Revista Ciência Aberta Lusófona
V.2, 2026
E-ISSN: 3051-7842

Pecha Kucha

Visual Analytics e Qualidade de Dados no BrCris: Uma Infraestrutura para a Ciência Aberta

Visual Analytics and Data Quality in BrCris: An Infrastructure for Open Science

Análisis visual y calidad de datos en BrCris: una infraestructura para la ciencia abierta

Thiago Magela Rodrigues Dias*
Centro Federal de Educação Tecnológica de Minas Gerais (CEFET-MG), Brasil
ORCID: https://orcid.org/0000-0001-5057-9936
Lattes: http://lattes.cnpq.br/4687858846001290
thiagomagela@cefetmg.br

Marcel Garcia de Souza
Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT), Brasil
ORCID: https://orcid.org/0000-0003-2255-199X
Lattes: http://lattes.cnpq.br/9517728665816047
marcelsouza@ibict.br

Washington L. R. Carvalho Segundo
Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT), Brasil
ORCID: https://orcid.org/0000-0003-3635-9384
Lattes: http://lattes.cnpq.br/9453481318889500
washingtonsegundo@ibict.br

* Autor Correspondente

Resumo

A crescente complexidade do ecossistema científico digital impõe desafios relacionados à integração, confiabilidade e exploração de dados científicos. Nesse contexto, o BrCris (Brazilian

CC BY

Current Research Information System) tem se consolidado como uma infraestrutura nacional estratégica para a gestão da informação científica no Brasil. Este artigo apresenta a evolução recente da plataforma, destacando avanços nos processos de curadoria de dados, incluindo deduplicação, desambiguação e validação de metadados e a modernização das interfaces de visualização e exploração analítica. A metodologia baseia-se na análise documental e técnica da arquitetura do sistema, com ênfase nas estratégias de qualificação de dados e nos mecanismos de visual analytics. Os resultados evidenciam melhorias significativas na confiabilidade dos dados, na interoperabilidade com infraestruturas internacionais e na experiência do usuário por meio de dashboards interativos e navegação entre entidades. Conclui-se que o BrCris reforça seu papel como infraestrutura alinhada aos princípios FAIR e à Ciência Aberta, ampliando a transparência, a reutilização de dados e a capacidade analítica sobre a ciência brasileira.

Palavras-chave:

BrCris; Ciência Aberta; Qualidade de Dados; Visualização Científica; FAIR; CRIS.

Abstract

The increasing complexity of the digital scientific ecosystem poses challenges related to the integration, reliability, and exploitation of scientific data. In this context, BrCris (Brazilian Current Research Information System) has consolidated itself as a strategic national infrastructure for the management of scientific information in Brazil. This article presents the recent evolution of the platform, highlighting advances in data curation processes, including deduplication, disambiguation, and metadata validation, and the modernization of visualization and analytical exploration interfaces. The methodology is based on the documentary and technical analysis of the system's architecture, with emphasis on data qualification strategies and visual analytics mechanisms. The results show significant improvements in data reliability, interoperability with international infrastructures, and user experience through interactive dashboards and navigation between entities. It is concluded that BrCris reinforces its role as an infrastructure aligned with the FAIR principles and Open Science, expanding transparency, data reuse, and analytical capacity regarding Brazilian science.

Keywords:

BrCris; Open Science; Data Quality; Scientific Visualization; FAIR; CRIS.

Resumen

La creciente complejidad del ecosistema científico digital plantea desafíos relacionados con la integración, confiabilidad y explotación de datos científicos. En este contexto, BrCris (Sistema

Brasileño de Información sobre Investigación Actual) se ha consolidado como una infraestructura nacional estratégica para la gestión de información científica en Brasil. Este artículo presenta la evolución reciente de la plataforma, destacando los avances en los procesos de curación de datos, incluyendo la deduplicación, la desambiguación y la validación de metadatos, así como la modernización de las interfaces de visualización y exploración analítica. La metodología se basa en el análisis documental y técnico de la arquitectura del sistema, con énfasis en las estrategias de cualificación de datos y los mecanismos de análisis visual. Los resultados muestran mejoras significativas en la confiabilidad de los datos, la interoperabilidad con infraestructuras internacionales y la experiencia del usuario a través de paneles interactivos y la navegación entre entidades. Se concluye que BrCris refuerza su rol como infraestructura alineada con los principios FAIR y la Ciencia Abierta, ampliando la transparencia, la reutilización de datos y la capacidad analítica en relación con la ciencia brasileña.

