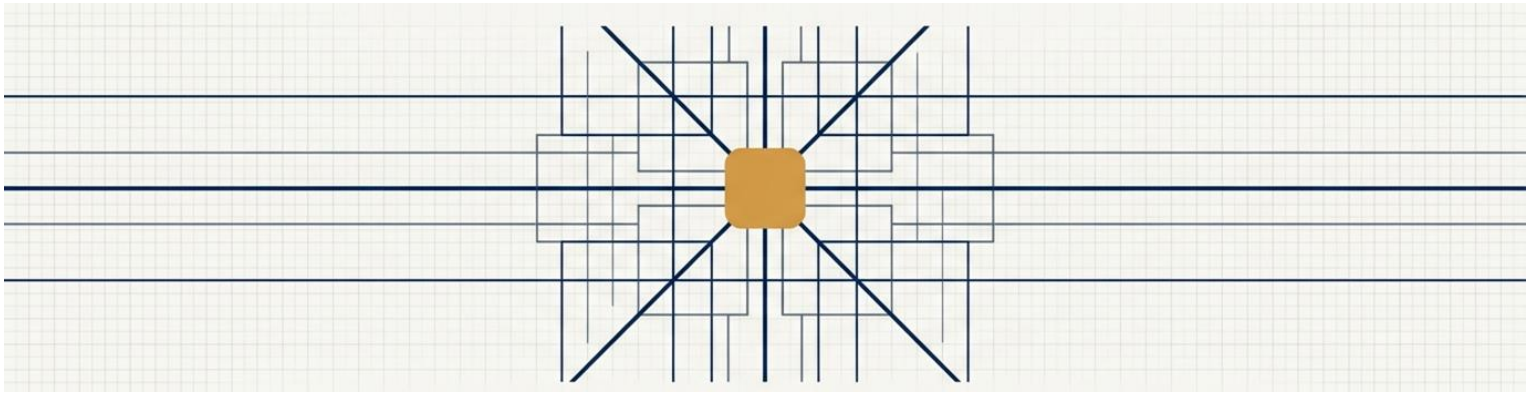
Visual Analytics for National CRIS

Exploring Brazilian research data through the **BrCris platform**

A national infrastructure integrating, certifying, and disseminating scientific, technological, and innovation information in Brazil.



CRIS 2026

17th CRIS Conference on Current Research Information Systems

May 20-22, 2026

ibict



Thiago Dias



Marcel Souza



Jabson Dias



Washington Segundo

Brazil's Fragmented Research Landscape

The Problem

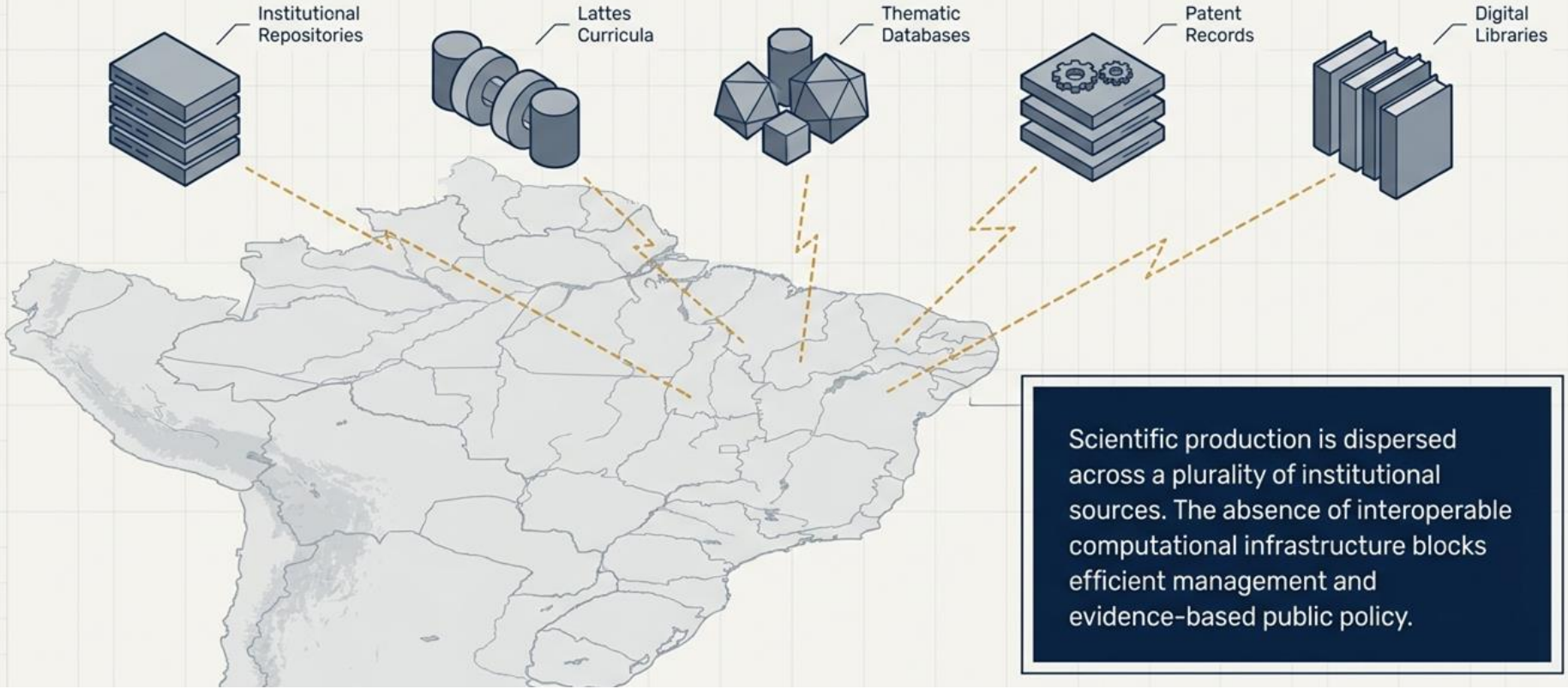
Brazil's rich academic production is dispersed across institutional sources, digital repositories, thematic databases, researchers' curricula, and other recording instances — with no unified infrastructure for systemic analysis.

