

BRCRIS: PLATAFORMA PARA INTEGRAÇÃO, ANÁLISES E VISUALIZAÇÃO DE DADOS TÉCNICOS-CIENTÍFICOS

BRCRIS: PLATFORM FOR INTEGRATION, ANALYSIS AND VISUALIZATION OF TECHNICAL-SCIENTIFIC DATA

Thiago Magela Rodrigues Dias^a
Jesus Pascual Mena Chalco^b
Washington Luís Ribeiro de Carvalho Segundo^c
Adilson Luiz Pinto^d
Tales Henrique José Moreira^e

RESUMO

Objetivo: neste trabalho apresentamos o processo de desenvolvimento da Plataforma BrCris cujo objetivo é de fornecer ferramentas tecnológicas para munir à comunidade acadêmica brasileira com dados integrados e consolidados da produção científica do Brasil. **Metodologia:** inicialmente para se contemplar os objetivos propostos um conjunto de repositórios de dados foram selecionados, e após a extração destes conjuntos, os dados foram transformados em um modelo baseado em entidades e relacionamentos, sendo integrados para as análises e visualizações propostas. **Resultados:** Como resultado da construção da Plataforma BrCris foi possível obter um conjunto de dados técnicos-científicos consistentes para viabilizar diversos tipos de análises. **Conclusões:** Esta iniciativa se apresenta como mecanismo ímpar de agregação de dados, possibilitando visualizações e análises cientométricas sobre o presente, passado e futuro da produção científica brasileira.

Descritores: Acesso Aberto. Recuperação de Informação. Dados Científicos. Dados Abertos.

^a Doutor em Modelagem Matemática e Computacional. (CEFET/MG). Docente do Centro Federal de Educação Tecnológica de Minas Gerais (CEFET-MG), , Divinópolis, Brasil. E-mail: thiagomagela@gmail.com

^b Doutor em Ciências da Computação pela Universidade de São Paulo (USP). Docente da Universidade Federal do ABC (UFABC). Santo André, Brasil. Email: jmenac@gmail.com.

^c Doutor em Informática pela Universidade de Brasília (UnB), Brasília, Brasil. E-mail: washingtonsegundo@ibict.br

^d Doutor em Documentação pela Universidad Carlos III de Madrid. Docente do Departamento de Ciência da Informação da Universidade Federal de Santa Catarina (UFSC), Florianópolis, Brasil. E-mail: adilson.pinto@ufsc.br

^e Mestre em Modelagem Matemática e Computacional pelo Centro Federal de Educação Tecnológica de Minas Gerais (CEFET/MG), Belo Horizonte, Brasil. E-mail: tales.info@gmail.com

1 INTRODUÇÃO

O ecossistema da pesquisa envolve a presença de diversos atores que interagem entre si. Desde o financiamento obtido por meio de um projeto de pesquisa, passando pela figura do próprio pesquisador, que faz uso de infraestrutura para efetuar seus trabalhos, tais como laboratórios e equipamentos físicos (LEE; BOZEMAN, 2005). Os pesquisadores, por sua vez, estão associados e são mantidos por Instituições onde as pesquisas são desenvolvidas. Uma das principais contribuições do trabalho de um pesquisador é a produção de conhecimento, expressa em termos de artigos científicos que comumente são indexados nos repositórios acadêmicos como por exemplo Web of Science, Scopus e Dimensions (SINGH *et al.*, 2021).

Há ainda, produtos de pesquisa que geralmente não são visíveis nos repositórios acadêmicos tradicionais, como, por exemplo os produtos associados à literatura cinzenta representada por teses, dissertações, relatórios técnicos, banco de dados com resultados de experimentos, manuais, protocolos (PAEZ, 2017). Aqui é importante frisar que esses produtos, não tradicionais, são tão importantes quanto os artigos científicos dado que permitem observar a produção de todo o ecossistema de pesquisa (RAMSDEN, 1994).

Para o campo da Ciência da Informação, e em especial da Cientometria, quantificar essa produção é uma tarefa árdua pois a disponibilização de bases de dados abertas muitas vezes é restrita ou simplesmente inexistente em determinados contextos. Bases de dados proprietárias como a Scopus, Web of Science, Google Acadêmico e Microsoft Research Data permitem o acesso, mas este é sempre limitado ao número de registros que podem ser obtidos, contemplam poucos repositórios e periódicos nacionais e ainda existe o grave problema da opacidade dos algoritmos utilizados por estas plataformas que determinam o que é ou não relevante.

A partir desse cenário, começaram a surgir iniciativas que visavam a criação de sistemas que gerenciam a produção acadêmica de uma instituição, país ou área de conhecimento. Tais sistemas são conhecidos pela sigla CRIS (*Current Research Information Systems*) e têm como objetivo agregar

informações de bases de dados diversas com intuito de fornecer relatórios e dados consolidados para que pesquisadores da área possam analisar como se dá a produção em seus países ou áreas de conhecimento.

CRIS define um sistema de informação sobre todo o ecossistema do processo científico. São organizadas 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).

Neste contexto, a idealização do projeto do Sistema BrCris (PINTO *et al.*, 2021), 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 e parcerias interinstitucionais para a execução do projeto. Em 2020, houve a implementação formal de um projeto de pesquisa para a construção do BrCris. O intuito é fornecer ferramentas tecnológicas visando munir dados consolidados da produção científica brasileira para toda a comunidade acadêmica.

Logo, 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 e patentes (KONG *et al.*, 2019).

