



DADOS CIENTÍFICOS CONFIÁVEIS NA CIÊNCIA ABERTA:

Desafios e estratégias de gestão e governança no BrCris

Washington Luís Ribeiro de Carvalho Segundo

Instituto Brasileiro de Informação em Ciência e
Tecnologia (Ibict), Brasil

washingtonsegundo@ibict.br

 <https://orcid.org/0000-0003-3635-9384>

Marcel Garcia de Souza

Instituto Brasileiro de Informação em Ciência e
Tecnologia (Ibict), Brasil

marcelsoouza@ibict.br

 <https://orcid.org/0000-0003-2255-199X>

Thiago Magela Rodrigues Dias

Centro Federal de Educação Tecnológica de
Minas Gerais (CEFET-MG), Brasil

thiagomagela@cefetmg.br

 <https://orcid.org/0000-0002-3090-9413>

Fabio Lorensi do Canto

Universidade Federal de Santa Catarina (UFSC), Brasil

fabio.lc@ufsc.br

 <https://orcid.org/0000-0002-8338-1931>

Priscila Machado Borges Sena

Instituto Brasileiro de Informação em Ciência e
Tecnologia (Ibict); Universidade Federal do
Rio Grande do Sul (UFRGS), Brasil

washingtonsegundo@ibict.br

 <https://orcid.org/0000-0002-5612-4315>

DOI: 10.22477/xiv.biredial.405

EJE TEMÁICO: Qualidade de dados; Desambiguação;
Deduplicação; Bibliometria; Cientometria.

RESUMEN

A qualidade e a confiabilidade dos dados científicos são aspectos fundamentais para assegurar análises bibliométricas e cientométricas precisas e orientadas à gestão da ciência e tecnologia. Este trabalho analisa os principais desafios enfrentados pelo BrCris (Ecossistema de Informação sobre a Pesquisa Científica Brasileira) no processo de integração e gestão de dados provenientes de múltiplas fontes, como Plataforma Lattes, OpenAlex e OasisBr. O objetivo da pesquisa é identificar os problemas de duplicidade, inconsistência e incompletude dos registros científicos, avaliar os impactos dessas fragilidades na análise da produção acadêmica brasileira e apresentar estratégias tecnológicas e metodológicas de qualificação dos dados. Os procedimentos metodológicos adotados envolveram estudo exploratório e documental, com análise de registros integrados ao BrCris e identificação de padrões de inconsistência nos metadados. Como resultados, foram identificados problemas relacionados à ausência de identificadores persistentes (DOI, ORCID), erros de preenchimento e desatualização dos registros. Entre as soluções propostas destacam-se o uso de processos automatizados de deduplicação, desambiguação de autores e instituições, validação de metadados e integração com ferramentas de análise bibliométrica. Espera-se que essas estratégias contribuam para fortalecer a governança dos dados científicos, promover a integridade da informação e consolidar o BrCris como infraestrutura nacional de suporte à Ciência Aberta.

Palavras-clave: Qualidade de dados; Desambiguação; Deduplicação; Bibliometria; Cientometria.

ABSTRACT

The quality and reliability of scientific data are fundamental aspects for ensuring accurate bibliometric and scientometric analyses aimed at science and technology management. This work analyzes the main challenges faced by BrCris (Brazilian Scientific Research Information Ecosystem) in the process of integrating and managing data from multiple sources, such as the Lattes Platform, OpenAlex, and OasisBr. The research aims to identify issues of duplication, inconsistency, and incompleteness in scientific records, assess the impacts of these weaknesses on the analysis of Brazilian academic output, and present technological and methodological strategies for data qualification. The methodological procedures adopted involved exploratory and documentary study, with an analysis of records integrated into BrCris and the identification of inconsistency patterns in the metadata. The results identified problems related to the absence of persistent identifiers (DOI, ORCID), data entry errors, and outdated records. Among the proposed solutions, the use of automated processes for deduplication, author and institution disambiguation, metadata validation, and integration with bibliometric analysis tools stand out. It is expected that these strategies will contribute to strengthening the governance of scientific data, promoting information integrity, and consolidating BrCris as a national infrastructure supporting Open Science.

Keywords: Data quality; Disambiguation; Deduplication; Bibliometrics; Scientometrics.

INTRODUÇÃO

Atualmente, tudo que se conhece sobre o surgimento e o desenvolvimento das disciplinas, a difusão do conhecimento e a evolução da ciência e tecnologia é resultado, predominantemente, da análise de publicações científicas (De Meis *et al.*, 2003; Leta *et al.*, 2006), da análise da colaboração científica (Yoshikane & Kageura, 2004) e da análise de registros de patentes (Abbas *et al.*, 2014). Esses registros constituem, historicamente, as principais fontes de dados para a avaliação, gestão e formulação de políticas científicas. No entanto, a intensificação do ambiente informacional digital, caracterizado por novas dinâmicas de produção e circulação de dados trouxe desafios inéditos relacionados à confiabilidade, à integridade e à rastreabilidade das informações científicas (Oliveira, 2020; Silva, 2022).