The Consequence

The absence of interoperable computational infrastructure has hindered:

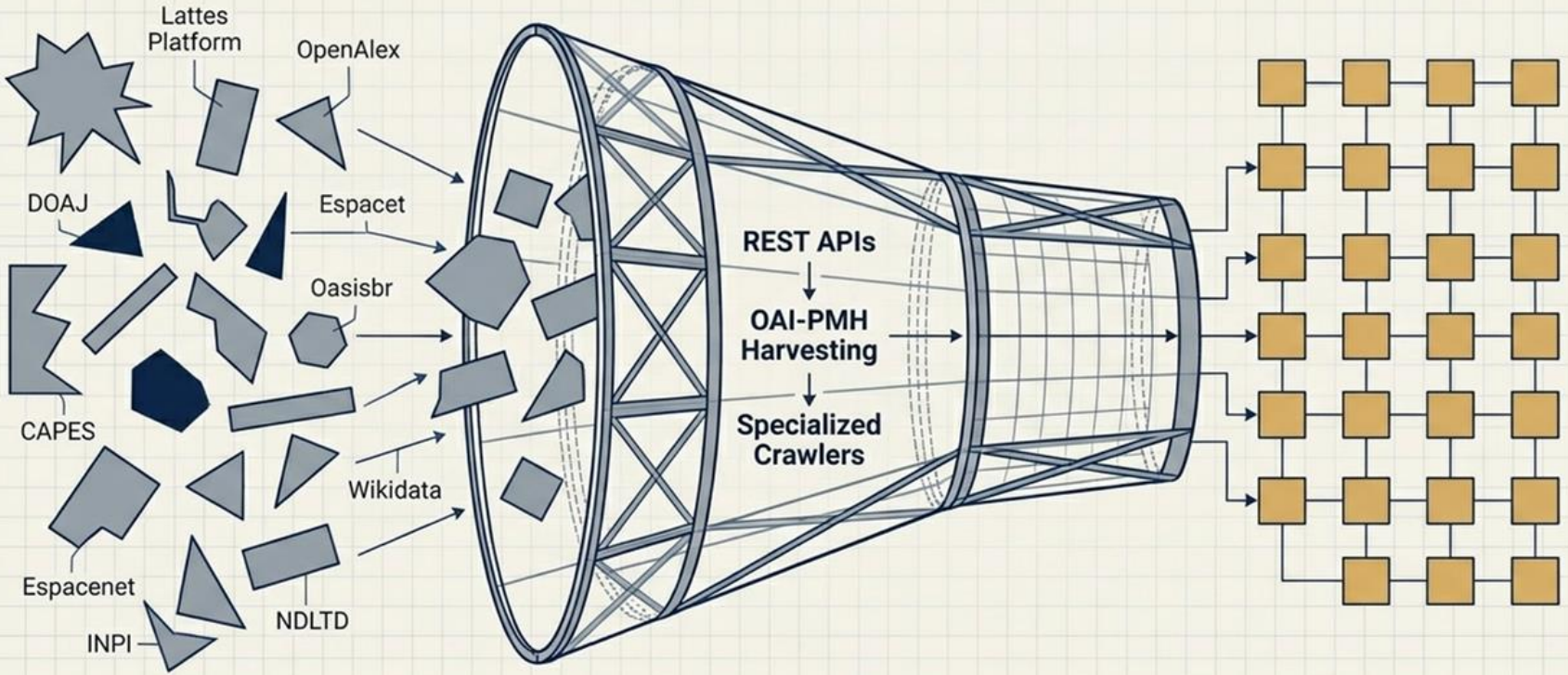
- Systemic analyses of scientific output
- Efficient management of scientific information
- Evidence-based public policy formulation

Fragmented Data Hinders Systemic Policy



Scientific production is dispersed across a plurality of institutional sources. The absence of interoperable computational infrastructure blocks efficient management and evidence-based public policy.

Ingesting Heterogeneous Open Data



Introducing BrCris

The **BrCris Platform**, coordinated by IBICT, responds to Brazil's data fragmentation gap by offering an ecosystem of structured, integrated data on Brazilian scientific, technical, and technological production.

