

ECOSSISTEMA DE INFORMAÇÃO DA PESQUISA CIENTÍFICA BRASILEIRA (BrCris)

Thiago Magela Rodrigues Dias
Washington Luís R. de Carvalho Segundo
Marcel Garcia de Souza



APRESENTAÇÃO

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 (Saraiva; Miranda, 2004). Destaca-se pela sua implementação de plataformas digitais de registro nacional do atuar de seus pesquisadores.

O destaque internacional é a Plataforma Lattes, em que se observa a importância estratégica de ter as informações científicas curriculares disponíveis de forma ampla (Lane, 2010). Similar à Plataforma Lattes existem outras plataformas nacionais que registram parte do atuar acadêmico e tecnológicos, como o Sucupira da Capes, o Banco de teses e dissertações do Ibict, o Banco de propriedade industrial do INPI e Portal Transparência do Governo Federal, por exemplo. Embora todas as informações dessas fontes de dados sejam abertas e livres para consulta, elas estão restritas a suas bases de dados, não existindo uma integração de forma a possibilitar a exploração dessas informações de forma ampla a ciência e a sociedade do Brasil.

Neste contexto, é importante destacar que a informação proveniente de pesquisas em Ciência, Tecnologia e Inovação (CT&I) representa um dos principais pilares para o desenvolvimento econômico e social de um país. Essa informação constitui-se como um elemento orientador fundamental para a geração sistemática de conhecimento, tanto teórico quanto aplicado.

A informação gerada no âmbito de Ciência, Tecnologia e Inovação (CT&I) é altamente especializada, distinguindo-se de outros tipos de informação. Sua produção é fundamentada em um método específico - o método científico - e a divulgação de seus resultados seguem procedimentos distintos, abrangendo avaliação, validação, publicação e acesso por meio de fontes especializadas (Meadows, 1999).

A diversidade de fontes de informação, juntamente com a variedade de seus modelos de dados (metadados), resultante das pesquisas em Ciência, Tecnologia e de Inovação (CT&I) são frequentemente armazenados em bases de dados e repositórios que tem estruturas de dados distintas, visando assegurar a preservação, visibilidade e recuperação eficaz das informações contidas, e em poucos casos objetivam a integração.

Complementando as fontes de informações brasileiras, existem fontes com abrangência internacional, que possibilitam a troca de dados tanto para seres humanos como para computadores. Dentre essas bases de dados pode-se citar o Wikidata, o Crossref, OpenCitations, OpenAIRE Research Graph, o Latindex e o DOAJ entre outras fontes abertas de informação em CT&I.

Reunir e integrar os dados dessas e de outras fontes é um desafio que exige um grande esforço intelectual e um grande poder computacional. A capacidade de armazenamento necessária ultrapassa a que um sistema gerenciador de banco de dados tradicional consegue suportar, exigindo soluções tecnológicas avançadas, provenientes da chamada Web 4.0, bem como, de análises que envolvam técnicas de inteligência computacional.

A partir desse cenário, surgiram iniciativas voltadas para a criação de sistemas destinados a gerenciar a produção acadêmica, sejam eles institucionais, nacionais ou temáticos. Esses sistemas são conhecidos pela sigla CRIS. CRIS (*Current Research Information System*) define um sistema de informação abrangente sobre todo o ecossistema do processo científico.

O modelo CRIS organiza em um só lugar todas as informações do ciclo da pesquisa Científica, desde o Fomento, passando pelos projetos, pesquisadores, instituições de pesquisa e laboratórios, até os outputs de uma pesquisa científica, tais como artigos científicos, teses, dissertações, livros, capítulos de livro, patentes e conjuntos de dados científicos (Sivertsen, 2019).

Iniciativas de mapeamento do ecossistema ocorrem na Europa com o *Directory of Research Information Systems* (DRIS) executado pela euroCRIS (Eurocris, 2023), no Brasil, o BrCris é o mapeamento do ecossistema de informação da pesquisa científica brasileira. Sua concretude se faz com uma plataforma agregadora de diversas fontes de informações, o que permite recuperar, certificar e visualizar dados e informações relativas aos diversos atores que atuam na pesquisa científica do contexto brasileiro. Dentre as principais fontes estão os dados curriculares de indivíduos, organizações, programas de pós-graduação, publicações, orientações acadêmicas, revistas científicas, patentes, grupos de pesquisa, softwares e outras fontes que ainda serão agregadas (Dias; et al., 2022).

