



Comunicação

DOI: [10.21680/2447-7842.2023v9n2ID33838](https://doi.org/10.21680/2447-7842.2023v9n2ID33838)

O BrCris como ferramenta para acesso e avaliação da ciência

The BrCris as a tool for accessing and assessing science

Thiago Magela Rodrigues Dias ¹

Vivian Santos Silva ²

Jesiel Viana da Silva ³

Rene Faustino Gabriel Junior ⁴

Tales Henrique José Moreira ⁵

Tiago Alves De Oliveira ⁶

Jesús P. Mena-Chalco ⁷

Washington L. R. Carvalho Segundo ⁸

Submetido em: 17/04/2023

Data de aprovação na ConfOA: 14/06/2023

Publicado em: 18/10/2023

Resumo: Atualmente o Brasil conta com diversas plataformas institucionais, como a Plataforma Lattes, Plataforma Sucupira, Portal da Transparência, Catálogo / Biblioteca de teses e dissertações, e Banco de propriedade industrial, que nasceram

¹Doutor e Mestre em Modelagem Matemática e Computacional pelo Centro Federal de Educação Tecnológica de Minas Gerais.

²Doutora em Ciência da Computação pela Universität Passau e Mestre em Informática pela Universidade Federal do Rio de Janeiro.

³ Mestre em Engenharia de Software pelo Centro de Estudos e Sistemas Avançados do Recife.

⁴Doutor em Ciência da Informação pela Universidade Estadual Paulista Júlio de Mesquita Filho e Mestre em Ciência, Gestão e Tecnologia da Informação pela Universidade Federal do Paraná.

⁵ Mestre em Modelagem Matemática e Computacional pelo Centro Federal de Educação Tecnológica de Minas Gerais.

⁶Doutorando em Bioengenharia na Universidade Federal de São João del Rei e Mestre em Informática pela Universidade Federal do Espírito Santo.

⁷ Doutor e Mestre em Ciências da Computação pela Universidade de São Paulo.

⁸ Doutor e Mestre em Informática pela Universidade de Brasília, com Estágio de Doutorado Sanduíche no King's College London.



com o objetivo de fornecer dados de forma centralizada e atualizada sobre o atuar acadêmico, científico e tecnológico de seus pesquisadores. Logo, o Projeto BrCris tem por objetivo estabelecer um modelo único de organização da informação científica de todo o ecossistema da pesquisa brasileira. Entre os agentes deste ecossistema estão os pesquisadores, os projetos, infraestruturas, laboratórios e instituições de pesquisa, os financiadores, além dos resultados da pesquisa expressos principalmente por publicações científicas, teses, dissertações, conjuntos de dados científicos, software e patentes. Diante disso, este trabalho visa apresentar como o BrCris tem contribuído no processo de agregação de plataformas e repositórios de dados, no intuito de fomentar o acesso e a avaliação da ciência. Como resultado, é possível acessar, exportar e avaliar os dados integrados de diversas fontes, possibilitando a certificação de informações científicas e acadêmicas, bem como a análise dos dados agregados sob diversas perspectivas.

Palavras-chave: acesso e avaliação da ciência; BrCris; gestão da informação; ciência aberta.

Abstract: Currently, Brazil has several institutional platforms, such as the Lattes Platform, Sucupira Platform, Transparency Portal, Catalog / Library of theses and dissertations, and Industrial Property Bank, which were born with the objective of providing centralized and updated data on the academic, scientific and technological performance of its researchers. Therefore, the BrCris Project aims to establish a unique model for organizing scientific information from the entire Brazilian research ecosystem. Among the agents of this ecosystem are researchers, projects, infrastructures, laboratories and research institutions, funders, in addition to research results expressed mainly by scientific publications, theses, dissertations, scientific datasets, software and patents. Therefore, this work aims to present how BrCris has contributed to the process of aggregating platforms and data repositories, in order to encourage access and evaluation of science. As a result, it is possible to access, export and evaluate integrated data from different sources, enabling the certification



of scientific and academic information, as well as the analysis of aggregated data from different perspectives.

Keywords: access and evaluation of science; BrCris; information management; open science.

1 INTRODUÇÃO

A humanidade demonstra ser capaz de enfrentar seus mais importantes desafios, em diferentes contextos, que impactam seu modo de vida e impedem seu desenvolvimento. Contemporaneamente, assistimos ao surgimento de novos tratamentos, novas drogas e vacinas para as mais diversas enfermidades, novos processos de produção de alimentos que garantem a perpetuação da espécie e novas tecnologias que transformam a vida cotidiana das pessoas. Estes avanços possuem uma origem comum, ou seja, são resultados da aplicação prática de alguma descoberta científica teórica que, por sua vez, derivam do esforço de pesquisa que é empreendido por pesquisadores nas mais diversas áreas do conhecimento (Tang *et al.*, 2008).