Open Science Principles

Built on open access data and transparent methodologies

CERIF & VIVO Standards

Follows international interoperability standards for global comparability

Data Certification

Includes deduplication, disambiguation, and persistent identifier validation

Traditional Reporting



- **Structure:** Siloed & Isolated



- **Data Quality:** Uncertified & Duplicated



- **Analysis:** Static & Closed Reports



- **Standards:** Localized Formats

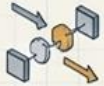
The BrCris Visual Analytics Approach



- **Structure:** Unified Ecosystem (CERIF aligned)



- **Data Quality:** Disambiguated & Certified (Persistent IDs)



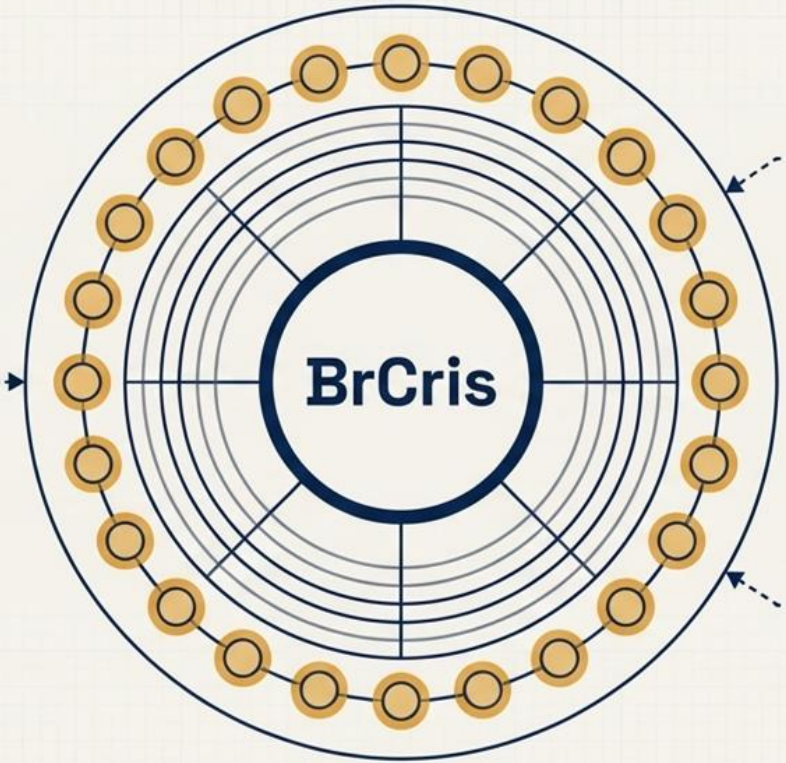
- **Analysis:** Dynamic Filtering & Visual Exploration



- **Standards:** Open Export (GraphML, RDF, CSV)

Moving from static aggregation to dynamic, visual exploration transforms complex information into actionable knowledge.

BrCris Unifies the National Research Ecosystem



Open Science:
Adheres to strict open access principles.

CERIF Standards:
Built on the Common European Research Information Format.

VIVO Ontology:
Semantic structuring for complex relationships.

BrCris Technical Challenges

1 **Heterogeneous Data at Scale**

Handling large volumes of data from diverse sources with different formats and standards

2 **Entity Disambiguation**

Identifying and disambiguating authors, publications, and institutions across sources

3 **Ontology & Vocabulary Adherence**

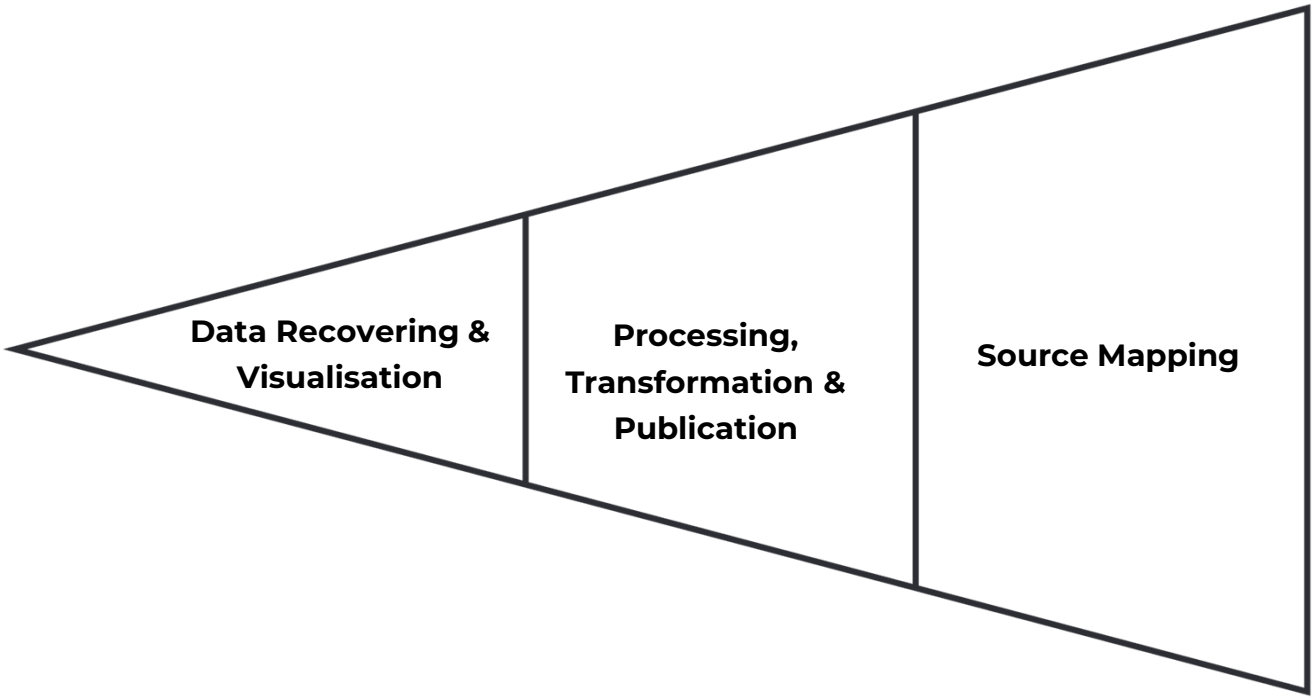
Maintaining complex relationships and adhering to internationally established vocabularies

