

O BrCris como ferramenta de apoio à Ciência Aberta

Washington Segundo

IBICT, Brasil

Thiago Magela Rodrigues Dias

CEFET-MG, Brasil

Tales Moreira

IBICT, Brasil

Adilson Luiz Pinto

UFSC, Brasil

Vivian Santos Silva

IBICT, Brasil

Josir Gomes

IBICT, Brasil

Luc Quoniam

UFMS, Brasil

Lautaro Matas

LA Referencia, Brasil

Ary Gabriel Dias

IBICT, Brasil

Juliana Schneider

IBICT, Brasil

CRIS

BrCris

Interoperabilidade

Ciência Aberta

Dados Abertos

Métricas Abertas

A proposta tem como objetivo a apresentação do sistema BrCris como ferramenta de apoio à Ciência Aberta. São alvos da pesquisa a organização de informações em uma só Plataforma dos dados coletados sobre o Ecossistema da Pesquisa Brasileira, que tem como metas a construção de serviços de visualização e recuperação da informação, além da construção de interfaces de interoperabilidade entre sistemas. Os primeiros resultados da execução do Projeto já incluem o desenvolvimento da arquitetura e o mapeamento de modelo de dados relacional e semântico, o mapeamento das fontes de dados a serem agregadas pelo sistema, a implementação de provas de agregação dos recursos mapeados e a definição e realização de testes de serviços a serem disponibilizados. Utilizando a mesma infraestrutura da Plataforma LA Referencia, o projeto agregou os dados da Plataforma Lattes, o conjunto de dados de patentes, os dados de programas de pós-graduação da CAPES e as bases internacionais Wikidata, OpenAIRE Research Graph e DOAJ. Sequencialmente, foi criado pela Ontologia VIVO as entidades, classes e categorias dos dados indexados na proposta. Finalmente, para a sua representação/visualização, foram construídos *dashboards* para apresentar as estatísticas dos conjuntos de dados já carregados, bem como os gráficos para simplificação da visualização e recuperação das informações agregadas, também por meio do *software* VIVO.

BrCris as a tool to support Open Science

Information literacy

Learning strategies

Higher education

Teaching-learning

Reflexive practice

ACRL Framework

The proposal aims to present the BrCris system as a support tool for Open Science. The research targets are the organization of information in a single Platform of data collected on the Brazilian Research Ecosystem, whose goals are the construction of visualization and information retrieval services, in addition to the construction of interoperability interfaces between systems. The first results of the execution of the Project already include the development of the architecture and mapping of the relational and semantic data model, mapping of the data sources to be aggregated by the system, the implementation of aggregation tests for the mapped resources and the definition and execution of tests of services to be made available. Using the same infrastructure as the LA Referencia Platform, the project aggregated the data from the Lattes Platform, the patent dataset, the data from graduate programs and the international Wikidata, OpenAIRE Research Graph and DOAJ databases. Sequentially, the entities, classes and categories of the data indexed in the proposal were created by VIVO Ontology. Finally, for its representation/visualization, dashboards were built to present the statistics of the datasets already loaded, as well as graphics to simplify the visualization and retrieval of the aggregated information through the VIVO software.

INTRODUÇÃO

CRIS é um acrônimo para *Current Research Information System* (SIVERTSEN, 2019) que define um sistema de informação que busca agregar dados recentes sobre o processo de desenvolvimento científico de uma organização ou região. Tem o objetivo de organizar em um só lugar todas informações do ciclo da pesquisa científica, desde o Fomento, passando pelos Projetos, Pesquisadores, Instituições de Pesquisa e Laboratórios, até os *outputs* (resultados) de uma pesquisa científica, tais como: artigos científicos; teses; dissertações; livros; capítulos de livro; patentes; *software* e conjuntos de dados de pesquisa.

A concepção do Projeto do Sistema BrCris, que é o CRIS no contexto da Ciência Brasileira, data de 2014, quando 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 sequência de estudos (PEREIRA *et al*, 2019) e parcerias interinstitucionais para a execução deste Projeto. Em 2020, houve a implementação formal de um Projeto de Pesquisa para a construção do BrCris.

A proposta fundamental do BrCris é o desenvolvimento de um ecossistema integrado e interoperável, de forma a se obter painéis de destaque (*dashboards*), que trabalham com a dinâmica interativa conforme a busca e as atribuições de análise que o usuário solicita. Este sistema procura também viabilizar o intercâmbio de dados entre os diferentes atores do ecossistema da pesquisa científica brasileira, envolvendo o sistema de financiamento, os conteúdos científicos gerados, os dados tecnológicos, os repositórios de teses e dissertações, bem como os demais conteúdos científicos abertos pelas instituições brasileiras de ensino e pesquisa, mas também gerando *interfaces* de recuperação e tratamento dos dados.

O BrCris é uma proposta que toma por base a interoperabilidade entre sistemas de informação e tem como objetivos armazenar, gerenciar e intercambiar metadados de natureza científico-tecnológica, nas esferas organizacional, governamental e internacional. Esta interoperabilidade segue quatro vertentes de disposição e representação dos dados agregados. A primeira é voltada à indexação de pesquisadores brasileiros nos sistemas de informação e suas especialidades, contando com um conjunto de dados da Plataforma Lattes (CNPQ, 2021a) e Portal de Dados Abertos da CAPES (CAPES, 2021b).

