Os dados científicos e os dados de pesquisa na Ciência da Informação Brasileira: uma análise a partir da BRAPCI

Ananda Fernanda de Jesus[a], *William Pires de Castro*[b] *e José Eduardo Santarém Segundo*[c]

1. Introdução

Como resultado dos avanços científicos e tecnológicos no início do século XX, ocorreu uma revolução que influenciou o desenvolvimento de novos suportes informacionais e o aumento da produção tecnocientífica com o estabelecimento das publicações seriadas. (SARACEVIC, 2009).

Esse aumento alterou a relação entre pesquisadores e informações pertinentes para o desenvolvimento de suas pesquisas, tornando inviável para muitos campos do conhecimento a obtenção e tratamento de todo material publicado a respeito das temáticas de interesse.

Nesse cenário, agravado com o estabelecimento do ambiente Web, os dados adquirem um papel de destaque no desenvolvimento das pesquisas, seja influenciando nos processos

a Doutoranda em Ciência da Informação pela Universidade Estadual Paulista – UNESP. E-mail: af.jesus@unesp.br.

b Graduando em Ciência da Computação pelo Centro Universitário Eurípedes de Marília – UNIVEM. E-mail: wpdcastro@gmail.com.

c Doutor em Ciência da Informação pela Universidade Estadual Paulista – UNESP. Professor na Universidade de São Paulo – USP. E-mail: santarem@usp.br. ORCID: https://orcid.org/0000-0003-3360-7872.

de recuperação e seleção de referências, seja atuando como insumo no desenvolvimento das pesquisas. Com base no cenário apresentado, questiona-se: Como os dados científicos e os dados de pesquisa têm sido abordados nos artigos brasileiros da ciência da informação?

Neste sentido, objetiva-se discutir como a comunidade brasileira da ciência da informação tem abordado os dados científicos e os dados de pesquisa, buscando compreender a relação entre os dois conceitos, definições conceituais, identificando as principais temáticas abordadas e as lacunas de pesquisa.

Como método optou-se por um mapeamento sistemático dos artigos nacionais publicados a respeito da temática. Na próxima seção são apresentados os procedimentos metodológicos adotados para o desenvolvimento da pesquisa.

2. Procedimentos metodológicos

Caracterizada como exploratória, descritiva e quali-quantitativa, essa pesquisa adotou como método de coleta o Mapeamento Sistemático da Literatura (MSL). Segundo Kitchenham et al. (2010), os MSL são realizados com uma abordagem ampla, sendo o principal objetivo classificar os estudos de uma temática específica em categorias bem definidas, que permitam um olhar estruturado para a temática em busca de assuntos comuns e assuntos pouco explorados.

Para orientar as buscas, a seleção dos artigos e a análise dos resultados foi elaborado o seguinte protocolo de busca ilustrado no Quadro 1.

Quadro 1 – Protocolo de busca do mapeamento sistemático

Elemento	**Protocolo de busca**
Bases de dados	BRAPCI
Idiomas	Português
Critérios de Inclusão	(I) Discute dados científicos
Critérios de exclusão	(E) Não está em português; (E) Não está no formato estabelecido (artigo); "(E) Apenas mencionam a temática de interesse; (E) Apenas mencionam a temática de interesse (E) Não é possível ter acesso ao artigo completo (E) Aborda o uso de dados para validação de resultados
Categorias de análise	(1) Enfoque do artigo (2) Definições de dados científicos (3) Definição de dados de pesquisa (4) Lacunas de pesquisa
Data da coleta	08/05/2021
Forma de análise dos resultados	Os documentos foram selecionados de acordo com os critérios estabelecidos, sendo divididos os resultados entre os autores responsáveis pela coleta dos dados, em caso de dúvida quanto a seleção do artigo esse foi discutido em conjunto entre os autores. Acoleta nos documentos aceitos foi realizada seguindo as categorias de análise apresentadas. Foi elaborado um quadro resumo contendo autor, título, ano da publicação, e os resultados de cada categoria. Os resultados foram categorizados, sendo identificados padrões de temáticas de nos documentos, levantadas as lacunas e agrupadas as definições.

Fonte: Autores.

