



INFRAESTRUTURAS ABERTAS E IMPACTO ACADÊMICO: o papel dos repositórios digitais nas citações científicas

Washington Luís Ribeiro de Carvalho Segundo

Instituto Brasileiro de Informação em Ciência e
Tecnologia (Ibict), Brasil

washingtonsegundo@ibict.br

 <https://orcid.org/0000-0003-3635-9384>

Marcel Garcia de Souza

Instituto Brasileiro de Informação em Ciência e
Tecnologia (Ibict),

marcelsouza@ibict.br

 <https://orcid.org/XXXX-XXXX-XXXX-XXXX>

Thiago Magela Rodrigues Dias

Centro Federal de Educação Tecnológica de
Minas Gerais (CEFET-MG), Brasil

thiagomagela@cefetmg.br

 <https://orcid.org/0000-0002-3090-9413>

Fabio Lorensi do Canto

Universidade Federal de Santa Catarina
(UFSC), Brasil

fabio.lc@ufsc.br

 <https://orcid.org/XXXX-XXXX-XXXX-XXXX>

DOI: 10.22477/xiv.biredial.408

EJE TEMÁICO: Comunicación académica, científica
y cultural en abierto

RESUMEN

O estudo, desenvolvido no âmbito dos projetos e portal BrCris, Laguna e Oasisbr, investigou o impacto dos repositórios digitais no número de citações de artigos científicos brasileiros. O portal Oasisbr, que agrupa mais de 5 milhões de objetos digitais, foi utilizado para coletar dados de 421.243 artigos de revista depositados em repositórios brasileiros. Em paralelo, a base OpenAlex forneceu 1.901.216 registros de artigos exclusivamente vinculados a instituições brasileiras. A análise comparativa entre artigos presentes simultaneamente em revistas científicas e repositórios digitais (Classe 1) e aqueles presentes apenas em revistas (Classe 2) revelou que os primeiros recebem, em média, 83,10% mais citações. Este crescimento varia conforme ano de publicação, percentil do h-index das revistas e áreas do conhecimento, destacando-se Ciências Sociais (+206,02%) e apresentando queda em campos como Engenharia e Ciência da Computação (-8,21%). A discussão sugere que a maior visibilidade promovida pelos repositórios, especialmente em práticas de compartilhamento aberto, é determinante para o aumento das citações. Os repositórios facilitam a disseminação dos artigos em plataformas como Google Scholar, aumentando a probabilidade de citação, mesmo em revistas de menor impacto. Conclui-se que os repositórios são ferramentas estratégicas para ampliar o impacto e o reuso do conhecimento científico.

Palabras-clave: Repositórios digitais; Oasisbr; OpenAlex; BrCris, Projeto Laguna.



ABSTRACT

(ver resumo em inglês com autores)

Keywords: Digital repositories; Oasisbr; OpenAlex; BrCris, Laguna Project.

INTRODUÇÃO

A Ciência Aberta propõe transformações significativas nos modos de produção, disseminação e avaliação do conhecimento científico, promovendo maior transparência, colaboração e acesso à informação. Nesse contexto, os repositórios digitais de acesso aberto têm desempenhado papel crucial ao garantir a preservação e a ampla circulação dos resultados de pesquisa.

O Brasil tem se destacado na construção de ecossistemas digitais que viabilizam essas práticas, como demonstram os projetos BrCris, Laguna e o portal Oasisbr. No entanto, ainda são necessárias evidências sobre os reais impactos dessas infraestruturas. Este estudo visa preencher essa lacuna ao analisar como a presença em repositórios influencia o número de citações recebidas por artigos científicos brasileiros.

O presente trabalho foi desenvolvido no âmbito dos Projetos BrCris (Ecossistema de Informação da Pesquisa Científica Brasileira (Silva *et al.*, 2023), Laguna (Infraestrutura Informatacional Aberta e Lago de Dados, (Neubert *et al.*, 2024) e do Portal Agregador Oasisbr (Portal Brasileiro de Publicações e Dados Científicos em Acesso Aberto (Morais *et al.*, 2024). Este portal agrupa registros de mais de 5 Milhões de objetos digitais, disponibilizados por aproximadamente 1.600 fontes de acesso aberto. Entre estas, 125 são repositórios digitais brasileiros e 17 são repositórios de dados de pesquisa. Já a base OpenAlex (Priem, J., Piwowar & Orr, 2022) provê um conjunto representativo de publicações brasileiras, com metadados importantes relacionados às afiliações dos autores e uma classificação de áreas do conhecimento refinada em 4 níveis (4 domínios, 26 campos, 252 subcampos e 4.516 tópicos (Van Eck & Waltman, 2024).

