

Pecha kucha

Fortalecimento do ecossistema das revistas científicas brasileiras: interoperabilidade do Diretório Miguilim com o Projeto Laguna

Strengthening the Brazilian scientific journals ecosystem: interoperability of the Miguilim Directory with the Laguna Project

Fortalecimiento del ecosistema de las revistas científicas brasileñas: interoperabilidad del Directorio Miguilim con el Proyecto Laguna

Phillipe de Freitas Campos*

Mestrando em Ciência da Informação

Universidade de Brasília (UnB) | Instituto Brasileiro de Informação em Ciência e Tecnologia (Ibict)

ORCID: <https://orcid.org/0000-0002-7093-703X>

Lattes: <http://lattes.cnpq.br/2076669848354453>

E-mail: phillipecampos@ibict.br

Denise Aparecida Freitas de Andrade

Graduada em Biblioteconomia

Instituto Brasileiro de Informação em Ciência e Tecnologia (Ibict)

ORCID: <https://orcid.org/0000-0003-3988-5929>

Lattes: <http://lattes.cnpq.br/6698900487294293>

E-mail: deniseandrade@ibict.br

Fabio Lorensi do Canto

Doutor em Ciência da Informação

Universidade Federal de Santa Catarina (UFSC) | Instituto Brasileiro de Informação em Ciência e Tecnologia (Ibict)

ORCID: <https://orcid.org/0000-0002-8338-1931>

Lattes: <http://lattes.cnpq.br/5914776544385758>

E-mail: fabio.lc@ufsc.br

Francisco da Costa Carvalho

Graduado em Análise e Desenvolvimento de Sistemas

Instituto Brasileiro de Informação em Ciência e Tecnologia (Ibict)

ORCID: <https://orcid.org/0009-0002-0988-9435>

Lattes: <http://lattes.cnpq.br/4996291063035894>

E-mail: franciscocarvalho@ibict.br

Bianca Amaro

Doutora em Linguística Aplicada

Ministério da Ciência, Tecnologia e Inovação (MCTI)

ORCID: <https://orcid.org/0000-0002-4703-8992>

Lattes: <http://lattes.cnpq.br/1445782939373313>

E-mail: bianca@ibict.br

Resumo

Este artigo descreve o processo de interoperabilidade de dados entre o Diretório Miguilim e o Projeto Laguna, ambos coordenados pelo Ibict. O Miguilim é um diretório que reúne informações bibliográficas e editoriais sobre revistas científicas brasileiras, enquanto o Projeto Laguna opera como um data lake que consolida e organiza dados sobre o ecossistema de ciência, tecnologia e inovação no Brasil. A interoperabilidade foi implementada com o objetivo de enriquecer os metadados disponíveis no Diretório Miguilim, especialmente no que diz respeito à indexação em bases como OpenAlex, SciELO, Scopus, Web of Science, DOAJ, Google Scholar Metrics e Latindex. O processo foi realizado em três etapas: análise e definição de metadados relevantes, extração e processamento dos dados no ambiente do Projeto Laguna e carga automatizada no Miguilim. Como resultado, houve um aumento significativo na completude e consistência dos registros das revistas, melhorando a transparência e confiabilidade das informações disponíveis. A iniciativa contribui para a padronização e eficiência na gestão de dados das revistas científicas, ao mesmo tempo em que promove maior visibilidade das publicações brasileiras em diferentes serviços de indexação. Esta proposta destaca a importância de estratégias de interoperabilidade como ferramenta essencial para o fortalecimento da comunicação científica aberta e para o desenvolvimento do ecossistema de ciência e tecnologia nacional.

Palavras-chave: Diretório Miguilim; Projeto Laguna; Revistas científicas; Interoperabilidade.

