



FORMAÇÃO E INVESTIGAÇÃO EM
CIÊNCIA DA INFORMAÇÃO

OPORTUNIDADES E DESAFIOS

10, 11 e 12 de novembro de 2025

POLITÉCNICO DO PORTO / ISCAP
PORTO - PORTUGAL



A PROBLEMÁTICA EXISTENTE ENTRE SILOS DE DADOS E O COMPARTILHAMENTO DE DADOS: UM ESTUDO SOBRE O SciELO DATA

Durval Vieira Pereira, Universidade Federal do Estado do Rio de Janeiro,
<https://orcid.org/0000-0001-6165-4938>, Brasil, durval.pereira@unirio.br

Fernanda do Valle Galvão Debetto, Universidade Federal do Estado do Rio de Janeiro,
<https://orcid.org/0000-0002-4156-027X>, Brasil, fernandavalle@unirio.br

Cláudio José Silva Ribeiro, Universidade Federal do Estado do Rio de Janeiro¹,
<https://orcid.org/0000-0002-9571-1707>, Brasil, claudio.ribeiro@unirio.br

Exo: Impacto das Tecnologias de Informação e Comunicação

1 Introdução

We must all accept that Science is data and that data are science, and thus provide for, and justify the need for the support of, much--improved data curation. (Hanson, Sugden, Alberts, 2011).

Ciência Aberta é definida pela UNESCO (2023) como um “conjunto de princípios e práticas que visam tornar a pesquisa científica em todas as áreas acessível a todos, para o benefício dos cientistas e da sociedade como um todo”. Em outras palavras, a Ciência Aberta visa a tornar públicos os diferentes processos de produção de dados que atravessam o fazer científico, tais como dados brutos, dados tratados e analisados, bem como pareceres entre pares. A ideia é que esse gesto de transparência também estimule o reuso dos dados e a colaboração entre pesquisadores e campos do conhecimento.

Temáticas com foco em produção e disseminação da informação, descritas por Gáal e Pereira (2023, pp. 16), se aproximam da Ciência da Informação pela via conceitual e metodológica do “acesso aberto”, da “comunicação científica”, do “gerenciamento de dados de pesquisa” e de “repositórios institucionais”. Sobre esse aspecto, Souza e Freitas (2024, pp. 1) afirmam que a Ciência

Aberta visa à disseminação de informações científicas por meio de colaborações facilitadas pela disponibilização de informação em repositórios digitais públicos contendo três elementos: “Acesso Aberto (publicações), Dados Abertos (incluindo dados brutos, modelos, especificações e documentação) e Processos Computacionais Abertos (como software e algoritmos)”.

Esse cenário nos conduz ao entrecruzamento teórico, metodológico e temático com outra área do conhecimento, a Ciência de dados. De acordo com Wang (2018), a Ciência de Dados utiliza a análise de dados para contribuir com outros campos do saber.

“The foundation of data science lies in the core statistical and analytical approach to data. Data science functions to support other subjects by analyzing and exploiting the data sets in subject fields and extracting knowledge”. (Wang, 2018, pp. 1244).

Conforme apontam Sayão e Sales (2016a), o conceito de “dado de pesquisa” possui um significado abrangente e varia conforme os domínios científicos, objetos de estudo, metodologias de geração e coleta de dados, entre outros fatores. Os autores destacam que esse termo pode referir-se ao resultado de experimentos conduzidos em ambientes laboratoriais controlados, estudos empíricos

nas ciências sociais ou observações de fenômenos culturais. No entanto, considera-se que a expressão “dados científicos” seja a mais adequada para abranger tais situações. O dilema terminológico sobre dado científico acompanha a história da Ciência da Informação desde a definição de informação e, mais especificamente, de informação científica. Schrader (1983) lembra que, no momento de institucionalização do campo, onde foi evidenciado o diálogo entre informação, ciência e tecnologia, não houve clareza ao delimitar se informação científica seria a informação necessária à atividade científica ou seria o resultado da atividade científica. Passadas seis décadas de institucionalização como área do saber, ainda não há consenso teórico na Ciência da Infprmação sobre as noções de dado, informação e conhecimento. Mas, a despeito das diferenças, desde os seus antecedentes históricos, tais como a Bibliografia, a Documentação, a Biblioteconomia e os *Science Services* (Araújo, 2018), tem-se na preservação e no compartilhamento a tônica para a construção dos saberes.

Nesse contexto, para além da preservação do suporte e do conteúdo informacional, a Ciência Aberta suscita o compartilhamento da infraestrutura responsável pela coleta e produção, isto é, destaca a importância do conhecimento do percurso de construção, não somente do conhecimento sobre os resultados de pesquisa. Assim, parte-se da ideia de que o compartilhamento de dados de pesquisa também impulsiona a produção de novos conhecimentos ao possibilitar a validação científica com mais acurácia. A necessidade de organização, guarda, preservação, disseminação e reuso dos dados brutos, as anotações em cadernos de laboratórios, de imagens e de vídeos produzidos durante a realização do estudo, têm valor agregado com potencial contribuição para discussões mais amplas e inovadoras. Diante disso, fica evidente a necessidade de zelar com mais vigor e rigor nos critérios e processos de armazenamento dos dados, buscando esta integração entre eles e seu reuso.

Em 2016, a *Scientific Data* publicou os “Princípios Orientadores FAIR para Gestão e Administração de Dados Científicos”, cujas diretrizes desejam melhorar as seguintes dimensões na produção, disseminação e reuso: *Findable* (Localizáveis); *Accessible* (Acessíveis); *Interoperable* (Interoperáveis); e *Reusable* (Reutilizáveis) de ativos digitais². Desses quatro princípios, destaca-se aqui o terceiro. A demanda por dados interoperáveis exige que o pesquisador utilize protocolos universais ao planejar e gerar seus dados. Dessa forma, podem ser comparados e integrados, caso estejam disponíveis publicamente, com outras bases de dados desenvolvidas em diferentes contextos e grupos de pesquisa. No entanto, nem sempre os repositórios de dados se conectam, funcionando como armazéns de grãos isolados, pois, como ressaltam Veiga et al. (2024), os princípios FAIR se constituem como diretrizes para boas práticas de gestão de dados, mas não oferecem detalhamento de como podem ser implementadas. Para as autorias, o desafio da implementação exige “contínuo investimento em capacitação, desenvolvimento de ferramentas e criação de políticas adequadas” (Veiga et al., 2024, pp. 281).