2 DESENVOLVIMENTO

O BrCris concentra um amplo ecossistema de dados, de diversas fontes, como por exemplo, dados curriculares de indivíduos, sobre organizações, programas de pós-graduação, publicações, orientações acadêmicas, revistas científicas, dentre outros, sendo necessário todo um esforço para tratamento dos dados de interesse. Neste contexto, tendo em vista as diversas fontes de dados que irão compor o BrCris se faz necessário a transformação dos dados em formato padronizado. Em se tratando do modelo de dados do BrCris, iniciou-se pela adoção de nove entidades de dados, seguindo padrões amplamente utilizados na comunidade científica internacional:

- *Project*: projetos de pesquisa executados, ou em execução;
- *Service*: revistas científicas, repositórios digitais, bibliotecas digitais e outras fontes de informação científica;
- *Program*: programas de pós-graduação brasileiros;
- *Course*: cursos de pós-graduação stricto ou lato sensu;
- *OrgUnit*: instituições, faculdades, departamentos de pesquisa;
- *Person*: pesquisadores, assistentes de pesquisa e de apoio técnico à pesquisa;
- *Patent*: patentes como resultado da pesquisa;
- *Dataset*: conjuntos de dados de pesquisa de um projeto ou pesquisa científica;
- *Publication*: artigos científicos, teses, dissertações, livros, capítulos de livro.

O modelo de dados é definido por um conjunto de entidades e relações, que por sua vez possuem identificadores e atributos pré-definidos. A utilização de um descritivo visa facilitar a identificação de atributos de cada entidade (Figura 1) e suas relações (Figura 2), possibilitando que o modelo possa incorporar todas as mudanças realizadas diretamente no modelo. Esta estratégia visa facilitar de forma significativa a incorporação de novos atributos e relações, sem a necessidade de alterações diretamente no modelo de dados.

Figura 1 – Descritivo de Entidades do BrCris

Field	Description	Diadorim	DOAJ
identifier diadorim	id diadorim	id	
identifier latinindex	id LatinDex		
identifier ulrichsweb	id Ulrichsweb		
identifier issn	id issn	dc.identifier.issn[pt_BR]	Journal EISSN (online version)
identifier issnI	id issnI	dc.identifier.issn[pt_BR]	Journal ISSN (print version)
identifier oai	id oai	dc.identifier.oai	
identifier uri	id uri	dc.identifier.uri	Journal URL
identifier brcris	hash gerado com título + publisher	dc.title[pt_BR] + dc.publisher.name[pt_BR]	Journal title + Publisher
identifier eissn	(Será considerado como um issn)		Journal EISSN (online version)
identifier opendoarid			
identifier other			
compatibility			
acronym			
status	status da revista (active / inactive)	dc.relation.situation[pt_BR]	
accessType	tipo (direito) de acesso	dc.rights.access[pt_BR]	Permite acesso ao texto completo
ccLicence	Permissões	dc.rights.cc[pt_BR]	Journal license
rightsType		dc.rights.type[pt_BR]	
researchArea	área de pesquisa		

Fonte: Os autores (2022)

Figura 2 – Descritivo das Relações entre Entidades do BrCris

Name	From Entity	From Label	To Entity	To Label	Description
Affiliation	OrgUnit	hasMember	Person	isMemberOf	Is a relation between Person and OrgUnit
IsUnitOf	OrgUnit	isUnitOf	OrgUnit	hasUnit	The unit related to an organization.
OrgUnitProgram	OrgUnit	hasProgram	Program	isProgramOf	The program related to an organization.
OrgUnitProject	OrgUnit	hasProject	Project	isProjectOf	Is a relation between Project and OrgUnit.
ThesisSponsorship	OrgUnit	sponsors	Publication	isSponsoredBy	Is a relation between a publication of type thesis and a sponsor OrgUnit.
CourseOrgUnit	OrgUnit	hasCourse	Course	isProvidedBy	The OrgUnit(s) that provided the Course.
Authorship	Publication	hasCreator	Person	isCreatorOf	The author of this content or rating.
Advisoring	Publication	hasAdvisor	Person	isAdvisorOf	The advisor of this content or rating.
CoAdvisoring	Publication	hasCoAdvisor	Person	isCoAdvisorOf	The coadvisor of this content or rating.
Referee	Publication	hasReferee	Person	isRefereeOf	The advisor of this content or rating.
Publisher	Publication	hasPublisher	Service	isPublishedIn	The editor of this content or rating.
PartOf	Publication	hasPartOf	Publication	isPartOf	The publisher of this content or rating.
ProgramThesis	Publication	hasPublication	Program	isProgramOf	Is a relation between two publications.
ServiceOrgUnit	Service	hasOrgUnit	OrgUnit	isOrgUnitOf	The publication (thesis) related to an program.
Editing	Service	hasEditor	Person	isEditorOf	The Course that is associated with a thesis (publication).
CourseThesis	Course	isAssociatedTo	Publication	hasAssociationWith	Is a relation between Patent and Person.
Inventor	Patent	hasInventor	Person	isInventorOf	Is a relation related to an OrgUnit of Patent.
PatentOrgUnit	Patent	isPatent	OrgUnit	hasPatent	Is a relation between Project consortium and member.

