atentamente a documentação, incluindo metodologia, amostragem e definição precisa das variáveis e em muitas vezes essas informações não estavam centralizadas ou não estavam disponíveis (Krämer *et al.*, 2021).

Muitos participantes desejavam mais apoio durante o processo de busca, por exemplo, ligação consistente entre literatura e conjuntos de dados e mais explicações para as variáveis. A alfabetização de pesquisa de conjuntos de dados foi considerada baixa. Embora a pesquisa de literatura seja frequentemente ensinada nas universidades, cursos semelhantes para manipulação de dados são muito mais raros. Ainda assim, há uma falta de base comum ou práticas recomendadas para a pesquisa de conjuntos de dados (Krämer *et al.*, 2021).

A partir do estudo do Krämer *et al.* (2021), pode-se verificar que os pesquisadores carecem de apoio na mediação das informações sobre os dados, suporte por parte das bibliotecas, com isso vislumbra-se as oportunidades de pesquisa para verificar qual o suporte da biblioteca para busca e compreensão dos dados de pesquisa, quais serviços ela pode oferecer, identificar a competência dos bibliotecários na busca por dados e no trabalho com eles.

Estudos sobre a competência crítica da informação não foram identificados, considerando que ela busca inserir o sujeito nos contextos históricos, políticos e sociais nos quais se dão suas relações e suas habilidades para lidar com informação por meio de um juízo de valor (Araújo, 2018), abre-se uma importante lacuna de pesquisa, relacionar o reúso de dados de pesquisa com a competência crítica.

Ainda em relação as oportunidades, destacam-se: como desenvolver a *data literacy* nos pesquisadores, usuários e bibliotecários? Quais serviços podem ser desenvolvidos em torno do reúso de dados nas bibliotecas e nos arquivos? Como avaliar a confiança nos dados de pesquisa?

### 3.3 Estudos métricos da informação

De acordo com Araújo (2018), os estudos métricos têm buscado inserir os resultados dos estudos quantitativos em quadros explicativos mais amplos, em busca de entendimentos mais globais dos fenômenos estudados considerando principalmente o caráter coletivo de construção da ciência, os estudos de He e Nahar (2016), Park e Wolfram (2017) e Womack (2015) e vão ao encontro desta perspectiva.

Buscando entender o reúso de dados como um fenômeno do fazer da Ciência Aberta, a partir da métrica de citação, He e Nahar (2016) buscaram identificar como e por que os pesquisadores reutilizaram dados

compartilhados por outros, em diferentes campos de pesquisa. Womack (2015) buscou identificar práticas de citação de dados, compartilhamento de dados, reprodutibilidade, publicação de dados em revistas científicas das disciplinas de Biologia, Química, Matemática e Física e Park e Wolfram (2017) buscaram examinar as características de compartilhamento de dados e reutilização de dados em Genética e Hereditariedade.

Neste íterim, os estudos de citação de dados, são entendidos como um link direto ou referência a uma fonte de dados (Womack, 2015), no entanto, ainda não se consolidaram como uma forma de incentivar o reuso. Park e Wolfram (2017) destacam que a ausência de padrões universalmente aceitos para citação de dados pode ser um problema para incentivar e avaliar o reuso dos dados de pesquisa.

Os referidos autores perceberam ainda, que as práticas de reuso de dados, muitas vezes não foram reconhecidas nas listas de referência, mas, sim ocultas na representação dos dados, sendo que muitas vezes, elas apareceram seção de metodologia, como coleta de dados ou análise de dados, com os termos indicadores e/ou frases para reutilização de dados incluem "doados de/por", "obtidos de", "comprados de", "doado de" (Park; Wolfram, 2017). Na estrutura das referências, quando estavam reconhecidos, Womack (2015) percebeu que o Digital Object Identifier (DOIs) era citado separado dos dados, os links diretos para os dados eram raros, faltava explicitamente a identificação de autoria, não havia qualquer precisão sobre a localização dos dados, data de acesso, proveniência dos dados utilizados, ou qualquer outro detalhe.