Campos, Campos e Barbosa (2020) apresentam os desafios para uma interoperabilidade semântica de dados.

These repositories of scientific research data, despite being a major advance in the description, management and sharing of research data, confront a major problem. The extreme complexity of the data models involved and the enormous diversity of objects and domains to be represented form an almost insurmountable barrier to the efficient management of these environments, preventing its main purpose, which is, sharing, discovery and reuse of these data to be performed properly. It is a clear example that only the computer and computing infrastructure that forms the basis of the global information network are not enough to solve the

existing problem. (Campos, Campos e Barbosa, 2020, pp. 683).

A metáfora dos grãos isolados foi utilizada para elaboração da expressão “silos de dados”, que representam, conforme dito, um desafio para o avanço da Ciência Aberta. A expressão aparece, inicialmente, no contexto das bibliotecas ligados ao contexto de metadados.

We can also appreciate from a library perspective the historical path which has led to independent “data silos.” In the past the need for such a level of inter-connectivity was neither apparent nor easy to achieve. Today the growth of the Web has highlighted the need for metadata which can “travel” across these standards and systems barriers (Macewan, 2004, pp. 432).

A descrição dos metadados é prática antiga na Biblioteconomia no contexto da catalogação, atividade herdada pela Ciência da Informação, especialmente vista na subárea de Representação da Informação e nas iniciativas de Organização do Conhecimento. Para realizar a organização do acervo, faz-se o controle bibliográfico que consiste em processos de representação descritiva e temática. Na etapa de representação descritiva, que visa à individualização do documento, o objeto a ser documentado é descrito a partir de características físicas e/ou temporais, seguindo protocolos internacionais. No caso dos livros, utiliza-se o *International Bibliographical Standard Description* (ISBD), definido pela *International Federation of Library Associations and Institutions* (IFLA). Já na representação temática, faz-se a análise do conteúdo informacional, comumente o seu assunto, representado por palavras-chave, que podem ser controladas por linguagens de especialidade, as chamadas linguagens documentárias. O tratamento temático por meio de metadados, que refletem temas e/ou conceitos, tem a finalidade de organizar os documentos por semelhanças e diferenças de assunto, uso e/ou função. Nesse sentido, para a integração de acervos, discute-se as particularidades terminológicas de seus

usuários, as pessoas buscadoras de informação, o que demanda diferentes formas de representação de um mesmo recurso, de um mesmo conjunto de metadados. Serra e Santarém Segundo (2021) destacam que

Ao copiar ou produzir registros para os seus catálogos, a biblioteca está criando um silo de dados, onde os metadados não possuem conexão com os demais dados presentes no catálogo, com os catálogos de outras bibliotecas ou com a Web (Serra; Segundo, 2021, pp. 626).

Nessa direção, Ribeiro e Ulhôa (2025, no prelo) explicam que, em silos, os dados são publicados em diferentes plataformas, formatos diversos e com aplicativos produzidos em momentos distintos, noção que nos conduz a outros conceitos relativos ao universo da informação digitalizada, como a *web semântica*.

A *Web Semântica* surge com tecnologias como o *Linked Data*, capazes de possibilitar a integração de dados até então isolados. Byrne e Goddard (2010) afirmam que “*the semantic web is about breaking down silos, not building better ones*”. Muitos repositórios de dados de grande porte permanecem isolados do exterior, gerando o que tem sido denominado como o problema do silo de dados. (Kim, 2016, pp. 1).

Tendo em vista a complexidade para efetivação de práticas desejadas de compartilhamento e reuso de dados com qualidade e ética, a proposta deste estudo é refletir sobre o problema dos silos de dados no compartilhamento de dados em repositório de dados abertos de pesquisa. A finalidade é discutir os impactos dos silos de dados no compartilhamento de dados científicos, tendo como campo empírico a análise do SciELO Data.

Acredita-se que a temática, aqui abordada, contribua para a compreensão dos desafios na formação em Ciência da Informação, pois discute sobre como melhorar as questões que envolvem compartilhamento de dados em repositórios de dados e como os silos de dados podem impactar de forma negativa no processo.

2 Referencial Teórico

Este estudo propõe um diálogo da Ciência da Informação com outros conceitos atualmente em destaque nos estudos informacionais contemporâneos, como a Ciência de dados (Davenport; Patil, 2012) e a Inteligência Artificial (Gómez Mont, 2020). Jain (2003) comenta sobre a mudança conceitual e de aplicação que o dado enfrenta nas últimas décadas.

Data has changed significantly over the last few decades. Computing systems that initially dealt with data and computation rapidly moved to information and communication. The next step on the evolutionary scale is insight and experience. Applications are demanding the use of live, spatio-temporal, heterogeneous data. (Jain, 2003, pp. 8).

Repositórios de dados de pesquisa, tal qual o SciELO Data, na visão de Sayão e Sales (2016b, pp. 96), são essenciais para reutilização dos dados, acelerando avanços científicos e economizando recursos, além de ampliar novos temas de pesquisa. Para Wang, Kung e Byrd (2018), a concepção de dados envolve aspectos históricos e pragmáticos; além do cuidado com precisão, acurácia e credibilidade dos dados; os processos e funções tradicionais da Biblioteconomia e as teorias da documentação.

Uma preocupação que existe sobre repositórios, em especial sobre repositórios de dados de pesquisa, é a questão da interoperabilidade, que acaba por aproximar os estudos das áreas da Ciência da Informação e da Ciência da Computação. (Campos, Campos e Barbosa, 2020). A interoperabilidade é conceituada pelo *Institute of Electrical and Electronic Engineering* (IEEE) como a habilidade de dois ou mais sistemas ou componentes trocarem informações e utilizarem as informações trocadas. (IEEE, 1990). Esta definição é complementada por Waters, Powers e Ceruti (2009, pp. 1158) ao afirmarem que a Interoperabilidade “*is a concept that is predicated on the assumption that the semantic integration, frameworks and*

standards that support information exchange, and advances in science and technology can enable information-systems interoperability for many diverse users”.

