



BASE DE DADOS EM DIREITO:

proposta de uma base de dados de publicações em acesso aberto para o domínio do Direito

Paula Carina de Araújo

 <https://orcid.org/0000-0003-4608-752X>.

✉ paulacarina@ufpr.br.

🏢 Universidade Federal do Paraná (UFPR) |

ROR: <https://ror.org/05syd6y78> | Curitiba, Brasil.

Rene Faustino Gabriel Junior

 <https://orcid.org/0000-0003-1021-3360>.

✉ renefgj@gmail.com.

🏢 Universidade Federal do Rio Grande do Sul (UFRGS) |

ROR: <https://ror.org/041yk2d64> | Porto Alegre, Brasil.

Patrick Fernandes Ribeiro Rezende

 <https://orcid.org/0000-0002-5973-1110>.

✉ patrick.ribeiro@ufpr.br.

🏢 Universidade Federal do Paraná (UFPR) |

ROR: <https://ror.org/05syd6y78> | Curitiba, Brasil.

Gustavo Gabriel Ripka

 <https://orcid.org/0009-0004-8933-0184>.

✉ gustavo.ripka@ufpr.br.

🏢 Universidade Federal do Paraná (UFPR) |

ROR: <https://ror.org/05syd6y78> | Curitiba, Brasil.

Eixo temático: Bases e Fontes de Dados

Modalidade: Resumo expandido

DOI: 10.22477/ix.ebbc.389

Resumo: A criação de bases de dados temáticas desempenham importante papel no avanço do conhecimento e da ciência aberta. Esta pesquisa objetiva apresentar a descrição técnica para a criação da Base de Dados em Direito. A construção da base encontra-se na etapa de prova de conceito, com testes de coleta e processamento de metadados em um estudo exploratório. A base está na fase de refinamento da estrutura de dados para recuperação e produção de indicadores, bem como o desenvolvimento da sua política de informação. A primeira versão da base será disponibilizada para consultas até o final de 2024 e proporcionará intensificar os estudos de comunicação científica da área do Direito brasileiro, com dados de produção e coautoria, objetivando subsidiar estudos com informação de qualidade.

Palavras-Chave: Base de dados. Fonte de Informação. Comunicação científica. Periódico científico. Direito.

1 INTRODUÇÃO

A década de 1990 marca o surgimento de movimentos de reação ao cerceamento da informação e do conhecimento científico e em prol da sua disseminação aberta e pública. São exemplos o movimento pelos códigos e softwares livres e também o movimento de acesso aberto à produção científica. Segundo Appel (2019), o direito de propriedade perde a centralidade na defesa de interesses comerciais e de garantia de lucros nesse contexto, para tornar-se, quase inteiramente, uma questão de respeito à autoria do trabalho original (Appel, 2019).

As bases de dados referenciais e de texto completo são importantes fontes de informação que cumprem o papel de organizar e disseminar a informação em diferentes áreas do conhecimento e também de forma multidisciplinar. Web of Science, Scopus e SciELO são alguns dos exemplos de bases de dados multidisciplinares mais conhecidos. Essas fontes de informação também funcionam como índices de citação, que são bases de dados que geram ligações entre documentos por meio das citações.

Os movimentos de acesso aberto e mais recentemente da ciência aberta impulsionaram o desenvolvimento de iniciativas que representem de forma mais efetiva a produção científica e bibliográfica regional e também de determinados domínios do conhecimento. A SciELO, citada anteriormente, é uma das principais iniciativas de acesso aberto regionais que ganhou visibilidade mundial e hoje está presente em vários países (Packer; Cop; Santos, 2014). Também há iniciativas de áreas do conhecimento específicas que fortalecem o movimento de ciência aberta no Brasil como a Base de Dados em Ciência da Informação (BRAPCI) (Bufrem *et al.*, 2010), a *Scientific Periodicals Electronic Library* (SPELL) (Rossoni; Rosa, 2023), a Biblioteca Digital Jurídica (BDJur) (Basevi, 2017; Vieira, 2017; Vieira; Lima; Neri, 2019).