Os resultados apresentados demonstram que o depósito de artigos em repositórios digitais abertos está associado ao aumento do número médio de citações, evidenciando o papel estratégico dessas infraestruturas na amplificação do impacto científico.

METODOLOGIA

A fim de avaliar a hipótese de que artigos depositados em repositórios digitais abertos recebem mais citações, delineou-se uma metodologia baseada na integração e análise de grandes volumes de dados extraídos do portal Oasisbr e da base OpenAlex. A estratégia de cruzamento entre essas fontes permitiu identificar, com critérios rigorosos de correspondência, os artigos com e sem depósito em repositórios.

Foram coletados do Portal Oasisbr, em dezembro de 2024, 500.132 registros de artigos



de revista que estão depositados em repositórios digitais brasileiros. Considerando o período de publicação entre 2000 e 2024, os registros foram deduplicados com base no ano de publicação e na similaridade entre títulos (após a limpeza de caracteres especiais e medição da distância de Levenshtein, considerou-se títulos semelhantes aqueles que possuíam uma distância de edição menor ou igual a 5), o número de registros mapeados foi reduzido a 421.243.

Do outro lado, foram coletados na base *OpenAlex*, também em dezembro de 2024, 2.834.342 registros de artigos de revista, publicados entre 2000 e 2024, que possuíam como autor pelo menos um pesquisador com afiliação em uma instituição brasileira. Deste conjunto, foram selecionados 1.901.216 registros, que foram identificados como tendo afiliação exclusiva de instituições brasileiras. A escolha por considerar apenas artigos com afiliação exclusiva a instituições brasileiras buscou evitar possíveis distorções causadas por colaborações internacionais, as quais poderiam resultar no depósito dos artigos em repositórios estrangeiros, dificultando a correta identificação do impacto dos repositórios nacionais.

Por meio da comparação entre os registros com mesmo ano de publicação e título semelhante entre o conjunto de registros provindo dos repositórios e do conjunto de registros selecionados na base *OpenAlex*, foram identificados 153.671 artigos que estão presentes nos dois conjuntos.

Assim, sendo a Classe 1 os registros presentes em revistas científicas e repositórios digitais, e a Classe 2 os artigos presentes somente em revistas científicas, e dado que α , β , n , m representam, respectivamente: o número total de citações recebido por um conjunto de registros dentro da Classe 1; o número total de citações recebido por um conjunto de registros dentro da Classe 2; o número total de artigos da Classe 1; e o número total de artigos da Classe 2; tem-se que as médias de citações em conjuntos das Classes 1 e 2 são dadas, respectivamente, por $A = \alpha / n$ e $B = \beta / m$.

Finalmente, o valor crescimento observado da média de citações para artigos da Classe 2 é obtido pela fórmula $(A - B) / B$, em cada um dos seguintes casos analisados: a) por ano da publicação; b) por percentil do *h-index* da revista onde o artigo foi publicado (O *h-index* da revista foi obtido por meio de conjunto de registros de revistas coletado, também em dezembro de 2024, da base *OpenAlex*); c) pela classificação de domínio da base *OpenAlex*; e d) pela classificação de campo da base *OpenAlex*.

