Advancing research information interoperability: Practices, challenges, and opportunities in modern CRIS ecosystems

Abstract
This paper examines the evolving role of Current Research Information Systems (CRIS) in enhancing research information quality, interoperability, and evaluation in the context of modern trends such as Artificial Intelligence (AI), open infrastructures, and open-source solutions. Using the BrCris as a case study, we analyze the integration of national and international data sources and highlight strategies for improving metadata enrichment, data harmonization, and research assessment practices. The study identifies key challenges in achieving FAIR principles and presents practical recommendations to support more trustworthy and interoperable research information ecosystems. This contribution aligns with the CRIS2026 themes by discussing the potential of AI-driven processes, open infrastructures, and governance approaches that strengthen research information management across institutional and national boundaries.

Keywords

Access and Evaluation of Science, BrCris, Information Management, Open Science.

Introduction

Currently, much of what is known about the emergence and development of disciplines, the dissemination of knowledge, and the evolution of science and technology is predominantly the result of the analysis of scientific publications (Meis et al., 2003; Leta et al., 2006), the analysis of scientific collaboration (Yoshikane & Kageura, 2004), and the analysis of patent records (Abbas et al., 2014). These records have historically constituted the main sources of data for the evaluation, management, and formulation of scientific policies. However, the intensification of the digital informational environment, characterized by new dynamics of data production and circulation, has brought unprecedented challenges related to the reliability, integrity, and traceability of scientific information (Oliveira, 2020; Silva, 2022).

In the Latin American context, Brazil has a significant share in international scientific production, especially in strategic niches associated with its economy, biodiversity, and public health. The country is recognized as a leader in knowledge production in Latin America (Collazo-Reyes, 2014), in addition to acting as a talent attractor and regional research center (Saravia & Miranda, 2004). This leading role has been accompanied by structuring initiatives that seek to improve the governance of scientific information, with emphasis on the development of national digital platforms for recording academic activity.

The increasing digitization of scientific production has resulted in a significant expansion of data available for bibliometric and scientometric analyses and for the evaluation of national science, technology, and innovation capabilities. In this scenario, CRIS (Current Research Information Systems) have become key infrastructures, designed not only to integrate information dispersed across multiple databases, but also to ensure its standardization, interoperability, and governance (Sivertsen, 2019).

The importance of these systems transcends administrative management and positions them as critical infrastructures for combating scientific misinformation, a phenomenon that threatens the credibility of institutions, distorts indicators, and compromises the construction of evidence-based public policies (ABC, 2024; UN, 2024).

In this sense, CRIS systems assume a strategic role by promoting a robust informational ecosystem that allows for the monitoring, certification, and validation of scientific data, reducing the risks of distortion, manipulation, and circulation of false information (Silva, 2022; Mancoso et al., 2023).

In Brazil, the Brazilian Scientific Research Information Ecosystem, BrCris (https://brcris.ibict.br/), was developed to unify and systematize information on national scientific production, including researchers,

institutions, publications, projects, and other outputs of the research cycle (Carvalho-Segundo, Dias & Souza, 2025; Pinto et al., 2021). Its ability to export and integrate with other databases such as the Lattes Platform, OpenAlex, Oasisbr, among others, enhances bibliometric and scientometric analyses of both broad and detailed scope. However, the diversity of sources and the heterogeneity of the data included generate critical challenges related to the quality, consistency, updating, and harmonization of the integrated information.

Recurring problems, such as duplicate records, lack of persistent identifiers (DOI, ORCID), and manual data entry errors, create informational vulnerabilities that can be inadvertently or intentionally exploited in disinformation practices (Oliveira, 2020; Silva, 2022). These distortions compromise not only academic analyses but also the governance of data-driven public policies, with direct impacts on the reliability of academic performance indicators and science and technology diagnoses.

This article seeks, therefore, to explore the main challenges and present technological and methodological strategies applied in BrCris to ensure the quality, integrity, and reliability of integrated scientific data. Practices and tools aimed at disambiguating authors and institutions, deduplicating records, certifying metadata, and governing data are highlighted, fundamental elements for consolidating BrCris as a strategic national infrastructure in promoting Open Science and addressing contemporary informational risks.