A produção do conhecimento científico é um processo que leva tempo e é incremental (Huang, Glänzel, & Zhang, 2021). Pesquisadores buscam as bases para suas pesquisas no desenvolvimento observado da sua área do conhecimento (estado da arte). A identificação do estado da arte de uma determinada área pode ser feita principalmente pela análise de publicações que concentram as informações a respeito daquele tópico do conhecimento.

Pesquisadores que fazem seu trabalho de forma colaborativa, comumente, publicam os resultados de suas pesquisas da mesma forma. A identificação das relações de coautoria em trabalhos científicos/tecnológicos e projetos de pesquisa possibilita a estruturação de redes de pesquisadores e instituições que são interligados por suas produções acadêmicas (Abbasi, Altmann, & Hossain, 2011).

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 (Meis *et al.*, 2003; Leta, Glänzel, & Thijs, 2006), da análise da colaboração científica (Yoshikane & Kageura, 2004) e da análise de registros de patentes (Abbas, Zhang, & Khan, 2014).

O Brasil tem uma parcela relevante na produção científica internacional principalmente em nichos específicos ao atuar em sua economia. O país é líder na produção do conhecimento no contexto de América Latina (Collazo-Reyes, 2014) e um atrator de talentos no contexto regional (Saravia & Miranda, 2004). O Brasil tem se destacado pela sua implementação de plataformas digitais de registro nacional do atuar de seus pesquisadores.

Logo, O Projeto BrCris que visa coletar dados abertos sobre a pesquisa científica em âmbito nacional teve início em abril de 2020, em que a equipe liderada pela Coordenação de Tratamento, Análise e Disseminação da Informação Científica do IBICT gestou a construção de um sistema que abarcasse registros de todo o ecossistema da pesquisa brasileira. Este sistema foi chamado de BrCris, em alusão aos sistemas do tipo CRIS, acrônimo para *Current Research Information System*.

2 DESENVOLVIMENTO

O BrCris tem por objetivo estabelecer um modelo único de organização da informação científica de todo o ecossistema da pesquisa brasileira. Entre os agentes deste ecossistema estão os pesquisadores, os projetos, infraestruturas, laboratórios e instituições de pesquisa, os financiadores, além dos resultados da pesquisa expressos principalmente por publicações científicas, teses, dissertações, conjuntos de dados científicos, software e patentes.

Com a integração dos dados em um repositório de dados padronizado, o uso do dashboard em forma de visualização elucidada alguns benefícios, como a redução de complexidade de dados, auxilia na percepção das propriedades existentes, ajuda na detecção de erros aparentes, consegue englobar a representação em pouco conteúdo, amplia a percepção cognitiva, dentre outros.



Um dos maiores desafios está em conseguir estratégias computacionais viáveis para que os diversos repositórios de dados a serem coletados possam ser integrados de forma consistente, sem ambiguidade nos registros armazenados. Existem na literatura, diversas outras iniciativas que fazem a integração de conjuntos de dados de diversas fontes, como por exemplo, o OpenAIRE Research Graph. Diante disso, utilizar as melhores práticas para a integração também com estes repositórios passa a ser um desafio.

Em se tratando do modelo de dados para este projeto, iniciou-se pela adoção de nove entidades de dados, seguindo padrões amplamente utilizados na comunidade científica internacional. Após todo o processo de tratamento, independentemente da fonte de dados, os dados gerados como saída são importados em um único banco de dados, e utilizando-se dos identificadores únicos gerados ou identificados, os conjuntos de dados são vinculados e deduplicados, viabilizando dessa forma a interoperabilidade dos dados, independentemente de sua fonte e formato.

Além disso, a aderência a padrões internacionais de modelagem e representação de dados é um dos pontos fundamentais para possibilitar o compartilhamento efetivo de informações. O BrCris adota um esquema de representação de dados em dois níveis. O primeiro é o nível lógico, materializado como um modelo de entidades e relacionamentos baseado no CERIF (Jörg, 2010), também chamado de metamodelo. Este modelo é traduzido para um esquema relacional físico na plataforma LA Referencia, onde os dados são carregados e processados. Trata-se, portanto, de um modelo que atende às necessidades internas de armazenamento de dados, sendo responsável pela organização e integração das informações coletadas para que se tornem insumos para os objetivos do projeto.