RESULTADOS

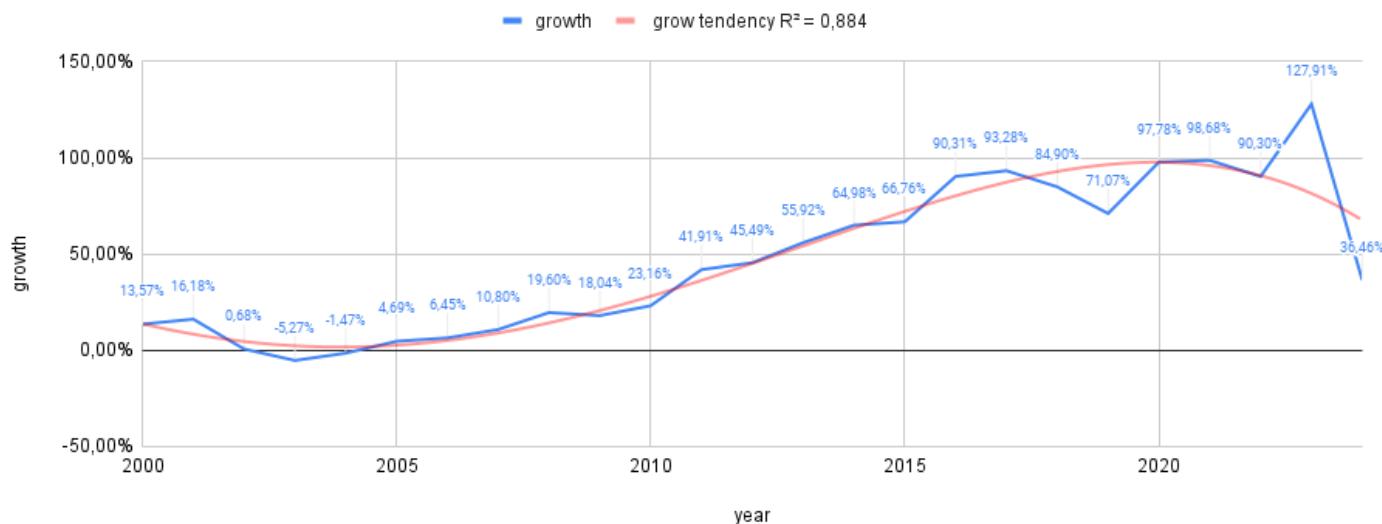
A análise comparativa entre os dois grupos de artigos revela tendências significativas na média de citações. Os dados obtidos foram desagregados por ano de publicação, percentil do *h-index* das revistas e área do conhecimento, buscando identificar padrões e variações no impacto da presença em repositórios sobre o número de citações.



Verifica-se que de forma geral o crescimento da média de citações de artigos que estão em repositórios e em revistas é de +83,10%, em comparação àqueles que estão somente em revistas. O que simboliza que artigos de repositórios recebem em média quase o dobro de citações.

Ao se calcular a variação do crescimento conforme o ano de publicação (Figura 1), observa-se que há um aumento crescente do valor observado nos anos mais recentes, com exceção dos anos de 2002 a 2004, e 2024, onde ocorrem quedas de crescimento, chegando a valores negativos em 2003 e 2004.

Figura 1 - Variação do crescimento pelo ano de publicação



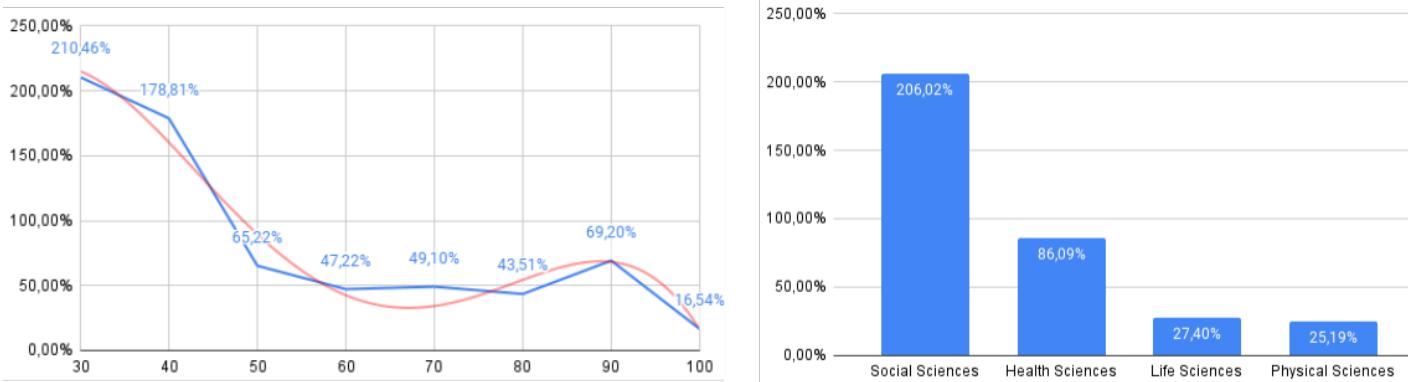
Fonte: Os autores (2025).