Development

Ensuring the reliability of the information used in bibliometric and scientometric analyses is not only a technical condition for the proper functioning of BrCris, but also an informational, political, and ethical requirement for the production of fair and evidence-based diagnoses of Brazilian science. The systematic adoption of persistent identifiers, the strengthening of interoperability with other scientific infrastructures, and integration with bibliometric data analysis and visualization tools represent essential steps for consolidating BrCris as a reliable, transparent informational ecosystem aligned with the principles of Open Science and information integrity.

The complexity and heterogeneity of the data sources integrated into BrCris require the adoption of an articulated set of technological and methodological strategies aimed at qualifying, standardizing, and certifying information. These procedures not only aim to correct inconsistencies and eliminate structural errors, but also constitute fundamental practices for strengthening data integrity and ensuring the reliability of bibliometric and scientometric analyses performed using the system.

***Record Deduplication**: Record deduplication is a central process in BrCris data management, aiming to ensure that each publication, author, or institution is represented uniquely and accurately. This is achieved through the application of machine learning algorithms, clustering techniques, and heuristic rules that identify similarities and detect duplicate records, even in contexts with spelling variations or minor inconsistencies in data entry. The use of specialized tools, such as OpenRefine, has proven efficient in the metadata cleaning and normalization stage, contributing to the consolidation of a more consistent and reliable database.*

***Researcher and Institution Disambiguation**: Disambiguation of researchers and institutions is a central challenge in BrCris. The implementation of persistent identifiers, such as ORCID for authors and ROR for institutions, minimizes ambiguities and allows for better tracking of academic output.*

A successful example is the integration between Lattes and OpenAlex, which demonstrated that cross-referencing data between different platforms significantly increases the accuracy in identifying authors and institutional affiliations. By applying techniques to correlate name variants and institutional links, it is possible to refine the database and eliminate latent or non-obvious duplicate records.

To ensure the standardization and consistency of integrated information, BrCris adopts controlled dictionaries for specific datasets, such as names of institutions and graduate programs. These dictionaries are composed of nomenclatures previously validated in other projects and reference databases, and are

updated semi-automatically through data verification and curation processes. Before being inserted into the BrCris database, records undergo a validation step, in which they are compared with these standardized dictionaries, minimizing spelling variations, inconsistent abbreviations, and data entry errors. This approach significantly reduces data fragmentation and improves interoperability between different scientific information systems, ensuring greater accuracy and reliability in analyses performed using BrCris.

***Data Validation and Certification**: Another fundamental strategy to ensure data quality in BrCris is the adoption of international metadata standards, such as CERIF (Common European Research Information Format), which guide the modeling, enrichment, and export of records. The use of APIs (Application Programming Interface) and automated cross-validation processes of metadata from reliable sources such as OasisBr and OpenAIRE Research Graph enables the detection of inconsistencies and the automatic correction of errors, increasing the reliability and interoperability of the system.*

In this context, data certification goes beyond a purely technical procedure and assumes a strategic role and assumes a strategic role in the governance of scientific information, contributing to the integrity of the indicators produced and to the transparency of science management processes.

***Integration with Bibliometric Tools**: As part of the strategies to improve data quality, BrCris also seeks to expand its integration with specialized bibliometric and scientometric analysis tools, such as VOSviewer, Visão, and Gephi. Exporting data for external analysis allows for the identification of inconsistent patterns that can be corrected within the system itself. Furthermore, the use of interactive dashboards to monitor data quality facilitates anomaly detection and data-driven decision-making.*

Results

With the integrated adoption of these technological and methodological strategies, BrCris has the potential to consolidate itself as a robust and reliable national infrastructure for the management of Brazilian scientific production. More than just an information repository, the system establishes itself as a strategic instrument for promoting the transparency, integrity, and traceability of scientific data, qualifying bibliometric and scientometric analyses and strengthening the country's capacity to produce accurate and well-founded informational diagnoses.