Para a realização do mapeamento sistemático seguiu-se as seguintes etapas: (i) Planejamento: Preenchimento do protocolo de busca que orientará a pesquisa; (ii) Execução: Busca nas bases de dados e seleção dos documentos; e (iii) Sumarização: agrupamento dos documentos por semelhança, criação de

categorias para classificação dos resultados, sistematização das informações de interesse em imagens e quadros.

Após a seleção os documentos foram classificados em categorias temáticas. Essas categorias foram estabelecidas a posteriori, levando em consideração principalmente os objetivos e os problemas de pesquisa apresentados pelos autores dos documentos, também foram checados os resultados obtidos como forma de verificar se os objetivos eram condizentes com os resultados apresentados, visando assim a inclusão dos documentos em categorias representativas. A nomenclatura aplicada na criação das categorias segue terminologia adotada pelos autores. Após estabelecimento das categorias, buscou-se realizar o agrupamento de categorias semelhantes, sempre levando em consideração a nomenclatura utilizada pelos próprios autores.

3. Resultados e discussões

As buscas na BRAPCI retornaram 392 documentos, sendo aceitos 51 para a análise. Ao observar os documentos rejeitados é possível identificar os fatores que levam ao alto número de rejeição. Os documentos rejeitados com base no critério "(E) aborda o uso de dados para validação de resultados" mencionam o uso de bases de dados ou conjuntos de dados científicos que atuaram como insumo para o desenvolvimento da pesquisa, como a identificação de bases de dados úteis a determinadas áreas, ou o levantamento de um conjunto de dados para solucionar uma questão de pesquisa. Esses documentos foram

recusados pois não abordaram ou discutiram o conceito ou a importância dos dados científicos ou dos dados de pesquisa, a sua necessidade de registro e possibilidade de reutilização, sendo esses dados utilizados de maneira mais operacional.

Já os documentos classificados no critério "(E) apenas mencionam a temática de interesse" reúnem artigos que citaram os dados científicos como um entre outros temas de interesse da CI, ou de alguma forma citaram os dados científicos para fins de exemplo ou contextualização, mas não se aprofundam na discussão da temática. A partir da análise dos documentos aceitos foram estabelecidas 15 categorias que refletem como o tema de dados pesquisa tem sido abordado nos artigos científicos da ciência da informação no Brasil. O quadro 2 apresenta essas categorias e o número de documentos que discutiram a temática, podendo um documento ser enquadrado em mais de uma categoria.

Quadro 2 – Categorias temáticas

Categoria	Quantidade de artigos	Definição
Gestão de dados de pesquisa	19	Documentos que buscam levantar as etapas, selecionar e aplicar ciclos de vida de dado de pesquisa, discutir aspectos de produção, compartilhamento e reutilização, de dados de pesquisa. Incluí ainda artigos que discutem os impactos no avanço da ciência, governança de dados, relação entre gestão de dados e fundamentação teórica da ciência da informação.

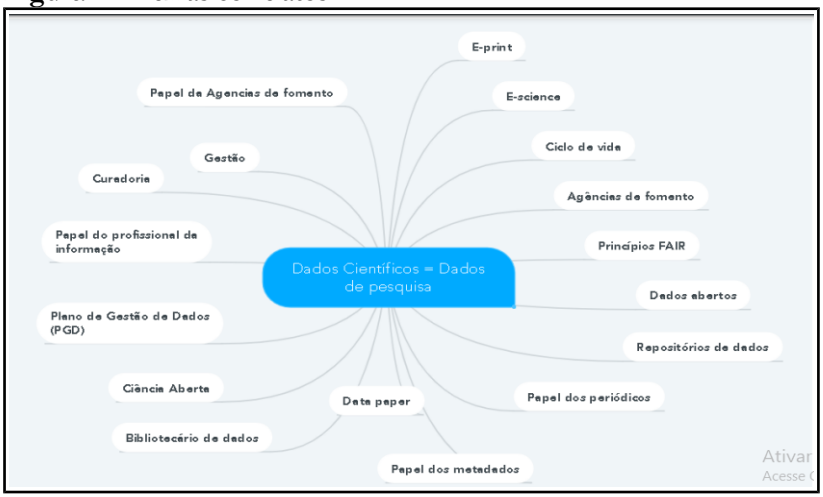
Categoria	Quantidade de artigos	Definição
Curadoria de dados de pesquisa	9	Discutem a concepção, elaboração, preservação e supervisão de dados científicos, sendo em ambientes físicos ou digitais.
Boas práticas, políticas e princípios para a gestão de dados de pesquisa	4	Pesquisas que abordam a necessidade de estabelecimento de princípios, boas práticas e políticas para a gestão de dados, que fazem o levantamento de documentos como esses disponíveis e os benefícios da aplicação de determinado instrumento, como por exemplo os princípios FAIR, Linked Data, o movimento de dados abertos.
Percepção e comportamento dos pesquisadores acerca de dados de pesquisa	4	Documentos que buscam analisar o comportamento dos pesquisadores na gestão e disponibilização dos dados derivados de suas pesquisas, bem como a utilização de dados elaborados por outros pesquisadores.
Elaboração do Plano de Gestão de Dados de Pesquisa (PGD)	1	Pesquisas teóricas acerca da importância do PGD, bem como pesquisas mais aplicadas, discutindo a elaboração e aplicação do plano.
Papel dos profissionais da informação na gestão de dados de pesquisa	3	Documentos que buscam identificar como bibliotecários, arquivistas e cientistas da informação podem auxiliar no processo de gestão de dados de pesquisa e quais as competências que precisam ser aprendidas para que esses profissionais possam atuar de maneira eficiente nesse mercado.
Armazenamento dos dados de pesquisa /repositório de dados	2	Documentos que discutem a importância e os desafios do estabelecimento de repositórios, que buscam mapear repositórios úteis para determinado