Considerando o contexto apresentado, este trabalho tem como objetivo geral apresentar a descrição técnica para a criação de uma Base de Dados em Direito. Destaca-se que esta pesquisa faz parte do projeto de pesquisa “Gestão da Informação Científica no Contexto da Ciência Aberta” que tem como um de seus resultados esperados a proposição dessa base de dados, cuja política de informação está em processo de construção pela equipe do projeto de pesquisa.

A política de informação da Base de Dados em Direito está em processo de construção pela equipe do projeto de pesquisa e a primeira versão da base apresentará a indexação de 120 periódicos científicos em acesso aberto editados por programas de pós-graduação em direito brasileiros. Apesar de iniciar com foco na produção bibliográfica em periódicos científicos em acesso aberto do domínio do direito, a política de informação também irá prever a indexação da produção bibliográfica em periódicos técnicos em acesso aberto desse domínio.

O trabalho aqui apresentado partiu de uma pesquisa exploratória e desenvolveu as seguintes etapas: a) identificação das bases de dados regionais de temáticas brasileiras; b) benchmarking com a Base de Dados em Ciência da Informação (BRAPCI); c) escolha do conjunto de periódicos científicos para compor a coleção inicial da base de dados; d) mapeamento e descrição das características técnicas necessárias para a criação da Base de Dados em Direito.

O referencial teórico sobre bases de dados regionais e nacionais é apresentado na próxima seção. A seção 3 apresenta a descrição técnica para a criação da Base de Dados em Direito e a seção 4 descre-

ve as considerações finais. Por fim, são apresentadas as referências.

2 BASES DE DADOS REGIONAIS E NACIONAIS

Apesar do seu pioneirismo e importância para a disseminação da informação científica quando da criação de bases de dados como a Web of Science (1997) e Scopus (2004), fortes críticas são feitas ao longo dos anos com relação aos limites e potencialidades da cobertura dessas fontes de informação que afirmam representar o núcleo central da produção científica mundial (Aguado-López; Rogel-Salazar; Becerril-García, 2010; Packer *et al.*, 2014)

Nesse contexto, sabe-se que as bases de dados regionais e nacionais desempenham um papel fundamental na pesquisa científica, fornecendo acesso a informações bibliográficas essenciais e facilitando a avaliação do conhecimento científico em uma determinada região, possibilidade limitada quando se pensa nas bases de dados multidisciplinares internacionais (Santini; Caregnato, 2018, p. 54).

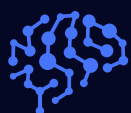
A Recomendação da UNESCO sobre a ciência aberta (UNESCO, 2022) aponta que, para atingir seu propósito, entre os principais objetivos e áreas de ação está o investimento em plataformas e repositórios abertos de publicações científicas como um dos “componentes cruciais das infraestruturas científicas abertas, que oferecem serviços abertos e padronizados essenciais para gerenciar e fornecer acesso, portabilidade, análise e agregação de dados, literatura científica, prioridades científicas temáticas ou envolvimento da comunidade” (UNESCO, 2022, p. 12).

Ressalta-se também que a criação de índices de citação nacionais e regionais é crucial para garantir um retrato multidimensional da ciência em regiões periféricas, como a América Latina. Esses índices permitem acompanhar a produção científica local e regional, bem como seu impacto, contribuindo para a retroalimentação das atividades de pesquisa e comunicação. Além disso, a existência de bases de dados exaustivas, sem viés de idioma e com cobertura apropriada das revistas, é essencial para uma avaliação adequada da ciência nacional ou regional. (Santini; Caregnato, 2018)

A importância das bases de dados locais e regionais foi destacada por Garfield (1995) como um meio crucial de comunicação e disseminação da pesquisa científica em nível local e regional. Ele ressaltou que é importante reconhecer e apoiar os pesquisadores de destaque que contribuem significativamente para o avanço do conhecimento, bem como a necessidade de considerar os interesses e necessidades regionais específicas para fortalecer a base científica de uma nação.