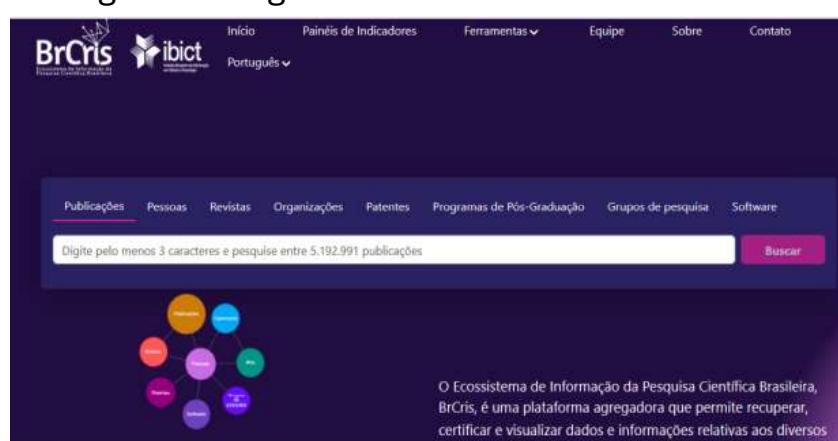
Logo, o BrCris tem por objetivo estabelecer um modelo único de organização da informação científica de todo o ecossistema de pesquisa brasileira. Entre os agentes

deste ecossistema estão os pesquisadores, os projetos, as infraestruturas, os laboratórios e as instituições de pesquisa, os financiadores, além dos resultados da pesquisa expressos principalmente por publicações científicas, teses, dissertações, conjuntos de dados científicos, software e patentes.

HISTÓRICO E DESENVOLVIMENTO

O BrCris oferece uma interface unificada de busca de informações, a visualização de redes de colaboração e painéis de indicadores em ciência, tecnologia e inovação (Figura 1). E estabelece um modelo único de organização da informação científica de todo o ecossistema da pesquisa brasileira (Dias; et al., 2022). 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 e patentes (Kong, et al., 2019).

Figura 1 - Página Inicial da Plataforma BrCris



Fonte: Ibict (2024).

A idealização do Projeto do Sistema BrCris, que é o CRIS no contexto da ciência brasileira, data de 2014. Inspirado no modelo proposto por Portugal de um CRIS nacional (o PTCRIS - <https://ptcris.pt>), o Instituto Brasileiro de Informação em Ciência e Tecnologia (Ibict) iniciou uma série de estudos e parcerias interinstitucionais para a execução do Projeto. Em 2020, houve a implementação formal de um Projeto de Pesquisa para a construção do BrCris, com o intuito de fornecer ferramentas tecnológicas para a comunidade acadêmica brasileira com dados consolidados da produção científica nacional, tomando como base outros projetos CRIS e padrões internacionais disponibilizados pelo OpenAire e COAR.

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 (Pinto *et al.*, 2021).

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.

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.

Atualmente, o BrCris oferece três métodos distintos para navegação e visualização de seus dados. O primeiro envolve a utilização de sua interface de busca, que permite

pesquisas abrangentes em todas as entidades da plataforma, inclusive com a aplicação de conectores lógicos para buscas mais específicas. Além disso, a visualização e exportação dos dados é facilitada pelos *dashboards* atualmente disponíveis. Por fim, o BrCris também possibilita a navegação entre as entidades utilizando a plataforma VIVO.

Como parte do sistema de recuperação de informações, foi desenvolvida uma interface gráfica web baseada na *Search-UI* da *Elastic*. A *Search-UI* é uma biblioteca de código aberto escrita em Typescript, que fornece uma variedade de componentes web personalizáveis compatíveis com aplicativos desktop e móveis. Essa biblioteca integra-se ao Elasticsearch, oferecendo uma interface abrangente para busca e visualização de informações na web.