O reuso de dados e a necessidade de sistemas interoperáveis são temas que estão em expansão em investigação científicas, pois “*trending topics, such as “smart cities,” “smart mobility,” or “smart homes,” are based, at their core, on connected physical devices that can exchange data automatically*”. (Hodapp; Hanelt, 2022, pp. 407). Por essa razão, cientistas continuam desenvolvendo pesquisas voltadas para a revisão, análise e aplicação de técnicas sobre interoperabilidade, a fim de esclarecer o que causa a baixa interoperabilidade, quais mecanismos influenciam a interoperabilidade e quais são os potenciais resultados desses mecanismos. É preciso fomentar o debate sobre interoperabilidade na inovação digital e revelar também deficiências no discurso atual sobre interoperabilidade e sua contribuição para uma inovação digital.

Maricato et al. (2023, pp. 1) explicam que a capacidade de produção e coleta de dados por meio de infraestruturas cirométricas de larga escala aumentaram por meio da publicação digital e do processamento automatizado de dados. No entanto, os autores, apoiados nos estudos de Krüger (2020), chamam atenção para a necessidade de uma compreensão melhor sobre os critérios, abrangência e prática de pesquisa capturados nessas infraestruturas.

No entendimento de Semeler e Pinto (2019, pp. 113), os dados de pesquisa são o resultado de qualquer investigação sistemática que envolva processos de observação, experimentação ou simulação de procedimentos de pesquisa científica. Tornando-se produtos tão importantes da ciência, quanto as publicações científicas. (Marcondes, 2022).

A solução para um melhor aproveitamento da potencialidade dos dados de pesquisa seria possibilitar as melhores condições para seu compartilhamento e, consequentemente, reuso de dados, aproximando-se das condicionantes política e econômica, além da

tecnológica, que atravessam o campo informacional (Iafrate 2015, pp. 3). Autores como Verhulst (2023) e Curtyÿ (2019); Borgman (2011), participam do pensamento de que o compartilhamento de dados científicos é uma prática fundamental na ciência, sempre presente nas interações entre pesquisadores e na dinâmica da produção científica, variando conforme a área do conhecimento e a natureza dos dados. Contudo, a formalização e sistematização dos processos para disponibilização de dados, juntamente com o amplo e aberto acesso a coleções de dados por meio de repositórios *online*, promovidos por órgãos governamentais, agências de fomento, entidades científicas e editoriais, têm contribuído para a maior legitimação do reuso de dados no contexto científico. Em particular, à medida que vários tipos de dados são acumulados em bancos de dados individuais, a necessidade de análise integrativa de bancos de dados separados, porém relevantes, torna-se mais importante. (Kim, 2016, pp. 743).

Há cada vez mais a necessidade de curadoria de dados interoperáveis, que exige do pesquisador a utilização de protocolos universais ao planejar e produzir seus dados científicos, permitindo que sejam comparados e agregados a outros bancos de dados provenientes de diferentes contextos e grupos, quando houver acesso livre.

Nessa direção, os silos de dados aparecem como um empecilho para que a integração entre grupos de dados e comunicação entre eles no universo dos repositórios de dados. “Silos de dados”, ou para alguns autores “silos de informação”, podem ser traduzidos por: impossibilidade de comunicação; o escopo e cobertura com o compartilhamento e/ou integração de grupos de dados; desenvolvimento de software em diferentes momentos ao longo de décadas, utilizando soluções de diferentes fabricantes; utilização de métodos e ambientes não compatíveis; ausência de interfaces padronizadas; falta de protocolos de comunicação; formato de transmissão de dados e regras de resposta uniformes (Zhang, Yuan e Xiong, 2009; Patel, 2019; ZHANG, et al. 2025).

Uma preocupação presente na ciência, atualmente, é a curadoria de dados. De acordo com CoreTrustSeal (2021), curadoria de dados pode ser entendida como:

The activity of managing and promoting the use of data from their point of creation to ensure that they are fit for contemporary purpose and available for discovery and reuse. For dynamic datasets this may mean continuous enrichment or updating to keep them fit for purpose. Higher levels of curation will also involve links with annotation and with other published materials.

Seria possível dizer, então, que a curadoria de dados abrange todo o ciclo de vida dos dados, desde o planejamento da coleta até a preservação para acesso e reutilização a longo prazo, definindo-se como as verificações e ações realizadas por curadores com o objetivo de garantir que o conjunto de dados esteja estruturado e documentado de maneira completa e conforme as melhores práticas.

De acordo com o pensamento de Borgman (2012, pp. 22), as maiores vantagens do compartilhamento de dados podem estar na combinação de dados de múltiplas fontes, comparados ou misturados de maneiras inovadoras. A autora acrescenta que os dados são integrados de forma mais confiável quando coletados e processados sistematicamente, de maneira que suportem os padrões de grandes comunidades.

Curtyÿ (2019, pp. 191) lembra que, no contexto científico atual, o compartilhamento sistemático de dados de pesquisa visando seu potencial de reuso futuro é fundamental. O reuso de dados, no entanto, não pode ser considerado uma simples consequência do compartilhamento e requer estratégias e incentivos para ocorrer conforme os princípios da ciência. Portanto, questões relativas ao reuso e à reusabilidade dos dados, tanto conceituais quanto práticas, precisam ser amplamente discutidas.

Um dos pontos a serem discutidos é a limitação do acesso aos dados, que pode impedir que a Ciência Aberta alcance seu potencial de beneficiar a sociedade como um todo, principalmente nas questões que

envolvem guarda, recuperação, integração e reuso de dados. Dessa forma, o estabelecimento de políticas e as reflexões sobre finalidade e indicadores de reuso tornam-se tão relevantes quanto às políticas para abertura dos dados de pesquisa. Para além da dimensão técnica dos dados abertos digitais e disponíveis em sistemas, há a produção fragmentada, à espera de análise, curadoria e instrumentos reguladores. Como exemplos, em janeiro de 2025, o presidente do Conselho Nacional de Desenvolvimento Científico e Tecnológico do Brasil (CNPq), Ricardo Galvão, publicou em coautoria com a professora Débora Menezes, da Universidade Federal de Santa Catarina (UFSC), um artigo intitulado “Ciência Aberta: uma visão desapaixonada” (Menezes; Galvão, 2025), onde destaca as fragilidades operacionais para gestão e governança de dados abertos, como equipe multidisciplinar defasada e particularidades das áreas do conhecimento e objetos de pesquisa. Em contraponto, a Rede Brasileira de Reprodutibilidade (2025) respondeu o artigo com uma carta pública salientando que, a despeito das concordantes limitações de infraestrutura, há soluções possíveis de baixo custo e as limitações apontadas não se aplicariam à maioria dos campos disciplinares. No contexto das ciências sociais, Melo e Rockembach (2025) destacam a relevância da cooperação entre pares e a transparência dos dados, mas afirmam ser insuficientes os mecanismos de garantias da privacidade dos sujeitos participantes de uma pesquisa científica no horizonte da abertura e reuso dos dados, pois