A criação do metamodelo de entidades e relacionamento levou em consideração as características dos diversos conjuntos de dados coletados, visando a facilitar as rotinas de tratamento dos dados, em especial a deduplicação de entidades. Porém, neste formato, as informações resultantes não seriam facilmente acessadas e reutilizadas por agentes externos. Para resolver este problema, foi



desenvolvido o segundo nível de representação, o nível semântico, implementado como uma ontologia para permitir a visualização e navegação dos dados como um grafo de conhecimento.

Na criação do modelo semântico foram adotadas como premissas a maximização do reuso de recursos existentes e a utilização de padrões internacionais para representação de dados na área, para que o BrCris fosse compatível com sistemas similares ao redor do mundo.

O modelo semântico do BrCris é composto por um subconjunto da ontologia VIVO, composta pelas classes e propriedades equivalentes às entidades, atributos e relacionamentos do metamodelo lógico, acrescido de uma extensão local que cobre informações específicas do contexto brasileiro.

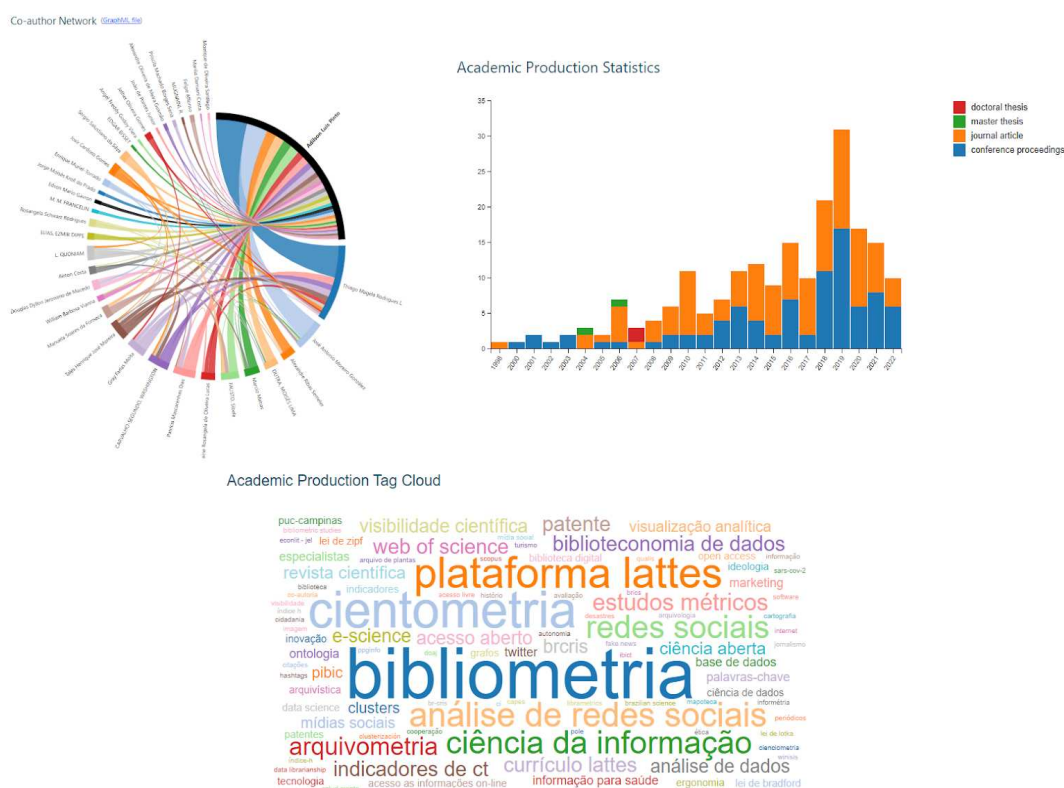
3 RESULTADOS

Após todas as etapas de coleta, tratamento e integração dos dados, é possível ter acesso ao conjunto de dados com auxílio de interfaces gráficas que facilitam o acesso e certificação dos conjuntos. A ontologia VIVO, em particular, possibilita que os dados sejam visualizados na Plataforma VIVO, uma ferramenta para navegação de dados do domínio acadêmico que permite que o BrCris sirva Linked Open Data a agentes externos, além de facilitar a exploração do grafo de conhecimento. Outro recurso importante oferecido pela plataforma VIVO são as visualizações gráficas, que fornecem um panorama mais amplo sobre um determinado indivíduo.

Além das visualizações pré-definidas, também é possível implementar e incluir na interface de forma simples gráficos customizados. A Figura 1 ilustra algumas destas visualizações: a rede de coautoria de um pesquisador, já disponível de forma nativa na plataforma, e dois gráficos customizados implementados para o BrCris: o total de publicações por tipo e a nuvem de palavras dada pelas palavras-chave das publicações de um determinado pesquisador.