Palabras clave:

BrCris; Ciencia Abierta; Calidad de Datos; Visualización Científica; FAIR; CRIS.

Introdução

A integração de dados provenientes de diferentes sistemas, como currículos acadêmicos, bases bibliográficas, repositórios institucionais e plataformas internacionais, frequentemente resulta em problemas como duplicidade de registros, ambiguidade na identificação de autores e instituições, inconsistências na representação de metadados e lacunas informacionais (Oliveira, 2020; Silva, 2022). Esses problemas comprometem não apenas a qualidade das análises realizadas, mas também a confiabilidade dos indicadores utilizados para avaliação científica, formulação de políticas públicas e tomada de decisão estratégica (ABC, 2024; ONU, 2024).

Nesse contexto, os sistemas CRIS (Current Research Information Systems) assumem um papel central na organização e governança da informação científica, ao oferecerem infraestruturas capazes de integrar, padronizar e disponibilizar dados sobre a atividade de pesquisa (Sivertsen, 2019). Mais do que ferramentas de gestão, esses sistemas configuram-se como infraestruturas sociotécnicas que sustentam a produção, disseminação e avaliação do conhecimento científico em escala institucional, nacional e internacional.

No Brasil, o BrCris emerge como uma iniciativa estratégica voltada à consolidação de um ecossistema nacional de informação científica, integrando dados provenientes de múltiplas fontes e promovendo sua organização em um ambiente interoperável e orientado à análise. A plataforma adota padrões internacionais de modelagem e interoperabilidade, como a ontologia VIVO e o modelo CERIF, além de incorporar identificadores persistentes como DOI, ORCID e ROR, elementos fundamentais para a rastreabilidade e integração dos dados científicos.

A evolução recente do BrCris representa um avanço significativo em direção à melhoria da qualidade e confiabilidade dos dados integrados, bem como à ampliação das capacidades analíticas da plataforma. A nova versão incorpora mecanismos mais sofisticados de curadoria de dados, incluindo processos automatizados de deduplicação e desambiguação, além de interfaces redesenhadas que favorecem a exploração interativa das informações científicas.

Além disso, o BrCris alinha-se aos princípios da Ciência Aberta e às diretrizes FAIR, promovendo dados mais acessíveis, interoperáveis e reutilizáveis. Esse alinhamento reforça o papel da plataforma como infraestrutura estratégica para a governança da informação científica no Brasil, contribuindo para a transparência, a reprodutibilidade e o reuso dos dados em diferentes contextos de pesquisa e gestão.

Diante desse cenário, este artigo tem como objetivo apresentar e analisar os avanços recentes do BrCris, com ênfase nos processos de curadoria de dados e nas novas interfaces de visualização, discutindo sua contribuição para o fortalecimento da Ciência Aberta e para a consolidação de uma infraestrutura nacional de dados científicos confiáveis.

Metodologia

O processo metodológico do BrCris pode ser compreendido como um pipeline estruturado de integração e curadoria de dados científicos, composto por etapas sequenciais e interdependentes (Figura 1).

Figura 1

Pipeline de Integração,Qualificação e Visualização dos Dados no BrCris



Nota. Os autores (2026).

Após a coleta, os dados passam por um processo de transformação e padronização, no qual são convertidos para um modelo comum de representação. Essa etapa inclui a normalização de campos textuais, a harmonização de identificadores e a aplicação de vocabulários controlados para reduzir inconsistências na representação de entidades científicas. A padronização é

fundamental para garantir a interoperabilidade entre diferentes fontes e permitir a consolidação dos dados em um ambiente integrado.