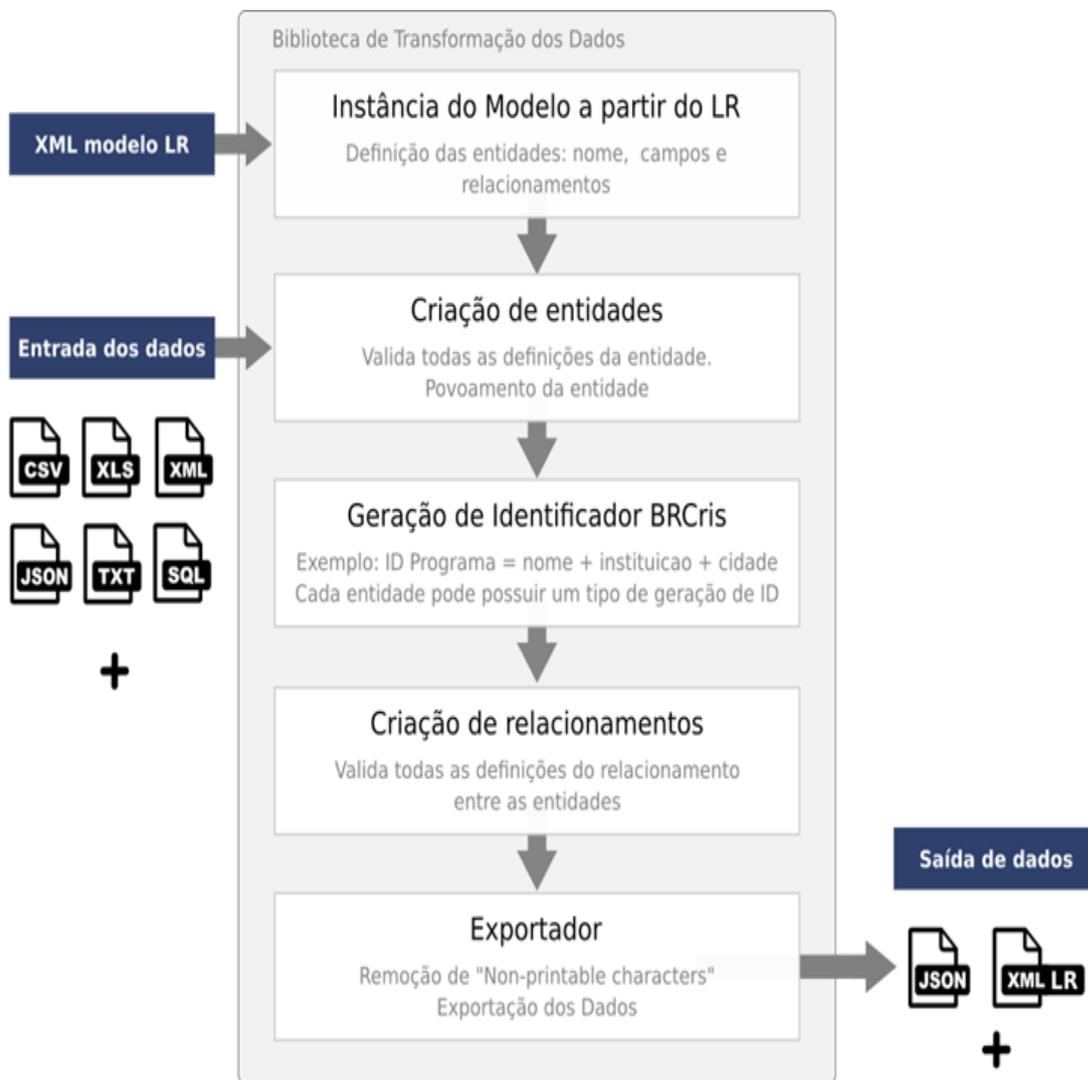
Fonte: Os autores (2022)

Como pode ser observado, o descritivo de uma Entidade apresenta inicialmente o atributo definido no Modelo de Dados bem como sua respectiva descrição. Logo, para cada conjunto de dados que é fonte de informações para a Entidade, são descritos os atributos dos conjuntos de dados relacionados aos do modelo.

Para o tratamento dos dados foi desenvolvida uma biblioteca computacional contendo uma estrutura de dados preparada para facilitar o processamento de dados originários de todas as fontes para o formato exigido pela plataforma LA Referencia (Figura 3). Nesse sentido, a biblioteca desenvolvida é responsável por toda a transformação e exportação dos dados,

utilizando como base o “Modelo de Dados” da plataforma LA Referencia, validando as entidades, campos e relacionamentos aceitos pelo modelo.

Figura 3 – Biblioteca para transformação dos dados



Fonte: Os autores (2022)

Como pode ser observado, independentemente do formato de dados que seja importado pela biblioteca, todo um conjunto de tratamento nos dados são aplicados, com rotinas computacionais amplamente otimizadas, possibilitando com auxílio do descritivo de dados, identificar as entidades, bem como os seus relacionamentos e gerar um arquivo de saída, padronizado para que este possa ser importado pela plataforma La referencia, ou outra ferramenta de armazenamento.

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.

Dentre as fontes de dados coletadas, estão:

- Plataforma Lattes do CNPq: Possui um conjunto de dados curriculares, sendo a principal fonte de dados do BrCris, ela agrega dados pessoais dos indivíduos, instituições de vínculo, produção científica e tecnológica (artigos em periódicos, artigos em congressos, livros, capítulos de livros, patentes, dentre outros), participação e organização de eventos, formação acadêmica, orientações, e participação em bancas. Além dos currículos, também são agregados dados de instituições de ensino e pesquisa. Todos estes dados, foram transformados em entidades do modelo BrCris, sendo elas, Person, OrgUnit, Publication, Event, Book, Program, Course e Service respectivamente;

- OasisBR: Mantido pelo IBICT, fornece dados confiáveis sobre publicações científicas em acesso aberto. Os dados foram mapeados para as entidades Publication, Service, Person.

- BDTD: A exemplo do OasisBR, a BDTD também é mantida pelo IBICT. Fornece dados confiáveis sobre teses e dissertações brasileiras. Os dados foram mapeados para as entidades Publication, Course e Person.

- Plataforma Sucupira: Concentra dados dos Programas de Pós-graduação do Brasil, fornecendo um conjunto de informações sobre os programas e cursos de pós-graduação. Todos os dados dos programas foram mapeados para as entidades Program, Course e OrgUnit;

- Dados Abertos da CAPES: Fornece dados como publicações, orientações, entre outros que são mapeados para diversas entidades, como Person, OrgUnit, Publication, Program e Course.

Diante do exposto, é possível verificar como ocorre o processo de interoperabilidade dos dados no BrCris e também como estes foram selecionados e processados para uma integração, que engloba alguns elementos para possíveis desambiguações, utilizando para isso, identificadores implícitos ou gerados no processo de tratamento dos dados. Logo, com os dados agregados, diversas análises são viabilizadas, proporcionando um maior conjunto de análises bibliométricas.

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* (<https://www.elastic.co/pt/what-is/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.