Categoria	Quantidade de artigos	Definição
		domínio, que analisa a viabilidade e necessidade de criação de um repositório institucional de dados de pesquisa.
Influências externas na disponibilização de dados	2	Documentos que buscam relacionar o impacto que influências externas, tais como buscadores, agências de fomento, periódicos científicos e instituições de ensino e pesquisa, têm sobre as práticas de gestão e
Relação de dados de pesquisa com e-Science	2	Documentos que abordam o papel dos dados de pesquisa no contexto da e-Science.
Proteção autoral no reúso de dados de pesquisa	2	Documento que propõe soluções para garantir que a autoria dos conjuntos de dados de pesquisa seja respeitada.
Elaboração de dados de pesquisa	2	Documento que aborda a criação de dados de pesquisa
Papel dos metadados na recuperação dos dados científicos	1	Documento que discute a importância da adoção de metadados consistentes e estruturados para a recuperação dos dados de pesquisa.

Fonte: Autores.

Além dos temas centrais tratados nos documentos, buscou-se mapear assuntos que aparecem com frequência nas discussões, embora não sejam o objetivo central dos documentos. Os temas mapeados são apresentados na Figura 1.

Figura 1 – Temas correlatos



Fonte: Autores.

A primeira observação traçada em relação à figura diz respeito aos termos "dados científicos" e "dados de pesquisa". Essa pesquisa partiu da hipótese de que "dados científicos" constituiria um termo mais amplo, incluindo os dados de pesquisa, mas também os metadados de publicações científicas, dados dos pesquisadores (como os disponíveis no currículo Lattes), e indicadores de produção. Entretanto, a análise dos artigos nacionais não confirmou essa hipótese. Na literatura levantada o termo "dados científicos" é abordado como sinônimo de "dados de pesquisa".

Em relação aos temas que circundam a discussão de dados de pesquisa, um grande destaque é o conceito de e-Science, caracterizando um novo paradigma do desenvolvimento de

pesquisas, baseado no uso de tecnologias da informação e comunicação e na preservação, curadoria e compartilhamento de dados de pesquisa (ANJOS, DIAS, 2019; GARU, 2005 *apud* SALES et al., 2013).