Abstract

This paper describes the data interoperability process between the Miguilim Directory and the Laguna Project, both coordinated by IBICT. Miguilim is a directory that gathers bibliographic and editorial information about Brazilian scientific journals, while the Laguna Project operates as a data lake that consolidates and organizes data about the science, technology, and innovation ecosystem in Brazil. Interoperability was implemented with the aim of enriching the metadata available in the Miguilim Directory, especially regarding indexing in databases such as OpenAlex, SciELO, Scopus, Web of Science, DOAJ, Google Scholar Metrics, and Latindex. The process was carried out in three stages: analysis and definition of relevant metadata, extraction and processing of data within the Laguna Project environment,

and automated loading into Miguilim. As a result, there was a significant increase in the completeness and consistency of journal records, improving the transparency and reliability of the available information. The initiative contributes to the standardization and efficiency of scientific journal data management, while also promoting greater visibility of Brazilian publications in different indexing services. This proposal highlights the importance of interoperability strategies as an essential tool for strengthening open scientific communication and fostering the development of the national science and technology ecosystem.

Keywords: Miguilim Directory; Laguna Project; Scientific Journals; Interoperability.

Resumen

Este trabajo describe el proceso de interoperabilidad de datos entre el Directorio Miguilim y el Proyecto Laguna, ambos coordinados por el Ibict. Miguilim es un directorio que reúne información bibliográfica y editorial sobre revistas científicas brasileñas, mientras que el Proyecto Laguna opera como un lago de datos que consolida y organiza datos sobre el ecosistema de ciencia, tecnología e innovación en Brasil. La interoperabilidad fue implementada con el objetivo de enriquecer los metadatos disponibles en el Directorio Miguilim, especialmente en lo que respecta a la indexación en bases como OpenAlex, SciELO, Scopus, Web of Science, DOAJ, Google Scholar Metrics y Latindex. El proceso se realizó en tres etapas: análisis y definición de metadatos relevantes, extracción y procesamiento de los datos en el entorno del Proyecto Laguna, y carga automatizada en Miguilim. Como resultado, se produjo un aumento significativo en la completitud y consistencia de los registros de las revistas, mejorando la transparencia y confiabilidad de la información disponible. La iniciativa contribuye a la estandarización y eficiencia en la gestión de datos de revistas científicas, al tiempo que promueve una mayor visibilidad de las publicaciones brasileñas en distintos servicios de indexación. Esta propuesta destaca la importancia de las estrategias de interoperabilidad como una herramienta esencial para el fortalecimiento de la comunicación científica abierta y el desarrollo del ecosistema nacional de ciencia y tecnología.

Palabras clave: Directorio Miguilim; Proyecto Laguna; Revistas científicas; Interoperabilidad.

Introdução

Há mais de duas décadas, o Instituto Brasileiro de Informação em Ciência e Tecnologia (Ibict), unidade de pesquisa vinculada ao Ministério da Ciência, Tecnologia e Inovação (MCTI), vem desenvolvendo serviços de informação científica e executando projetos de pesquisa alinhados às premissas do Movimento de Acesso Aberto à Informação Científica (MAA) e, mais recentemente, aos princípios da Ciência Aberta (CA).

Campos et al. (2023) apontam que o protagonismo do Instituto remonta à década de 1950, ainda sob sua denominação original, Instituto Brasileiro de Bibliografia e Documentação (IBBD). Desde então, o Instituto tem desempenhado fundamental papel na formulação de políticas, no desenvolvimento de infraestruturas informacionais e na oferta de produtos e serviços voltados para a informação científica e tecnológica e para a sociedade.

Dentre os diversos serviços e projetos de pesquisa conduzidos pelo Ibict, destacam-se, no contexto deste estudo, o Diretório das revistas científicas eletrônicas brasileiras (Miguilim)¹ e o Projeto Laguna. Enquanto o Miguilim é “[...] um serviço de informação [...] que tem como um de seus principais objetivos agregar dados das revistas científicas e dos portais de revistas brasileiros que se encontravam dispersas em diferentes plataformas” (Campos et al., 2025b), o Projeto Laguna opera como um data lake, que “[...] tem por objetivo a criação de uma infraestrutura informacional aberta para dados do ecossistema de informações brasileiro em Ciência, Tecnologia e Inovação” (Carvalho Segundo et al., 2023). Embora com propostas distintas, ambos têm em comum a coleta, o tratamento e a disponibilização de informações sobre as revistas científicas brasileiras.

