

Visual analytics for national CRIS: exploring brazilian research data through the BrCris platform

Abstract

This paper presents an analysis of the data visualization and exploration capabilities of the Brazilian Current Research Information System (BrCris), a national platform designed to integrate, certify and disseminate information on scientific, technological and innovation activities in Brazil. Grounded in international CRIS standards such as CERIF and VIVO, BrCris aggregates heterogeneous open data sources and provides interactive visual interfaces that support data discovery, filtering, analysis and reuse. The study highlights how visual analytics mechanisms embedded in BrCris enhance access to research information, facilitate evidence-based analysis and foster interoperability within the context of Open Science. By examining the platform's search interface, dashboards and export functionalities, the paper discusses the role of visualization as a central component for understanding national research ecosystems and supporting decision-making processes.

Keywords

Data Visualization, BrCris, Visual Interfaces, Open Science.

Introduction

In recent decades, the increasing volume and heterogeneity of research-related data have intensified the demand for infrastructures capable not only of integrating information, but also of enabling its effective exploration and interpretation. In this context, data visualization has emerged as a key element for transforming complex research information into actionable knowledge, supporting analytical reasoning, transparency and evidence-based decision-making. Current Research Information Systems (CRIS) play a central role in this process by combining structured data integration with visual interfaces that allow users to navigate, filter and analyze research ecosystems at multiple levels.

The production of scientific knowledge is a time-consuming and incremental process (HUANG et al., 2021). Identifying the state of the art in a given area of knowledge is largely conducted through the careful analysis of specialized publications that concentrate relevant information on the topic in question.

Currently, everything that 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 (DE MEIS et al. 2003; LETA et al., 2006), the analysis of scientific collaboration (YOSHIKANE and KAGEURA, 2004), and the analysis of patent records (ABBAS et al., 2014).

The production and dissemination of scientific knowledge are central activities for the social, economic, and technological development of any nation. In the Brazilian context, the richness of academic production is dispersed across a plurality of institutional sources, digital repositories, thematic databases, researchers' curricula, and other instances of recording and disseminating scientific activity. Despite this plurality, the absence of a computational infrastructure capable of integrating this information in an interoperable way has hindered systemic analyses, the efficient management of scientific information, and the formulation of evidence-based public policies.

From this scenario, initiatives began to emerge aimed at creating systems that manage the academic output of an institution, country, or area of knowledge. These systems are known by the acronym CRIS (Current Research Information Systems) and aim to aggregate information from various databases in order to provide consolidated reports and data so that researchers in the field can analyze how production occurs in their countries or areas of expertise.

CRIS defines an information system about the entire ecosystem of scientific production. All information about the scientific research cycle is organized in one place, from dissemination, through Projects, Researchers, Research Institutions and Laboratories, to the outputs of scientific research, such as scientific articles, theses, dissertations, books, book chapters, patents and scientific datasets (SIVERTSEN, 2019).

The BrCris Platform, coordinated by the Brazilian Institute of Information in Science and Technology (IBICT), emerges as a response to this gap, offering an ecosystem of structured and integrated data on Brazilian scientific, technical, and technological production. The platform adopts the principles of Open Science and follows international interoperability standards such as CERIF (Common European Research Information Format) and the VIVO ontology, promoting not only data integration but also certification, deduplication, and availability in reusable formats.

Among the challenges faced in computer science applied to this context are: (1) handling large volumes of heterogeneous data; (2) identifying and disambiguating entities such as authors, publications, and institutions; (3) maintaining complex relationships between data objects; (4) adhering to internationally established vocabularies and ontologies; (5) exporting data in formats compatible with widely used analytical tools.

The BrCris Platform has advantages over other solutions for managing and analyzing scientific and technological information. Its main feature is its ability to integrate, in a unified and standardized environment, data from multiple reliable sources such as curricula from the Lattes Platform, publication databases, and institutional records, allowing a comprehensive, up-to-date, and contextualized view of national scientific production. In addition, BrCris adopts international interoperability standards (such as the CERIF model), which facilitates the export, reuse, and comparison of data at an international level.

Development

The methodological architecture of the BrCris Platform combines strategies for data collection, integration, semantic transformation, ontological modeling, visualization, and export. The project is based on the premise that scientific data is generated by multiple actors and deposited in multiple repositories, with different standards and formats. Therefore, a pipeline capable of handling this heterogeneity with technical rigor and adherence to standards is necessary.

Initially, more than 15 data sources of interest to BrCris were mapped, including: Lattes Platform, OpenAlex, DOAJ, Oasisbr, Espacenet, CAPES, Wikidata, NDLTD, INPI, among others. Data collection was carried out via REST APIs whenever available, or through OAI-PMH harvesting and specialized crawlers in cases where APIs were not available. In all cases, open access data was chosen, in accordance with the principles of Open Science.