No contexto latino-americano, o Brasil tem uma parcela relevante na produção científica internacional, especialmente em nichos estratégicos associados à sua economia, biodiversidade e saúde pública. O país é reconhecido como líder na produção de conhecimento na América Latina (Collazo-Reyes, 2014), além de atuar como atrator de talentos e centro regional de pesquisa (Saravia & Miranda, 2004). Tal protagonismo foi acompanhado por iniciativas estruturantes que buscam aprimorar a governança da informação científica, com destaque para o desenvolvimento de plataformas digitais nacionais de registro da atividade acadêmica.

A crescente digitalização da produção científica resultou na ampliação expressiva de dados disponíveis para análises bibliométricas, cientométricas e para a avaliação das capacidades nacionais de ciência, tecnologia e inovação. Neste cenário emergem os sistemas CRIS (*Current Research Information Systems*), concebidos não apenas para integrar informações dispersas em múltiplas bases de dados, mas também para assegurar sua padronização, interoperabilidade e governança (Sivertsen, 2019). A importância desses sistemas transcende a gestão



administrativa e os posiciona como infraestruturas críticas de enfrentamento à desinformação científica, fenômeno que ameaça a credibilidade das instituições, distorce indicadores e compromete a construção de políticas públicas baseadas em evidências (Academia Brasileira de Ciências [ABC], 2024; Organização das Nações Unidas [ONU], 2024).

A Ciência Aberta, enquanto paradigma que defende a transparência, o acesso aberto e a ética na produção e disseminação do conhecimento, requer não apenas dados disponíveis, mas também dados confiáveis e verificáveis (Oliveira, 2020). Nesse sentido, os sistemas CRIS assumem um papel estratégico ao promover um ecossistema informacional robusto, que permita o monitoramento, a certificação e a validação dos dados científicos, reduzindo riscos de distorção, manipulação e circulação de informações falsas (Silva, 2022; Mancoso *et al.*, 2023).

No Brasil, o Ecossistema de Informação da Pesquisa Científica Brasileira, BrCris¹, foi desenvolvido para unificar e sistematizar informações sobre a produção científica nacional, incluindo pesquisadores, instituições, publicações, projetos e demais outputs do ciclo de pesquisa (Carvalho-Segundo *et al.*, 2024; Pinto *et al.*, 2021). Sua capacidade de exportação e integração com outras bases de dados como a Plataforma Lattes², OpenAlex³, Oasisbr⁴, entre outras, potencializa análises bibliométricas e cientométricas de caráter ampliado e detalhado. No entanto, a diversidade de fontes e a heterogeneidade dos dados inseridos geram desafios críticos relacionados à qualidade, consistência, atualização e harmonização das informações integradas.

Problemas recorrentes, como duplicidade de registros, ausência de identificadores persistentes (DOI, ORCID) e erros de preenchimento manual, criam vulnerabilidades informacionais que podem ser exploradas inadvertidamente ou intencionalmente em práticas desinformativas (Oliveira, 2020; Silva, 2022). Essas distorções comprometem não apenas análises acadêmicas, mas também a governança de políticas públicas baseadas em dados, com impactos diretos sobre a confiabilidade dos indicadores de desempenho acadêmico e dos diagnósticos de ciência e tecnologia.

Este artigo busca, assim, explorar os principais desafios e apresentar estratégias tecnológicas e metodológicas aplicadas no BrCris para assegurar a qualidade, a integridade e a confiabilidade dos dados científicos integrados. São destacadas práticas e ferramentas voltadas à desambiguação de autores e instituições, deduplicação de registros, certificação de metadados e governança dos dados, elementos fundamentais para consolidar o BrCris como infraestrutura nacional estratégica na promoção da Ciência Aberta e no enfrentamento aos riscos informacionais contemporâneos.

Ressalta-se que, a pertinência desta proposta está diretamente relacionada ao tema central da 14^a Conferência Internacional BIREDIAL-ISTEC 2025 “La desinformación: un nuevo

¹ Disponível em: <https://brcris.ibict.br/>

² Disponível em: <https://lattes.cnpq.br/>

³ Disponível em: <https://openalex.org/>

⁴ Disponível em: <https://www.oasisbr.ibict.br/>



desafío para la Ciencia Abierta”, ao tratar dos riscos e desafios informacionais decorrentes da baixa qualidade dos dados científicos em ambientes digitais e da necessidade de governança de dados confiáveis para combater a desinformação acadêmica. Inserido no eixo temático Infraestrutura tecnológica, o trabalho contribui para o debate sobre o desenvolvimento e a gestão de sistemas CRIS como ferramentas essenciais na garantia da integridade dos dados, na promoção da transparência e no fortalecimento das práticas de Ciência Aberta na América Latina.