A segunda está destinada à organização de dados sobre as revistas científicas que compõem o universo brasileiro de publicação e disseminação de pesquisas científicas, coletados das seguintes fontes:

- a. Diadorim: <https://diadorim.ibict.br/>
- b. Latindex: <https://www.latindex.org/>
- c. SherpaRomeo: <https://v2.sherpa.ac.uk/romeo/>
- d. Portal ISSN: <https://portal.issn.org/>
- e. DOAJ: <https://doaj.org/>
- f. UlrichsWeb: <https://ulrichsweb.serialssolutions.com/>
- g. Plataforma Lattes: <https://lattes.cnpq.br/>
- h. Google Scholar Metrics: https://scholar.google.com/citations?view_op=top_venues
- i. Dimensions: <https://www.dimensions.ai/>
- j. OpenAIRE Research Graph: <https://graph.openaire.eu/>
- k. Wikidata: <https://www.wikidata.org/>

A terceira via está destinada às patentes e seus dados relacionados, tais como citações, referências, informações sobre o pagamento de anuidade, da família da patente, e se esta é triádica (depósitos nos Estados Unidos da América, Europa e Japão). Para a reunião deste conjunto específico de dados, foram realizadas coletas das declarações de registro de patente

da Plataforma Lattes, do INPI (INPI, 2021) e da base de dados internacional Espacenet (EPO, 2021).

A quarta e última está voltada para a consolidação de um sistema de recomendação de especialistas tomando como base inicial o conjunto de dados das defesas de teses e dissertações, e tendo como fonte, novamente, a Plataforma Lattes (em suas sessões de formação, orientações concluídas e participações em bancas de defesa); a Biblioteca Digital Brasileira de Teses e Dissertações (IBICT, 2021a); o Portal Brasileiro de Publicações e Dados (IBICT, 2021c); e o

Catálogo de Teses e Dissertações da CAPES (CAPES, 2021a).

Por esta via, o desenvolvimento de um ecossistema de informação que permita que as várias fontes existentes trabalhem de forma interoperável à nível nacional faz paralelo a outras iniciativas existentes na América Latina e no mundo (BEVAN; HARRINGTON, 2011; JETTEN *et al*, 2019). Cabe ressaltar a importância dos sistemas do tipo CRIS para a América Latina, no que se refere aos mecanismos alternativos e abertos de avaliação da Ciência (FOLEC, 2021).

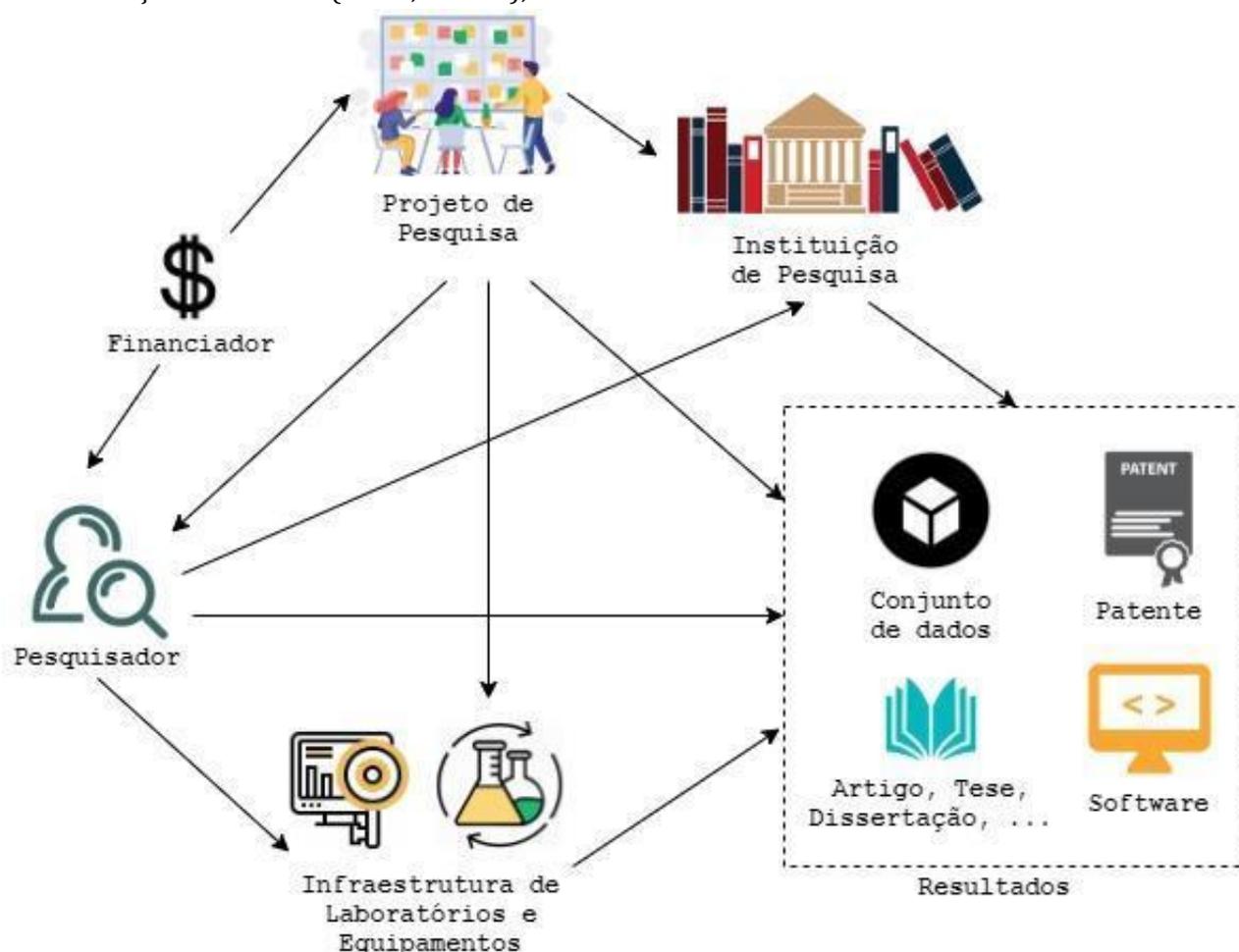


Figura 1 - Ecossistema do universo de desenvolvimento do BrCris

Fonte: Dados da pesquisa.