Apresenta-se a evolução do crescimento percentual da média de citações dos artigos depositados em repositórios ao longo dos anos de publicação. Nota-se uma tendência de aumento contínuo no impacto a partir de 2005, coincidindo com o fortalecimento de algumas políticas de acesso aberto no país.

Já a Figura 2 mostra a variação do crescimento de acordo com as faixas de percentil do *h-index* (nos intervalos (30, 40] a (90, 100]) da revista onde os artigos foram publicados. A curva indica um crescimento de +210,46% para artigos publicados em revistas no intervalo de percentil (20, 30). Posteriormente, observa-se uma queda entre os intervalos (30, 40] a (50, 60], mantendo-se estável nos intervalos de (50, 60] a (70, 80], com aumento no intervalo (80, 90] e nova queda na última faixa de percentil (90, 100], sendo que o valor mínimo observado de +16,54%.



Figura 2 - Variações do crescimento pelo percentil do *h*-index e pela classificação de domínio



Fonte: Os autores (2025).

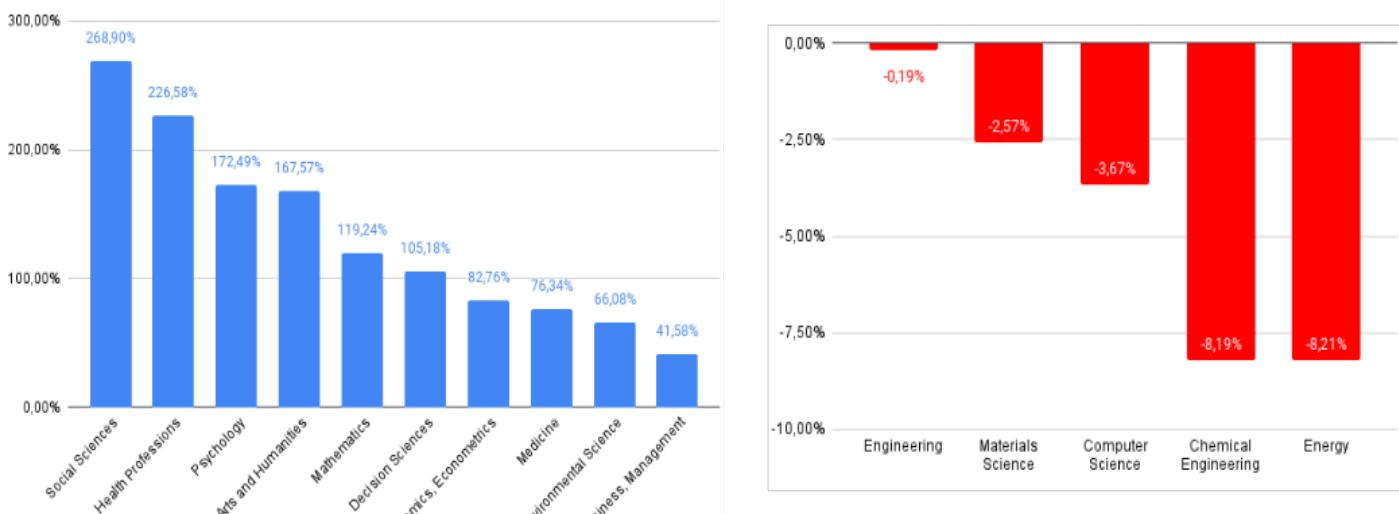
Mostra-se como a presença em repositórios impacta diferentemente os artigos de acordo com o percentil do *h*-index das revistas. Observa-se que o efeito positivo dos repositórios é mais pronunciado em revistas de menor impacto (percentis mais baixos), sugerindo que o acesso aberto funciona como um mecanismo de compensação de visibilidade, especialmente relevante para periódicos com menor alcance

Evidencia-se que o maior crescimento (+206,02%) ocorre no domínio de Ciências Sociais, enquanto Ciências Físicas ficam com o menor valor (+25,19%). Na variação de crescimento de acordo com o campo de conhecimento, os Gráficos 4 e 5, mostram, respectivamente os top 10 que possuem maiores valores, que vão de +268,90% para o campo de Ciências Sociais a +41,58% para Business, Management and Accounting, em contraposição ao Gráfico 5 que exibe os menores valores (negativos) para os campos de Engineering, Materials Science, Computer Science, Chemical Engineering e Energy, chegando ao patamar mínimo de -8,21%.