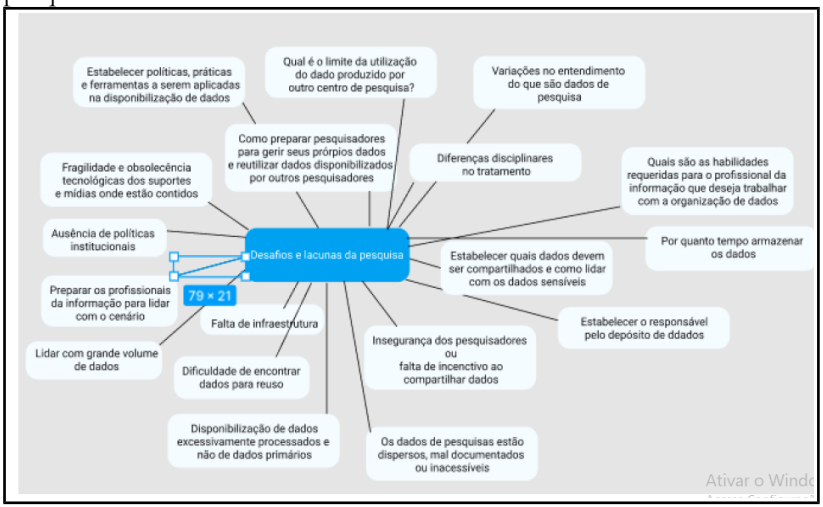
Muitos documentos trazem a preocupação com a curadoria, gestão, estabelecimento dos metadados, escolha da base de dados e disponibilização, aparece de maneira secundária se preocupando piamente em estabelecer o ciclo de vida dos dados como norte.

Sant'Ana (2016, p.119) define o ciclo de vida dos dados como "uma delimitação de fases (momentos em que distintas necessidades e competências são necessárias) envolvidas no acesso e uso dos dados", o autor descreve as fases como: processo de coleta, recuperação, armazenamento e descarte.

Muito se discute nesses documentos acerca das influências externas, o papel das agências de fomento, das instituições de ensino e pesquisa, dos periódicos científicos no incentivo formal ao compartilhamento de dados de pesquisa, no estabelecimento de políticas e de diretrizes que apoiem o compartilhamento dos dados. Também se destaca nos documentos a preocupação com a ciência aberta e com os movimentos de disponibilização aberta dos dados. Existe um destaque para o entendimento de que quando produzidos com financiamento público, esses dados constituem um bem público que deve ser organizado e disponibilizadode maneira estruturada, seguindo licenças de acesso aberto, preferencialmente a partir de formatos não proprietários.

Além dos temas correlatos foram levantados as dificuldades e lacunas apontadas pelos documentos a respeito da temática, apresentadas na Figura 2.

Figura 2 – Lacunas e desafios relacionados à temática de dados de pesquisa



Fonte: Autores.

Nota-se que a atuação dos pesquisadores na gestão dos dados é uma grande preocupação quando se discute dados de pesquisa. É possível observar, nos artigos incluídos na categoria "percepção e comportamento dos pesquisadores acerca de dados de pesquisa" que os pesquisadores entendem a importância de disponibilizar esses dados, mas, na prática, seguem não compartilhando-os. Alguns dos motivos para o não compartilhamento desses dados aparecem como lacunas de

pesquisa, tais como as preocupações como perder a vantagem competitiva e os direitos autorais desses dados e ainda a falta de conhecimento acerca do procedimento de disponibilização.

Essa dificuldade tem início logo na definição de dados de pesquisa, que varia entre as áreas do conhecimento e entre os próprios pesquisadores, existindo uma tendência de entender como dados de pesquisa apenas os dados que são excessivamente processados e tratados. Também existe uma preocupação sobre o processo de disponibilização desses dados, levando a questões como: Quais dados devem ser disponibilizados? Como lidar com dados sensíveis? Por quanto tempo armazenar esses dados? Onde disponibilizar? Como descrever esses dados?

Ainda em relação ao papel dos pesquisadores, também se identifica uma preocupação em como preparar esses pesquisadores para reutilizar esses dados, com dúvidas como: onde encontrar dados de interesse? Como selecionar esses dados? Como lidar com grandes volumes de dados? Qual o limite para a reutilização de dados de outros pesquisadores no estabelecimento de uma nova pesquisa?

Também se destaca entre os desafios a falta de infraestrutura e de diretrizes para auxiliá-los nesse processo, a gestão dos dados de pesquisa demanda muita atenção e tempo dos pesquisadores, que esperam que as instituições de pesquisa providenciem a infraestrutura pessoal e tecnológica necessária para apoiar esse processo. Também se espera que essas instituições disponibilizem diretrizes para embasar a gestão dos dados de pesquisa. Nota-se

que algumas instituições têm se movimentado para discutir e estabelecer essas diretrizes, mas que elas em sua maioria ainda não estão estabelecidas e divulgadas.