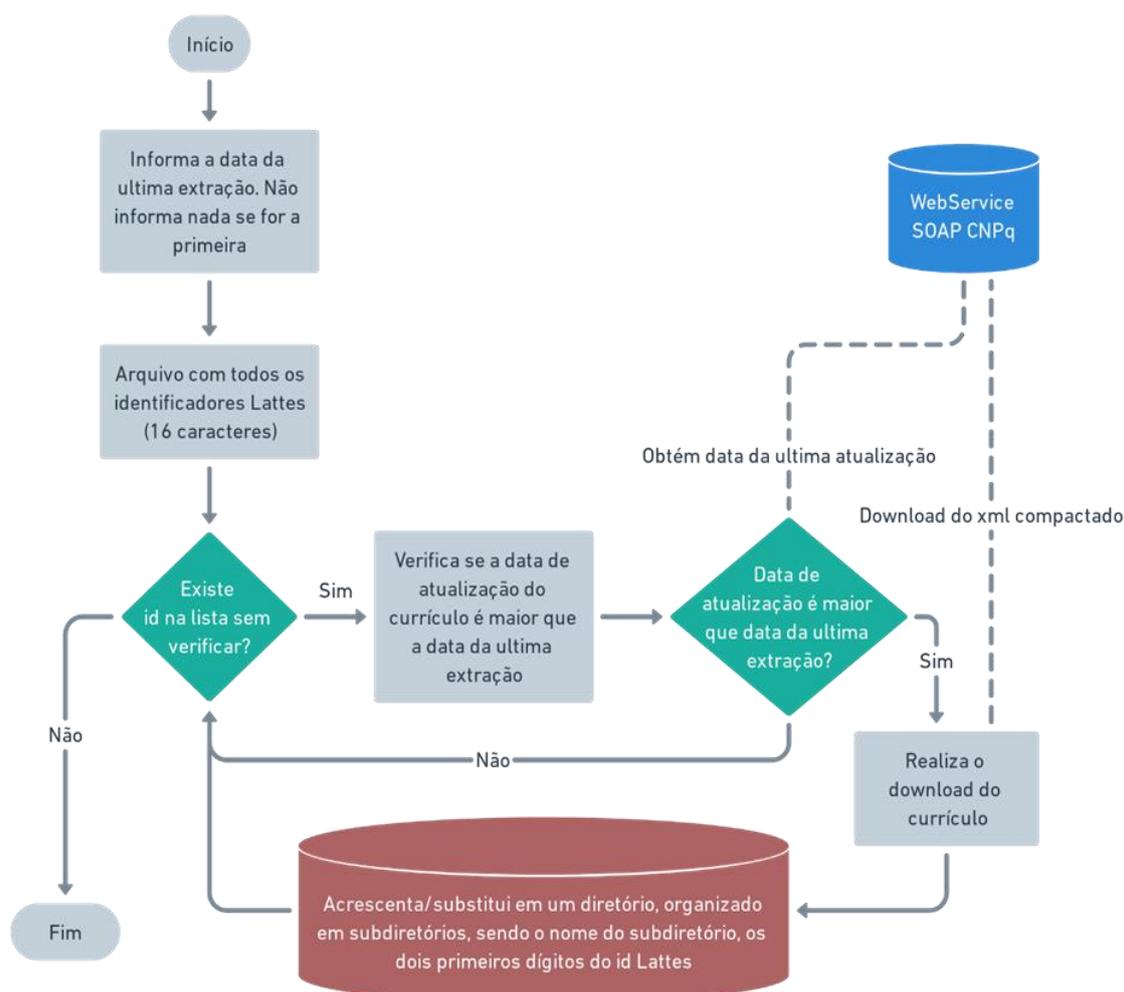
3 RESULTADOS

O principal repositório de dados para o BrCris são os currículos cadastrados na Plataforma Lattes do CNPq. Acessível em <<http://lattes.cnpq.br/>>, a Plataforma Lattes foi criada e é mantida pelo CNPq, contando atualmente com mais de 7,8 milhões de currículos cadastrados, além de grupos de pesquisa e diretórios de instituições.

Diante da necessidade de manter bases constantemente atualizadas para processamento em diversas frentes, foi implementado um sistema para manter uma base local atualizada e automatizada através de serviços disponibilizados pela Plataforma Lattes. A Plataforma Lattes, através do CNPq, disponibiliza um *webservice* SOAP para acesso aos currículos, podendo ser usado por instituições previamente credenciadas. A partir desse acesso, é possível obter dados como data da última atualização dos currículos, e o XML contendo todos os dados curriculares.

Diante disso, o extrator incremental (Figura 4) desenvolvido acessa o *webservice*, a fim de obter a data de atualização de todos os currículos para decidir se o mesmo será atualizado. Caso a data da atualização seja maior que a data da última coleta, ou seja, se depois da última execução da atualização incremental, algum currículo foi modificado, estes são baixados e substituídos em disco, organizados em uma hierarquia de diretórios, a fim de evitar grandes leituras em um mesmo diretório.