Aguado-López; Rogel-Salazar; Becerril-García (2010) afirmaram que a escassez da produção científica da América Latina e do Caribe nas bases de dados mainstream estimulou projetos que recuperem a dinâmica da produção científica da região. Portanto, são citados exemplos como Redalyc e SciELO que têm concedido visibilidade à produção científica da região e também de outras partes do mundo.

Segundo Santini e Caregnato (2018) a criação de bases de dados regionais e nacionais pode impulsionar a produção científica na região por meio da: 1) Visibilidade da Produção Científica Local; 2) Avaliação e Monitoramento da Produção Científica; 3) Estímulo à Colaboração e Integração Regional; 4) Identificação de Tendências e Demandas Locais; 5) Incentivo à Qualidade e Relevância da Produção



Científica. Tais premissas se relacionam com a recente [Declaração em Apoio à Ciência Aberta com IDEIA - Impacto, Diversidade, Equidade, Inclusão e Acessibilidade](#) e com o [Manifiesto sobre la Ciencia como Bien Público Global: Acceso Abierto No Comercial](#). Dessa forma, reafirma-se que a criação de bases de dados regionais e nacionais têm um papel crucial no fortalecimento da produção científica na região, aumentando a visibilidade, a cooperação e a qualidade das pesquisas realizadas na América Latina.

3 A PROPOSTA DE UMA BASE DE DADOS EM DIREITO

Um dos principais benefícios da criação de uma base de dados em direito é possibilitar, além do acesso à informação, um reconhecimento de autores, revistas e temas do interesse da área, e que não são contempladas pelas bases de dados mainstream. Desta forma para a construção de uma base de dados agregadora é necessário contar com uma equipe multidisciplinar.

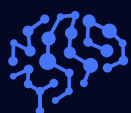
O benchmarking com a BRAPCI envolveu a análise da plataforma da base de dados, uma reunião técnica com o pesquisador que mantém a base e também orientações contínuas da sua equipe. O cumprimento desse objetivo específico levou à definição dos requisitos necessários ao desenvolvimento técnico da Base de Dados em Direito, que conta com uma equipe formada por dois pesquisadores, um técnico e três bolsistas de iniciação científica, sendo um da área de Ciência da Computação e dois da área de Gestão da Informação.

Foram definidos os critérios para a formação da coleção de teste da base, a saber: ser uma publicação periódica científica, ter o International Standard Serial Number (ISSN), ter o foco e o escopo no domínio do direito de Programa de Pós-graduação em Direito do Brasil, atender o protocolo OAI-PMH para coleta de metadados; disponibilizar o acesso para coleta, atribuir uma licença Creative Commons à publicação.

A proposição da base de dados envolve o planejamento da infraestrutura técnica e também informacional dessa fonte de informação. Faz parte do escopo desta pesquisa apresentar uma proposta técnica para a implementação da Base de Dados em Direito. Para a execução técnica da base de dados serão utilizados softwares livres, iniciando do sistema operacional baseado no Ubuntu Server 22.03 LTS, servidor web Apache2, e armazenamento de dados em MySQL, além de disponibilizar 1 TeraByte de armazenamento.

O banco de dados foi pensado para armazenar tanto informações das revistas quanto informações das publicações. Logo, sua implementação acontece por etapas, em que primeiro foram inseridos os dados das revistas em uma única tabela (Journals), depois uma a uma serão inseridos os dados das publicações, sendo os dados de publicações de uma única revista em uma única tabela (Papers_Journal A, B, C, etc).