The implementation of quality control and record validation mechanisms supported by controlled dictionaries, automated deduplication and disambiguation processes, and the systematic adoption of persistent identifiers represents a significant advance in research data infrastructure in Brazil. These practices not only improve the accuracy of analyses but also expand the interoperability of BrCris with other national and international informational ecosystems.

However, ensuring the sustainability of data quality in a dynamic and constantly expanding system requires a continuous effort of technological, methodological, and institutional improvement. Beyond technical resources, the effectiveness of BrCris as an Open Science infrastructure and in combating scientific misinformation depends on the ongoing engagement of the academic community, science and technology managers, and research institutions in updating, curating, and validating records. In this sense, the governance of scientific information is a shared responsibility, essential for ensuring a more transparent, ethical, and socially committed information ecosystem.

The application of structured data governance practices within BrCris has resulted in qualitative improvements in data consistency and interpretability across integrated sources. The combined use of automated deduplication, persistent identifiers, and controlled vocabularies has reduced ambiguities in author and institutional records, enabling more coherent aggregation of research outputs. These improvements are particularly relevant for longitudinal and comparative analyses, where inconsistencies in metadata can significantly distort indicators related to productivity, collaboration, and institutional performance.

Another relevant outcome observed is the enhancement of interoperability between BrCris and external research information infrastructures. Alignment with international standards and metadata models facilitates data exchange with global platforms, supporting broader visibility of national research outputs and improving compatibility with international evaluation frameworks. As a result, BrCris contributes not only to internal management and policy formulation but also to positioning Brazilian research within transnational analytical contexts, reinforcing the role of CRIS as mediators between local data production and global research assessment systems.

Despite these advances, the results also reveal structural limitations inherent to large-scale, multi-source research information systems. Variations in data completeness, delays in updates, and uneven adoption of persistent identifiers across institutions continue to affect the uniformity of records. Nevertheless, the identification of these limitations constitutes an important outcome in itself, as it informs continuous improvement strategies and supports the development of monitoring mechanisms aimed at strengthening data quality over time. In this sense, BrCris demonstrates the capacity not only to aggregate data, but also to expose informational gaps that are critical for evidence-based governance.

Considerations

The challenges faced by BrCris regarding the quality and reliability of scientific data reveal a scenario that goes beyond the limits of a traditional technological infrastructure. The system is embedded in a contemporary informational logic, marked not only by the intensification of data production and circulation in digital environments, but also by the need to address scientific misinformation and the weaknesses of informational ecosystems.

The analysis carried out in this study demonstrated that the inconsistencies observed, such as duplicate records, data entry errors, lack of persistent identifiers, and outdated data, directly impact the reliability of bibliometric and scientometric analyses, which are fundamental to the processes of evaluation, planning, and management of science, technology, and innovation in Brazil.

The strategies implemented by BrCris, especially those related to record deduplication, disambiguation of authors and institutions, data certification and validation, as well as integration with international standards and infrastructures, represent significant advances in the consolidation of a more transparent, interoperable, and reliable information system. However, maintaining data quality is an ongoing process that demands not only the improvement of technological solutions, but, above all, the strengthening of institutional practices for the governance of scientific information.

From a broader perspective, this study reinforces the understanding of CRIS as socio-technical infrastructures whose effectiveness depends on both technological robustness and institutional commitment. The experience of BrCris illustrates how data quality, interoperability, and evaluation practices are deeply intertwined with governance models, community engagement, and policy frameworks. By articulating technical solutions with institutional processes, CRIS can evolve from passive repositories into active instruments for research transparency and responsible evaluation.

Looking ahead, the findings suggest that future developments in CRIS should increasingly incorporate AI-assisted processes, open-source architectures, and participatory governance models to address scalability and sustainability challenges. Strengthening collaboration between national and international research information infrastructures will be essential to support FAIR principles and trustworthy metrics.

References

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.

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

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

Leta, J., Glänzel, W., & Thijs, B. (2006). Science in Brazil. Part 2: Sectoral and institutional research profiles. Scientometrics, 67(1), 87–105. https://doi.org/10.1007/s11192-006-0058-5

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/

Meis, L.de, 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/S0100-879X2003000900011

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/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 (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. Glänzel, 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.