A *Search-UI* disponibiliza um conjunto de componentes de busca configuráveis, simplificando a criação de interfaces altamente personalizáveis com funcionalidades avançadas de filtragem. Entre essas funcionalidades estão recursos como paginação, digitação preditiva, autocompletar, filtros, classificação e geração de gráficos de indicadores. Essa abordagem visa facilitar aos usuários a localização precisa do que necessitam, promovendo uma experiência de busca eficiente e aprimorada (Figura 2).

A interface de busca do BrCris oferece uma série de funcionalidades avançadas para a pesquisa e visualização de informações. Ao acessar a interface, os usuários podem realizar buscas detalhadas utilizando a caixa de pesquisa,

que aceita termos específicos ou operadores lógicos para refinar os resultados.

Figura 2 - Interface de Busca de Publicações do BrCris



Fonte: Ibict (2024).

A página também permite a personalização da quantidade de resultados exibidos por página e a ordenação das publicações por diferentes critérios, como a data de publicação em ordem decrescente, garantindo que as informações mais recentes sejam facilmente acessíveis.

Além das funcionalidades de busca, a página de publicações do BrCris apresenta opções avançadas de filtragem e visualização. Os usuários podem aplicar filtros para restringir os resultados a critérios específicos, como autor, instituição ou tipo de documento.

A interface também inclui a opção de exportar os dados das publicações, permitindo que os usuários baixem as informações em formatos compatíveis para análise externa. A apresentação dos resultados é clara e organizada, facilitando a navegação entre as diferentes

publicações e proporcionando uma experiência de usuário eficiente e intuitiva.

A implementação de diversas visualizações em um conjunto de *dashboards* se caracteriza como um importante mecanismo de exportação de dados, oferecendo uma perspectiva abrangente sobre a produção científica nacional.

Essas visualizações permitem a aplicação de filtros e outros métodos de personalização em todos os atributos de uma entidade específica, proporcionando aos usuários a capacidade de realizar análises detalhadas tanto em termos temporais quanto em áreas específicas ou tópicos de interesse (Figura 3).

Além disso, a aplicação desses filtros gera um conjunto de dados específicos, resultantes do processo de filtragem, que podem ser exportados em formatos padronizados, como .csv. Isso facilita a manipulação e análise dos dados fora da plataforma, permitindo que pesquisadores e analistas explorem as informações de maneira mais aprofundada e integrada com outras ferramentas e sistemas.

Com essa funcionalidade, o BrCris se destaca como uma ferramenta flexível para o mapeamento e análise da produção científica, atendendo às necessidades de uma comunidade científica diversificada e em constante evolução.

Os painéis de indicadores do BrCris proporcionam uma visualização detalhada e interativa da produção científica nacional, oferecendo diversas possibilidades de análise e personalização. Os usuários podem explorar os dados por meio de gráficos, tabelas e outras

representações visuais que facilitam a compreensão de padrões e tendências. A interface permite a aplicação de múltiplos filtros, como por período de tempo, instituição, área do conhecimento e tipo de publicação, possibilitando uma análise refinada e direcionada conforme os interesses específicos de cada usuário.

Figura 3 - Painéis de Visualização e Exportação de Dados



Fonte: Ibict (2024).

Além das opções de visualização, os painéis de indicadores do BrCris oferecem funcionalidades de exportação de dados. Os usuários podem exportar os conjuntos de dados filtrados em formatos padronizados, como .csv, o que permite uma análise mais aprofundada, utilizando outras ferramentas e softwares de análise. Essa capacidade de exportação facilita a integração dos dados com diferentes sistemas e apoia a realização de pesquisas e estudos mais detalhados, tornando o BrCris uma ferramenta essencial para pesquisadores, gestores e formuladores de políticas na área de ciência e tecnologia.