4 **Interoperable Export**

Exporting data in formats compatible with widely used analytical tools globally

Data Sources & Collection Pipeline

Over **15 data sources** were mapped for BrCris, including Lattes Platform, OpenAlex, DOAJ, Oasibr, Espacenet, CAPES, Wikidata, NDLTD, and INPI. Collection uses REST APIs, OAI-PMH harvesting, and specialized crawlers, all prioritizing open access data.



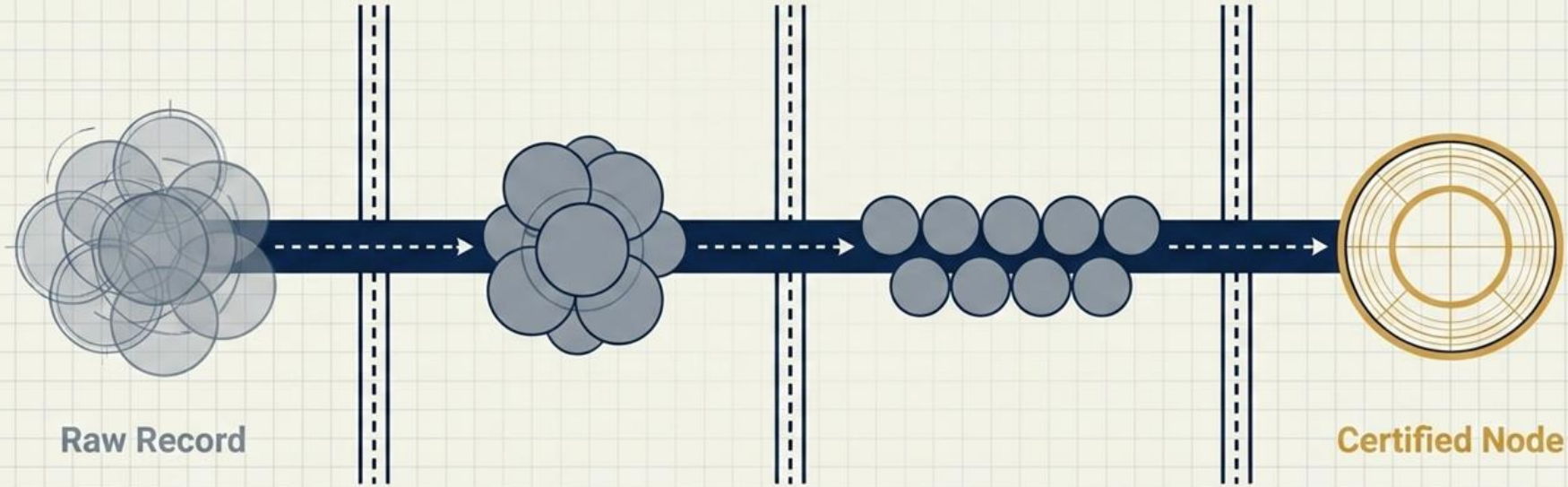
This pipeline ensures technical rigor and standards adherence across heterogeneous data sources.