Pensando em uma audiência interessada na segmentação de aplicação e de estudos, o BrCris está voltado principalmente aos gestores de informação em órgãos públicos, mas também contempla audiência para os profissionais de informática e gestores de dados por trabalhar com raspagem, extração,

tratamento e disseminação da informação. Finalmente, contempla ainda a audiência de profissionais da informação (bibliotecários, arquivistas e documentalistas) e gestores de conteúdos editoriais científicos, por realizar extração, tratamento, armazenamento, organização e disponibilização de conteúdos

científicos, de acesso aberto, para a sociedade em geral.

Voltado ao acesso aberto de dados de investigação, a pesquisa está voltada a sistemas, políticas e práticas, no qual tem a função de:

- a. Repositórios digitais – institucionais, temáticos, de dados de investigação ou de patrimônio cultural;
- b. Revistas científicas de Acesso Aberto;
- c. Inovação na comunicação científica para a Ciência Aberta;
- d. Publicação institucional em Acesso Aberto;
- e. Gestão de Dados de Pesquisa e dados FAIR;
- f. Definição, análise e avaliação de políticas institucionais e de financiadores;
- g. Modelos e padrões de metadados; e
- h. Acesso Aberto.

O BrCris também está voltado para a geração de conhecimento aberto, visando identificar com a proposta:

- a. Ética, Integridade da Investigação e RRI (Responsible Research and Innovation /Investigação e Inovação Responsáveis);
- b. A avaliação da pesquisa e dos pesquisadores na transição para Ciência Aberta;
- c. Dados governamentais abertos;
- d. Desenvolvimento e comunidades de *software* livre para a promoção do Acesso Aberto e da Ciência Aberta, e;
- e. Outras práticas de conhecimento aberto tais como o *software* livre e a educação aberta.

Finalmente, pensando o BrCris do ponto de vista da gestão de informação de ciência e tecnologia, a proposta se enquadra em:

- a. Interoperabilidade entre sistemas de informação de apoio à atividade científica e acadêmica;

- b. Normas e diretrizes, e; Identificadores persistentes.

OBJETIVOS

A presente proposta tem como objetivo principal a apresentação do sistema BrCris, implementado via Projeto de Pesquisa recém-iniciado em 2020, como ferramenta de apoio à Ciência Aberta. São alvos da pesquisa a organização em uma só Plataforma das informações do Ecossistema da Pesquisa Brasileira. Esta organização se desdobra nas metas de construção de serviços de visualização e recuperação da informação, além da construção de *interfaces* de interoperabilidade entre sistemas.

DESENVOLVIMENTO

O desenvolvimento se deu em duas etapas: modelagem e implementação. A etapa de modelagem buscou levantar os conjuntos de dados disponíveis e, a partir daí, construir um modelo que integrasse as heterogeneidades de tais conjuntos e permitisse, assim, construir um sistema viável e intuitivo para os usuários. A etapa de implementação visa coletar e integrar o conjunto de dados qualificados para que estes possam servir para a construção de sistemas de recomendações.

Partindo da visão geral, temos o que se considera um modelo ideal para representar informações estratégicas que possam ser úteis para um grande número de usuários. O modelo inicial foi baseado em Schafer, Konstan e Riedl (2001) e adaptado para a realidade brasileira com o intuito de priorizar a construção de sistemas de recomendações.

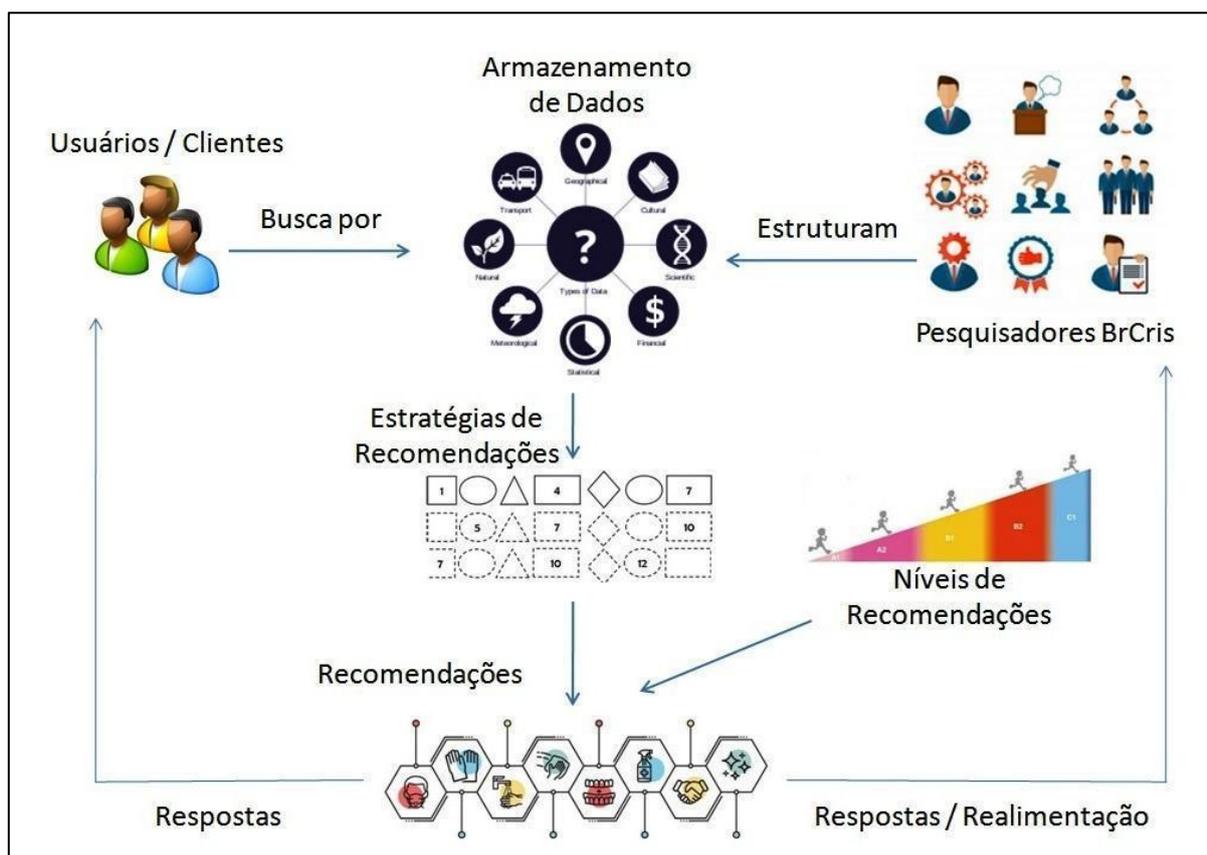


Figura 2 – Modelo representativo do Sistema de Recomendações a partir do BrCris
 Fonte: Dados da pesquisa baseada em Schafer, Konstan e Riedl (2001).