A tabela *Journals* armazena os seguintes dados: título do periódico ou revista científica, URL (endereço web) do periódico, quando o periódico começou a publicar todo o conteúdo utilizando uma licença aberta, título alternativo do periódico, ISSN da versão impressa do periódico, E-ISSN da versão online do periódico, palavras-chave associadas ao periódico, idiomas nos quais o periódico aceita manuscritos, editora responsável pelo periódico, país da editora, licença do periódico, atributos da licença, URL dos



termos da licença, informações de licenciamento em formato legível por máquina incorporadas ou exibidas nos artigos, URL para uma página de exemplo com informações de licenciamento incorporadas, se o autor detém os direitos autorais sem restrições, URL para informações de direitos autorais, se a Revista apresenta taxa de processamento de artigo (Article Processing Charge), URL para informações sobre preservação, diretório de políticas de depósito,.

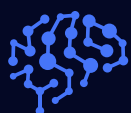
Já a tabela `Papers_Journal_X` armazena: títulos associados ao artigo, autores do artigo, palavras-chave associadas ao artigo, resumos associados ao artigo, editora responsável pela publicação do artigo, data de publicação do artigo, formato do arquivo PDF do artigo, formato do arquivo XML do artigo, link para visualizar o artigo na web, Identificador de Objeto Digital (DOI) atribuído ao artigo, E-ISSN eletrônico associado ao artigo, idiomas em que o artigo está disponível, link para visualizar o artigo em formato PDF, link para visualizar o artigo em formato XML, informações de direitos autorais associadas ao artigo, processo de avaliação por pares (*peer review*) do artigo, número de páginas do artigo, mês de publicação do artigo, ano de publicação do artigo, volume do periódico em que o artigo foi publicado, número do periódico em que o artigo foi publicado e licença sob a qual o artigo está disponível para uso ou distribuição.

Em relação à segurança e ao armazenamento físico dos dados, serão implementadas medidas adicionais para garantir a integridade e confidencialidade das informações com HDs físicos, visando garantir a disponibilidade e recuperação dos dados em caso de falhas ou incidentes. Além disso, uma cópia dos arquivos PDFs serão salvos no armazenamento em nuvem de forma gratuita através do Google Drive API. Essa API, possibilita interagir com o servidor web de forma programática, basta criar credenciais usando o método de autenticação OAuth2.0, garantindo segurança e integridade dos arquivos.

O projeto está na fase de prova de conceito (*proof of concept*) com testes de coleta (*harvesting* OAI-PMH) de metadados e padronização de autoridades está sendo desenvolvida em Python, nesta construção está se estabelecendo o backend da base. Nesta primeira fase busca-se testar as possibilidades técnicas de coleta e organização para construção da base de dados. Em paralelo às políticas de acesso, incorporação, processamento e gerenciamento dos dados a serem coletados estão sendo desenvolvidas. Além da definição dos critérios das revistas que poderão ser indexadas na base.

Com a camada *backend* em andamento, parte do projeto que envolve programação de scripts para acesso via API, pode-se desenvolver a interface do usuário, definida como frontend. Para este fim, será utilizado o Angular, em uma estrutura *Single Page Application* (SPA), o que possibilita um acesso rápido e seguro aos dados, com uma complexidade menor na construção das telas de acesso. As interfaces serão projetadas de acordo com as necessidades dos usuários, seguindo as melhores práticas de usabilidade e design.

A base encontra-se em processo de desenvolvimento, com algumas revistas coletadas para a prova de conceito, a saber, 120 periódicos científicos em acesso aberto editados por programas de pós-graduação em Direito brasileiros. De acordo com o Qualis Capes foram avaliadas um total de 2.326 revistas na área de Direito no último quadriênio, das quais apenas 27 estão indexadas na Scopus. Estima-se que entre 80% dessas revistas são brasileiras, ou seja, há a possibilidade de indexar 1.860 revistas. Entretanto existe a necessidade de avaliação dos critérios de inclusão, e a disponibilidade de coleta de metadados por meio do OAI-PMH.