Park e Wolfram (2017) afirmam que a falta de padrões de citação dificulta a identificação da recompensa e crédito para os autores dos dados, a implementação de padrões de citação de dados para trabalhos acadêmicos é, então, um aspecto importante para garantir que os dados publicados relevantes sejam citados e que a citação seja benéfica para quem os publica. Logo, a citação de dados pode vir a representar uma nova forma de crédito que pesquisadores que publicam seus dados recebem quando são requisitados por grandes agências de fomento e/ou políticas de periódicos influentes para compartilhar seus dados (Park; Wolfram, 2017).

De maneira prática, He e Nahar (2016) ao examinar as práticas de citação no repositório Dryad Digital Repository (DDR) perceberam que a áreas que mais reutilizam dados são as Ciências Agrícolas e Biológicas, Ciências Ambientais e Medicina. Em geral, essas citações são caracterizadas como autocitações para demonstrar e apoiar seus argumentos ou para aumentar a confiança do leitor na confiabilidade da pesquisa, além de cumprir com a obrigação de depositar os dados de pesquisa.

Como oportunidades de pesquisa, vislumbra-se aquelas voltadas para a identificação dos estudos sobre as práticas citação, como referenciar os dados de pesquisa? Como a ciência vem se construindo de maneira coletiva? Como identificar as relações interdisciplinares na Ciência Aberta?

## 4 Conclusões

O objetivo deste artigo foi correlacionar o reuso de dados de pesquisa na agenda das teorias contemporâneas da Ciência da Informação, foi verificado que os estudos se concentraram nas teorias voltadas para a Produção e Comunicação Científica, seguida daquelas que envolvem os Estudos sobre os Sujeitos e Estudos Métricos da Informação.

Na Produção e Comunicação Científica, os estudos se voltaram para entendimento dos aspectos que norteiam o conceito de reuso de dados de pesquisa, para a identificação dos fatores que limitam o reuso em pesquisas qualitativas e na identificação dos aspectos relacionados a curadoria de dados de pesquisa

Nos estudos sobre os Sujeitos, os artigos se voltaram para identificar quais as barreiras para o compartilhamento e reuso de dados, os incentivos que podem estimular os pesquisadores a incentivar e reusar dados de pesquisa e na identificação das características envolvidas na busca de dados de pesquisa.

Nos estudos Métricos da Informação foi possível identificar a métrica de citação como a mais utilizada para identificar o reuso dos dados e as dificuldades para mensurá-lo.

Algumas oportunidades de pesquisa nessas áreas foram identificadas e observou-se que estudos nas subáreas de Representação e Organização da Informação, Gestão da Informação, Economia Política da Informação e estudos sobre Memória, Patrimônio e Documento ainda carecem de posicionamento na agenda de teorias contemporâneas em relação ao reuso de dados.

A limitação desse estudo, se concentra no posicionamento voltado às teorias contemporâneas propostas por Carlos Alberto Araújo, em 2018, portanto, este estudo foca na perspectiva brasileira das lacunas de estudos na Ciência da Informação.

Este artigo mostrou que o paradigma da Ciência da Informação vem sendo deslocado para um eixo de estudos sobre dados de pesquisa. Ainda não trata de uma substituição no paradigma vigente, mas mostra que a Ciência da Dados pode trabalhar de maneira conjunta com os paradigmas existentes. O desafio para a Ciência da Informação neste contexto consiste relacionar os fatores técnicos tratados no reúso de dados de pesquisa com uma Ciência da Informação mais humana, voltada para a participação dos cidadãos e para a superação de problemas sociais. Portanto, neste contexto, desafios também são lançados às bibliotecas e aos bibliotecários, uma vez que esses precisam desenvolver sua *data literacy*, para depois fornecer produtos e serviços que opóiem as competências para curadoria e reúso dos dados pelos pesquisadores, além de potenciais reúsos pela comunidade do entorno, não pertecente diretamente ao ecossistema da pesquisa.