A implementação, que ainda está em curso, foi estruturada a partir da plataforma LA Referencia (LA REFERENCIA, 2021). A escolha por tal plataforma se deu em função da parceria entre o IBICT e a rede, e por conta da estabilidade de tal plataforma e de seu uso por outros CRIS. A LA Referencia permite que se tenha visibilidade para a produção científica do ensino superior e instituições de pesquisas da América Latina, promovendo acesso livre e gratuito ao texto integral, privilegiando os trabalhos financiados pelo setor público e em acesso aberto.

No contexto do BrCris, a Plataforma LA Referencia é utilizada primariamente para armazenar os dados coletados e serve também como ponto de acesso único a todos os dados que são armazenados. Por meio de um módulo versátil, construído na linguagem Python, e baseado em padrões abertos, a LA Referencia fornece um padrão de interoperabilidade com outras plataformas que venham a utilizar o sistema.

A partir da API de carga de dados disponibilizada pela LA Referencia, já foram construídas ou estão em fase final de construção interfaces para as seguintes fontes de dados: Oasisbr/IBICT (IBICT, 2021a), BDTD/IBICT (IBICT, 2021b), Espacenet, WikiData (WIKIDATA, 2021), Diadorim/IBICT, Latindex, DOAJ (DOAJ, 2021), UlrichsWeb, Plataforma Lattes, Plataforma Sucupira (CAPES, 2021a), Portal de Dados Abertos da CAPES, OpenAIRE Research Graph (OPENAIRE, 2021; LA BRUZZO, 2019) e Google Scholar Metrics. Neste conjunto, algumas plataformas servem para a coleta primária, outras para homogeneização e deduplicação dos dados (MANGHI, 2020) e outras para certificar os dados registrados nas demais fontes.

Para cada uma das fontes de dados citadas acima, foram e estão sendo construídos scripts na linguagem Python que transformam os dados brutos em um formato que pode ser utilizado pela API. O resultado do processamento destes scripts são arquivos XML que representam os dados de cada um dos

elementos que serão importados. Desta forma, é possível de forma prática e rápida a realização de testes de integridade, duplicar as bases para testes integrados e realizar também testes de performance em diferentes cenários e configurações de *hardware* e *software*.

O passo final é permitir a visualização dos dados por parte de pesquisadores que irão validar e analisar os dados coletados. Para tal, foram construídas inicialmente dois tipos de visualizações: uma sob forma de *dashboards* com os dados agregados utilizando o *software* Elasticsearch/Kibana e a segunda foi a construção de um modelo semântico que poderá exportar os dados no formato RDF e também permitir que os dados individuais possam ser visualizados com o *software* VIVO. A partir dos dados já coletados já é possível visualizar os projetos de pesquisa, programas de pós-graduação, cursos, organizações, pesquisadores e suas publicações sob a forma de artigos científicos em periódicos, livros, patentes e dados de pesquisa.

Os principais resultados esperados são o sistema de busca, bem como um painel integrado de visualização de dados (*dashboard*) que tenha a capacidade de consolidar os dados em uma visão macro sob forma de métricas e indicadores, quantificando assim a presença das organizações científicas e tecnológicas brasileiras e de sua produção.

Para questões metodológicas da proposta, ela é de natureza aplicada por estar reunindo e testando uma série de conteúdos/dados e aplicações de uso; a abordagem é tanto quantitativa como qualitativa por estar reunindo e criando um mecanismo de interoperabilidade e até mesmo de visualização e aplicação de métricas; em relação aos objetivos, podemos tê-los como sendo de base exploratória, em função dos testes e até mesmo da criação de um sistema inovador, que acreditamos ser único, na forma em que está construído, no cenário mundial; quanto aos procedimentos, caracteriza-se como experimental, levando justamente por conta o fato de se estar testando tecnologias e conjuntos de dados novos, e ao mesmo tempo se estar reunindo uma vasta gama de dados heterogêneos e de naturezas distintas.

DESENVOLVIMENTO

1 - Arquitetura: Os primeiros resultados da execução do Projeto já incluem o desenvolvimento da arquitetura do BrCris (ver *Figura 1*), o mapeamento das fontes de dados a serem agregadas pelo Sistema, a implementação de provas de agregação dos recursos mapeados e a definição e realização de testes de serviços a serem disponibilizados.



Figura 3 – Arquitetura do BrCris

Fonte: Os Autores

Como esquematiza a *Figura 3*, o coletador da LA Referencia (LA REFERENCIA, 2021b), um *software* desenvolvido de forma colaborativa, no âmbito da Rede LA Referencia, que aparece como plataforma agregadora e organizadora da informação coletada das diversas fontes mapeadas em âmbito nacional e internacional. A Plataforma LA Referencia é flexível quanto ao modelo de dados adotado para armazenamento das informações.

2 - Modelo semântico: Para consulta externa, os dados coletados no âmbito do projeto BrCris serão também representados em um modelo semântico baseado na Ontologia VIVO (VIVO, 2021), um modelo bastante utilizado internacionalmente por diversos CRISs. A representação em RDF com base em uma ontologia permite que o BrCris disponibilize Linked Open Data (LOD), tornando os dados não só abertos, mas também acessíveis e interoperáveis (SILVA *et al.*, 2021), seguindo, desta forma, as boas práticas disciplinadas pelos princípios FAIR (WILKINSON *et al.*, 2016).