E ainda, a plataforma VIVO oferece diversas visualizações que desempenham um papel crucial na exploração detalhada dos dados. Essas visualizações não apenas facilitam a navegação pelas informações de uma entidade, mas também revelam seus relacionamentos com outras entidades, proporcionando uma análise interativa dos dados. Através da plataforma, é possível acessar todas as informações pertinentes à entidade de interesse e seus vínculos. Detalhes como publicações, áreas de atuação e redes de coautoria de um pesquisador específico são prontamente acessíveis, permitindo uma visão abrangente de sua contribuição acadêmica e colaborativa (Figura 4).

Figura 4 - Visualização de uma entidade na Plataforma VIVO

Thiago Magela Rodrigues Dias

Doutor em Modelagem Matemática e Computacional pelo CEFET-MG (2016) tendo trabalhado com Bibliometria, Extração de Dados Científicos e Análise de Redes de Colaboração Científica. Mestre em Modelagem Matemática e Computacional pelo CEFET-MG (2008), trabalhando com Arquitetura Orientada a Serviços e Web Mining com o tema: Uma Arquitetura Orientada a Serviços para emprego em Sistemas de Mineração de Dados na Web. Possui graduação em Ciência da Computação pelo Centro Universitário de Formiga ([... more](#))

PhD in Mathematical and Computational Modeling from CEFET-MG (2016) having worked with Bibliometrics, Scientific Data Extraction and Analysis of Scientific Collaboration Networks. Master in Mathematical and Computational Modeling from CEFET-MG (2008), working with Service Oriented Architecture and Web Mining with the theme: A Service Oriented Architecture for use in Data Mining Systems on the Web. Graduated in Computer Science from Centro Universitário de Formiga - UNIFOR (2004), in addit ([... more](#))

Áreas De Investigação

Banco de Dados | Ciéncia da Computação | Engenharia de Software | Métodos Quantitativos. Bibliometria |
Sistemas de Informação | Técnicas de Recuperação de Informação |



Fonte: Ibict (2024).

A visualização oferecida pela plataforma VIVO é importante para a análise e compreensão da produção científica, tecnológica e das redes de colaboração científica. Ela permite obter uma visão detalhada e integrada das contribuições individuais e coletivas dentro do ambiente acadêmico. Ao apresentar informações de forma gráfica e interativa, a plataforma facilita a

identificação de padrões, tendências e conexões que, de outra forma, não poderiam ser facilmente identificadas, principalmente em detrimento de análises tradicionais baseadas apenas em dados tabulares.

A representação gráfica de dados, como coautoria e áreas de atuação, não só revela as interações existentes, mas também pode sugerir novas oportunidades de parceria e colaboração. Isso é particularmente valioso em um contexto de pesquisa cada vez mais globalizado e interconectado, onde a sinergia entre diferentes áreas do conhecimento pode levar a inovações significativas. Assim, a plataforma VIVO não apenas melhora a transparência e a acessibilidade das informações científicas, mas também potencializa o avanço do conhecimento ao fomentar um ambiente colaborativo e bem informado.

Um exemplo de análise específica proporcionada pela plataforma VIVO é a exploração da rede de colaboração de um pesquisador, onde informações detalhadas sobre a rede são apresentadas, acompanhadas por uma representação gráfica em forma de grafo (Figura 5).

Figura 5 - Rede de Colaboração Científica de um Pesquisador



Fonte: Ibict (2024).

Esse grafo visualiza as conexões entre pesquisadores, destacando colaborações e coautorias, o que facilita a identificação de parcerias. Além disso, essas visualizações interativas permitem aos usuários explorar dados de forma intuitiva, ajustando os filtros e parâmetros conforme necessário para obter uma compreensão mais profunda e personalizada das redes de pesquisa e dos impactos acadêmicos. Assim, a plataforma VIVO se torna uma ferramenta poderosa para pesquisadores, gestores e instituições acadêmicas na análise e valorização da produção científica.

CONSIDERAÇÕES FINAIS

O BrCris configura-se como um espaço vital para a 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, bem como na área de informação para a inovação. Esta plataforma é uma iniciativa que coleta e enriquece dados de repositórios e bases de dados abertos, sendo uma proposta singular no mundo. Ela facilita a obtenção de um panorama abrangente da produção e atuação dos atores acadêmicos e científicos brasileiros. No entanto, a operação do BrCris requer significativos recursos computacionais e humanos para o tratamento e padronização desses dados.