[...] o termo de consentimento informado, cujo objetivo é prover ao participante a autonomia de tomar decisões sobre os riscos que a pesquisa oferece à sua privacidade, não é capaz de informar os possíveis usos posteriores dos dados, simplesmente porque um dos princípios da ciência aberta repousa justamente sobre o livre (Melo; Rockembach, 2025, pp. 2).

À luz do exposto, quando dados são retidos em silos no digital, seu potencial impacto social é abalado e oportunidades de inovação

e descobertas científicas são perdidas, afetando diretamente áreas como: saúde, meio ambiente, tecnologia dentre outras.

Information systems, in contrast, create data silos. The metadata is defined and introduced for a data of a particular type, which is indexed and neatly stashed in its own place. [...] The challenge to the database community, then, is to break down these silos to unify information. This requires more out-of-the-box thinking because most data sources are designed to behave like independent silos. Their creators assume that after the integration system analyzes the silos and extracts their metadata, it will somehow combine the metadata to provide correct results. Indeed, many current research efforts are aimed at this kind of solution. (Jain, 2003, pp. 14).

No caso de repositórios de dados, os silos aparecem não só por razões técnicas, como citadas na seção 2, mas também por questões de gestão. Zhang et al. (2025) convalidam a nossa afirmação quando apontam que os silos não existem apenas por questões técnicas, mas também por decisões de implementação e implantação sem fazer uso de protocolos padronizados integrativos.

Not only because of different manufacturer, different methods of development, different development environment, but also as a result of these systems have no uniform standard interface, communication protocols, and data transmission format and response rules. So the exchange of information between them becomes very difficult, which leads to the resources cannot be shared, and this is the so-called information island. (Zhang et al., 2025, pp. 2905).

Esses autores apontam que a solução para uma efetiva integração de dados está na eliminação de silos de dados. Portanto, percebe-se que os silos de dados têm um impacto negativo no compartilhamento e na integração de dados ao dificultar a

colaboração, reduzir a transparência e aumentar a ineficiência. Eles limitam o acesso aos dados, impedem que pesquisadores de diferentes áreas trabalhem juntos e podem levar a resultados científicos inconsistentes.

Com base na literatura analisada, pode-se afirmar que:

- Os silos de dados prejudicam a qualidade da pesquisa científica ao limitar o acesso a informações essenciais para a validação e replicação de estudos. A falta de dados compartilhados pode resultar em conclusões enviesadas e menos robustas.
- A transparência é um dos pilares da Ciência Aberta. Silos de dados dificultam a verificação independente de resultados, tornando o processo científico menos transparente e aumentando o risco de fraudes e erros.
- O impacto das descobertas científicas é reduzido quando os dados não são compartilhados. Soluções com potencialidades de inovação para problemas urgentes da sociedade podem não ser descobertas devido à falta de integração de informações.
- Se os livros compõem uma parte dos dados de pesquisa utilizados, a partir da digitalização dos processos biblioteconômicos, os dados bibliográficos administrados pelas bibliotecas também podem se tornar silos dentro de silos de dados, como destacam Serra e Santarém Segundo (2021);
- A existência de boas práticas de gestão, como os Princípios FAIR, não é suficiente. É preciso compreender as particularidades dos domínios e as necessidades técnicas e profissionais para sua implementação.

Diante do exposto, a discussão sobre a gestão e o reuso de dados de pesquisa no contexto da Ciência Aberta evidencia a crescente relevância dos repositórios de dados como o SciELO Data. A literatura aqui apresentada salienta que o compartilhamento sistemático

e a curadoria eficiente dos dados são fundamentais para ampliar o impacto da produção científica, acelerar descobertas e promover a transparência nas práticas de pesquisa. No entanto, os chamados silos de dados — causados tanto por fatores técnicos, quanto por decisões de gestão — constituem um obstáculo significativo para a interoperabilidade entre repositórios e a integração dos dados. Esses silos comprometem ainda a replicabilidade das pesquisas e a reutilização dos seus dados. Assim posto, superar os desafios impostos pelos silos de dados é um passo essencial para consolidar a Ciência Aberta como um modelo capaz de ampliar o acesso ao conhecimento, fortalecer a cooperação científica e gerar benefícios concretos à sociedade.

3 Procedimentos Metodológicos

Trata-se de uma pesquisa descritiva, aplicada, de abordagem qualitativa, do tipo exploratória; de carácter analítico experimental (Flick, 2013). Utiliza pesquisa bibliográfica para levantamento de estudos capazes de fundamentar a argumentação apresentada. Realiza um estudo de caso sobre o portal de pesquisa do SciELO Data a partir da análise da documentação disponibilizada *online*, com a finalidade de identificar possíveis questões associadas ao conceito de silo de dados, conforme apresentado na seção 2. A escolha do repositório de dados se justifica pela sua importância para preservação, guarda e disseminação de dados de pesquisa em diferentes países e abrigando um quantitativo expressivo de periódicos associados.

4 Resultados

O repositório SciELO Data, lançado em 2020, está definido, de acordo com o site oficial, como “[...] um repositório multidisciplinar para depósito, preservação e disseminação de dados de pesquisa de artigos submetidos e aprovados para publicação, ou já publicados em periódicos da Rede SciELO, ou depositados

no SciELO Preprints” (SciELO, 2025). Para tal, utiliza o repositório Dataverse (SciELO, 2023), que busca infraestrutura se apoia em normas, padrões e protocolos que se propõem a garantir a interoperabilidade dos diversos ambientes, sistemas e dados heterogêneos.

o SciELO Data é ao mesmo tempo um conjunto de repositórios específicos de cada periódico e um repositório geral que agrupa todos os conteúdos. Esta solução segue o princípio de trabalho em rede, pois dota cada periódico com autonomia de contar com repositório próprio e realizar sua gestão de acordo com as especificidades das políticas de gestão de dados de pesquisa. (Packer e Santos, 2022, pp. 100-101).