The data is transformed according to two complementary models: a logical model based on Entity-Relationship (E-R), aimed at relational storage and visualization via dashboards; and a semantic model based on the VIVO ontology, aimed at semantic interoperability, the provision of Linked Open Data, and graph-based navigation.

The semantic integration process includes author disambiguation, record deduplication, and enrichment with metadata from external sources. In addition, a data certification mechanism was implemented, based on source reliability, internal consistency, and the presence of persistent identifiers (e.g., ORCID, DOI, ROR).

Finally, the data is made available via three interfaces: (1) a search interface based on Search-UI and Elasticsearch; (2) interactive dashboards with real-time filtering; and (3) semantic visualization, with export in CSV, GraphML, and RDF.

Within this architectural and methodological framework, data visualization is not treated merely as a presentation layer, but as an integral analytical component of the BrCris platform. The combination of interoperable data models, certified records and interactive visual interfaces enables users to construct customized analytical views of the Brazilian research ecosystem. This approach supports exploratory analysis, comparative studies and the identification of patterns and trends, reinforcing the role of BrCris as a research intelligence infrastructure aligned with international CRIS practices.

Results

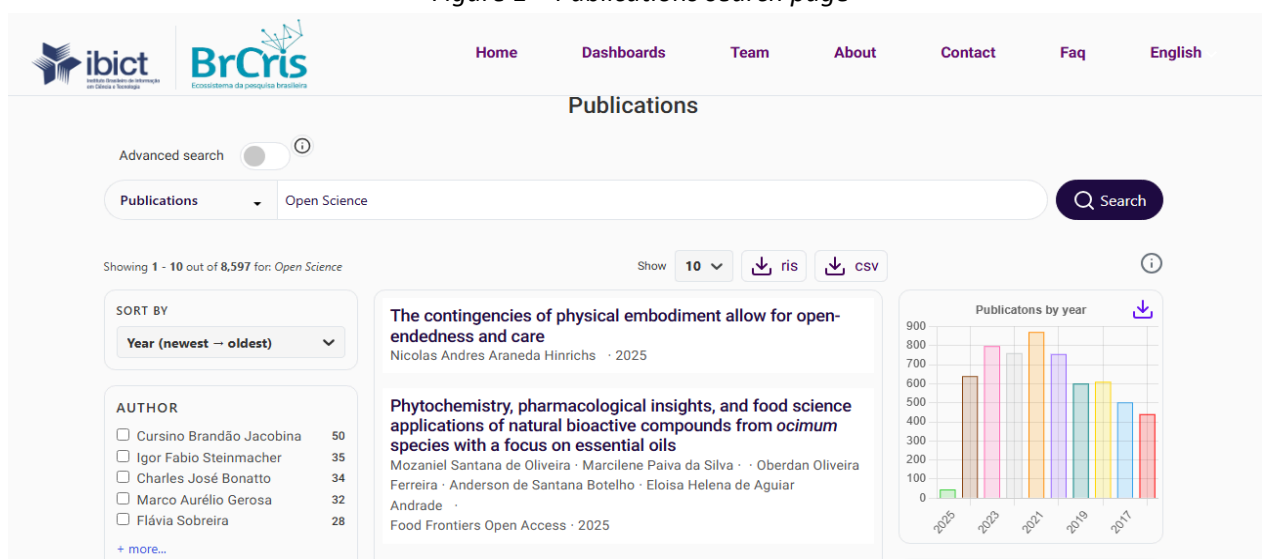
The results presented in this section focus on the visualization and interaction mechanisms offered by the BrCris platform, highlighting how these tools support the exploration, analysis and reuse of integrated

research data. Rather than emphasizing raw data aggregation, the analysis centers on the user-oriented visual interfaces that enable dynamic filtering, contextual navigation and the generation of analytical subsets, which are essential for understanding complex relationships within national research information systems.

All interfaces are specifically designed to simplify the process of accessing and certifying available datasets, providing a user-friendly and visually appealing experience, allowing for more intuitive interaction with the information contained in the datasets.

The search interface constitutes the primary entry point for exploring the BrCris data ecosystem and plays a fundamental role in supporting visual data discovery. Designed to provide intuitive access to certified and integrated research information, this interface allows users to perform refined searches across multiple entities, such as publications, researchers and institutions, while dynamically applying filters that immediately reshape the result set. Figure 1 illustrates the BrCris search interface for publications, highlighting its role as a visual and interactive mechanism for navigating large-scale research data.

Figure 1 – Publications search page



In the available visualizations, it is possible to apply various filters to virtually all fields that make up an entity. This allows the creation of specific visualizations, taking into account only the applied filters. These subsets of data can be analyzed from different perspectives and visualizations.

By accessing the "Dashboards and Indicators" option, it is possible to perform analyses of various entities and explore various metrics associated with them. These visualizations offer an interactive experience, allowing you to apply various filters and generate customized visualizations on the datasets (Figure 2).