4. Considerações finais

Essa pesquisa teve por objetivo discutir como a comunidade brasileira da ciência da informação tem abordado os dados científicos e os dados de pesquisa. Nesse sentido conclui-se que o termo dados científico tem sido abordado como um sinônimo de dados de pesquisa, e que a definição de dados de pesquisa é ampla, exigindo mais discussões.

Em relação a como a temática tem sido discutida, conclui-se que abordagens voltadas para gestão, curadoria e recuperação dos dados são assuntos de destaque, sendo abordados em grande parte dos documentos. É possível observar que entre as duas categorias de maior destaque, sendo elas "Gestão de dados de pesquisa" e "Curadoria de dados de pesquisa" existem muitas semelhanças, não sendo clara a diferenciação entre as duas abordagens. Optou-se por respeitar a nomenclatura adotada pelos autores para definir seus trabalhos, entretanto estudos futuros são necessários para discorrer sobre a relação entre os dois conceitos.

Outrora se sofria da escassez de dados para pesquisas, hoje o problema é o contrário, quantidades exorbitantes de dados espalhados de formas diversas e sem uma política de gestão de dados predefinida, fato esse que pode também ser responsável pela falta de compartilhamento de dados gerados pelos próprios

pesquisadores, criando mais um gargalo que é a recuperação e seleção de dados já compartilhados.

Todos os documentos que trazem os desafios de se lidar com dados científicos abordam em essência a necessidade de se preparar o pesquisador para disponibilizar seus dados e seguir com a utilização de dados de terceiros. Essa falta de compartilhamento gera gargalos nas provas de trabalhos e em continuidades dos mesmos, podendo gerar assim inconsistências veladas, trazendo lentidão ao avanço científico em uma área.

Para pesquisas futuras, propõe se um aprofundamento na discussão acerca do conceito de dados de científicos, analisando como incluir os demais dados provenientes do processo de comunicação científica, tais como os metadados descritivos, os dados dos pesquisadores, os indicadores de produção, levanto o papel e a importância exercido por esses dados como insumos para a compreensão do cenário científico. Propõe-se ainda: o aprofundamento acerca do conceito de dados de pesquisa, a ampliação do levantamento realizado, incluindo trabalho para eventos e monografia, para um panorama mais completo do cenário nacional e a ampliação para o cenário internacional, realizando uma comparação entre o desenvolvimento da temática na ciência da informação no Brasil e no exterior.

Nesse sentido conclui-se que ainda existem muitos aspectos a serem explorados a respeito da relação entre dados científicos e dados de pesquisa, e que esse trabalho pode atuar como um ponto de partida para o desenvolvimento de novas pesquisas.

Referências

KITCHENHAM, B. et al. Systematic literature reviews in software engineering: a tertiary study**. Information and Software Technology**, [S.l.], v. 52, p. 792–805, 2010.

SARACEVIC, T. Information science. **Encyclopedia of Library and Information Sciences**, [S.l.: s.v.: s.n.], p. 2570-2585, 17 dez. 2009. Disponível em: https://tefkos.comminfo.rutgers.edu/SaracevicInformationScienceELIS2009.pdf. Acesso em: 11 maio. 2021

ANJOS, R. L. D.; DIAS, G. A. Atuação dos profissionais da informação no ciclo de vida dos dados – dataone: um estudo comparado. **Informação & Informação**, v. 24, n. 1, p. 80-101, 2019. Disponível em: http://www.uel.br/revistas/uel/index.php/informacao/article/view/31085. Acesso em: 17 maio. 2021.

SANT'ANA, R. C. B.. Ciclo de vida dos dados: uma perspectiva a partir da ciência da informação. **Informação & Informação**, [S.L.], v. 21, n. 2, p. 116, 20 dez. 2016. Disponível em: http://www.uel.br/revistas/uel/index.php/informacao/article/view/27940/20124. Acesso em: 18 maio 2021.

SALES, L. F.; SAYÃO, L. F.; SOUZA, R. F. Publicações ampliadas: um novo modelo de publicação acadêmica para o ambiente de e-Science. In: Encontro Nacional de Pesquisa

em Ciência da Informação (ENANCIB), 14., 2013, Florianópolis. **Anais…** Florianópolis: ENANCIB, 2013. Disponível em: http://eprints.rclis.org/20666/1/APRESENTA%C3%87%C3%83O%20-%20ENANCIB%202013.pdf. Acesso em: 23 jun. 2017.