Conforme os autores acima mencionados, o SciELO Data desempenha, no contexto da Ciência Aberta, as funções de preservar os conteúdos das pesquisas subjacentes aos textos, promover a ampla disseminação dos resultados científicos, viabilizar o reuso de dados por novos estudos, oferecer suporte às análises, resultados, discussões e conclusões dos manuscritos, além de facilitar o processo de avaliação por pares e favorecer a reprodução e replicação das pesquisas (Packer e Santos, 2022, pp. 101).

Os arquivos armazenados no SciELO Data são formalmente reconhecidos como objetos de comunicação de pesquisas conforme o Modelo SciELO de Publicação. Conforme Packer & Santos (2022, pp. 101) afirmam que os registros recebem identificador DOI atribuído pelo *Datacite* e devem conter metadados que possibilitem citação, referenciamento, indexação e interoperabilidade abrangentes, com alinhamento progressivo aos princípios FAIR.

No contexto do compartilhamento de dados, a vantagem do uso desse repositório é o alinhamento com os princípios FAIR (encontrável, acessível, interoperável, reutilizável). (SciELO, 2023).

Ademais, o SciELO Data realiza o compartilhamento dos dados com princípios voltados para repositórios confiáveis. Ele presta serviço com base nos princípios TRUST: transparência (*transparency*),

responsabilidade (*responsibility*), foco no utilizador (*user focus*), sustentabilidade (*sustainability*) e tecnologia (*technology*) (Lin, 2020).

A proposta do SciELO Data é promover o acesso aberto aos dados de pesquisa, vinculando-os a artigos publicados na rede SciELO. Retomando a discussão conceitual sobre dados de pesquisa, para o SciELO, dados de pesquisa podem ser “arquivos de dados, de programas de computador e de outros conteúdos subjacentes aos textos” (SciELO, 2024a, pp. 6). Além disso, suporta a criação de metadados padronizados e DOI para conjuntos de dados e alinha sua prática com a política de dados abertos da Ciência Aberta. (SciELO, 2025).

Com base na definição de silos de dados apresentada na seção 2, foram estipulados critérios para análise do SciELO Data. Vejamos:

Quadro 1: Critérios analisados no SciELO Data

Possível causa do silo de dados	Critério analisado
Impossibilidade de comunicação	Interoperabilidade
Escopo e cobertura com o compartilhamento e/ou integração de grupos de dados	Escopo e cobertura
Desenvolvimento de software em diferentes momentos ao longo de décadas, utilizando soluções de diferentes fabricantes	Software consistente
Utilização de métodos e ambientes não compatíveis	Interoperabilidade
Ausência de interfaces padronizadas	Interoperabilidade
Falta de protocolos de comunicação	Interoperabilidade
Formato de transmissão de dados e regras de resposta uniformes	Interoperabilidade

Fonte: Elaboração própria (2025)

A partir desses critérios, foi realizada uma análise em sua base de busca (SciELO, 2025) e na documentação disponível (SciELO, 2023), no recorte adotado por esta investigação.

Quadro 2: Resultado da analise no SciELO Data

Critério analisado	Resultado
Interoperabilidade	utiliza métodos e ambientes compatíveis com os padrões e protocolos internacionais e com a Ciência Aberta em termos de infraestrutura, mas ainda apresenta lacunas funcionais e práticas que podem gerar incompatibilidades operacionais e semânticas com ambientes mais avançados ou integrados.
Escopo e cobertura	Possui 68 comunidades Dataverses; 769 conjunto de dados; 5.178 arquivos. Organizados em Dataverses individuais (por revistas ou instituições)
Software consistente	Caracteriza-se como repositório confiável. Utiliza o DataVERSE desenvolvida pela Universidade de Harvard, que suporta alguns protocolos padrão da Web científica.

Fonte: Elaboração própria (2025)

A documentação disponibilizada pelo SciELO (2023; 2025) apresenta as capacidades técnicas, habilidades funcionais e objetivos do Scielo Data, porém não foi encontrado menção aos estudos voltados para curadoria semântica (ontológica) ou de interoperabilidade com plataformas mais complexas como: European Open Science Cloud (EOSC)ⁱ ou OpenAIREⁱⁱ.

Podemos inferir que não há evidência direta ou documentada publicamente de que o SciELO Data apresente silos de dados estruturais, como ocorre em ambientes corporativos ou institucionais fechados.

O processo investigativo constatou que o SciELO Data atende aos requisitos analisados. Todavia, há indícios e limitações que podem ser interpretados como riscos potenciais à interoperabilidade e ao compartilhamento pleno, o que pode ser associado ao conceito de silos de dados. Apesar de utilizar métodos e ambientes compatíveis (ver quadro 3) com

os padrões internacionais e com a Ciência Aberta em termos de infraestrutura, ainda apresenta algumas lacunas funcionais e práticas que podem causar incompatibilidades operacionais e semânticas com ambientes mais avançados ou integrados. A intereoperabilidade é limitada, não se conectando a outras plataformas externas, como o Zenodo, Dryad e Figshare, impondo barreiras à curadoria internacional, bem como pouca visibilidade dos *endpoints* e documentação técnica. Da perspectiva do controle terminológico, não foram identificadas ontologias ou outro tipo de linguagem documentária, resultando em efeito direto na representação da informação e integração semântica. Ainda no contexto dos metadados, apesar do uso do padrão DataVERSE, muitos dados são depositados com metadados insuficientes ou pouco estruturados, dificultando o reuso automatizado — o que pode ser interpretado também como barreiras semânticas.

A problemática que aqui se apresenta discute como é possível contemplar a interoperabilidade semântica em repositório de dados, utilizando ambientes heterogêneos que não se beneficiam de uma linguagem universal de representação. Campos, Campos e Barbosa (2020, pp. 682) apontam que uma potencial solução seria que a *"interoperation between two systems seems to be possible only when we can overcome the linguistic, terminological, and verbal differences that are defined by the idiosyncratic conditions of each environment"*.