Figura 4 – Fluxograma do Extrator Incremental

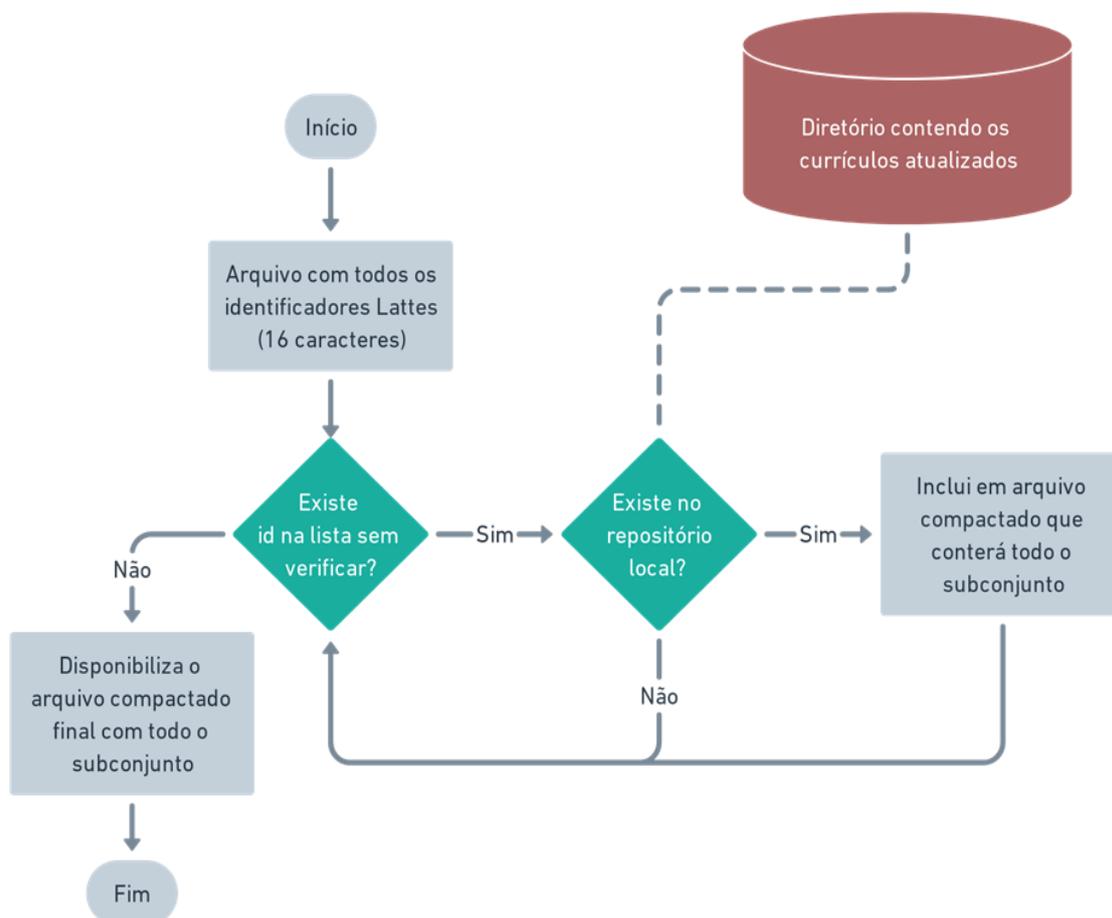


Fonte: Os autores (2022)

A estrutura de diretórios é formada contendo os dois primeiros dígitos do identificador Lattes, gerando uma gama de diretórios iniciam em 00 e terminam em 99. Cada diretório, conterà apenas currículos com identificadores correspondentes. Ou seja, se houvesse um currículo com identificador “123456789012345”, este seria armazenado dentro do diretório “12”.

A ferramenta ainda contempla a possibilidade de gerar pacotes com subconjuntos de currículos, a partir de uma lista de identificadores. Isso possibilita o processamento parcial da base de forma simplificada (Figura 5).

Figura 5 – Empacotamento para geração de subconjuntos



Fonte: Os autores (2022)

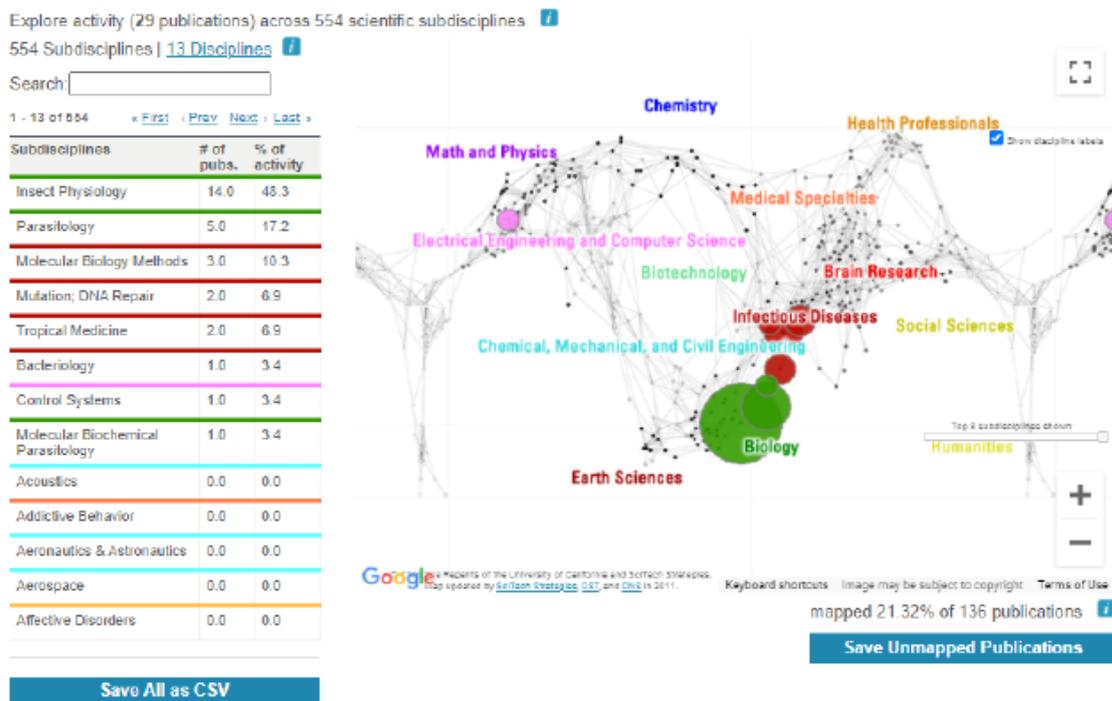
Como pode ser observado, as diversas fontes de dados mapeadas, se completam, possibilitando a criação de um conjunto padronizado e consistente, validado através de dados provenientes de diversas entidades brasileiras amplamente consolidadas e utilizadas. Ao se agregar todos os repositórios apresentados, é possível a adoção de técnicas que visam permitir a vinculação de conjuntos que inicialmente não era possíveis de se comunicarem, possibilitando dessa forma, a construção de um grande conjunto de dados, interligados, que facilitam a aplicação de consultas que inicialmente não seriam possíveis.