Figura 1 - Visualizações originadas da Plataforma VIVO



Fonte: Os autores (2023).

Como componente de recuperação de informação foi desenvolvido uma interface gráfica web (Figura 2) baseada na Search-UI da Elastic. Search-UI é uma biblioteca livre de código aberto, escrita em Typescript, que disponibiliza uma variedade de componentes web customizáveis e compatíveis com aplicativos desktop e aplicativos móveis, que se integra com o Elasticsearch para fornecer uma interface de busca e visualização de informações na web.

A Search-UI dispõe de um conjunto de componentes de busca configuráveis facilitando a construção de interfaces ricas de busca totalmente customizáveis com funcionalidades avançadas de filtragem que ajudam os usuários a encontrar exatamente o que eles precisam. Disponibilizando recursos de paginação, digitação preditiva, autocomplete, filtros, classificação e geração de gráficos de indicadores.

Figura 2 - Interface de busca e visualização do BrCris



The screenshot shows the BiblioCanto search interface. At the top, there is a navigation bar with 'BrCris' and 'ibict' logos, and links for 'Home', 'Team', and 'About'. There are also buttons for 'PKTI' and 'API'. Below the navigation bar, the search interface is annotated with red arrows and labels: 'Índice de busca' points to the search bar; 'Campos de busca' points to the search input field containing 'tecnologia'; 'Publications' points to the search results area; 'Ordenação e filtros' points to the sidebar filters; 'Resultados' points to the main results area; and 'Gráficos de indicadores' points to the charts on the right. The search results show a list of documents with details such as title, year, author(s), type, and journal. The charts show 'Documents by year' and 'Document type'.

Fonte: Os autores (2023).

Com a utilização da Search-UI foi possível desenvolver uma interface para realizar buscas por texto em linguagem natural com a adição de operadores booleanos. As solicitações de buscas dos usuários são cruzadas com um índice do Elasticsearch, possibilitando que a recuperação e o ranqueamento sejam realizados rapidamente.

Realizada a busca, a interface de recuperação do BrCris permite realizar classificação dos resultados (coluna da esquerda da Figura 2), tais como: classificação alfabética, classificação cronológica, bastante útil para pesquisadores, pois permite a seleção dos resultados mais recentes, classificação por relevância de acordo com o ranqueamento do Elasticsearch. Também é possível aplicar filtros de refinamento da busca por algum atributo dos documentos, como por exemplo: nome do autor, tipo de documento.

4 CONSIDERAÇÕES FINAIS



O BrCris se configura como um importante espaço de pesquisa e análise de dados. As informações agregadas e organizadas, segundo um modelo de dados semântico, permitem a geração de serviços para diversos atores, nos contextos de gestão e pesquisa acadêmica, assim como na área de informação para a inovação.

O BrCris é uma iniciativa que coleta e enriquece dados de repositórios e bases de dados abertas sendo uma proposta ímpar no mundo, que facilita obter um Panorama Brasileiro da Produção e Atuação de todos os seus atores acadêmicos/científicos.

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>

Abbasi, A., Altmann, J., & Hossain, L. (2011). Identifying the effects of co-authorship networks on the performance of scholars: A correlation and regression analysis of performance measures and social network analysis measures.

Journal of informetrics, 5(4), 594-607. <https://doi.org/10.1016/j.joi.2011.05.007>

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.

Meis, L. de, Velloso, A., Lannes, D., Carmo, M. S., & de Meis, C. de. (2003). The growing competition in Brazilian science: rites of passage, stress and burnout.



Huang, Y., Glänzel, W., & Zhang, L. (2021). Tracing the development of mapping knowledge domains. *Scientometrics*, 126(7), 6201-6224.

<https://doi.org/10.1007/s11192-020-03821-x>

Jörg, B. (2010). CERIF: The common European research information format model.

Data Science Journal, 9, CRIS24-CRIS31.

Leta, J., Glänzel, W., & Thijs, B. (2006). Science in Brazil. Part 2: Sectoral and institutional research profiles. *Scientometrics*, 67(1), 87-105.

Saravia, N. G., & Miranda, J. F. (2004). Plumbing the brain drain. *Bulletin of the World Health Organization*, 82, 608-615.

Tang, J., Zhang, J., Yao, L., Li, J., Zhang, L., & Su, Z. (2008, August). Arnetminer: extraction and mining of academic social networks. In *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 990-998). <https://doi.org/10.1145/1401890.1402008>

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