Quadro 3: Interface/Protocolo utilizado pelo SciELO Data

Interface/ Protocolo	Finalidade	Compatibilidade Padrão
OAI-PMH	Coleta de metadados por outros repositórios (harvesting)	Sim
API REST	Acesso programático para consulta e submissão de dados	Sim
DOI /	Identificador	Sim

DataCite	persistente e metadados padronizados	
Exportação de metadados	Exporta em formatos como DDI, Dublin Core, JSON	Sim

Fonte: Elaboração própria (2025)

Já a respeito do escopo e cobertura com o compartilhamento e/ou integração de grupos de dados, constatou-se:

- **Fragmentação entre coleções nacionais:** como a rede SciELO é organizada por país, há pouca interoperabilidade entre os sistemas SciELO Data de diferentes países – por exemplo: Brasil, Chile, África do Sul (SciElo, 2024b), o que pode configurar silos geográficos. Maricato et al., (2023, pp. 2) apontam que “*the underrepresentation of many countries in selective databases does not allow for the proper scientific evaluation of countries that are not part of the scientific mainstream, causing biases in the research evaluations elaborated from their data*”.
- **Adoção limitada por parte dos autores:** muitos pesquisadores ainda não depositam os dados brutos de suas pesquisas, por desconhecimento, conforme sugere a investigação de Rodrigues et. al. (2019). Segundo aos autores, que analisaram a percepção de bolsistas de produtividade nível A do CNPq, vinculados a instituições públicas brasileiras, sobre Ciência Aberta, dos 53 respondentes, 28,57% desconhecem as práticas de Ciência Aberta. Ou seja, os dados podem existir, mas nem sempre estão ou serão compartilhados.

O Programa SciELO opera o repositório SciELO Data, que permite a cada periódico operar seu próprio servidor de dados de acordo com os padrões internacionais de repositórios de dados e passível, portanto, de ser utilizado por todos os periódicos. (SciELO, 2024a, pp. 10).

Este repositório de dados foi implantado na plataforma Dataverse, que permite a criação

de múltiplos repositórios de dados de pesquisa, também chamados de Dataverses (um Dataverse pode conter outros Dataverses ou disponibilizar conjuntos de dados diretamente) (Caregnato; Rosa e Rodrigues, 2024, pp. 2).

Outro ponto é levantado por Mazoni et al. (2023). Estes autores criticam a falta de dados totalmente estruturados.

The SciELO database, although using metadata such as XML and JSON, does not yet have fully structured data and, consequently, does not have a robust API available. In bibliographic databases, in many cases, the definition of minimum requirements and steps is necessary for the available data to become useful. (Mazoni et al., 2023, pp. 3).

A adoção é opcional pelos autores e revistas de depósito dos dados de pesquisa, em outras palavras, o SciELO Data não impõe o depósito de dados, sua utilização é voluntária por parte dos autores, o que gera inconsistência na adoção dos métodos e inviabiliza fluxos sistemáticos de interoperabilidade. Na visão de Caregnato; Rosa e Rodrigues (2024, pp. 3-4) a “ausência de exigências para a publicação de dados em repositórios de acesso aberto, ou a definição de diferentes níveis de acesso, pode dificultar a divulgação de resultados de pesquisa”.

O “Guia de preparação de dados de pesquisa” disponibilizado pelo Scielo (2021) esclarece que o SciELO Data aceita qualquer tipo de arquivo, porém é recomendado o depósito de formatos: “Não proprietários; Abertos, com padrões internacionais documentados; com codificação de caracteres padrão, de preferência Unicode” SciELO (2021, pp. 3). Entretanto, é apenas uma recomendação e não condição obrigatória para aceitar o depósito dos dados, acarretando a presença de dados sem as definições, descrições e formatos necessário para serem considerados de acordo com os princípios FAIR, por exemplo, e não padronizando a entrada de arquivos de dados, conforme exigência da Ciência Aberta. (Maricato et al., 2023, pp. 3).

A carência de padronização no formato dos dados a serem

submetidos também foi observada. A falta de um formato normalizado e uma grande variação de formatação de arquivos pode dificultar a recuperação, disseminação e preservação desses dados. O pouco alcance da revisão por pares aos dados de pesquisa foi um dos pontos que mais chamou atenção durante a análise, já que essa prática seria extremamente benéfica ao assegurar a confiabilidade dos dados divulgados. (Caregnato; Rosa & Rodrigues, 2024, pp. 4).

O SciELO Data possui características analíticas únicas que precisam ser valorizadas, possuindo uma fonte relevante para publicação científica em acesso aberto nas regiões que abrange. É também uma plataforma voltada para aumentar a visibilidade e a internacionalização da produção científica dos países (Maricato, 2023, pp. 8). Todavia, esses autores chamam a atenção para pontos a serem aprimorados:

we would like to remark the need to seek ways to improve the large-scale availability of SciELO data for researchers, policy makers, librarians, and any other interested stakeholders. Such large-scale availability may represent the tipping point for the SciELO database to become an open research scientometric infrastructure on its own. (Maricato, 2023, pp. 8).

Logo, podemos inferir que há uma necessidade de destacar a importância de buscar maneiras de melhorar a interoperabilidade dos dados depositados, de forma a aumentar a disponibilidade em larga escala do SciElo Data para pesquisadores, formuladores de políticas, bibliotecários e quaisquer outras partes interessadas.

5 Algumas considerações

Os silos de dados são um obstáculo significativo para o compartilhamento de dados e a Ciência Aberta, prejudicando a qualidade, a transparência e o impacto da pesquisa científica. A superação desses desafios é essencial para o avanço do conhecimento e o desenvolvimento de

soluções para os problemas mais urgentes da sociedade.

Após a análise realizada, pode-se afirmar que o SciELO Data não é um silo de dados por definição, mas ainda apresenta barreiras operacionais, técnicas e culturais que podem afetar a fluidez do compartilhamento em larga escala. Tais barreiras se apresentam como campo fértil para desdobramentos desta pesquisa a partir da contribuição da Ciência da Informação, com foco em interoperabilidade, curadoria de dados, metadados e políticas informacionais.