PROBLEMAS COMUNS NA QUALIDADE DOS DADOS NO BRCRIS

A operacionalização do BrCris enquanto infraestrutura tecnológica de integração de dados científicos brasileiros envolve um ecossistema complexo de fontes heterogêneas e padrões de metadados divergentes. Essa realidade impõe uma série de desafios que ultrapassam as questões puramente técnicas, evidenciando fragilidades informacionais que impactam diretamente a governança da ciência em ambientes digitais abertos.

A qualidade dos dados é um fator decisivo não apenas para a realização de análises bibliométricas e cientométricas mais precisas, mas também para garantir a confiabilidade das informações disponibilizadas publicamente, especialmente em um cenário de intensificação da desinformação científica e de disputas informacionais que ameaçam a credibilidade das instituições de ciência e tecnologia (ABC, 2024; Oliveira, 2020; ONU, 2024).

No âmbito do BrCris, os principais problemas de qualidade dos dados podem ser agrupados em quatro categorias recorrentes:

- a) Duplicidade de Registros:** A duplicidade de registros decorre da integração de dados provenientes de fontes diversas, como a Plataforma Lattes, o OpenAlex, o Oasis-br, entre outras. Variações na grafia de nomes, ausência de identificadores persistentes ou inconsistências nos metadados favorecem a ocorrência de múltiplos registros para um mesmo autor, instituição ou produção científica. Tal situação distorce redes de colaboração, inflaciona métricas de produtividade e compromete a representatividade das análises.
- b) Inconsistência de Identificadores Persistentes:** O uso inadequado ou a ausência de identificadores persistentes, como DOI (*Digital Object Identifier*), ORCID (*Open Researcher and Contributor ID*) e ISSN (*International Standard Serial Number*), constitui um dos principais entraves à normalização dos dados integrados no BrCris. Esses identificadores desempenham um papel fundamental na garantia da precisão e da rastreabilidade das informações científicas, uma vez que possibilitam a correta vinculação entre autores, instituições e publicações em diferentes sistemas de informação.



No entanto, a realidade brasileira ainda revela baixa adesão a essas práticas. Estima-se que apenas 6% dos pesquisadores cadastrados na Plataforma Lattes possuem ORCID informado em seus currículos, o que evidencia uma lacuna significativa na adoção desses padrões e limita a interoperabilidade entre sistemas nacionais e internacionais. A ausência ou o uso inconsistente desses identificadores compromete a integridade dos dados e pode ocasionar diversos problemas, tais como: atribuição incorreta de citações e colaborações acadêmicas; dificuldades na verificação do impacto das publicações; além de entraves na integração e no cruzamento de dados provenientes de múltiplas bases.

c) Erros de Preenchimento e Padronização: A utilização de campos textuais preenchidos manualmente nas diferentes fontes de dados integradas ao BrCris, especialmente nos registros oriundos da Plataforma Lattes têm se configurado como um dos principais fatores de inconsistência informacional. Nesses campos, elementos fundamentais como nomes de autores, instituições e programas de pós-graduação podem apresentar variações ortográficas, abreviações não padronizadas e erros tipográficos, dificultando substancialmente os processos de normalização e agregação dos registros.

Essas inconsistências impactam diretamente a confiabilidade das análises derivadas dos dados, uma vez que registros de uma mesma instituição ou de um mesmo pesquisador acabam fragmentados e tratados como entidades distintas em estudos de redes de colaboração, métricas de produtividade ou visualizações de dados.

Um exemplo recorrente desse problema refere-se à diversidade de formas de representação de instituições consolidadas no cenário científico brasileiro. A Universidade Federal de Minas Gerais (UFMG), por exemplo, pode aparecer nos registros de dados com diferentes grafias, tais como “Universidade Federal de Minas Gerais”, “Univ. Federal de Minas Gerais”, “UFMG” ou “Universidade Federal de MG”, o que compromete a unificação dos registros e a acurácia das análises.

d) Desatualização e Incompletude dos Registros: A desatualização dos registros científicos e a incompletude dos metadados representam um desafio estrutural significativo para a integridade dos dados integrados ao BrCris. Esse problema decorre, em grande parte, da baixa frequência com que muitos pesquisadores atualizam seus perfis na Plataforma Lattes, assim como da ausência de práticas consolidadas de vinculação de suas produções acadêmicas a identificadores persistentes, como DOI, ORCID e ISSN.