Apresenta-se na Figura 3 a variação do crescimento da média de citações para: a) os diferentes domínios do conhecimento de acordo com a tabela de classificação utilizada pela base *OpenAlex*; b) os diferentes campos do conhecimento também de acordo com a tabela da *OpenAlex*.



Figura 3 - Variações de crescimento pelo campo de conhecimento



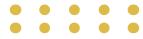
Fonte: Os autores (2025).

Apresenta-se o crescimento percentual da média de citações em diferentes domínios e campos do conhecimento, segundo a classificação da *OpenAlex*. As Ciências Sociais destacam-se como o campo com maior ganho, sugerindo uma forte aderência dessas áreas às práticas de Ciência Aberta. Por outro lado, áreas como Engenharia e Computação demonstram crescimento negativo, refletindo uma possível resistência ou menor penetração dos repositórios nesses campos.

Os resultados obtidos confirmam que a presença em repositórios digitais está associada a um aumento significativo na média de citações, embora com variações expressivas entre áreas do conhecimento e características das revistas. Essas evidências reforçam o papel estratégico dos repositórios no ecossistema da Ciência Aberta e indicam caminhos para o fortalecimento de políticas institucionais voltadas ao depósito sistemático de publicações científicas.

CONSIDERAÇÕES

Observa-se que, salvo algumas exceções, os artigos de revista que foram depositados em repositórios digitais abertos possuem um crescimento positivo na média de citações recebidas. No caso da variação por ano de publicação há uma tendência para que artigos mais recentes tenham um maior crescimento. Uma hipótese para explicar esse fenômeno é a de que, no Brasil, repositórios digitais surgiram principalmente no final da década de 2010, tendo portanto maior influência na disseminação dos artigos que foram neles depositados a partir desse período.



A queda de crescimento em 2024 poderá ser explicada pelo fato de que há sempre um gap temporal entre a publicação do artigo, seu depósito em um repositório e sua aparição na base *OpenAlex*. Na variação de acordo com o intervalo de percentil do *h-index* da revista onde o artigo foi publicado, verifica-se que os repositórios têm uma relação mais forte com o crescimento do número médio de citações nas revistas em percentis menores de *h-index*, de forma que um artigo publicado em uma revista com maior percentil e não depositado em um repositório se equipara em número médio de citações com um artigo depositado em repositório e publicado em uma revista com menor percentil.

Já as análises por domínio e campo do conhecimento demonstraram que o crescimento de citações é bastante elevado em áreas como Ciências Sociais, mas chega a ser negativo em áreas como Engenharias, Ciência da Computação, Ciência dos Materiais e Energia, exemplificando o fenômeno de que algumas áreas são mais sensíveis e afetas às práticas de ciência aberta do que outras. Por fim, conclui-se que há uma evidência de aumento de citações em artigos publicados em revistas que são depositados em repositórios digitais, de forma a sugerir que o aumento da visibilidade promovido por essas infraestruturas é real e deve ser considerado.

Esse aumento pode ser atribuído à capacidade dos repositórios, quando bem estruturados e interoperáveis, de ampliar a disseminação dos artigos por meio de plataformas como o Oasisbr, Google Scholar, Red LA Referencia, Portal RCAAP e OpenAIRE. Dessa forma, observa-se um ganho significativo no reuso do conhecimento científico, com maior probabilidade de citações, mesmo para aqueles artigos publicados em revistas científicas de baixo ou médio impacto aferido.

REFERÊNCIAS

Silva, V. D. S., Silva, J. V. D., Dias, T. M. R., Gabriel, R. F., & Segundo, W. L. R. de C. (2023). BrCris: Tools for treatment, analysis, and dissemination of scientific information in support of open science in Brazil. *RDBCi: Revista Digital de Biblioteconomia e Ciência da Informação*, 21, e023027. <https://doi.org/10.20396/rdbcii.v21i00.8673171en>

Neubert, P., Canto, F., Pinto, A. L., Lima, D., & Silva, F. (2024, June). OpenAlex como fonte de dados para sistemas nacionais de informação científica: A experiência do projeto Lagona. *VII Workshop de Informação, Dados e Tecnologia – WIDaT 2024*, 7, e184–e184. <https://doi.org/10.22477/vii.widat.184>

