



Comunicação

CRIAÇÃO DE INDICADORES CIENTOMETRICOS: Estudo de caso da integração do Observatório Ciência, Tecnologia & Inovação em Saúde da Fiocruz ao Sistema BrCris/IBICT

Waldeyr Mendes¹, Rafaela Lora Grandó¹, Thiago Magela Rodrigues Dias², Vanessa de Arruda Jorge¹ e Washington Luís Ribeiro de Carvalho Segundo³

¹*Fundação Oswaldo Cruz (Fiocruz), Brasil*

²*Centro Federal de Tecnologias educacionais de Minas Gerais (CEFETMG), Brasil*

³*Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT), Brasil*

RESUMO: O Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT) vem desenvolvendo o Projeto BrCris (Current Research Information System), um ecossistema informacional que agrega e combina informações da pesquisa brasileira através da integração de bancos de dados públicos e privados. Neste contexto, o Observatório em Ciência, Tecnologia e Inovação em Saúde da Fiocruz firmou uma parceria com o IBICT para, simultaneamente, gerar maior robustez na coleta de seus indicadores e colaborar nas atividades de desenvolvimento do BrCris. Desta forma, em uma estratégia de mútuo benefício, o IBICT e a Fiocruz têm buscado novas abordagens de avaliação dos impactos da pesquisa, ampliando sua capacidade de percepção dos benefícios gerados para a sociedade, além de subsidiar a gestão e tomada de decisões institucionais. Tomando os resultados desta parceria como estudo de caso, o IBICT poderá testar as funcionalidades criadas no sistema BrCris, a partir do vasto e rico arcabouço de produção científica da Fiocruz, na homologação da sua versão beta. Este trabalho apresenta, portanto, o BrCris como objeto desta parceria, seu processo de desenvolvimento e os resultados parciais da missão de agregar e combinar as informações de pesquisa da Fiocruz.

Palavras-chave: Ciência & Tecnologia, Cientometria, Indicadores de Inovação.

SCIENTOMETRIC INDICATORS: Case study of the integration of the Fiocruz Science, Technology & Innovation in Health Observatory to the BrCris System/IBICT

ABSTRACT: The Brazilian Institute of Information in Science and Technology (IBICT) has been developing the BrCris Project (Current Research Information System). This information ecosystem aggregates and combines information from Brazilian research by integrating public and private databases. In this context, Fiocruz's Observatory on Science, Technology and Innovation in Health signed a partnership with IBICT to simultaneously generate more robustness in collecting its indicators and collaborating in the BrCris development activities. Thus, in a strategy of mutual benefit, IBICT and Fiocruz have sought new approaches to assess the impacts of research, expanding their capacity to perceive the benefits generated for society and supporting institutional management and decision-making. Taking the results of this partnership as a case study, BrCris may have its functionalities tested with a view to the approval of beta versions based on Fiocruz's vast and rich scientific production framework. Therefore, this work presents the BrCris as the object of this partnership, its development process and the partial results of the mission to aggregate and combine Fiocruz's research information.

Keywords: Science & Technology, scientometry, Innovation Indicators

Correspondência para: (correspondence to:) washingtonsegundo@ibict.br

INTRODUÇÃO

O mundo globalizado exige cada vez mais que pesquisadores e gestores estejam preparados para enfrentar novos desafios. No caso de instituições de ensino e pesquisa com natureza social, como a Fundação Oswaldo Cruz (Fiocruz), é preciso ainda atender às demandas provenientes do SUS e da sociedade de uma forma geral. Acompanhar os avanços científicos e tecnológicos é uma tarefa laboriosa e custosa para a gestão do conhecimento e da informação. Estima-se que apenas 0.5% da produção científica global seja analisada (Gantz e Reinsel, 2012). Portanto, e sobretudo nas instituições de ensino e pesquisa, analisar essa gama de informação tem importância estratégica para desenvolver indicadores primordiais e traçar cenários para a tomada de decisões. Avançar quantitativa e qualitativamente na análise de informações científicas implica em desenvolver instrumentos que facilitem a coleta e o processamento dos dados científicos para a organização e publicidade da produção científica.