Dentro da estrutura de organização e gerenciamento de informação a base de dados está sendo estruturada de forma a atender tanto pesquisas acadêmicas quanto pesquisas profissionais. Além do processo de consulta, o projeto da base prevê a geração de indicadores para mapeamento de temas, autores e instituições com o objetivo de possibilitar o uso da ferramenta para pesquisas sobre a área. Entretanto, uma das maiores dificuldades está no reconhecimento dos critérios de inclusão das revistas na base, tanto pela quantidade de revistas existentes, quanto pela sua classificação como revista profissional, acadêmica ou de divulgação.

4 CONSIDERAÇÕES FINAIS

A área do Direito tem muitas especificidades e elas variam conforme os países, uma vez que dependem do sistema de leis adotado. Ainda que seja importante o reconhecimento da legislação internacional e a comparação do Direito brasileiro com outros países, o acesso de forma rápida e precisa à literatura nacional desse domínio é fundamental para atender as necessidades de pesquisadores da área e também juristas. Fato é que não há no Brasil uma base de dados que possibilite o acesso à produção científica e técnicas em acesso aberto do domínio do direito de forma sistematizada e, de modo a reconhecê-lo mapeá-lo a partir dos seus indicadores bibliométricos e cientométricos.

O desenvolvimento da Base de Dados em Direito possibilitará a disseminação, a análise, bem como a geração de indicadores sobre a produção científica na área do Direito no Brasil. Alguns desafios encontrados são a constituição de uma equipe multidisciplinar, a obtenção de recursos financeiros e a garantia da sustentabilidade do projeto, especialmente, do seu produto principal. Pesquisas futuras podem envolver o estudo para a constituição de uma política de informação, a constituição das diretrizes da organização da informação, a proposição da interface e a análise da sua usabilidade, por exemplo. A parte empírica da base de dados encontra-se na prova de conceito e a previsão é que até o final de 2024 já se tenha a primeira versão disponibilizada para consultas, estudos e validações.

AGRADECIMENTOS

Agradecemos o financiamento do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) e da Universidade Federal do Paraná (UFPR).

REFERÊNCIAS

AGUADO-LÓPEZ, Eduardo; ROGEL-SALAZAR, Rosario; BECERRIL-GARCÍA, Arianna. Limites e potencialidades da avaliação científica: crítica epistemológica à cobertura de bases de dados e à construção de indicadores. In: FERREIRA, Sueli Mara Soares Pinto; TARGINO, Maria das Graças (Orgs.). **Acessibilidade e visibilidade de revista científica**. São Paulo, SP: Editora Senac São Paulo, 2010. Disponível em: <https://bibbase.org/network/publication/aguadolpez-rogelsalazar-becerrilgarca-limitesepotencialidades-daavaliacientificacrticaepistemolgicacoberturadebasesdedadoseconstruodeindicadores-2010>. Acesso em: 20 maio 2024.