Certifying the Scientific Record

Source Reliability
Cross-referencing institutional datasets

Internal Consistency
Deduplication & entity disambiguation

Persistent Identifiers
Validation via ORCID, DOI, ROR



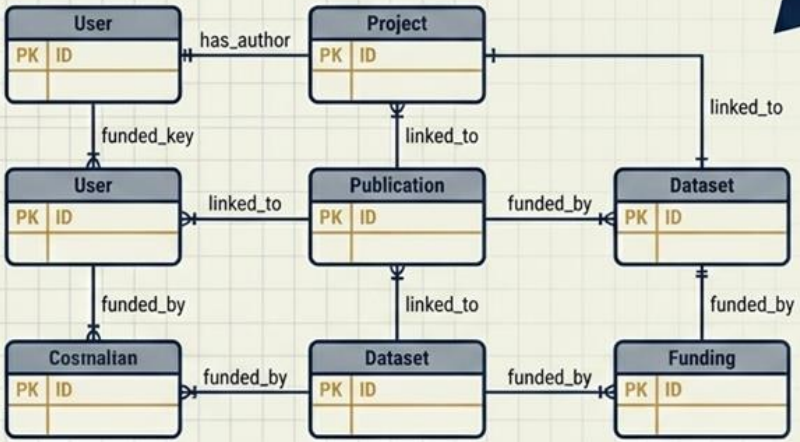
Raw Record

Certified Node

Two Models Power One Unified Ecosystem

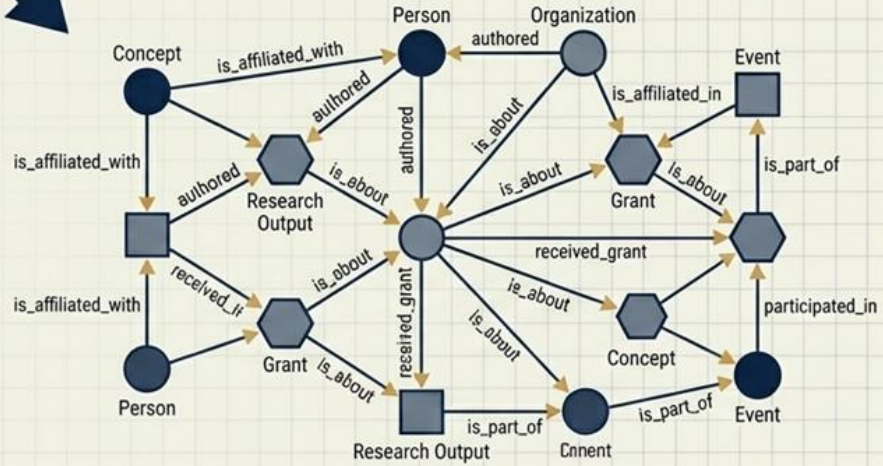
Certified Data Stream

Logical Model (E-R)



Relational storage designed for high-speed, interactive visual dashboards and real-time filtering.

Semantic Model (VIVO)



Designed for semantic interoperability, Linked Open Data provision, and graph-based navigation.

Three Interface Layers



Search Interface

Built on Search-UI and Elasticsearch, enabling refined searches across publications, researchers, and institutions with dynamic real-time filtering.



Interactive Dashboards

Real-time filtering and customizable visualizations across multiple entities and metrics, supporting exploratory and comparative analysis.



Semantic Visualization & Export

Graph-based navigation with export in CSV, GraphML, and RDF formats for use in VOSviewer, Gephi, SPARQL endpoints, and other tools.

Search Interface: Visual Data Discovery

The search interface is the primary entry point for the BrCris data ecosystem. It allows users to perform refined searches across entities — publications, researchers, institutions — while dynamically applying filters that immediately reshape result sets. Below, the Publications search page illustrates this interactive mechanism.



Figure 1 – BrCris Publications search page, showing 24,037 results for "Open Science" with author filters and a Publications by Year bar chart.

Dynamic Filtering & Dashboards

The **"Dashboards and Indicators"** section enables analysis of various entities and exploration of associated metrics. Users apply filters interactively to generate customized visualizations. The Research Groups dashboard below demonstrates keyword frequency analysis across 409,223 groups.



Figure 2 – Research Groups dashboard with keyword frequency list and word cloud. Top keywords: Educação (2,375), Políticas Públicas (1,681), Cultura (1,353), Sustentabilidade (1,324), Saúde (1,123).

Entity Viewing & Navigation

BrCris provides graphical visualizations for each entity, offering a broader overview of individual records and enabling navigation through all associated relationships. Predefined and customizable graphs are available for each researcher profile.