Diante do exposto, os resultados obtidos com o BrCris não apenas complementam, mas também expandem significativamente a compreensão da produção científica brasileira. Destaca-se a relevância e o potencial

do BrCris na promoção de uma visão abrangente e aprofundada da pesquisa no Brasil, evidenciando como suas funcionalidades avançadas de visualização e exportação de dados possibilitam análises detalhadas e personalizadas. Ao integrar diversas fontes de dados e oferecer ferramentas para a exploração e interpretação dessas informações, o BrCris se consolida como uma importante plataforma, contribuindo para um entendimento mais completo e estratégico da ciência e tecnologia no país.

REFERÊNCIAS

COLLAZO-REYES, Francisco. Growth of the number of indexed journals of Latin America and the Caribbean: the effect on the impact of each country. **Scientometrics**, Budapest, v. 98, p. 197-209, 2014.

DIAS, Thiago Magela Rodrigues et al. Brccris: plataforma para integração, análises e visualização de dados técnicos-científicos. p. 622-638, **Informação e Informação**, Londrina, v. 27, n. 3, 2022. DOI: 10.5433/1981-8920.2022v27n3p622.

EUROCRIS. **Directory of Research Information Systems (DRIS)**. Nijmegen, Netherlands. Disponível em: <https://eurocris.org/services/dris>. Acesso em: 23 out. 2023.

INSTITUTO BRASILEIRO DE INFORMAÇÃO EM CIÊNCIA E TECNOLOGIA (IBICT). **Ecossistema de Informação da Pesquisa Científica Brasileira, BrCris**. Disponível em: <https://brccris.ibict.br>. Acesso em: 23 out. 2024.

JÖRG, Brigitte. CERIF: The common European research information format model. **Data Science Journal**, London, v. 9, p. CRIS24-CRIS31, 2010.

- KONG, Xiangjie *et al.* Academic social networks: Modeling, analysis, mining and applications. **Journal of Network and Computer Applications**, London, v. 132, p. 86-103, 2019.
- LANE, Julia. Let's make science metrics more scientific. **Nature**, London, v. 464, n. 7288, p. 488-489, 2010.
- MEADOWS, Arthur Jack. **A comunicação científica**. Trad. A. A. B. de Lemos. Brasília: Briquet de Lemos, 1999.
- PINTO, Adilson Luiz *et al.* The Brazilian current research information system: BrCris. *In:* SILVA, Carlos Guardado da Silva; REVEZ, Jorge; CORUJO, Luis (coord.). **Organização do conhecimento no horizonte 2030: desenvolvimento sustentável e saúde**. Lisboa: Universidade de Lisboa, 2021. ISBN 9789895661374. p. 319-330. DOI: 10.51427/10451/50067.
- SARAVIA, Nancy Gore; MIRANDA, Juan Francisco. Plumbing the brain drain. **Bull World Health Organ**, Genebra, v. 82, n. 8, set. 2004.
- SIVERTSEN, Gunnar. Developing Current Research Information Systems (CRIS) as data sources for studies of research. *In:* GLÄNZEL, Wolfgang *et al.* (ed.). **Springer handbook of science and technology indicators**. Amsterdã: Springer, Cham. 2019. p. 667-683. DOI: 10.1007/978-3-030-02511-3_25.
- Como citar este capítulo
- DIAS, Thiago Magela Rodrigues; CARVALHO SEGUNDO, Washington Luís R. de; SOUZA, Marcel Garcia de. Ecossistema de Informação da Pesquisa Científica Brasileira (BrCris). *In:* AMARO, Bianca; CAMPOS, Fhillipe de Freitas; BARCELOS, Janinne. (org.). **Infraestruturas de Ciência e de Acesso Aberto no Brasil: iniciativas do Instituto Brasileiro de Informação em Ciência e Tecnologia**. Brasília, DF: Editora Ibit, 2025. Cap. 14, p. 233-248. DOI: 10.22477/9788570132543.cap14