Os resultados da execução do projeto já incluem o desenvolvimento da

arquitetura do BrCris, o mapeamento das fontes de dados a serem agregadas pelo Sistema, a implementação de provas de agregação dos recursos mapeados, a definição e realização de testes de serviços a serem disponibilizados.

Diversas visualizações já estão implementadas o que viabiliza obter retratos da produção científica nacional de forma inédita. Tais visualizações possibilitam a utilização de filtros e outros métodos para agregar elementos de personalização como por exemplo análises temporais ou por áreas (ver Figura 2).

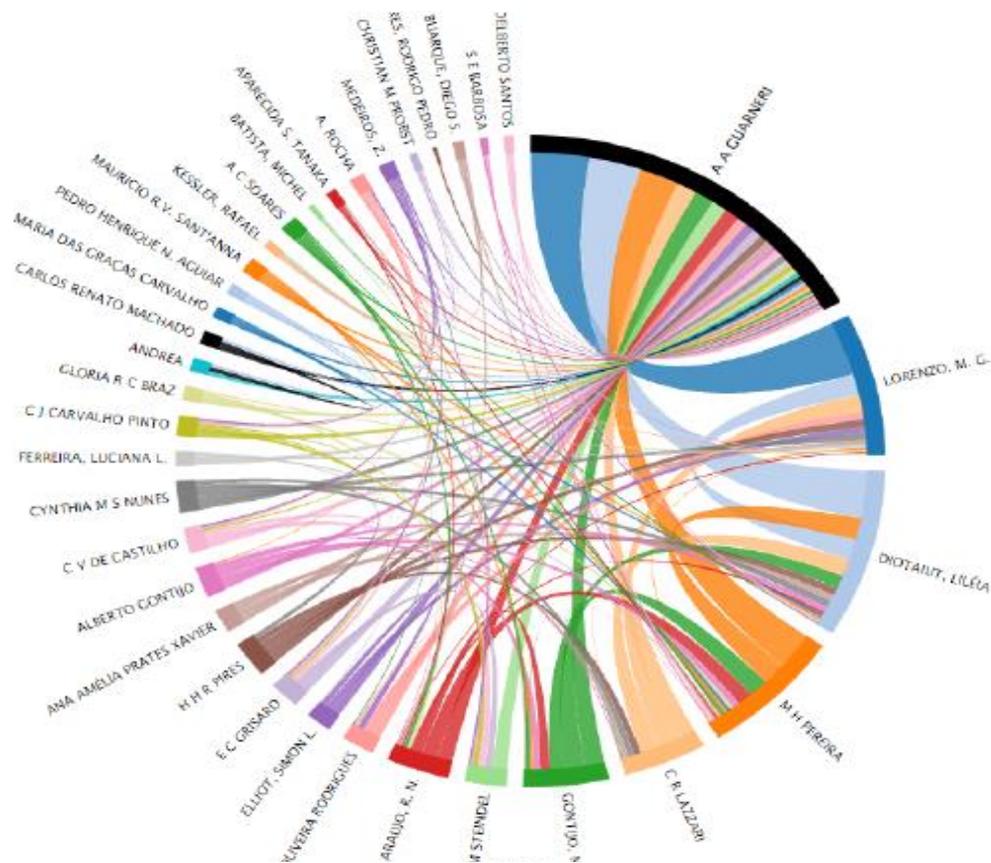
Figura 6 – Dashboard para Análises da Produção Científica



Fonte: Os autores (2022)

Também é possível realizar a visualização de redes de colaboração o que viabiliza diversos tipos de análise sobre como tem ocorrido o processo de coautoria nas diversas áreas do conhecimento (Figura 7).

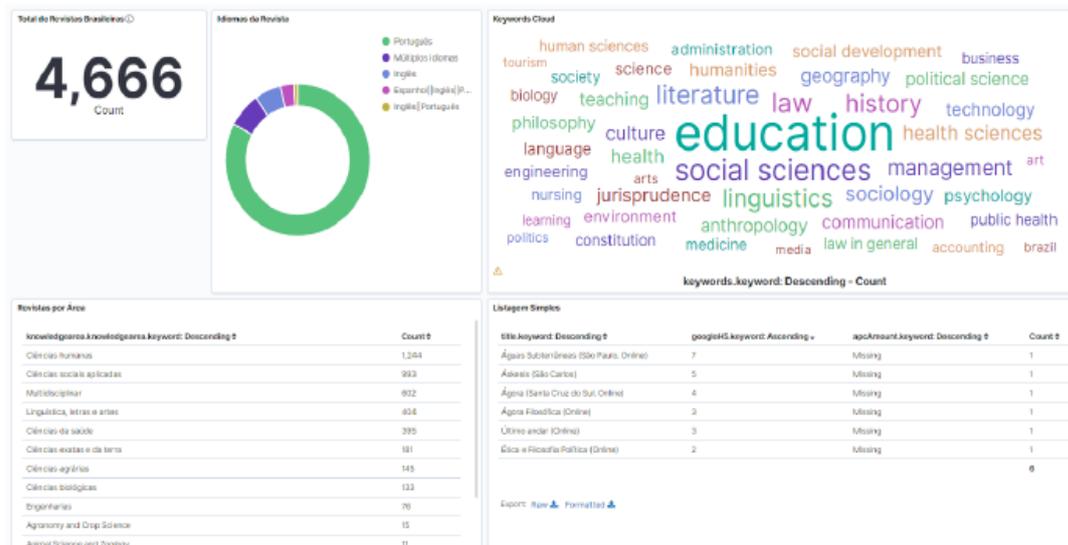
Figura 7 – Grafo de Colaboração Científica



Fonte: Os autores (2022)

Além disso visualizações que integram grandes conjuntos de dados também já estão disponíveis como por exemplo, de periódicos científicos (Figura 8).