A ontologia VIVO é um modelo que descreve o domínio de pesquisa acadêmica, e é composta por classes e propriedades para representar uma rede de pesquisadores, as instituições e projetos com as quais estão envolvidos e as publicações e demais produtos resultantes de suas pesquisas. Sua principal vantagem é reutilizar diversas outras ontologias já bastante consolidadas, entre as quais se destaca a *Basic Formal Ontology* (BFO), uma ontologia de fundamentação que fornece uma sólida base conceitual às classes e propriedades do modelo.

Outra vantagem da utilização de ontologias é o fato de estes recursos serem flexíveis e

extensíveis, permitindo a introdução de novos elementos (classes e propriedades) que se integram aos elementos já existentes sem, entretanto, descaracterizar o modelo original. Desta forma, o modelo semântico do BrCris foi desenvolvido através do mapeamento entre os elementos do modelo de entidades e relacionamento no qual os dados são armazenados internamente, na plataforma LA Referencia, e as classes e propriedades da ontologia VIVO. Quando uma classe ou propriedade equivalente não foi encontrada na ontologia, ou em outras ontologias existentes, um novo elemento foi criado, dando origem a uma extensão local da ontologia.

O modelo semântico final é, portanto, composto por um subconjunto da ontologia VIVO, acrescido de uma extensão local que engloba as classes e propriedades destinadas a representar informações específicas do projeto BrCris. Este modelo é composto por 18 classes, sendo que 16 delas foram reutilizadas da ontologia VIVO, e apenas duas criadas na extensão local, seguindo os princípios de maximização do reuso e minimização de customizações no modelo, para que ele seja tão aderente aos padrões internacionais quanto possível. Completam o modelo 78 propriedades de objeto e 133 propriedades de dados. A *Figura 4* ilustra dois recortes do modelo, com foco em Organizações (representadas pela classe foaf:Organization) e Publicações (classe bibo:Document). As classes representadas em verde pertencem à extensão local, e se relacionam com as classes da ontologia VIVO (em amarelo) tanto através de novas propriedades (em verde) quanto utilizando-se propriedades já existentes (em azul).

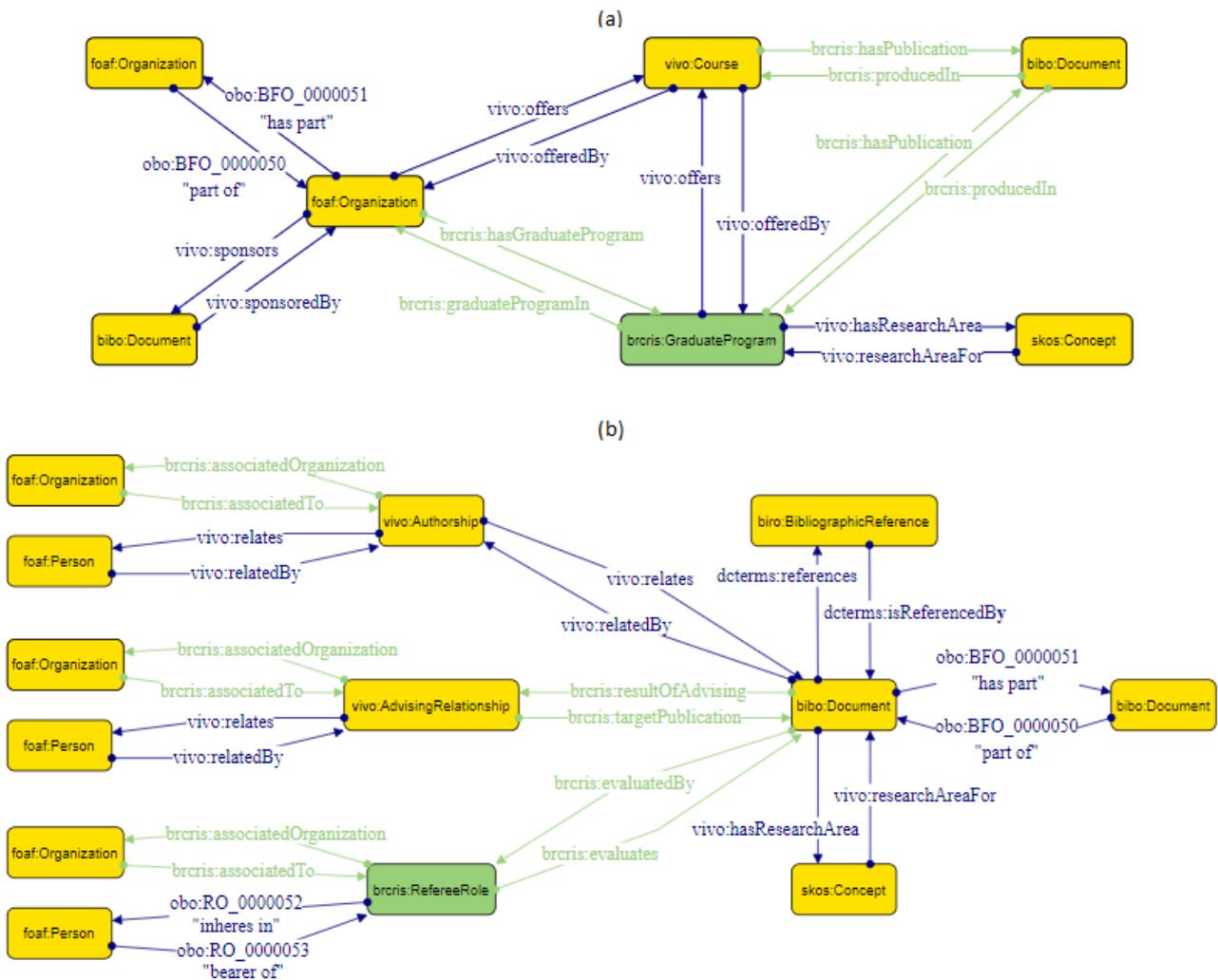


Figura 4 – Recortes do modelo semântico do BrCris para (a) Organizações e (b) Publicações
Fonte: Os Autores

3 – Recuperação e visualização dos dados:

Todas as ações de mapeio, coleta, transformação entre formatos e carga na Plataforma LA Referencia são executadas por módulos desenvolvidos em Python, o qual foi denominado Módulo BrCris, que se caracteriza como o orquestrador do Sistema. Dos dados organizados na Plataforma LA Referencia, é possível indexar as informações agregadas no motor de busca Elasticsearch. Esta ferramenta é um mecanismo de busca de texto completo de código aberto escrito em Java que foi projetado para receber grandes volumes de dados além de ser distributivo e escalável.