- APPEL, André Luiz. **Dimensões tecnopolíticas e econômicas da comunicação científica em transformação**. 2019. 172 f. Tese (Doutorado em Ciência da Informação) - Escola da Comunicação, Universidade Federal do Rio de Janeiro, Instituto Brasileiro de Informação em Ciência e Tecnologia, Programa de Pós-Graduação em Ciência da Informação, 2019. Disponível em: <http://ridi.ibict.br/handle/123456789/1024>. Acesso em: 4 nov. 2021.
- BASEVI, Teresa Helena da Rocha. O uso do software livre no STJ: experiência na implantação da BDJur. **Cadernos de Informação Jurídica (Cajur)**, v. 4, n. 2, 2017. Disponível em: <https://www.cajur.com.br/index.php/cajur/article/view/158>. Acesso em: 7 mar. 2024.
- BUFREM, Leilah Santiago *et al.* Modelizando práticas para a socialização de informações: a construção de saberes no ensino superior. **Perspectivas em Ciência da Informação**, v. 15, n. 2, 2010. DOI: <https://doi.org/10.1590/S1413-99362010000200003>. Disponível em: <https://www.scielo.br/j/pci/a/DHV9RQMtPKMrRRrdDMzMVvh/?lang=pt>. Acesso em: 7 mar. 2024.
- GARFIELD, Eugene. Quantitative analysis of the scientific literature and its implications for science policymaking in Latin America and the Caribbean. **Bulletin of the Pan American Health Organization**, v. 29, n. 1, p. 87–95, 1995. Disponível em: <https://iris.paho.org/handle/10665.2/26904>. Acesso em 7 fev. 2024.
- PACKER, Abel L. *et al.* (org). **SciELO – 15 Anos de Acesso Aberto**: um estudo analítico sobre Acesso Aberto e comunicação científica. Paris: UNESCO, 2014, 188 p. ISBN 978-923701-237-6. DOI: <http://dx.doi.org/10.7476/9789237012376>. Disponível em: <https://www.scielo15.org/livro-scielo-15>. Acesso em: 7 mar. 2023.
- PACKER, Abel L.; COP, Nicholas; SANTOS, Solange. A Rede SciELO em perspectiva. In: PACKER, Abel L. *et al.* (org). **SciELO – 15 Anos de Acesso Aberto**: um estudo analítico sobre Acesso Aberto e comunicação científica. Paris: UNESCO, 2014, 188 p. ISBN 978-923701-237-6. DOI: <http://dx.doi.org/10.7476/9789237012376>. Disponível em: <https://www.scielo15.org/livro-scielo-15>. Acesso em: 7 mar. 2023.
- ROSSONI, Luciano; ROSA, **Rodrigo Assunção**. **Reducing the inequality of citations based on local indexers**: the Brazilian Spell (Scientific Periodicals Electronic Library) Case [preprint]. Preprints.org, 2023. DOI: <https://doi.org/10.20944/preprints202311.0755.v1>. Disponível em: <https://www.preprints.org/manuscript/202311.0755/v1>. Acesso em: 7 mar. 2024.
- SANTIN, Dirce Maria; CAREGNATO, Sônia Elisa. Índices de citação nacionais e regionais: importância, experiências e perspectivas para a América Latina. In: ENCONTRO BRASILEIRO DE BIBLIOMETRIA E CIENTOMETRIA, 6., 2018. **Anais eletrônicos [...]**. Rio de Janeiro: UFRJ, 2018. Disponível em: <https://lume.ufrgs.br/bitstream/handle/10183/183984/001075906.pdf?sequence=1>. Acesso em: 7 fev. 2024.
- UNESCO. **Recomendação da UNESCO sobre Ciência Aberta**. Paris: UNESCO, 2022. DOI: <https://doi.org/10.54677/XFFX3334>. Disponível em: https://unesdoc.unesco.org/ark:/48223/pf0000379949_por. Acesso em: 5 fev. 2024.
- VIEIRA, José Ronaldo. Biblioteca Digital Jurídica do Superior Tribunal de Justiça – BDJur: análises e perspectivas. **Cadernos de Informação Jurídica (Cajur)**, v. 4, n. 1, 2017. Disponível em: <https://www.cajur.com.br>.



[com.br/index.php/cajur/article/view/109](https://www.cajur.com.br/index.php/cajur/article/view/109). Acesso em: 7 mar. 2023.

VIEIRA, José Ronaldo; LIMA, Arlan Moraes de; NERI, Vandeilson Soares. Biblioteca Digital do Superior Tribunal de Justiça (BDJur): relato de experiência. **Cadernos de Informação Jurídica (Cajur)**, v. 6, n. 1, 2019. Disponível em: <https://www.cajur.com.br/index.php/cajur/article/view/227>. Acesso em: 7 mar. 2023.