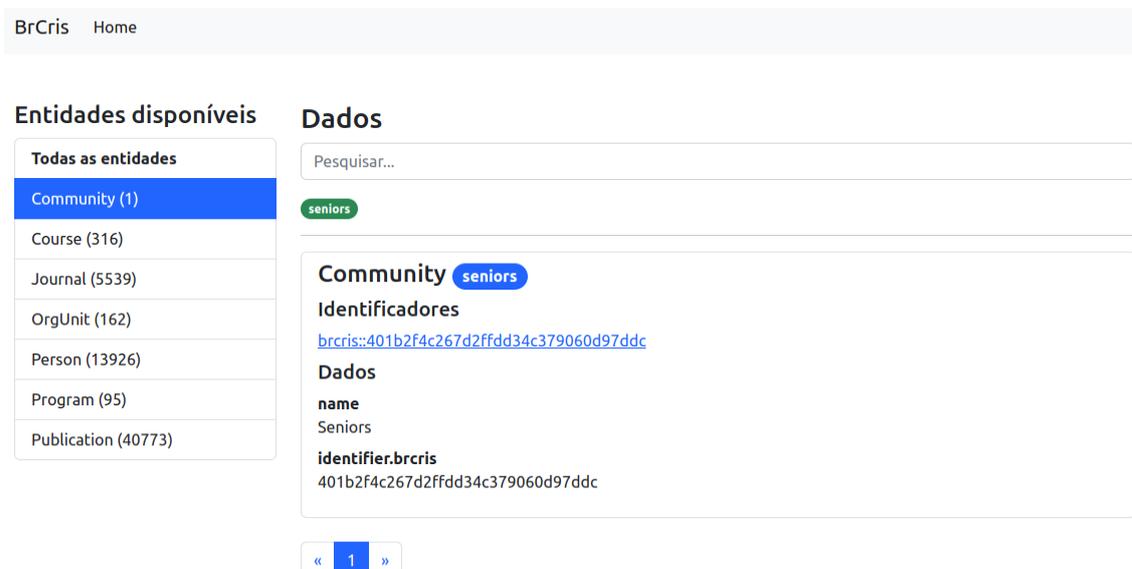
Figura 8 – Dashboard para Análises sobre Periódicos Científicos



Fonte: Os autores (2022)

Objetivando a possibilidade de análises de grupos específicos de indivíduos, é possível realizar o agrupamento utilizando a estratégia de comunidades. Uma comunidade pode agrupar indivíduos baseado em critérios de classificação (Figura 9). Logo, grupos específicos como os docentes de um programa de pós-graduação ou pesquisadores de uma determinada áreas podem ser agrupados e posteriormente analisados.

Figura 9 – Caracterização de comunidades



Fonte: Os autores (2022)

Com as comunidades caracterizadas, diversas visualizações com os dados do conjunto podem ser realizadas, como por exemplo seus dados pessoais (Figura 10), instituições de vínculo (Figura 11), bem como suas publicações científicas (Figura 12).

Figura 10 – Dados pessoais dos integrantes da comunidade

The screenshot shows the BrCris interface with a search bar and a list of entities on the left. The 'Person' entity is selected, showing details for two individuals. The first individual is identified by the Lattes ID [lattes::1251759109510185](#) and has the citation name 'YOCCOZ, J.C.'. The second individual is identified by the Lattes ID [lattes::7514918598084999](#) and has the citation name 'Os autores (2022)'.

BrCris Home

Entidades disponíveis

- Todas as entidades
- Community (1)
- Course (316)
- Journal (5539)
- OrgUnit (162)
- Person (13926)**
- Program (95)
- Publication (40773)

Dados

Pesquisar...

seniors

Person seniors

Identificadores

[lattes::1251759109510185](#)

Dados

citationName
YOCCOZ, J.C.

identifier.lattes
1251759109510185

name
J.C. Yoccoz

Person seniors

Identificadores

[lattes::7514918598084999](#)

Dados

citationName
Os autores (2022)

Fonte: Os autores (2022)

Figura 11 – Instituições de vínculo dos integrantes da comunidade

The screenshot shows the BrCris interface with a search bar and a list of entities on the left. The 'OrgUnit' entity is selected, showing details for two institutions. The first institution is identified by the BrCris ID [brcris::5f20ae4fb19ea55af73b36d0e6f9739a](#) and has the name 'INSTITUTO DE ASTRONOMIA GEOFÍSICA E CIÊNCIAS ATMOSFÉRICAS (USP)'. The second institution is identified by the BrCris ID [brcris::08b80b72e0bfcf3f24bf5f1333475110](#) and has the name 'Os autores (2022)'.

BrCris Home

Entidades disponíveis

- Todas as entidades
- Community (1)
- Course (316)
- Journal (5539)
- OrgUnit (162)
- Person (13926)
- Program (95)
- Publication (40773)

Dados

Pesquisar...

seniors

OrgUnit seniors

Identificadores

[brcris::5f20ae4fb19ea55af73b36d0e6f9739a](#)

Dados

identifier.brcris
5f20ae4fb19ea55af73b36d0e6f9739a

name
INSTITUTO DE ASTRONOMIA GEOFÍSICA E CIÊNCIAS ATMOSFÉRICAS (USP)

OrgUnit seniors

Identificadores

[brcris::08b80b72e0bfcf3f24bf5f1333475110](#)

Dados

identifier.brcris
08b80b72e0bfcf3f24bf5f1333475110

Fonte: Os autores (2022)

Figura 12 – Detalhes da produção técnica-científica dos integrantes da comunidade

Entidades disponíveis

- Todas as entidades
- Community (1)
- Course (316)
- Journal (5539)
- OrgUnit (162)
- Person (13926)
- Program (95)
- Publication (40773)**

Dados

Pesquisar...

seniors

Publication seniors

Identificadores
[brcris:9f4ebb747eaa8c0fa458e1064f5f22ae](https://brcris.org/9f4ebb747eaa8c0fa458e1064f5f22ae)

Dados

startPage
99

language
Inglês

type
journal article

title
General solution of one-dimensional approximations to the transport equation

researchArea.cnpq
Princípios Variacionais e Métodos Numéricos
Fenômenos de Transporte

identificador.brcris
9f4ebb747eaa8c0fa458e1064f5f22ae

publicationDate
1998

Fonte: Os autores (2022)

Todas as visualizações podem ser manipuladas de forma flexível e ainda, com a possibilidade de exportação dos dados em formatos tabulares ou consumidos por *API*'s.