Figure 2 – Filtering and Export Panel



A major advantage of the platform is the ability to export data subsets in standardized formats, such as .csv files. These files can be easily imported into various other data analysis and visualization tools, providing greater flexibility and allowing for more in-depth analysis.

Another important feature offered by the BrCris platform is the graphical visualizations for each entity, which provide a broader overview of a given record. It also allows navigation through all the relationships that a given entity has. In addition to the predefined visualizations, it is also possible to easily implement and include customized graphs in the interface (Figure 3 and Figure 4).

Regarding export, the data can be exported in CSV format for use in spreadsheets or bibliometric tools such as VOSviewer; in GraphML format for use in Gephi and network analysis tools; and in RDF format for use in SPARQL endpoints and semantic systems. A REST API for programmable queries is also under development.

Figure 3 – Entity Viewing and Navigation Page

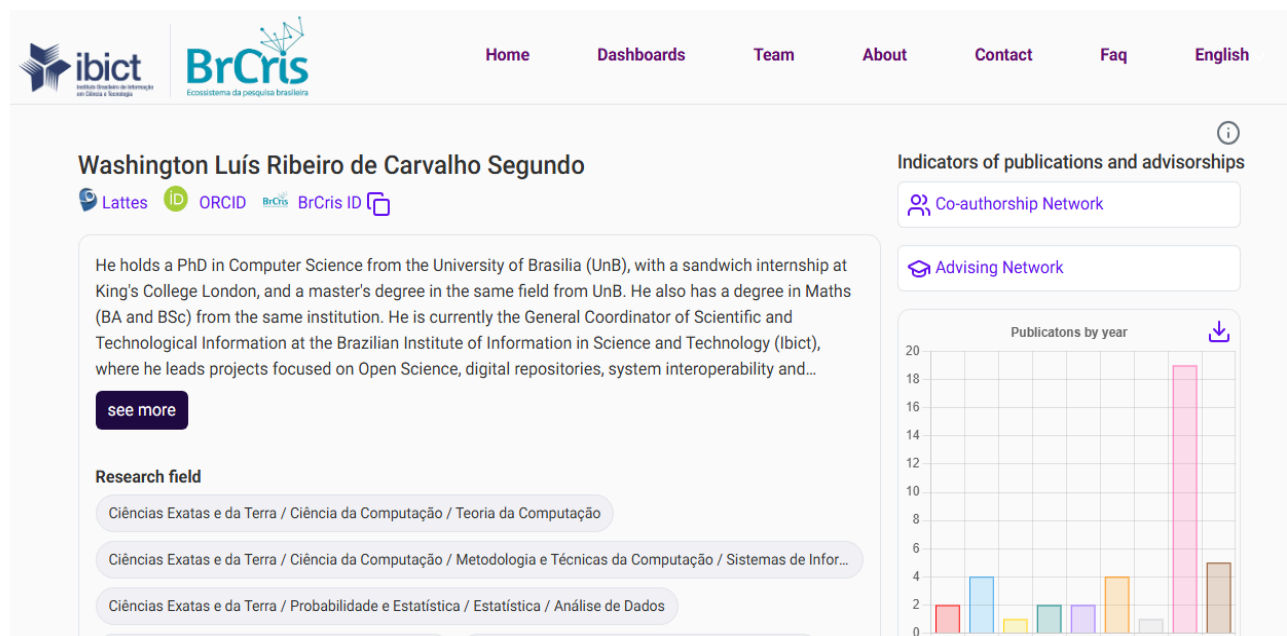
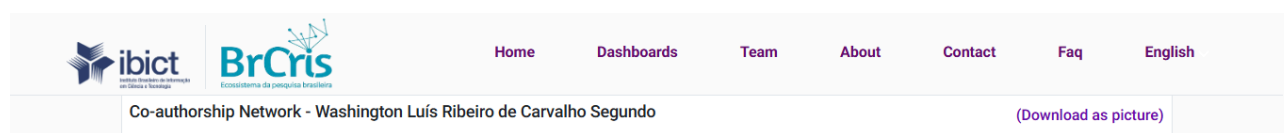
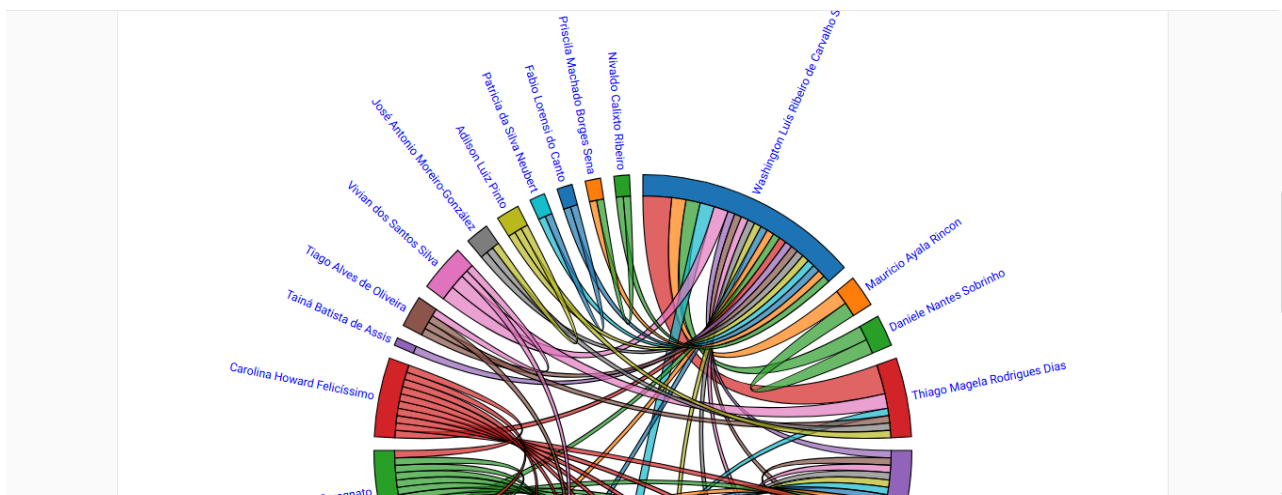