A fragmentação e a incompletude dos dados afetam diretamente a visibilidade, o impacto e a rastreabilidade da produção científica nacional. Além dos registros desatualizados, é comum que publicações depositadas em repositórios institucionais ou



sistemas locais de gestão da informação não estejam devidamente indexadas, seja pela ausência de metadados mínimos obrigatórios, seja pela adoção de padrões heterogêneos de catalogação.

Essas lacunas prejudicam não apenas a representação adequada da atividade científica brasileira, mas também comprometem a geração de indicadores utilizados em processos de avaliação acadêmica, financiamento de pesquisa e formulação de políticas públicas baseadas em evidências.

Nesse cenário, a implementação de mecanismos automatizados para a atualização de registros, bem como a associação sistemática de identificadores persistentes às produções científicas, configura-se como uma estratégia fundamental para qualificar a base de dados do BrCris. A combinação de tecnologias de verificação automática com práticas institucionais de curadoria ativa dos registros contribui para a construção de um ecossistema de informação mais completo, confiável e alinhado aos princípios da Ciência Aberta e da integridade da informação.

Diante desses desafios, torna-se evidente que a qualidade dos dados científicos integrados ao BrCris desempenha um papel crucial na confiabilidade das análises bibliométricas e cientométricas, que dependem diretamente da precisão, completude e padronização dos registros disponíveis. Em um contexto de Ciência Aberta e crescente valorização de práticas baseadas em dados, garantir a integridade das informações é uma condição indispensável não apenas para aferir com rigor a produção e o desempenho da ciência brasileira, mas também para sustentar decisões estratégicas em ciência, tecnologia e inovação. Assim, o enfrentamento dos problemas de duplicidade, inconsistência, erros de preenchimento e desatualização dos registros não constitui apenas um requisito técnico, mas uma exigência informacional e ética, que impacta a transparência, a credibilidade e a capacidade do BrCris de se consolidar como infraestrutura nacional de governança de dados científicos confiáveis.

IMPACTO NA ANÁLISE BIBLIOMÉTRICA E CIENTOMÉTRICA

Em um ecossistema informacional orientado por dados, como é o caso do BrCris, a qualidade dos registros integrados constitui um fator determinante para a confiabilidade das análises bibliométricas e cientométricas. Essas análises desempenham um papel estratégico na avaliação do desempenho científico de pesquisadores, instituições e áreas do conhecimento, sendo amplamente utilizadas em processos de distribuição de recursos, formulação de políticas públicas, avaliação institucional e reconhecimento acadêmico.

Entretanto, conforme discutido anteriormente, a presença de problemas de duplicidade, inconsistência, erros de preenchimento e desatualização dos dados compromete significativamente a veracidade das análises realizadas. A literatura e as evidências empíricas apontam



que esses problemas podem distorcer tanto a representação quantitativa quanto qualitativa da produção científica brasileira, gerando diagnósticos imprecisos que afetam decisões de grande relevância estratégica (ABC, 2024; Oliveira, 2020).

Erros na contagem de publicações e duplicações de registros, por exemplo, podem inflacionar artificialmente a produtividade de pesquisadores e instituições, comprometendo a equidade e a justiça nos processos de avaliação. Por outro lado, a ausência de padronização na grafia de nomes de autores, títulos de publicações e designações institucionais pode fragmentar a produção científica e invisibilizar parte relevante dos resultados de pesquisa, dificultando a correta aferição do impacto acadêmico.

As redes de coautoria, amplamente utilizadas em estudos de colaboração científica, também são diretamente afetadas pela ausência de desambiguação adequada de autores e instituições. Registros com variações ortográficas ou metadados incompletos podem fazer com que um mesmo pesquisador seja tratado como múltiplos indivíduos distintos, gerando redes fragmentadas, distorcidas e com baixa representatividade das relações colaborativas reais. Dados do BrCris já revelaram casos em que variações mínimas na grafia de nomes resultaram na subestimação da colaboração entre instituições brasileiras, afetando indicadores estratégicos de internacionalização e cooperação acadêmica.

Além disso, as análises bibliométricas frequentemente se apoiam em métricas como número de citações, fator de impacto, altmetria e indicadores de visibilidade em redes sociais científicas. A inconsistência na indexação de artigos e a ausência de metadados completos comprometem diretamente essas análises, prejudicando a correta atribuição de citações e a avaliação do alcance e da relevância de determinadas publicações.