4 CONSIDERAÇÕES FINAIS

A produção científica brasileira tem crescido expressivamente e, em perspectiva às especificidades de campos disciplinares distintos, heterogênea quanto à tipificação de sua produção tanto em termos quantitativos como qualitativos. E o resultado desta produção se materializa em forma de artigos em periódicos e em anais de eventos, teses e dissertações, além de produtos diversos como: softwares, patentes, obras e instalações artísticas, entrevistas e projetos cinematográficos.

O BrCris é um importante espaço para 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, que

pretende ser o alvo da proposta apresentada. 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

KONG, X.; SHI, Y.; YU, S.; LIU, J.; XIA, F. Academic social networks: Modeling, analysis, mining and applications. **Journal of Network and Computer Applications**, [S.l.], v. 132, p. 86-103, 2019. Disponível em: <https://dl.acm.org/doi/abs/10.1016/j.jnca.2019.01.029>. Acesso em: 27 mar. 2023.

LEE, S.; BOZEMAN, B. The impact of research collaboration on scientific productivity. **Social Studies of Science**, v. 35, n. 5, p.673-702, 2005. Disponível em: <https://elibrary.ru/item.asp?id=11423996>. Acesso em: 27 mar. 2023.

PAEZ, A. Gray literature: An important resource in systematic reviews. **Journal of Evidence-Based Medicine**, v. 10, n. 3, p. 233-240, 2017. Disponível em: <https://onlinelibrary.wiley.com/doi/abs/10.1111/jebm.12266>. Acesso em: 27 mar. 2023.

PINTO, A. L.; DIAS, T. M. R.; SILVA, V. S.; GOMES, J. C.; CARVALHO SEGUNDO, W. L. R. de. BrCris como um sistema de recomendação científico-tecnológica. *In*: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 21., 2021. Rio de Janeiro (online). **Anais [...]**. Rio de Janeiro: Ancib, 2021. Disponível em: <https://enancib.ancib.org/index.php/enancib/xxienancib/paper/view/335/456>. Acesso em: 27 mar. 2023.

RAMSDEN, P. Describing and explaining research productivity. **Higher education**, v. 28, n. 2, p. 207-226, 1994. Disponível em: <https://www.jstor.org/stable/3447753>. Acesso em: 27 mar. 2023.

SINGH, V. K.; SINGH, P.; KARMAKAR, M.; LETA, J.; MAYR, P. The journal coverage of web of science, scopus and dimensions: A comparative analysis. **Scientometrics**, [S.l.], v. 126, p. 1-30, 2021. Disponível em: https://ideas.repec.org/a/spr/scient/v126y2021i6d10.1007_s11192-021-03948-5.html. Acesso em: 27 mar. 2023.

SIVERTSEN, G. Developing Current Research Information Systems (CRIS) as data sources for studies of research. *In*: GLÄNZEL, W.; MOED, H. F.; SCHMOCH, U.; THELWALL, M. (ed.). **Springer Handbook of Science and Technology Indicators**. Springer Handbooks. Springer, Cham. p. 667-683,

2019. Disponível em: https://doi.org/10.1007/978-3-030-02511-3_25. Acesso em: 27 mar. 2023.

BRCRIS: PLATFORM FOR INTEGRATION, ANALYSIS AND VISUALIZATION OF TECHNICAL-SCIENTIFIC DATA

ABSTRACT

Objective: In this work, we present the BrCris Platform development process, whose objective is to provide technological tools to provide the Brazilian academic community with integrated and consolidated data on scientific production in Brazil. **Methodology:** initially, in order to contemplate the proposed objectives, a set of data repositories were selected, and after extracting these sets, the data were transformed into a model based on entities and relationships, being integrated for the proposed analyzes and visualizations. **Results:** As a result of the construction of the BrCris Platform, it was possible to obtain a set of consistent technical-scientific data to enable different types of analyses. **Conclusions:** This initiative presents itself as a unique data aggregation mechanism, enabling scientometric visualizations and analyzes on the present, past and future of Brazilian scientific production.

Descriptors: Open Access. Information Retrieval. Scientific Data. Open Data.

BRCRIS: PLATAFORMA DE INTEGRACIÓN, ANÁLISIS Y VISUALIZACIÓN DE DATOS TÉCNICO-CIENTÍFICOS

RESUMEN

Objetivo: En este trabajo, presentamos el proceso de desarrollo de la Plataforma BrCris, cuyo objetivo es proporcionar herramientas tecnológicas para proporcionar a la comunidad académica brasileña datos integrados y consolidados sobre la producción científica en Brasil. **Metodología:** inicialmente, para contemplar los objetivos propuestos, se seleccionaron un conjunto de repositorios de datos, y luego de extraer estos conjuntos, los datos se transformaron en un modelo basado en entidades y relaciones, siendo integrados para los análisis y visualizaciones propuestas. **Resultados:** Como resultado de la construcción de la Plataforma BrCris, fue posible obtener un conjunto de datos técnico-científicos consistentes para permitir diferentes tipos de análisis. **Conclusiones:** Esta iniciativa se presenta como un mecanismo único de agregación de datos, posibilitando visualizaciones y análisis cuantitativos sobre el presente, pasado y futuro de la producción científica brasileña.

Descriptores: Acceso Abierto. Recuperación de información. Datos científicos. Información abierta.

Recebido em: 20.12.2022

Aceito em: 24.03.2023