Um *Current Research Information System* (CRIS) consiste em um sistema agregador de informação em C&T que pode servir a muitos propósitos a partir da padronização, documentação, comunicação e administração da pesquisa (SIVERTSEN, 2019). O desenvolvimento de um CRIS, no entanto, é permeado por desafios como o rápido crescimento e disponibilidade de dados em toda sua multidimensionalidade e complexidade, a possibilidade de estudos altimétricos para além dos estudos bibliométricos, a proliferação de indicadores sem propósitos claros, e a óbvia demanda por políticas de estado e de mercado (DARAI, 2016).

O Projeto BrCris (*Current Research Information System*) é uma iniciativa do Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT). Seu objetivo é ser um

ecossistema informacional que agrega e combina informações sobre a produção científica brasileira corrente e histórica a partir de múltiplas fontes de dados (PINTO, 2021a). O BrCris é um ecossistema de sistemas, e, portanto, a interoperabilidade é um ponto central para agregar os dados da produção científica brasileira, o que é uma tarefa complexa por vários motivos como a dispersão dos dados em várias plataformas, repositórios e outros sistemas, que são desenvolvidos e indexados de forma heterogênea e, frequentemente, com acesso difícil ou limitado (PINTO, 2021b). Diante desse cenário, a ambiciosa proposta do BrCris é justamente reunir os dados e metadados de pesquisa em Ciência e Tecnologia (C&T) do Brasil em um único espaço, estabelecendo relações e sistematização (PINTO, 2021b). Como estratégia de engenharia de software, com requisitos iniciais razoavelmente bem definidos, as versões beta do BrCris têm buscado agregar tais dados de forma incremental à medida que o sistema atinge sua maturidade (<https://brcris.ibict.br/vivo/>).

O Observatório em Ciência, Tecnologia e Inovação em Saúde da Fiocruz (<http://observatorio.fiocruz.br>) vem desenvolvendo este trabalho de agregação de dados em C&T para a Fiocruz através da geração de quatro indicadores: produção científica (GRANDO, 2021a), depósito de patentes (GRANDO, 2021b), fomento à pesquisa (GRANDO, 2021c) e perfil dos servidores (GRANDO, 2021d). A partir da parceria estabelecida neste ano de 2022 entre o IBICT e a Fiocruz, os esforços e recursos de ambos os lados foram direcionados para que o BrCris avance para uma versão que disponibilize dados de pesquisa em C&T da Fiocruz. Tal coleção de dados permitirá ao observatório maior dinamismo e riqueza na coleta de dados para atualização dos indicadores já existentes além da geração de novos indicadores.

Desta forma a Fiocruz, que vem produzindo C&T no Brasil desde os anos 1900, espera expandir e atualizar seus mecanismos para conhecer e avaliar os impactos do conhecimento produzido na instituição, e assim aproveitar o pleno potencial da Instituição nos avanços científicos na área da saúde. De outro lado, o BrCris, beneficiar-se-á da disponibilização de uma parcela significativa e importante da C&T brasileira produzida pela Fundação Oswaldo Cruz no período de 2008 a 2022.

Este trabalho está organizado em mais três seções, quais sejam: "Metodologia", que descreve como tem ocorrido o desenvolvimento das atividades da parceria Fiocruz/BrCris. A seção "Resultados Parciais e Perspectivas" apresenta um panorama das conquistas já alcançadas, as dificuldades encontradas e o promissor horizonte à frente. Finalmente, a seção "Considerações Finais" traz uma análise dos gestores do projeto no que se refere a integração deste sistema para formulação de novos indicadores.

METODOLOGIA

Os conjuntos de dados utilizados pela plataforma BrCris são provenientes de fontes diversas como repositórios institucionais, bibliotecas digitais de teses e dissertações, revistas eletrônicas de acesso aberto, repositórios de dados de pesquisa brasileiros reunidos nos portais da Biblioteca Digital de Teses e Dissertações (BDTD) e do Portal brasileiro de publicações científicas em acesso aberto (Oasisbr). Além destas fontes, fazem parte do arcabouço de dados, a Plataforma Lattes, a Plataforma Sucupira e o Portal de Dados Abertos da Capes e fontes internacionais abertas, tais como: Oasisbr/IBICT, BDTD/IBICT, Plataforma Lattes, Espacenet, WikiData, Diadorim/IBICT, Latindex, DOAJ, UlrichsWeb, ROR, Dimensions, Altmetric, OpenAIRE, DBpedia, OpenCitations, SherpaRomeo, Miguilim e Google Scholar

Metrics (PINTO 2021b).