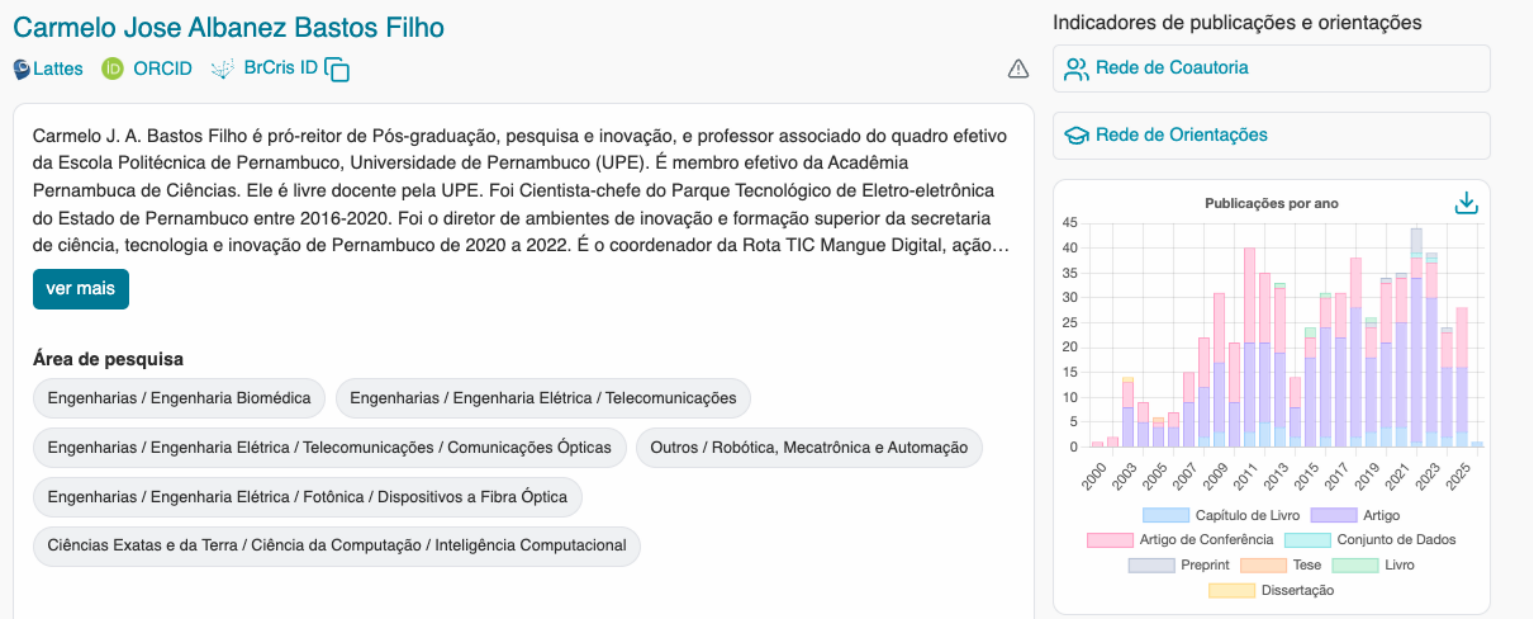


Figure 3 – Researcher profile page showing bio, research fields, co-authorship network link, and Publications by Year chart with a notable peak in 2019 (19 publications).

Co-authorship Network Visualization

BrCris generates interactive co-authorship network graphs for each researcher, revealing collaboration patterns and institutional connections across the Brazilian scientific ecosystem.

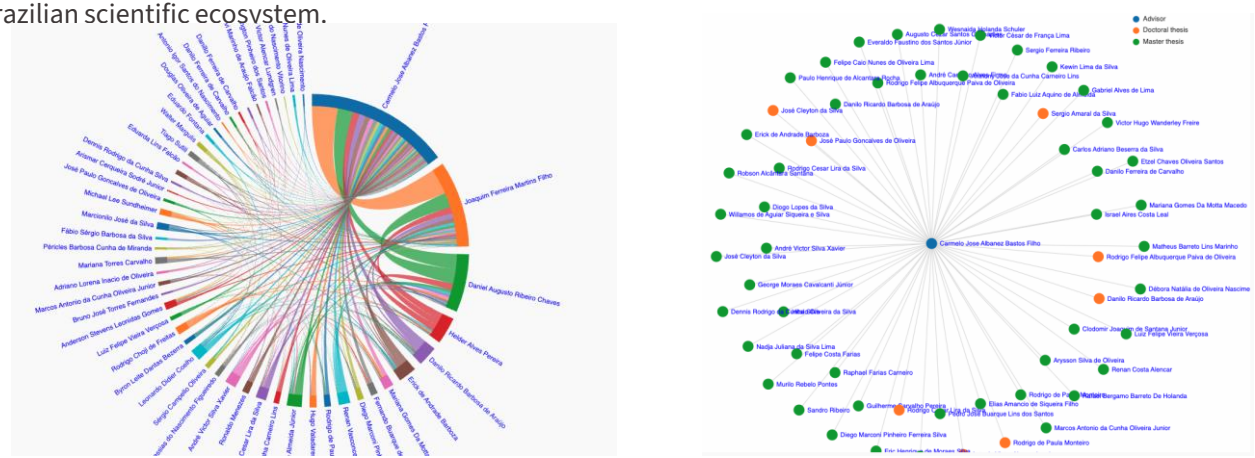


Figure 4 – Co-authorship network for Washington Luís Ribeiro de Carvalho Segundo, showing a chord diagram of researcher relationships. Networks can be downloaded as images.

Export Capabilities



CSV Format

For spreadsheets and bibliometric tools such as VOSviewer



GraphML Format

For Gephi and network analysis tools



RDF Format

For SPARQL endpoints and semantic systems



REST API

For programmable queries — currently under development

Data subsets can be exported after applying any combination of filters, enabling targeted analysis with external tools.

Key Results: What BrCris Enables

BrCris offers an integrated set of visualization and export mechanisms that significantly expand possibilities for analyzing the Brazilian scientific ecosystem. By making tools accessible and user-friendly, the platform allows users to:

Navigate complex data

Explore certified, integrated research information across multiple entity types

Identify trends and patterns

Generate customized analytical subsets through dynamic filtering

Promote new studies

Support scientometric research and evidence-based understanding of Brazilian science



Technical

International Interoperability

BrCris converts Lattes Platform metadata into structures compatible with OpenAIRE or ROR, enabling comparison of Brazilian scientific production with data from other countries and supporting transnational analyses.

Scalable & Standards-Compliant

A robust, scalable technical solution adhering to international best practices. By unifying historically dispersed data, BrCris enables a new generation of scientometric studies and a broader understanding of Brazilian science.

Implications for Data Science & AI

The consolidation of BrCris has important implications for data science applied to scientific information management. By providing a **unified, interoperable, and up-to-date repository**, the platform facilitates advanced analytical techniques.