Embora o BrCris tenha avançado significativamente ao integrar dados de múltiplas fontes como OpenAlex, [Oasis.br](#), Plataforma Lattes e repositórios institucionais, a interoperabilidade entre essas bases de dados apresenta desafios substanciais. Cada fonte opera com padrões próprios de metadados, terminologias e estruturas, o que gera divergências nas informações sobre afiliação institucional, autoria e vinculação de projetos, principalmente quando se trata de dados importados de fontes internacionais.

Esse cenário reforça a necessidade de estratégias robustas de gestão da qualidade dos dados, que devem combinar soluções tecnológicas de alto desempenho como algoritmos de deduplicação, desambiguação de autores e validação automatizada de metadados com práticas institucionais de curadoria ativa e engajamento da comunidade científica na atualização e enriquecimento dos registros.

Garantir a confiabilidade das informações utilizadas nas análises bibliométricas e ciptométricas não é apenas uma condição técnica para o funcionamento adequado do BrCris, mas uma exigência informacional, política e ética para a produção de diagnósticos justos e baseados em evidências sobre a ciência brasileira. A adoção sistemática de identificadores



persistentes, o fortalecimento da interoperabilidade com outras infraestruturas científicas e a integração com ferramentas de análise e visualização de dados bibliométricos representam passos essenciais para a consolidação do BrCris como um ecossistema informacional confiável, transparente e alinhado aos princípios da Ciência Aberta e da integridade da informação.

ESTRATÉGIAS PARA MELHORIA DA QUALIDADE DOS DADOS

A complexidade e a heterogeneidade das fontes de dados integradas ao BrCris exigem a adoção de um conjunto articulado de estratégias tecnológicas e metodológicas voltadas à qualificação, padronização e certificação das informações. Esses procedimentos não apenas visam corrigir inconsistências e eliminar erros estruturais, mas constituem práticas fundamentais para fortalecer a integridade dos dados e garantir a confiabilidade das análises bibliométricas e cientométricas realizadas a partir do sistema.

Deduplicação de Registros: A deduplicação de registros constitui um processo central na gestão de dados do BrCris, com vistas a assegurar que cada publicação, autor ou instituição seja representado de maneira única e precisa. Para isso, são aplicados algoritmos de *machine learning*, técnicas de *clustering* e regras heurísticas que permitem identificar similaridades e detectar registros duplicados, mesmo em contextos com variações ortográficas ou pequenas inconsistências de preenchimento. O uso de ferramentas especializadas, como o OpenRefine, tem se mostrado eficiente na etapa de limpeza e normalização dos metadados, contribuindo para a consolidação de um banco de dados mais consistente e confiável.

Desambiguação de Pesquisadores e Instituições: A desambiguação de pesquisadores e instituições é um desafio central no BrCris. A implementação de identificadores persistentes, como ORCID para autores e ROR para instituições, minimiza ambiguidades e permite um melhor rastreamento da produção acadêmica.

Um exemplo bem-sucedido é a integração entre Lattes e OpenAlex, que demonstrou que o cruzamento de dados entre diferentes plataformas aumenta significativamente a precisão na identificação de autores e afiliações institucionais. Ao aplicar técnicas para correlacionar variantes de nomes e vinculações institucionais, é possível refinar a base de dados e eliminar duplicidades ocultas.

Para garantir a padronização e a consistência das informações integradas, o BrCris adota dicionários controlados para conjuntos de dados específicos, como nomes de instituições e programas de pós-graduação. Esses dicionários são compostos por nomenclaturas previamente validadas em outros projetos e bases de referência, sendo atualizados de forma semiautomática por meio de processos de verificação e curadoria de dados. Antes da inserção no banco de dados do BrCris, os registros passam por uma etapa de validação, na qual são comparados com esses dicionários padronizados, minimizando variações ortográficas, abreviações incon-



sistentes e erros de preenchimento. Essa abordagem reduz significativamente a fragmentação dos dados e melhora a interoperabilidade entre diferentes sistemas de informação científica, assegurando maior precisão e confiabilidade nas análises realizadas a partir do BrCris.

Validação e Certificação de Dados: Outra estratégia fundamental para garantir a qualidade dos dados no BrCris consiste na adoção de padrões internacionais de metadados, como o CERIF (*Common European Research Information Format*), que orientam a modelagem, o enriquecimento e a exportação dos registros. O uso de APIs (*Application Programming Interface*) e de processos automatizados de validação cruzada dos metadados a partir de fontes confiáveis como o OasisBr e o *OpenAIRE Research Graph* possibilita a detecção de inconsistências e a correção automática de erros, aumentando a confiabilidade e a interoperabilidade do sistema.