Como desdobramento da presente investigação, vislumbramos a possibilidade de reaplicação da metodologia adotada em estudos comparativos com outros repositórios de dados de pesquisa, tendo como eixo analítico a problemática dos silos de dados. Outro potencial desdobramento consiste no aprofundamento das análises acerca dos princípios FAIR, com ênfase nos aspectos semânticos relacionados à descrição e à recuperação da informação. A partir da experiência acumulada, levanta-se a hipótese de que há uma lacuna significativa na literatura quanto à implementação dos processos de FAIRification e à consolidação de objetos FAIR, os quais despontam como estratégias promissoras para mitigar ou romper com as barreiras impostas pelos silos de dados.

Por fim, podemos afirmar que determinadas barreiras operacionais, técnicas e culturais revelam a complexidade dos processos de compartilhamento e integração de dados em larga escala. Desta forma, a Ciência da Informação possui interesses e estudos que podem contribuir com possíveis soluções voltadas à interoperabilidade, à curadoria qualificada, ao uso eficaz de metadados e à formulação de políticas informacionais. Suplantar essas lacunas não é apenas uma exigência técnica, mas um compromisso estratégico com a transparência, a colaboração e o impacto social da pesquisa científica.

Referências

- Araújo, C. A. A. (2018). *O que é ciência da informação*. Belo Horizonte: KMA.
- Borgman, C. L (2012). The conundrum of sharing research data. *Journal of the American Society for Information Science and Technology*, v. 63, n. 3, June. Disponível em: <https://doi.org/10.1002/asi.22634>
- Byrne, G.; Goddard, L. (2010). The strongest link: libraries and linked data. *D-Lib Magazine*, v. 16, n. 11/12, nov./dez. Disponível em: <http://www.dlib.org/dlib/november10/byrn/e/11byrne.html>.
- Campos, M. L. A.; Campos, L. M.; Barbosa, N. T. (2020). The Challenges of semantic interoperability in the Era of Escience on the Web. *Knowl. Org.*, v. 47, n. 8, pp. 680-695. Disponível em: <https://www.nomos-eibrary.de/10.5771/0943-7444-2020-8-680.pdf>
- Caregnato, S. E.; Rosa; J. A.; Rodrigues, P. H. S. (2024). Políticas editoriais sobre dados de pesquisa em periódicos no repositório SciELO Data. In: Encontro Brasileiro de Bibliometria e Cientometria (9. 2024. Brasília). *Anais...* Brasília: UNB: IBICT. Disponível em: <https://doi.org/10.22477/ix.ebbc.285>
- Core TrustSeal Standards and Certification Board. *CoreTrustSeal Trustworthy Data Repositories Requirements 2020–2022*.
- DataVerse Project (2025). *Dataverse documentation v.6.6*. Disponível em: <https://guides.dataverse.org/en/latest/>
- Davenport, T. H.; Patil, D. J. (2012). Data Scientist: the sexiest job of the 21st Century. *Harvard Business Review*, v. 90, n. 10, pp. 70–76, Oct. Disponível em: <https://hbr.org/2012/10/data-scientist-the-sexiest-job-of-the-21st-century>.
- Flick, U. (2013). *Introdução à metodologia de pesquisa*: um guia para iniciantes. Porto Alegre: Penso.
- Gaal, L. P. M.; Pereira, C. A. (2023). Colaboração científica sobre ciência aberta no campo da ciência da informação. *Revista digital de biblioteconomia & ciência da informação*; v. 21, pp. 1-26. Disponível em:
- <https://doi.org/10.20396/rdbc.v21i00.8673825>
- Gómez Mont, Constanza et al. (2020). *La inteligencia artificial al servicio del bien social en América Latina y el Caribe: Panorámica regional e instantáneas de doce países*. Washington: Banco Interamericano de Desarrollo. Disponível em: <http://dx.doi.org/10.18235/0002393>
- Hanson, B., Sugden, A.; Alberts, B. (2011). Making data maximally available. *Science*, v. 331, n. 6018). Disponível em: DOI: 10.1126/science.1203354
- Hodapp, D.; Hanelt, A. (2022). Interoperability in the era of digital innovation: An information systems research agenda. *Journal of Information Technology*, 37(4), 407-427. <https://doi.org/10.1177/02683962211064304>
- Iafrate, F. (2015). *Do Big Data ao Smart Data*. Londres: ISTE; Hoboken, NJ: John Wiley.
- IEEE (1990). *IEEE standard computer dictionary: a compilation of IEEE standard computer glossaries*. New York: IEEE.
- Jain, R. (2003). Out-of-the-box data engineering events in heterogeneous data environments. *Proceedings 19th International Conference on Data Engineering (Cat. No.03CH37405)*, Bangalore. Disponível em: doi: 10.1109/ICDE.2003.1260778.
- Kim, J. et al. Collaborative analytics for data silos. 2016. *IEEE 32nd International Conference on Data Engineering (ICDE)*, Helsinki, Finland, 2016, pp. 743-754, doi: 10.1109/ICDE.2016.7498286.
- Krüger, A. K. (2020). Quantification 2.0: Bibliometric infrastructures in academic evaluation. *Politics and Governance*, v. 8, n.2, pp. 58-67. Disponível em: <https://doi.org/10.17645/pag.v8i2.2575>
- Lin D. (2020). The TRUST Principles for digital repositories. *Sci Datav.* 7, n. 1. Disponível em: <https://doi.org/10.1038/s41597-0200486-7>
- Macewan, A. (2004). Project InterParty: from library authority files to e-commerce. *Cataloging & Classification Quarterly*, v.