Data Mining

Systematic extraction of patterns from large-scale research datasets

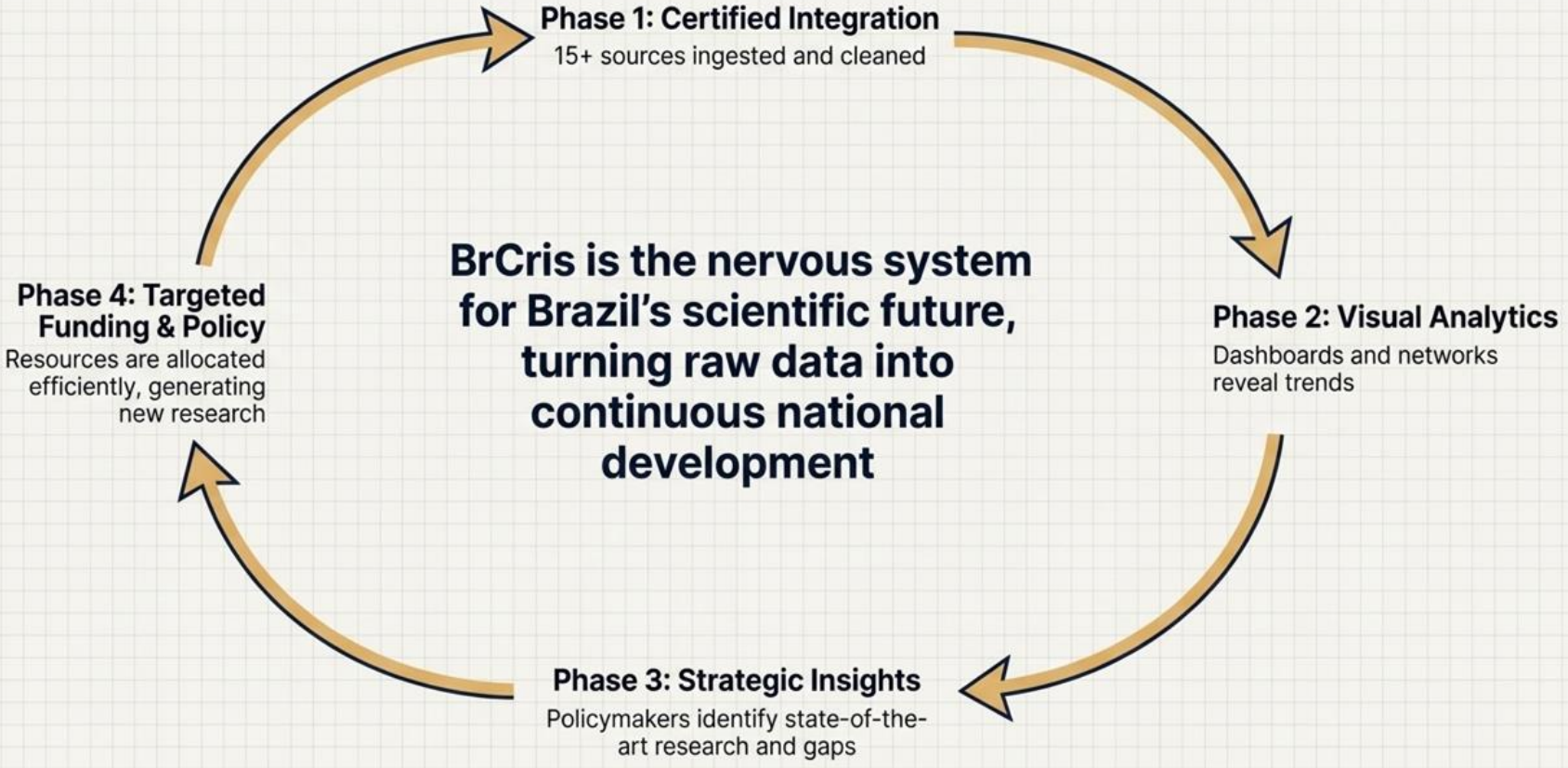
Machine Learning

Predictive models for research trend identification and impact assessment

Intelligent Dashboards

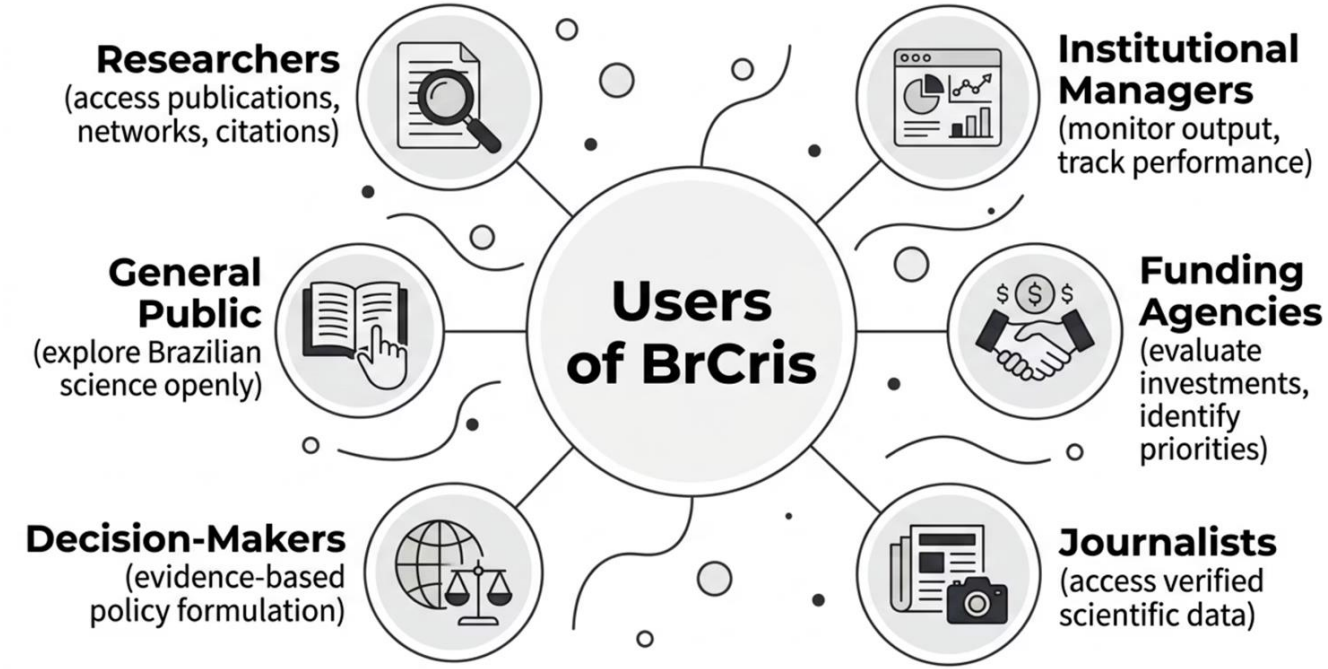
Alert systems for emerging trends and automated scientific recommendation services