A certificação de dados, nesse contexto, deixa de ser um procedimento meramente técnico e assume um papel estratégico na governança da informação científica, contribuindo para a integridade dos indicadores produzidos e para a transparência dos processos de gestão da ciência.

Integração com Ferramentas Bibliométricas: Como parte das estratégias de aprimoramento da qualidade dos dados, o BrCris também busca ampliar sua integração com ferramentas especializadas de análise bibliométrica e cientométrica, como VOSviewer⁵, Visão⁶ e Gephi⁷. A exportação de dados para análises externas permite identificar padrões inconsistentes que podem ser corrigidos no próprio sistema. Além disso, a utilização de *dashboards* interativos para monitorar a qualidade dos dados facilita a detecção de anomalias e a tomada de decisão baseada em dados.

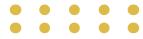
Com a adoção integrada dessas estratégias tecnológicas e metodológicas, o BrCris tem potencial para se consolidar como uma infraestrutura nacional robusta e confiável para a gestão da produção científica brasileira. Mais do que um repositório de informações, o sistema se afirma como um instrumento estratégico para a promoção da transparência, da integridade e da rastreabilidade dos dados científicos, qualificando as análises bibliométricas e cientométricas e fortalecendo a capacidade do país em produzir diagnósticos informacionais precisos e fundamentados.

A implementação de mecanismos de controle de qualidade e validação dos registros apoiados em dicionários controlados, processos automatizados de deduplicação e desambiguação, e na adoção sistemática de identificadores persistentes representa um avanço significativo na infraestrutura de dados de pesquisa no Brasil. Essas práticas não apenas aprimoram a precisão das análises, mas também ampliam a interoperabilidade do BrCris com outros ecossistemas informacionais nacionais e internacionais.

⁵ Disponível em: <https://www.vosviewer.com>

⁶ Disponível em: <https://visao.ibict.br>

⁷ Disponível em: <https://gephi.org>



Entretanto, garantir a sustentabilidade da qualidade dos dados em um sistema dinâmico e em constante expansão exige um esforço contínuo de aperfeiçoamento tecnológico, metodológico e institucional. Para além dos recursos técnicos, a efetividade do BrCris enquanto infraestrutura de Ciência Aberta e enfrentamento da desinformação científica depende do engajamento permanente da comunidade acadêmica, dos gestores de ciência e tecnologia, e das instituições de pesquisa na atualização, curadoria e validação dos registros. A governança da informação científica, nesse sentido, configura-se como uma responsabilidade compartilhada, essencial para assegurar um ecossistema informacional mais transparente, ético e socialmente comprometido.

CONSIDERAÇÕES FINAIS

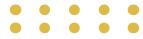
Os desafios enfrentados pelo BrCris em relação à qualidade e confiabilidade dos dados científicos revelam um cenário que ultrapassa os limites de uma infraestrutura tecnológica tradicional. O sistema se insere em uma lógica informacional contemporânea, marcada não apenas pela intensificação da produção e circulação de dados em ambientes digitais, mas também pela necessidade de enfrentamento à desinformação científica e às fragilidades dos ecossistemas informacionais.

A análise realizada neste estudo demonstrou que as inconsistências observadas tais como duplicidade de registros, erros de preenchimento, ausência de identificadores persistentes e desatualização dos dados impactam de maneira direta a fidedignidade das análises bibliométricas e cientométricas, fundamentais para os processos de avaliação, planejamento e gestão da ciência, tecnologia e inovação no Brasil.

As estratégias implementadas pelo BrCris especialmente aquelas relacionadas à deduplicação de registros, desambiguação de autores e instituições, certificação e validação de dados, além da integração com padrões e infraestruturas internacionais representam avanços relevantes na consolidação de um sistema informacional mais transparente, interoperável e confiável. No entanto, a manutenção da qualidade dos dados é um processo contínuo, que demanda não apenas o aprimoramento de soluções tecnológicas, mas, sobretudo, o fortalecimento de práticas institucionais de governança da informação científica.

O futuro do BrCris como infraestrutura estratégica da Ciência Aberta no Brasil depende de ações articuladas em múltiplas frentes: o desenvolvimento de políticas públicas que incentivem o uso ético e responsável dos identificadores persistentes; a criação de mecanismos de curadoria colaborativa que envolvam a comunidade acadêmica na validação dos registros; e o investimento em tecnologias emergentes, como inteligência artificial e machine learning, aplicadas à gestão e verificação automatizada dos dados.

Além disso, destaca-se que a construção de um ecossistema de dados científicos confiáveis é uma responsabilidade compartilhada, que requer o compromisso ativo de pesquisado-



res, gestores institucionais, agências de fomento e desenvolvedores de sistemas de informação. Apenas com esse esforço coletivo será possível consolidar o BrCris como uma referência internacional em governança de dados científicos, contribuindo não só para a valorização da ciência brasileira, mas também para o fortalecimento de práticas de integridade da informação e de combate à desinformação acadêmica em escala global.