## Referências

- ARAÚJO, C. A. A. O que é ciência da informação. Belo Horizonte: KMA, 2018.
- BHATTACHARYA, S. et al. ImmPort, toward repurposing of open access immunological assay data for translational and clinical research. *Scientific Data*, [s.l.], v.5, p. 1-9, 2018.
- BORGMAN, C. L.; SCHARNHORST, A.; GOLSHAN, M. S. Digital data archives as knowledge infrastructures: Mediating data sharing and reuse. *Journal of the Association for Information Science and Technology*, [s.l.], v. 70, n. 8, p. 888-904, 2019.
- CAPURRO, R. Epistemología y ciencia de la información. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO. 5., 2003, Belo Horizonte. Anais... Belo Horizonte: UFMG, 2003. Disponível em: [http://www.capurro.de/enancib\\_p.htm](http://www.capurro.de/enancib_p.htm). Acesso em: 27 set. 2022.
- CARROLL, S.R. et al. The CARE principles for indigenous data governance. *Data Science Journal*, [s.l.], v. 19, n. 43, p. 1-12, 2020.
- CHATFIELD, S. L. Recommendations for secondary analysis of qualitative data. *The Qualitative Report*, [s.l.], v. 25, n. 3, p. 833-842, 2020
- CHAUVETTE, A.; SCHICK-MAKAROFF, K.; MOLZAHN, A. E. Open data in qualitative research. *International Journal of Qualitative Methods*, [s.l.], v. 18, p. 1-6, 2019.
- CHILDS, S. et al. Opening research data: issues and opportunities. *Records Management Journal*, [s.l.], v. 24, n. 2, p. 142-162, 2014.
- CURTY, R.G. et al. Attitudes and norms affecting scientists' data reuse. *PLoS ONE*, [s.l.], v. 12, n. 12, 2017.
- DANCIU, I. et al. Secondary use of clinical data: The Vanderbilt approach. *Journal of Biomedical Informatics*, [s.l.], v.52, p. 28-35, 2014.
- ENSSLIN, L.; ENSSLIN, S. R.; e PINTO, H. M. Processo de investigação e análise bibliométrica: avaliação da qualidade dos serviços bancários. *Revista de Administração Contemporânea*, [s.l.], v. 17, n. 3, p. 325-349, 2013.
- FECHER, B.; FRIESIKE, S.; HEBING, M. What drives academic data sharing? *PLoS ONE*, [s.l.], v. 10, n. 2, 2015.
- FEDERER, L. M. et al. Biomedical data sharing and reuse: attitudes and practices of clinical and scientific research staff. *PLoS ONE*, [s.l.], v.10, n. 6, 2015.
- FOSTER. Open Science Definition. [s.l.], [202-]. Disponível em: <https://www.fosteropenscience.eu/taxonomy/term/100>. Acesso em: 05 set. 2021.
- GREGORY, K. M. A dataset describing data discovery and reuse practices in research. *Scientific Data*, [s.l.], v. 7, n. 232, 2020.
- GREGORY, K. M. et al. Understanding data search as a socio-technical practice. *Journal of Information Science*, [s.l.], v. 46, n. 4, p. 459-475, 2020
- HE, L.; NAHAR, V. Reuse of scientific data in academic publications: an investigation of Dryad Digital Repository. *Aslib Journal of Information Management*, [s.l.], v. 68, n. 4 p1-23, 2016.