Em seguida, são aplicados mecanismos de detecção de inconsistências, baseados em técnicas de similaridade textual, comparação de atributos e análise de metadados. Esses mecanismos permitem identificar registros duplicados, ambiguidades na identificação de autores e instituições, e inconsistências estruturais nos dados. A detecção dessas anomalias constitui uma etapa crítica para a qualidade da base integrada.

A etapa de deduplicação e desambiguação envolve a aplicação de algoritmos que combinam regras determinísticas, análise de similaridade e uso de identificadores persistentes. A deduplicação visa consolidar registros redundantes, enquanto a desambiguação busca resolver ambiguidades na identificação de entidades, especialmente autores, publicações e instituições. O uso de identificadores como DOI, ORCID e ROR desempenha um papel central nesse processo, contribuindo para a precisão e rastreabilidade dos dados.

Posteriormente, os dados passam por um processo de validação e enriquecimento, no qual são comparados com fontes externas e complementados com informações adicionais. Essa etapa amplia a consistência dos metadados e fortalece a interoperabilidade do sistema, permitindo a integração com infraestruturas internacionais.

Por fim, os dados qualificados são disponibilizados por meio de diferentes interfaces analíticas, incluindo mecanismos de busca, dashboards interativos e visualizações semânticas. Essa abordagem permite tratar a visualização não apenas como camada de apresentação, mas como componente central do processo analítico, possibilitando a exploração dinâmica e a geração de resultados sobre a produção científica.

Resultados

Os resultados obtidos a partir da implementação das estratégias de curadoria e visualização no BrCris evidenciam avanços significativos na qualidade, confiabilidade e capacidade analítica dos dados científicos integrados à plataforma. Esses avanços podem ser compreendidos a partir da análise de três dimensões principais: a melhoria da qualidade dos dados, a evolução das interfaces de visualização e o impacto na Ciência Aberta.

No que se refere à qualidade dos dados, a aplicação de técnicas de deduplicação e desambiguação resultou em uma redução expressiva da redundância informacional, permitindo a consolidação de registros anteriormente fragmentados. Essa melhoria contribui diretamente para a precisão das métricas de produção científica, evitando distorções decorrentes da contagem duplicada de publicações ou da fragmentação de perfis de pesquisadores. Além disso, a padronização dos

metadados e o uso de identificadores persistentes ampliaram a consistência e a rastreabilidade das informações, fortalecendo a confiabilidade das análises bibliométricas e cientométricas.

Outro resultado relevante refere-se ao enriquecimento dos dados por meio da integração com fontes externas, o que possibilitou a complementação de metadados e a validação cruzada das informações. Esse processo contribui para a construção de uma base de dados mais robusta e interoperável, alinhada às melhores práticas internacionais de gestão da informação científica.

No âmbito das interfaces, a nova versão do BrCris apresenta avanços significativos na forma como os dados são disponibilizados e explorados pelos usuários. A interface de busca foi aprimorada, incorporando mecanismos de relevância, filtros dinâmicos, permitindo consultas mais precisas e flexíveis.

A interface de busca do BrCris também se destaca pela incorporação de mecanismos avançados de recuperação da informação, que combinam técnicas de indexação eficiente. Diferentemente de sistemas tradicionais baseados apenas em correspondência lexical, o ambiente de busca avançada do BrCris permite recuperar resultados mais precisos ao considerar múltiplos atributos dos registros, como títulos, autores, instituições e ano de publicação.

A partir de um resultado de busca, é possível acessar diretamente perfis de pesquisadores, instituições, periódicos e redes de colaboração associadas aos pesquisadores, promovendo uma experiência de exploração conectada e contextualizada. Essa funcionalidade transforma a busca em um ponto de partida para análises mais aprofundadas, favorecendo a construção de percursos investigativos que articulam diferentes tipos de informação científica em uma mesma interface.

Adicionalmente, a interface de busca incorpora funcionalidades voltadas à exportação e reutilização dos dados recuperados, permitindo que os resultados sejam facilmente convertidos em formatos estruturados para uso em ferramentas externas de análise bibliométrica e cientométrica. Essa capacidade reforça o alinhamento do BrCris com os princípios da Ciência Aberta, ao facilitar o reuso dos dados e promover maior transparência nos processos analíticos. Ao integrar busca, análise e exportação em um único ambiente, a plataforma amplia seu potencial como ferramenta estratégica para pesquisadores e gestores interessados na exploração sistemática da produção científica.