Neste contexto, tendo em vista as diversas fontes de dados que irão compor o BrCris faz-se necessária a transformação dos dados para um modelo de dados capaz de combinar as informações das diferentes fontes de dados. 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 e suas relações, possibilitando que com o auxílio de uma rotina desenvolvida especificamente para esta funcionalidade, 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.

Para o tratamento dos dados foi desenvolvida uma biblioteca na linguagem de programação Python, contendo uma estrutura de dados preparada para facilitar o processamento de dados originários de todas as fontes para o formato especificado no modelo de dados. Logo, a biblioteca desenvolvida é responsável por toda a transformação e exportação dos dados, utilizando como base o "Modelo de Dados" previamente definido, validando as entidades, campos e relacionamentos aceitos pelo modelo.

Além disso, a biblioteca desenvolvida também é responsável por gerar Identificadores BrCris, criados com o intuito de realizar uma pré-desambiguação dos dados, evitando entidades duplicadas, com baixo custo computacional. A geração do Identificador BrCris é realizada de forma distinta para cada entidade, em geral realizando um *hash* de seus próprios campos de dados. Por fim, o ferramental proposto possibilita realizar a exportação dos dados originais no formato XML (Extensible

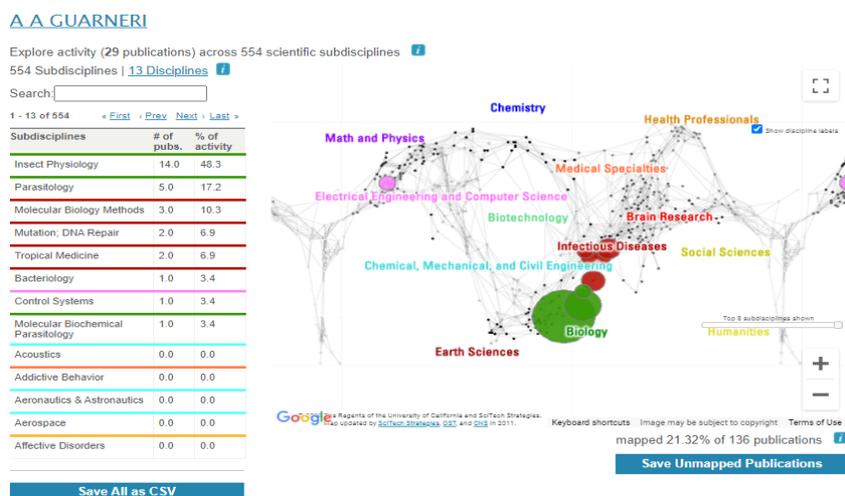


FIGURA 2: Distribuição de Publicações por áreas do conhecimento

DISCUSSÃO

O Observatório em Ciência, Tecnologia e Inovação em Saúde da Fiocruz visa contribuir para a gestão e formulação de políticas institucionais de Ciência, Tecnologia e Inovação, por meio da produção de indicadores, métricas, estudos e conteúdo de comunicação que apoiem aos gestores nos processos de tomada de decisão, aos pesquisadores com dados para embasar estudos e prospecções, além de trazer publicidade e transparência sobre a pesquisa e desenvolvimento tecnológico da Fiocruz para sociedade. Nesse contexto de colaboração IBICT/Fiocruz, o Observatório da Fiocruz têm cumprido seu papel para com a Fiocruz e com a sociedade, ampliando a prestação de contas e fomentando ações de participação cidadã, a partir do acesso democrático às informações e conhecimentos disponibilizados.

Entretanto, à medida que os dados crescem em volume e complexidade, novas estratégias tecnológicas que permitam escalar o processamento desses dados foram necessárias. Um ponto crucial para garantir o mencionado acesso democrático às informações de C&T, é que os dados estejam disponíveis de forma aderente aos princípios FAIR (*Findability, Accessibility,*

Interoperability, and Reusability) (WILKINSON, 2016), o que implica que os dados devem ser estruturados e disponibilizados com garantia de serem buscáveis, acessíveis, interoperáveis e reusáveis, buscando soluções sustentáveis para os desafios mencionados por Daraio (2016).