Figure 4 – Co-authorship networks of a researcher





The results show that the BrCris Platform offers an integrated set of visualization and export mechanisms that significantly expand the possibilities for analyzing the Brazilian scientific ecosystem. These tools aim to provide a significant contribution to the academic community, offering easy-to-use mechanisms and reliable data on the Brazilian scientific research ecosystem. This enables analyses that seek to understand science in Brazil more deeply and promote the conduct of new studies.

By making these tools available in an accessible and user-friendly way, the BrCris Platform offers resources that allow users to navigate the data, identify trends, patterns, and relevant insights. This provides a better understanding of Brazilian scientific research, contributing to the generation of knowledge and the promotion of innovative studies.

Considerations

This work contributes to the CRIS and research information management literature by demonstrating how data visualization can be systematically embedded as a core analytical layer within a national research information infrastructure. Using the BrCris platform as a case study, the paper advances the understanding of how interactive visual interfaces enhance access, interpretation and reuse of integrated research data, moving beyond traditional reporting approaches. The presented contributions highlight the role of visual analytics in supporting exploratory analysis, evidence-based decision-making and interoperability in large-scale CRIS environments, particularly within the context of Open Science and national research ecosystems.

One of the platform's main technical differentiators is its ability to adapt to different international standards through data transformation modules. For example, BrCris is able to convert metadata from the Lattes Platform into structures compatible with OpenAIRE or ROR, ensuring interoperability on a global scale. This allows, for example, data on Brazilian scientific production to be compared with data from other countries, supporting transnational analyses and studies of international scientific collaboration.

The development of the BrCris Platform represents an important contribution to the integration of scientific data in Brazil, offering a robust, scalable technical solution that adheres to international best practices. By bringing together historically dispersed data in a single environment, BrCris enables a new generation of scientometric studies, promoting a broader and more detailed understanding of Brazilian science.

The consolidation of BrCris also has important implications in the field of data science applied to the management of scientific information. By providing a unified, interoperable, and up-to-date repository, the platform facilitates the use of data mining, machine learning, and artificial intelligence techniques to generate strategic insights. This paves the way for the development of intelligent dashboards, alert systems for emerging trends, and automated scientific recommendation services.

As a computational infrastructure, BrCris meets the needs of different user profiles: researchers, institutional managers, funding agencies, journalists, decision-makers, and the general public. By enabling visual

navigation, data export, and reuse, the platform consolidates itself as a pillar of Open Science and data-driven management.

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.*

De Meis, L. et al. The growing competition in Brazilian science: rites of passage, stress and burnout. *Brazilian journal of medical and biological research*, v. 36, p. 1135-1141, 2003.

Huang, Y.; Glanzel, W.; Zhang, L.. Tracing the development of mapping knowledge domains. *Scientometrics*, v. 126, p. 6201-6224, 2021.

Leta, Jacqueline; Glanzel, W.; Thijs, B.. Science in Brazil. Part 2: Sectoral and institutional research profiles. *Scientometrics*, v. 67, n. 1, p. 87-105, 2006.

Sivertsen, G.. Developing Current Research Information Systems (CRIS) as data sources for studies of research. *Springer handbook of science and technology indicators*, p. 667-683, 2019.

Yoshikane, F.; Kageura, K.. Comparative analysis of coauthorship networks of different domains: The growth and change of networks. *Scientometrics*, v. 60, p. 435-446, 2004.