- 39, n. 1-2, pp. 429-442. Disponível em:
DOI: 10.1300/J104v39n01_11.
- Marcondes, C. H. (2022). The role of vocabularies in the age of data: the question of research data. *Knowl. Org.*, v. 49, n. 7, pp. 467-482. Disponível em: <https://doi.org/10.5771/0943-7444-2022-7-467>
- Maricato et al. (2023). SciELO as an open scientometric research infrastructure: general discussion of coverage in OpenAlex, WoS, Scopus and Dimensions. In: International Conference on Science, Technology and Innovation Indicators (27, 2023, Leiden, The Netherlands). *Proceedings... [s.l.]: International Conference on Science, Technology and Innovation Indicators*. Disponível em: <https://dapp.orvium.io/deposits/6442c231903ef57acd6dc640/view>
- Mazoni, A. F. et al. (2023). Challenges in cloud infrastructure and scientific data: technical proposals and tools applied to the SciELO database. *Anais do Workshop de Informação, Dados e Tecnologia - WIDat*, v. 6. DOI: 10.22477/vi.widat.44. Disponível em: <https://labcotec.ibict.br/widat/index.php/widat2023/article/view/44>.
- Melo, J. F.; Rockembach, M. (2025). Consentimento informado no contexto da ciência aberta: convergências, divergências e pontos de integração, *Ciência da Informação*, v. 54, n. 2. Disponível em: <https://cip.brapci.inf.br//download/346720>
- Menezes, D. P.; Galvão, R. (2025). Ciência Aberta: uma visão desapaixonada. *Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq)*, Atualidades, 16 jan. 2025. Disponível em: <https://www.gov.br/cnpq/pt-br/assuntos/noticias/atualidades/ciencia-aberta-uma-visao-desapaixonada>
- Packer, A. L.; Santos, S. M. dos. (2022). O SciELO como programa de Ciência Aberta. In: Príncipe, E.; Rode, S. de M. (org.) (2022). *Comunicação científica*. Rio de Janeiro: IBICT, 2022. Disponível em: <https://www1.abecbrasil.org.br/painel/u>
- [ploads/www/geral/E-book-Comunicacao-cientifica-aberta.pdf](https://www/geral/E-book-Comunicacao-cientifica-aberta.pdf).
- Patel, J. (2019). Bridging data silos using big data integration. *International Journal of Database Management Systems*, v. 11, n. 3, pp. 01-06.
- Rede Brasileira de Reprodutibilidade. (2025). Ciência aberta: outra visão desapaixonada. OSF. Disponível em: <https://osf.io/h2unb>
- Rodrigues, K. O. et al.. (2019). Percepção de pesquisadores de instituições públicas acerca da ciência aberta. *Ciência da Informação*, v. 48, n. 3. Disponível em: <https://brapci.inf.br/v/136333>
- Sayão, L. F.; Sales, L. F. (2016a). Algumas considerações sobre os repositórios digitais de dados de pesquisa. *Informação & Informação*, Londrina, v. 21, n. 2, pp. 90-115. Disponível em: <<http://www.uel.br/revistas/uel/index.php/informacao/article/view/27939/20122>>
- Sayão, L. F.; Sales, L. F. (2016b). Subsídios para a construção de um modelo de avaliação de sistemas de gestão de dados de pesquisa. *Ponto de Acesso*, Salvador, v. 12, n. 3, pp. 80-108. Disponível em: <https://doi.org/10.9771/rpa.v12i3.28965>
- SciELO (2021). *Guia de preparação de dados de pesquisa*. SciELO. Disponível em: https://wp.scielo.org/wp-content/uploads/Guia_preparacao_pt.pdf
- SciELO (2023). *Research data curation guidelines for editorial teams*. [s.l.]: SciELO. Disponível em: https://wp.scielo.org/wp-content/uploads/Guia_curadoria_en.pdf
- SciELO (2025). SciELO Data. Disponível em: <https://data.scielo.org/dataverse/scielodata>
- SciELO (2024a). *Critérios, política e procedimentos para a admissão e a permanência de periódicos na Coleção SciELO Brasil*. Scielo. Disponível em: <https://www.scielo.br/media/files/20240900-Criterios-SciELO-Brasil.pdf>
- SciELO (2024b). *SciELO NetWork*. Disponível

- em: <https://www.scielo.org/en/about-scielo/scielo-network/>
- Schrader, Alvin. (1983). *Toward a theory of Library and Information Science*. 1014f. Tese (PhD in Library and Information Science). Indiana University.
- Semeler, A. R., Pinto, A. L. (2019). Os diferentes conceitos de dados de pesquisa na abordagem da Biblioteconomia de dados. *Ciência da Informação*, v. 48, n. 1, pp. 113-129.
- Serra, L. G.; Segundo, J. E. S. (2021). Dos silos de dados à web dos dados: bibliotecas e o linked data. *Informação & Informação*, v. 26, n. 2.
- Souza, J. B.; Freitas, R. L. (2024). A publicação científica torna-se maior que o artigo - repositório de dados da BrJP. *BRJP*, v. 7, pp.1-2. Disponível em: <https://doi.org/10.5935/2595-0118.20240021-pt>
- UNESCO (2023). *Recomendação da UNESCO sobre Ciência Aberta*. Disponível em: <https://www.unesco.org/en/open-science/about>.
- Veiga, V.; et al..(2024). Um panorama dos princípios de dados fair: teoria, práticas e serviços. *Informação & Informação*, v. 29, n. 4.
- Wang, L. (2018). Twinning Data Science with Information Science in schools of library and Information Science. *Journal of Documentation*. V. 74, n. 6, pp. 1243–1257. Disponível em: <https://doi.org/10.1108/JD-02-2018-0036>
- Wang, Y.; Kung, L.; Byrd, T.A. (2018). Big Data analytics: understanding its capabilities and potential benefits for healthcare organizations. *Technological Forecasting and Social Change*, v. 126, pp. 3-13. <https://doi.org/10.1016/j.techfore.2015.12.019>
- Waters, J.; Powers, B. J.; Ceruti, M. G. (2009). Global interoperability using semantics, standards, science and technology (GIS3 T). *Computer Standards & Interfaces*, v. 31, pp. 1158–1166. Disponível em: <https://doi.org/10.1016/j.csi.2009.03.001>
- Zhang, H. e. at. (2025). AutoDDG: Automated Dataset Description Generation using Large Language Models. *Arxiv*. New York: Cornell University. Disponível em: <https://doi.org/10.48550/arXiv.2502.01050>.

NOTAS

¹ Pesquisador apoiado por Bolsa de Produtividade CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico – Brasil) e Bolsa CNE – Cientista do Nossa Estado/FAPERJ (Fundação Carlos Chagas Filho de Amparo à Pesquisa do Estado do Rio de Janeiro), Processo SEI E-26/204.049/2024.

² Para maiores informações, consultar: <https://www.go-fair.org/fair-principles/>.

³ Para maiores informações, consultar: https://research-and-innovation.ec.europa.eu/index_en.

⁴ Para maiores informações, consultar: <https://www.openaire.eu/>.

ⁱ Para maiores informações, consultar: https://research-and-innovation.ec.europa.eu/index_en.

ⁱⁱ⁴ Para maiores informações, consultar: <https://www.openaire.eu/>.