Para efeito de testes já foram dadas cargas de um volume de registros de Teses disponibilizadas no Portal de Dados Abertos da CAPES, além das publicações com DOI dos pesquisadores bolsistas de produtividade 1A do CNPq, obtidas por meio da Plataforma Lattes. A partir destas cargas, foram disponibilizados os primeiros *dashboards* na ferramenta Kibana (ELASTIC, 2021b), conectada automaticamente aos índices do Elasticsearch, para que outros pesquisadores possam realizar análises dos dados importados sem a necessidade de *softwares* fechados ou que dependam de licenças de uso proprietárias ou pagas (ver exemplo na Figura 5).

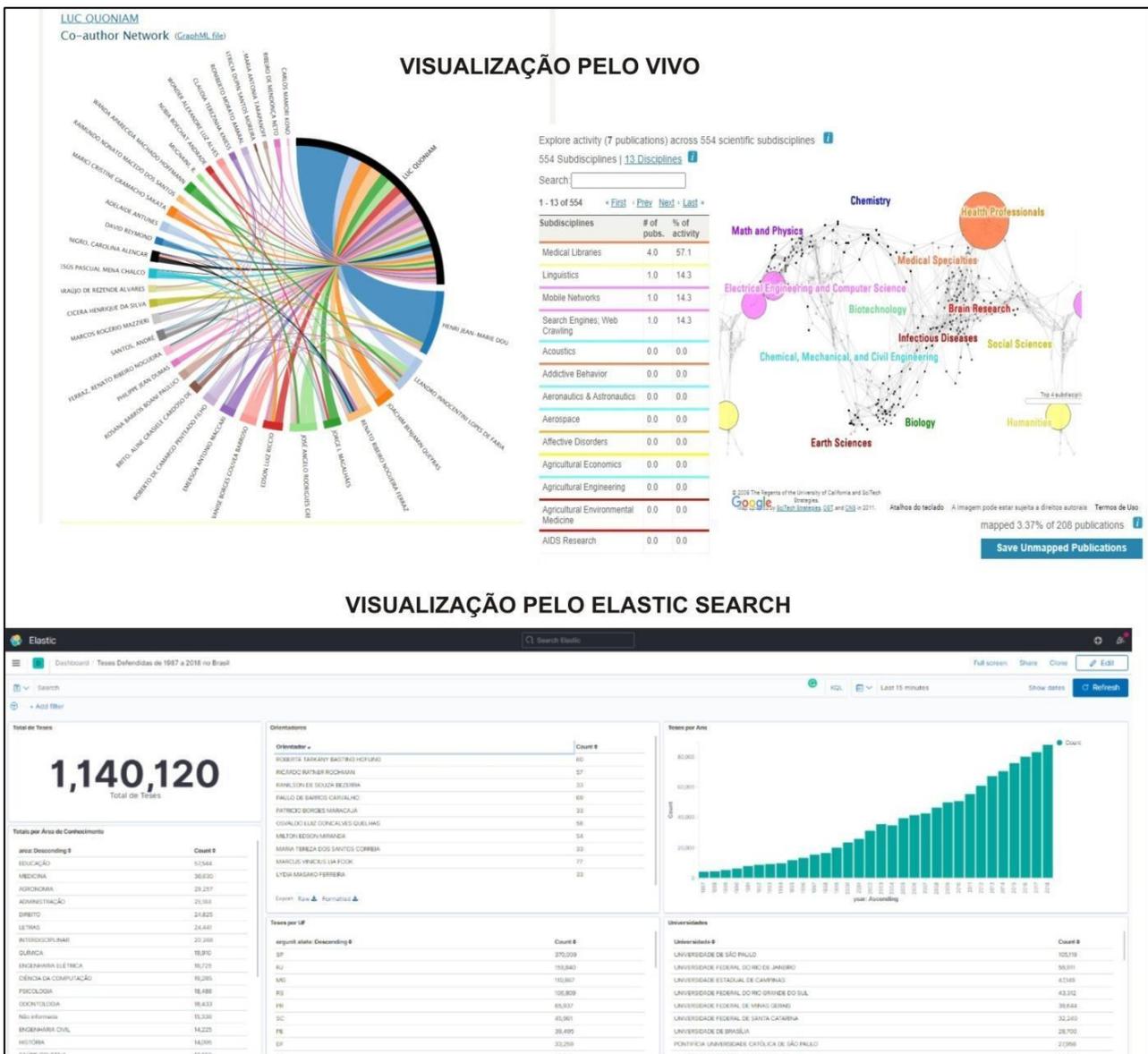


Figura 5 – Visualizações no VIVO e dashboard de Teses e Dissertações
 Fonte: Os Autores

Os principais resultados até o presente momento são: Tipos de informação abertas para construção de métricas e geração de conhecimento, voltados para:

- a. Teses e dissertações, com orientações, participações em bancas, atuação, vínculo institucional, genealogia, grupos de formações;
- b. Patentes por inovação frugal, valor de patente por citação, valor de patente por questões financeiras, valor da tríade dos depósitos de patentes (Estados Unidos da América, Japão e União Europeia), patentes de instituições pública e/ou empresarial, segmentação das patentes por sua classificação;

- c. Conjunto de dados sobre artigos científicos, visando dois aspectos, como autoridades produtores (pesquisadores e instituições), produtos produzidos como revistas e o impacto científico por métricas de impacto (H5 média capturado do Google Scholar Metrics), taxa de APC, tempo de submissão até à sua publicação, endogenia visando indicadores mais justos, fugindo dos padrões mundiais de quartis ou sistemas de classificação por indexação, exclusivamente, em base de dados comerciais. Um exemplo de dashboard gerado com o filtro de todos os periódicos pode ser visualizado na Figura 6.