AGRADECIMENTOS

Os autores agradecem a contribuição dos colaboradores: Tales Moreira (IBICT), Adilson Pinto (UFSC), Vivian Silva (IBICT), Luc Quoniam (UFMS), Lautaro Matas (LA Referencia), Jesús Mena-Chalco (UFABC), Tiago Oliveira (CEFET-MG), Mariana Salgueiro (IBICT) e José Estevam Silva (IBICT).

REFERÊNCIAS

- CÓRDOVA YAMAUCHI, C. G. **Plataforma PerúCRIS**. 2021. Disponível em: <https://perucris.concytec.gob.pe/>. Acesso em: 21/03/2023
- DARAIO, Cinzia; GLÄNZEL, Wolfgang. Grand challenges in data integration—State of the art and future perspectives: An introduction. *Scientometrics*, v. 108, n. 1, p. 391-400, 2016.
- GANTZ, J., & REINSEL, D. The Digital

Universe in 2020: Big Data, Bigger Digital Shadows, and Biggest Growth in the Far East, 2012. Disponível em: [https://www.scirp.org/\(S\(lz5mqp453ed%20np55rrgict55\)\)/reference/referencespapers.aspx?referenceid=2944497](https://www.scirp.org/(S(lz5mqp453ed%20np55rrgict55))/reference/referencespapers.aspx?referenceid=2944497). Acesso: 21/03/2023

Grando, Rafaela Lora; Fonseca, Fernanda; Silva, Waldeyr Mendes Cordeiro; Pereira-Silva, Marcus Vínicius, 2021a, "Indicadores da produção científica da Fundação Oswaldo Cruz (Fiocruz) em Ciência, Tecnologia e Inovação em Saúde no período de 2008 a 2020". Disponível em: <https://doi.org/10.35078/JEDCTR>, Arca Dados, V1. Acesso: 21/03/2023

Grando, Rafaela Lora; Fonseca, Fernanda; Silva, Waldeyr Mendes Cordeiro; Pereira-Silva, Marcus Vínicius, 2021b, "Indicadores de depósito de patentes da Fundação Oswaldo Cruz (Fiocruz) no contexto da ciência, tecnologia e inovação no período de 1988 a 2020". Disponível em: <https://doi.org/10.35078/YN2UTF>. Arca Dados, V2. Acesso: 21/03/2023

Grando, Rafaela Lora; Fonseca, Fernanda; Silva, Waldeyr Mendes Cordeiro; Pereira-Silva, Marcus Vínicius, 2021c, "Indicadores de financiamento de pesquisa científica da Fundação Oswaldo Cruz (Fiocruz) pelo Departamento de Ciência e Tecnologia da Secretaria de Ciência, Tecnologia e Insumo Estratégicos do Ministério da Saúde (Decit/SCTIE/MS) no período de 2002 a 2020". Disponível em: <https://doi.org/10.35078/ADK0NS>, Arca Dados, V1. Acesso: 21/03/2023

Grando, Rafaela Lora; Fonseca, Fernanda; Silva, Waldeyr Mendes Cordeiro; Pereira-Silva, Marcus Vínicius 2021d, "Perfil dos Servidores da Fundação Oswaldo Cruz (Fiocruz) no ano de 2020". Disponível em: <https://doi.org/10.35078/RCVfYZ>, Arca Dados, V1. Acesso em: 21/03/

MANGHI, Paolo et al. The OpenAIRE research graph data model. **Zenodo**, (2019). Disponível em: <https://zenodo.org/record/2643199#.ZBnTDnZv-3A> Acesso em: 21/03/2023

PINTO, Adilson Luiz et al. BrCris como um sistema de recomendação científico-tecnológica. In: **XXI Encontro Nacional de Pesquisa em Ciência da Informação**. 2021.

PINTO, Adilson Luiz et al. The Brazilian current research information system: BrCris. **Organização do Conhecimento no Horizonte 2030: Desenvolvimento Sustentável e Saúde: Atas do V Congresso ISKO Espanha-Portugal**. Coleção CA–Ciência Aberta, p. 319., 2021.

SIVERTSEN, G. Developing Current Research Information Systems (CRIS) as data sources for studies of research. In: **Springer handbook of science and technology indicators**. Springer, Cham, 2019. p. 667-683.

WILKINSON, Mark D. et al. The FAIR Guiding Principles for scientific data management and stewardship. **Scientific data**, v. 3, n. 1, p. 1-9, 2016.