- IMKER, H. J. et al. An examination of data reuse practices within highly cited articles of faculty at a research university. *The Journal of Academic Librarianship*, [s.l.], v. 47, p.1-11, 2021.
- IRWIN, S.; WINTERTON, M. Qualitative secondary analysis and social explanation. *Sociological research online*, [s.l.], v. 17, n. 2, p. 1-12, 2012.
- JOO, S.; KIM, S.; KIM, Y. An exploratory study of health scientists' data reuse behaviors: examining attitudinal, social, and resource factors. *Aslib Journal of Information Management*, [s.l.], v. 69, n. 4, p. 1-29, 2017.
- JOO, Y. K.; KIM, Y. Engineering researchers' data reuse behaviours: a structural equation modelling approach. *The Electronic Library*, [s.l.], v. 35, n. 6, 2017.
- KIM, Y.; NAH, S. Internet Researchers' data sharing behaviors: an integration of data reuse experience, attitudinal beliefs, social norms, and resource factors. *Online Information Review*, [s.l.], v. 42, n. 1, p. 1-31, 2018.
- KRÄMER, T. et al. Data-seeking behaviour in the Social Sciences. *International Journal on Digital Libraries*, [s.l.], v. 22, p. 175-195, 2021.
- LACERDA, R. T. O.; ENSSLIN, L. ENSSLIN, R. S. Uma análise bibliométrica da literatura sobre estratégia e avaliação de desempenho. *Gestão de Produção, São Carlos*, v. 19, n. 1, p. 59-78, 2012.
- PĂNESCU, A.; MANTA, V. Smart contracts for research data rights management over the Ethereum blockchain network. *Science & Technology Libraries*, [s.l.], p. 1-12, 2018
- PARK, H.; WOLFRAM, D. An examination of research data sharing and re-use: implications for data citation practice. *Scientometrics*, [s.l.], v. 111, p. 443-461, 2017.
- PARK, M. S.; PARK, H. An examination of metadata practices for research data reuse: Characteristics and predictive probability of metadata elements. *Malaysian Journal of Library & Information Science*, [s.l.], v. 24, n. 3, p. 61-75, 2019.
- PERRIER, L.; BLONDAL, E.; MACDONALD, H. The views, perspectives, and experiences of academic researchers with data sharing and reuse: a meta-synthesis. *PLoS ONE*, [s.l.], v. 15, n. 2, 2020.
- POOLE, A. H. The conceptual landscape of digital curation. *Journal of Documentation*, [s.l.], v. 72, n. 5, p. 961-986, 2016.
- SINACI, A. A. et al. From raw data to FAIR Data: the FAIRification workflow for health research. *Methods of Information in Medicine*, [s.l.], v. 59, p. e21–e32. 2020.
- TENOPIR, C. et al. Research data sharing: practices and attitudes of geophysicists. *Earth and Space Science*, [s.l.], v. 5, p. 891-902, 2018.
- THANOS, C. Research data reusability: conceptual foundations, barriers and enabling Technologies. *Publications*, [s.l.], v. 5, n. 2, p. 1-19, 2017.
- VAN DE SANDT, S. et al. The definition of reuse. *Data Science Journal*, [s.l.], v. 18, n. 22, p. 1-19, 2019.
- WOMACK, R. P. Research data in core journals in Biology, Chemistry, Mathematics, and Physics. *PLoS ONE*, [s.l.], v. 10, n. 12, 2015.
- YOON, A. Data reusers' trust development. *Journal of the Association for Information Science and Technology*, [s.l.], v. 68, n. 4, p. 946-956, 2017.
- YOON, A.; LEE, Y. Y. Factors of trust in data reuse. *Online Information Review*, [s.l.], v. 43, n. 7, p. 1245-1262, 2019.
- ZUIDERWIJKA, A.; SPIERS, H. Sharing and re-using open data: a case study of motivations in astrophysics. *International Journal of Information Management*, [s.l.], v. 49, p. 228-241, 2019.



## Dados dos autores

### Crislaine Zurilda Silveira

Doutoranda no Programa de Pós-Graduação em Ciência da Informação pela UFSC. Mestre em Gestão de Unidades de Informação pela UDESC (2017). Especialista em Docência no Ensino Superior pela PUC RS (2019) e Especialista em Gestão Pública pela Unisul (2015). Bacharel em Biblioteconomia com habilitação em Gestão da Informação pela UDESC (2011). Atualmente é Bibliotecária da Universidade Federal de Santa Catarina. Tem interesse nas temáticas de pesquisa Ciência da Informação ciência aberta, gestão de dados de pesquisa, gestão de bibliotecas universitárias e competência em informação.

<https://orcid.org/0000-0003-3081-9968>

[crislaine.silveira@ufsc.br](mailto:crislaine.silveira@ufsc.br)

### Thiago Magela Rodrigues Días

Docente do Programa de Pós-graduação em Ciência da Informação da Universidade Federal de Santa Catarina (UFSC) e do Programa de Pós-graduação em Modelagem Matemática e Computacional do CEFET-MG, Florianópolis, Santa Catarina, Brasil.

<https://orcid.org/0000-0001-5057-9936>

[thiagomagela@cefetmg.br](mailto:thiagomagela@cefetmg.br)

**Received:** 2023-03-17

**Accepted:** 2024-01-19



This work is licensed under a Creative Commons Attribution 4.0 United States License.



This journal is published by the [University Library System](#) of the [University of Pittsburgh](#) as part of its [D-Scribe Digital Publishing Program](#) and is cosponsored by the [University of Pittsburgh Press](#).