- d. Certificação dos dados recuperados para que tenham validação de uso e até mesmo o reuso de dados abertos, e;
- e. Por fim, geração de indicação de especialistas, tanto para solução de problemas, quanto para formação de comitê de avaliação, bem como um sistema de recomendações de periódicos, ganhando transparência nos outputs da pesquisa científica brasileira.

Especificamente, no tema das patentes agregadas pelo BrCris, cujos registros são coletados principalmente da Plataforma Lattes e da base de dados Espacenet. Estes produtos da pesquisa científica constituem uma

descrição completa da invenção, de tal modo que pode ser reproduzida por um profissional da área. Nas bases de patentes constam os pedidos de patentes (levadas a exame ou não), às patentes aceitas ou reprovadas, as patentes extintas (por tempo, falta de pagamento ou outra razão) e enfim as patentes em vigor. Só esta parte de patentes em vigor é que não se pode usar livremente, sem autorização do inventor ou da empresa que depositou a patente. As demais constituem um conhecimento que pode ser utilizado, até para fins comerciais (caso conhecido dos remédios genéricos), com aplicações à ciência aberta, e a tecnologias abertas.

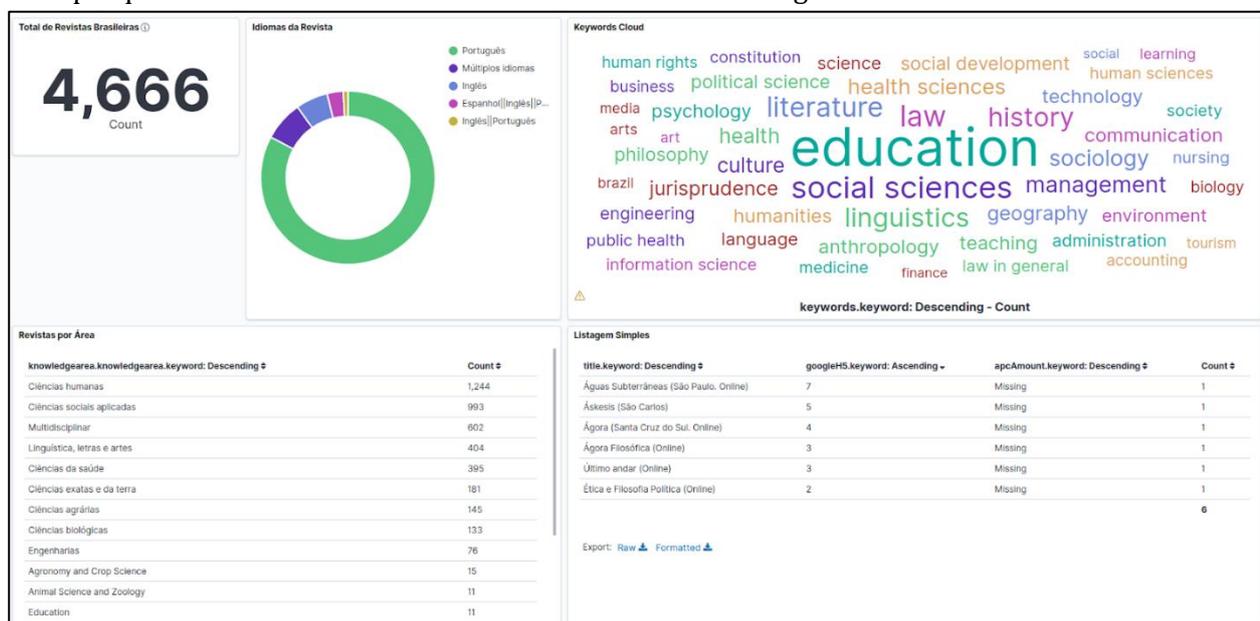


Figura 6 – Dashboard com os Periódicos Brasileiros

Fonte: Os Autores

CONCLUSÕES

O BrCris é uma iniciativa que coleta e enriquece dados de repositórios e bases de dados abertas utilizando a infra-estrutura do LA Referencia, protocolos OAI-PMH e múltiplos formatos de dados em XML, CSV e JSON.

Para esta primeira versão, o projeto está trabalhando com dados abertos da Plataforma Lattes, Bancos de patentes, Bancos de teses/dissertações (BDTD) e a base do OpenAIRE Research Graph. O foco do sistema está em processar os dados abertos relacionados à produção acadêmica, identificando entidades e seus atributos e

relacionamentos. Seu processo de armazenar entidades está caracterizado por sistemas de relacionamentos, com o objetivo de fornecer serviços de recomendações para a sociedade científica e tecnológica e também desenvolver um sistema de busca, bem como um padrão de visualização da informação por meio de *dashboards*.

A próxima etapa do projeto é a aplicação dos sistemas de recomendações através das métricas que podem ser explanadas em cada conjunto de dados, como: para periódicos científicos, com métricas de: (a) tempo de resposta da submissão até a publicação; (b) H5 média; (c) taxa de publicação; (d) comitê da

revista; (e) endogenia, e; (f) idiomas de publicação. Em se tratando das representações de orientações e formação, as métricas são orientadas para: (a) genealogia por até 11 níveis; (b) orientações por regiões, (c) país, (d) especialistas em temáticas em orientações ou por participação de bancas. Para as representações das patentes, pretende-se quantificar por: (a) citações recebidas, (b) citações realizadas, (c) concessão/prorrogação, (d) família de patentes, (e) concessão triádica, e (f) classificação.