Assim, este trabalho reforça que a qualidade dos dados não é um requisito meramente operacional, mas um compromisso ético e político com a transparência, a confiança e a justiça informacional em ambientes digitais. O BrCris, enquanto infraestrutura orientada pelos princípios da Ciência Aberta, tem um papel central na construção desse futuro.

REFERÊNCIAS

- Abbas, A., Zhang, L., & Khan, S. U. (2014). A literature review on the state-of-the-art in patent analysis. *World Patent Information*, 37, 3–13. <https://doi.org/10.1016/j.wpi.2013.12.006>
- Academia Brasileira de Ciências. (2024). *Desafios e estratégias na luta contra a desinformação científica*. Academia Brasileira de Ciências. https://www.abc.org.br/wp-content/uploads/2024/06/Livro_-Desinformacao-Cientifica_-ABC_Junho2024.pdf
- Collazo-Reyes, F. (2014). Growth of the number of indexed journals of Latin America and the Caribbean: The effect on the impact of each country. *Scientometrics*, 98, 197–209. <https://doi.org/10.1007/s11192-013-1036-2>
- Carvalho-Segundo, W. L. R., Dias, T. M. R., & Souza, M. G. (2025). BrCris: Facilitando o caminho para a Ciência Aberta através da Exportação de Conjuntos de Dados. *Ciência da Informação*, 54(2). <https://revista.ibict.br/ciinf/article/view/7205>
- De Meis, L., Velloso, A., Lannes, D., Carmo, M. S., & De Meis, C. (2003). The growing competition in Brazilian science: Rites of passage, stress and burnout. *Brazilian Journal of Medical and Biological Research*, 36, 1135–1141. <https://doi.org/10.1590/S0100-879X2003000900001>
- Leta, J., Gläzel, W., & Thijs, B. (2006). Science in Brazil. Part 2: sectoral and institutional research profiles. *Scientometrics*, 67(1), 87–105. <https://doi.org/10.1007/s11192-006-0051-y>
- Mancoso, K., Paes, A., Oliveira, T., & Massarani, L. (2023). Pesquisa em desinformação e divulgação científica: uma revisão da literatura latino-americana. *JCOM América Latina*, 6(1), A01. https://jcomal.sissa.it/article/pubid/JCOMAL_0601_2023_A01/
- Oliveira, T. M. (2020). Como enfrentar a desinformação científica? Desafios sociais, políticos e jurídicos intensificados no contexto da pandemia. *Liinc em Revista*, 16(2), e5374. <https://revisita.ibict.br/liinc/article/view/5374/5123>



Organização das Nações Unidas. (2024). *Princípios Globais das Nações Unidas para a Integridade da Informação*: recomendações para ação de múltiplas partes interessadas. Nações Unidas. https://brasil.un.org/sites/default/files/2024-07/ONU_PrincipiosGlobais_IntegridadeDaInformacao_20240624.pdf

Pinto, A. L., Dias, T. M. R., Silva, V. S., Gomes, J. C., Carvalho-Segundo, W. L. R. (2021). BrCris como um sistema de recomendação científico-tecnológica. In XXI Encontro Nacional de Pesquisa e Pós-Graduação em Ciência da Informação. *Anais do XXI ENANCIB*. Associação Nacional de Pesquisa e Pós-Graduação em Ciência da Informação (ANCIB). <https://enancib.ancib.org/index.php/enancib/xxienancib/paper/view/335>

Saravia, N. G., & Miranda, J. F. (2004). Plumbing the brain drain. *Bulletin of the World Health Organization*, 82, 608–615. <https://www.scielosp.org/article/bwho/2004.v82n8/608-615/>

Silva, M. M. (2022). *A desinformação científica no meio digital: dos mecanismos de funcionamento às estratégias de enfrentamento* [Tese de doutorado, Universidade Federal de Minas Gerais]. Repositório Institucional UFMG. <https://repositorio.ufmg.br/handle/1843/47615>

Sivertsen, G. (2019). Developing Current Research Information Systems (CRIS) as data sources for studies of research. In W. Gläzel, H. F. Moed, U. Schmoch, & M. Thelwall (Eds.), *Springer handbook of science and technology indicators* (pp. 667–683). https://doi.org/10.1007/978-3-030-02511-3_25

Yoshikane, F., & Kageura, K. (2004). Comparative analysis of coauthorship networks of different domains: The growth and change of networks. *Scientometrics*, 60, 435–446. <https://doi.org/10.1023/B:SCIE.0000034384.46076.34>