Morais, C. T. de, Segundo, W. L. R. de C., Sena, P. M. B., Campos, F. D. F., Amaro, B., Silva, T. G. M. D., & Sousa, J. A. G. D. (2024). Rede Brasileira de Repositórios Digitais (RBRD): Análise de

sua constituição e representatividade por meio do Portal Oasisbr. *Revista Integración y Conocimiento*. Advance online publication. <https://revistas.unc.edu.ar/index.php/integracionyconocimiento/article/view/44207>

Priem, J., Piwowar, H., & Orr, R. (2022). OpenAlex: A fully open index of scholarly works, authors, venues, institutions, and concepts. *arXiv*. <https://doi.org/10.48550/arXiv.2205.01833>

Van Eck, N. J., & Waltman, L. (2024, January 24). An open approach for classifying research publications. *Leiden Madtrics*. <https://www.leidenmadtrics.nl/articles/an-open-approach-for-classifying-research-publications>

ANEXO 1

RESUMEN BIOGRÁFICO DE LOS AUTORES

Washington L. R. Carvalho Segundo

É doutor em Informática pela Universidade de Brasília (UnB), com período sanduíche no King's College London, e mestre na mesma área pela UnB. Possui também formação em Matemática (bacharelado e licenciatura) pela mesma instituição. Atualmente, é Coordenador-Geral de Informação Científica e Tecnológica no Instituto Brasileiro de Informação em Ciência e Tecnologia (Ibict), onde lidera projetos voltados à Ciência Aberta, repositórios digitais, interoperabilidade de sistemas e gestão de dados científicos. Entre suas contribuições no Ibict, destaca-se a coordenação de iniciativas como o Oasisbr, um portal que agrupa e dissemina conteúdos científicos brasileiros de acesso aberto, e a Biblioteca Digital Brasileira de Teses e Dissertações (BDTD), que centraliza a produção acadêmica de programas de pós-graduação em todo o país. Ele também lidera esforços relacionados à Rede dARK, um sistema descentralizado de identificadores persistentes baseado em blockchain, que visa garantir a integridade e rastreabilidade de recursos digitais, promovendo inovação em gestão de dados científicos. Sua trajetória inclui o desenvolvimento do BrCris, um sistema de integração e visualização de dados científicos no Brasil, e o projeto Laguna, voltado à criação de um lago de dados científicos de acesso aberto.

Marcel Garcia de Souza

Doutorando em Ciência da Informação pela Universidade de Brasília. Mestre em Educação em Ciências pela Universidade Federal do Rio Grande do Sul (2016). Graduado em Psicologia pela Universidade Católica de Brasília (2005). Servidor público federal; Analista em Ciência e



Tecnologia no Instituto Brasileiro de Informação em Ciência e Tecnologia atuando como Coordenador de Tratamento, Análise e Disseminação da Informação Científica, além de coordenar pesquisas aplicadas voltadas à Ciência da Informação, Informação para Sustentabilidade, Avaliação do Ciclo de Vida, Informação Tecnológica.

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 - UNIFOR (2004), além de Especialização em Produção de Software - com Ênfase em Software Livre pela UFLA (2007) e Especialização em Melhoria do Processo de Software, UFLA (2007). Atua como Professor no Centro Federal de Educação Tecnológica de Minas Gerais (CEFET-MG). Professor do Programa de Pós-graduação em Modelagem Matemática e Computacional do CEFET-MG e Professor Permanente do Programa de Pós-graduação em Ciência da Informação da Universidade Federal de Santa Catarina (UFSC). Também tem atuação como Colaborador em Projetos no Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT).

Fabio Lorensi do Canto

É Doutor (2022) e Mestre (2018) em Ciência da Informação pela Universidade Federal de Santa Catarina - UFSC. É Graduado em Biblioteconomia - Gestão da Informação pela Universidade do Estado de Santa Catarina - UDESC (2005) e Bacharel em Direito pela Faculdade CESUSC (2012). Atua como Bibliotecário/Documentalista na Biblioteca Central da Universidade Federal de Santa Catarina - UFSC e como professor substituto do Departamento de Biblioteconomia da Universidade do Estado de Santa Catarina (UDESC). É bolsista do Projeto Laguna, coordenado pelo Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT).

ANEXO 2

REQUERIMENTOS DE EQUIPAMENTO TÉCNICO PARA A APRESENTAÇÃO

Computador e projetor.