REFERÊNCIAS

- BEVAN, Simon J.; HARRINGTON, John. Managing research publications: lessons learned from the implementation of a Current Research Information System. *Serials: The Journal for the Serials Community*. ISSN 0953-0460, 1475-3308. 24:1 (2011) 26–30. doi: 10.1629/2426.
- CAPEs. *Catálogo de teses & dissertações*. CAPEs. <https://catalogodeteses.capes.gov.br/catalogo-teses/#!/>.
- CAPEs. *Plataforma Sucupira*. <https://sucupira.capes.gov.br>.
- CAPEs. *Portal de Dados Abertos da Capes*. <https://dadosabertos.capes.gov.br/>.
- CNPQ. *Plataforma Lattes*. <https://lattes.cnpq.br/>.
- DOAJ. *The Directory of Open Access Journals*. <https://doaj.org>.
- ELASTIC. *What is Elastic Search?* <https://www.elastic.co/pt/what-is/elasticsearch>.
- ELASTIC. *Kibana*. <https://www.elastic.co/pt/kibana>.
- EPO. *Espacenet – patent search*. <https://worldwide.espacenet.com/>.
- FOLEC. *Hacia la transformación de los sistemas de evaluación en América Latina y el Caribe: herramientas para promover nuevas políticas evaluativas*. 2021.
- IBICT. *Biblioteca Digital Brasileira de Teses e Dissertações (BDTD)*. <http://bdttd.ibict.br>.
- IBICT. *Plataforma de Instituições de Ciência, Tecnologia e Inovação*. <http://BrCris.ibict.br/picti/api/icti/docs/>.
- IBICT. *Portal Brasileiro de Dados e Publicações Científicas (OasisBr)*. <http://oasisbr.ibict.br>.
- INPI. *Busca de Patentes*. <https://busca.inpi.gov.br/pePI/>.
- LA BRUZZO, Sandro; MANGHI, Paolo; MANNOCCI, Andrea. OpenAIRE's DOIBoost . Boosting Crossref for Research. In MANGHI, Paolo; CANDELA, Leonardo; SILVELLO, Gianmaria (Eds.). *Digital Libraries: Supporting Open Science. Communications in Computer and Information Science*. Cham : Springer International Publishing, 2019.
- LAREFERENCIA. *Quem somos*. <https://www.lareferencia.info/pt/institucional/quienes-somos>.
- LAREFERENCIA. *Plataforma de cosecha / transformación / validación / publicación – LAReferencia*. <https://github.com/lareferencia/lrharvester>.
- MANGHI, Paolo *et al.* Entity deduplication in big data graphs for scholarly communication. *Data Technologies and Applications*. ISSN 2514-9288. 54:4 (2020) 409–435. doi: 10.1108/DTA-09-2019-0163.
- OPENAIRE. *OpenAIRE Research Graph*. <https://graph.openaire.eu>.
- PEREIRA, Maria De Nazaré Freitas; CHAVES, Hélia de Sousa; ARAÚJO, Ronaldo Ferreira. *Dos padrões internacionais de estruturação da informação de pesquisa aos indicadores: primeira incursão na temática*. Brasília, DF : IBICT, 2019 <https://bit.ly/3gaZbTG>. ISBN 978-85-7013-153-9. p. 220.
- SCHAFER, J. Ben; KONSTAN, Joseph A.; RIEDL, John. E-commerce recommendation applications. *Data Mining and Knowledge Discovery*. ISSN 1573-756X. 5:1 (2001) 115–153. doi: 10.1023/A:1009804230409.
- SILVA, Vivian S., MOREIRA, Tales, DIAS, Thiago M., GOMES, Josir, SEGUNDO, Washington C. Um modelo semântico baseado em ontologia para o CRIS brasileiro. *Proceedings of the V Congress of the International Society for Knowledge Organization (ISKO)*. Lisboa, Portugal. 2021.
- SIVERTSEN, Gunnar. Developing Current Research Information Systems (CRIS) as Data Sources for Studies of Research. In GLÄNZEL, WOLFGANG *et al.* (Eds.). *Springer handbook of science and technology indicators*. Springer Handbooks. Cham: Springer International Publishing, 2019.
- WIKIMEDIA. *Portal Wikidata*. <https://www.wikidata.org>.
- WILKINSON, Mark D. *et al.* The FAIR guiding principles for scientific data management and stewardship. *Scientific Data*. ISSN 2052-4463. 3:1 (2016) 160018. doi: 10.1038/sdata.2016.18.
- VIVO. VIVO Ontology Domain Definition. <https://wiki.lyrasis.org/display/VIVODOC111x/VIVO+Ontology+Domain+Definition>.

Segundo, W.; Gomes, J.; Dias, T. M. R.; Quoniam, L.; Moreira, T.; Matas, L.; Pinto, A. L.; Dias, A. G.; Silva, V. S.; Schneider, J. (2021). O BrCris como ferramenta de apoio à Ciência Aberta. *Cadernos BAD*, n. 1-2.
<https://doi.org/10.48798/cadernosbad.2735>

Acesso e licença

Artigo em acesso aberto distribuído nos termos da licença Creative Commons Atribuição 4.0 Internacional (CC-by 4.0).

Revisão por Pares

Esta revista usa um sistema de revisão duplamente cega por pares assegurada pelo Conselho Científico da *Cadernos BAD*.

Conflitos de Interesse

Os autores declaram a inexistência de conflitos de interesse na realização do presente trabalho.

Confidencialidade dos Dados

Os autores declaram ter seguido os protocolos do RGPD.

Financiamento, apoio e patrocínios

Não existiram fontes externas de financiamento para a realização deste artigo.

Recebido

06/10/2021

Aceite

13/10/2021

Publicado

26/09/2022