ANEXO 1

RESUMO BIOGRÁFICO DOS AUTORES

Washington L. R. Carvalho Segundo

É doutor em Informática pela Universidade de Brasília (UnB), com período sanduíche no King's College London, e mestre na mesma área pela UnB. Possui também formação em Matemática (bacharelado e licenciatura) pela mesma instituição. Atualmente, é Coordenador-Geral de Informação Científica e Tecnológica no Instituto Brasileiro de Informação em Ciência e Tecnologia (Ibict), onde lidera projetos voltados à Ciência Aberta, repositórios digitais, interoperabilidade de sistemas e gestão de dados científicos. Entre suas contribuições no Ibict, destaca-se a coordenação de iniciativas como o Oasisbr, um portal que agrupa e dissemina conteúdos científicos brasileiros de acesso aberto, e a Biblioteca Digital Brasileira de Teses e Dissertações



(BDTD), que centraliza a produção acadêmica de programas de pós-graduação em todo o país. Ele também lidera esforços relacionados à Rede dARK, um sistema descentralizado de identificadores persistentes baseado em blockchain, que visa garantir a integridade e rastreabilidade de recursos digitais, promovendo inovação em gestão de dados científicos. Sua trajetória inclui o desenvolvimento do BrCris, um sistema de integração e visualização de dados científicos no Brasil, e o projeto Laguna, voltado à criação de um lago de dados científicos de acesso aberto.

Marcel Garcia de Souza

Doutorando em Ciência da Informação pela Universidade de Brasília. Mestre em Educação em Ciências pela Universidade Federal do Rio Grande do Sul (2016). Graduado em Psicologia pela Universidade Católica de Brasília (2005). Servidor público federal; Analista em Ciência e Tecnologia no Instituto Brasileiro de Informação em Ciência e Tecnologia atuando como Coordenador de Tratamento, Análise e Disseminação da Informação Científica, além de coordenar pesquisas aplicadas voltadas à Ciência da Informação, Informação para Sustentabilidade, Avaliação do Ciclo de Vida, Informação Tecnológica.

Thiago Magela Rodrigues Dias

Doutor em Modelagem Matemática e Computacional pelo CEFET-MG (2016) tendo trabalhado com Bibliometria, Extração de Dados Científicos e Análise de Redes de Colaboração Científica. Mestre em Modelagem Matemática e Computacional pelo CEFET-MG (2008), trabalhando com Arquitetura Orientada a Serviços e Web Mining com o tema: Uma Arquitetura Orientada a Serviços para emprego em Sistemas de Mineração de Dados na Web. Possui graduação em Ciência da Computação pelo Centro Universitário de Formiga - UNIFOR (2004), além de Especialização em Produção de Software - com Ênfase em Software Livre pela UFLA (2007) e Especialização em Melhoria do Processo de Software, UFLA (2007). Atua como Professor no Centro Federal de Educação Tecnológica de Minas Gerais (CEFET-MG). Professor do Programa de Pós-graduação em Modelagem Matemática e Computacional do CEFET-MG e Professor Permanente do Programa de Pós-graduação em Ciência da Informação da Universidade Federal de Santa Catarina (UFSC). Também tem atuação como Colaborador em Projetos no Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT).

Fabio Lorensi do Canto

É Doutor (2022) e Mestre (2018) em Ciência da Informação pela Universidade Federal de Santa Catarina - UFSC. É Graduado em Biblioteconomia - Gestão da Informação pela Universidade do Estado de Santa Catarina - UDESC (2005) e Bacharel em Direito pela Faculdade CESUSC



(2012). Atua como Bibliotecário/Documentalista na Biblioteca Central da Universidade Federal de Santa Catarina - UFSC e como professor substituto do Departamento de Biblioteconomia da Universidade do Estado de Santa Catarina (UDESC). É bolsista do Projeto Laguna, coordenado pelo Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT).

Priscila Machado Borges Sena

Professora Adjunta da Universidade Federal do Rio Grande do Sul (UFRGS). Pesquisadora no Instituto Brasileiro de Informação em Ciência e Tecnologia (Ibict). Professora Colaboradora no Programa de Pós-Graduação em Gestão da Informação (PPGInfo) da Universidade do Estado de Santa Catarina (Udesc). Doutora e Mestra em Ciência da Informação pela Universidade Federal de Santa Catarina (UFSC). Bibliotecária pela Universidade Federal de Mato Grosso (UFMT).

ANEXO 2

REQUERIMENTOS DE EQUIPAMENTO TÉCNICO PARA A APRESENTAÇÃO

Computador e projetor.