Adicionalmente, as visualizações em rede permitem analisar relações de colaboração científica, evidenciando conexões entre pesquisadores, instituições e áreas do conhecimento. Essas visualizações ampliam a compreensão das dinâmicas de produção científica e favorecem a identificação de padrões estruturais no ecossistema científico.

Outro recurso de destaque na plataforma BrCris consiste nas visualizações associadas a entidades individuais, como pesquisadores, instituições e publicações, que oferecem uma visão integrada e

contextualizada dos registros científicos. Essas interfaces permitem consolidar, em um único ambiente, informações provenientes de diferentes fontes, organizadas de forma estruturada e semanticamente relacionada. Ao reunir dados sobre produção científica, vínculos institucionais, colaborações e trajetórias acadêmicas, essas visualizações contribuem para uma compreensão mais abrangente e articulada das entidades analisadas, superando abordagens fragmentadas tradicionais.

Adicionalmente, essas funcionalidades assumem um papel fundamental na análise de redes de colaboração científica, ao evidenciar padrões estruturais e dinâmicas de produção do conhecimento. Essa perspectiva relacional não apenas enriquece a análise científica, mas também contribui para a formulação de políticas de ciência e tecnologia mais informadas, reforçando o papel do BrCris como uma infraestrutura analítica no contexto da Ciência Aberta.

Conclusões

A análise da evolução recente do BrCris evidencia a consolidação da plataforma como uma infraestrutura estratégica para a gestão e governança da informação científica no Brasil. A modernização das interfaces e a incorporação de recursos de visualização interativa representam um avanço significativo na forma como os dados científicos são explorados, permitindo uma interação mais intuitiva e orientada à descoberta de conhecimento. Essas melhorias ampliam o potencial do BrCris como ferramenta analítica, favorecendo sua utilização por diferentes perfis de usuários, incluindo pesquisadores, gestores e tomadores de decisão.

O alinhamento da plataforma aos princípios FAIR reforça seu papel como infraestrutura interoperável e orientada ao reuso, contribuindo para a integração do Brasil ao ecossistema global de Ciência Aberta. Nesse sentido, o BrCris não apenas integra dados, mas promove uma nova lógica de organização e governança da informação científica, baseada em transparência, acessibilidade e confiabilidade.

Adicionalmente, os avanços observados evidenciam o potencial do BrCris como instrumento de combate à desinformação científica, ao garantir a integridade e a rastreabilidade dos dados utilizados em análises e avaliações. Essa dimensão torna-se especialmente relevante em um contexto em que a qualidade da informação científica é fundamental para a credibilidade das instituições e para a tomada de decisões informadas.

Agradecimentos

Os autores agradecem ao Instituto Brasileiro de Informação em Ciência e Tecnologia - IBICT pelo suporte na Pesquisa.

Conflito de Interesses

Os autores declaram não haver conflitos de interesses.

Disponibilização dos Dados de Investigação

https://brcris.ibict.br/pt-BR

CRediT – Contribuições dos Autores

Washington L. R. Carvalho Segundo | Escrita – revisão e edição.
Marcel Garcia de Souza | Escrita – revisão e edição.
Thiago Magela Rodrigues Dias | Escrita – revisão e edição.

Referências

Academia Brasileira de Ciências (ABC). (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

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://revista.ibict.br/liinc/article/view/5374/5123

Organização das Nações Unidas (ONU). (2024). Princípios globais das Nações Unidas para a integridade da informação: recomendações para ação de múltiplas partes interessadas. Nações Unidas. https://brasil.un.org/pt-br/download/165221/274644

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. Glänzel, H. F. Moed, U. Schmoch, & M. Thelwall (Eds.), *Springer handbook of science and technology indicators* (pp. 667–683). Springer. https://doi.org/10.1007/978-3-030-02511-3_25