The Open Science Flywheel



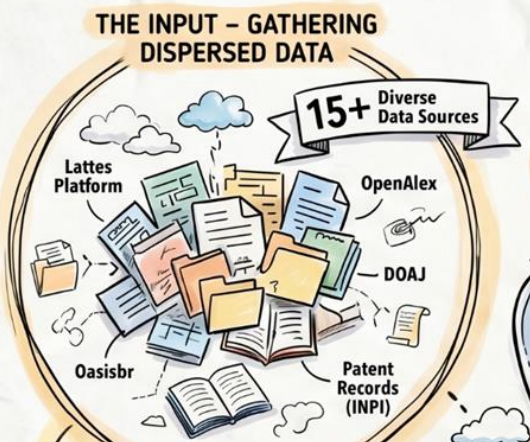
BrCris User Ecosystem

As a computational infrastructure, BrCris meets the needs of diverse user profiles by enabling visual navigation, data export, and reuse — consolidating itself as a pillar of Open Science and data-driven management.



BrCris: Visualizing the Brazilian Research Ecosystem

THE INPUT – GATHERING DISPERSED DATA

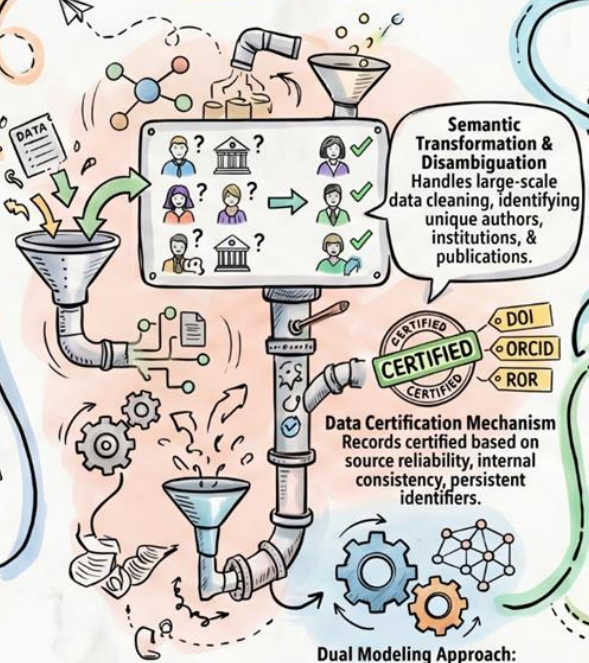


Bridging the "Fragmented" Gap:
Integrating isolated records, curricula, theses, scientific datasets.



Adherence to Open Science:
Priorities open-access & international interoperability standards.

THE PIPELINE – TURNING DATA INTO KNOWLEDGE



Data Certification Mechanism
Records certified based on source reliability, internal consistency, persistent identifiers.

Dual Modeling Approach:
Relational logical model (for dashboards) & Semantic model (for graph-based navigation & Linked Open Data)

THE INTERFACE – EXPLORING THE ECOSYSTEM



Intuitive Search & Discovery:
Refined searches across publications & researchers with real-time filtering.



Interactive Dashboards:
Analyze various entities & metrics via visual indicators for exploratory studies.



Co-authorship & Semantic Networks:
Navigate complex relationships, trends, & collaboration patterns.

THE OUTPUT – ACTIONABLE RESEARCH INTELLIGENCE





Multi-Format Export for Analysis



Global Interoperability:
BrCris converts local Lattes data to OpenAIRE/ROR structures for global comparison.



Evidence-Based Decision Making:
Reliable data for funding agencies & managers to formulate public policies & identify emerging trends.

Conclusions

BrCris demonstrates how **data visualization can be systematically embedded as a core analytical layer** within a national research information infrastructure — moving beyond traditional reporting approaches.

Visual Analytics

Supports exploratory analysis and evidence-based decision-making in large-scale CRIS environments

Open Science

Promotes interoperability, transparency, and data reuse aligned with international standards

National Impact

Enables a new generation of scientometric studies and deeper understanding of Brazilian science



washingtonsegundo@ibict.br

marcelsouza@ibict.br

Thank you!

CRIS 2026

17th CRIS Conference on Current Research Information Systems

May 20-22, 2026