Nesse contexto, identificou-se a possibilidade de promover a interoperabilidade entre os sistemas, com o objetivo de ampliar e qualificar as informações apresentadas no Diretório Miguilim. A interoperabilidade, nesse caso, refere-se à integração entre diferentes plataformas, de maneira não só a permitir o compartilhamento coordenado, seguro e funcional de dados, mas também a favorecer o intercâmbio de informações que fortaleçam o ecossistema nacional de comunicação científica, especialmente no que diz respeito às revistas científicas (Marcondes & Sayão, 2001).

Ante o exposto, este trabalho busca evidenciar como a interoperabilidade entre o Diretório Miguilim e o Projeto Laguna contribui para o fortalecimento do ecossistema das revistas científicas brasileiras, por meio da ampliação, qualificação e integração das informações sobre os periódicos científicos nacionais.

Procedimentos metodológicos

Trata-se de um relato de experiência do processo de interoperabilidade entre o Diretório Miguilim e os dados oriundos do Projeto Laguna. A integração dos dados ocorreu em três fases principais:

I) Seleção de metadados pela equipe do Miguilim: nesta etapa, foi realizado um mapeamento do padrão de metadados do Diretório, com o objetivo de avaliar o grau de completude das informações disponibilizadas. Embora o Miguilim já possuísse um conjunto amplo de metadados, identificou-se a ausência de informações mais detalhadas sobre a indexação² das revistas em outros serviços de informação científica. Diante disso, propôs-se a ampliação dos dados relativos à indexação por meio da inclusão de links diretos aos registros das revistas nas respectivas bases. As bases selecionadas foram: OpenAlex, SciELO, Scopus, Web of Science/Journal Citation Reports, DOAJ, Google Scholar Metrics e Latindex. Além disso, considerou-se a inclusão da área-mãe de avaliação do Qualis (2017-2020), ampliando e complementando os metadados já existentes sobre essa classificação;

¹ Disponível em: <https://miguilim.ibict.br/>.

² Para efeitos deste estudo, entende-se o termo indexação de maneira ampla, representando o registro, o cadastro, a coleta e a presença de dados das revistas científicas em diferentes sistemas.

II) Execução, tratamento e preparação dos dados pelo Projeto Laguna: nesta fase, utilizou-se a infraestrutura em nuvem da Amazon Web Services (AWS), com gerenciamento baseado em uma arquitetura serverless e x86_64, e processamento de dados realizado em cluster Spark por meio da ferramenta EMR. A ingestão dos dados selecionados pela equipe do Miguilim foi feita utilizando-se os mecanismos de extração disponíveis em cada fonte (download, API ou dumps). Em seguida, realizou-se o cruzamento dos registros por meio do número ISSN entre as revistas do Miguilim e as fontes selecionadas. No caso do Google Scholar Metrics, diante da inexistência de campo de ISSN e de recurso de extração de dados, foi utilizado o framework GSM_hdata (Canto et al., 2024), realizando-se o cruzamento pelos títulos das revistas. Ao final do processo, os dados foram tabulados e disponibilizados à equipe do Miguilim em arquivo CSV;

III) Carga dos dados provenientes do Projeto Laguna para os registros do Miguilim: os dados obtidos foram importados ao Miguilim de forma semiautomática, por meio da funcionalidade “Importar metadados” do software DSpace. Essa funcionalidade permite a carga de grandes conjuntos de dados em lote, utilizando o identificador de cada item como chave principal.

As três fases descritas permitiram a ampliação do conjunto de metadados disponíveis no Diretório Miguilim, além de fluxo de interoperabilidade que pode ser replicado em outras bases de dados. Na seção seguinte, apresentam-se os principais resultados obtidos com a integração.

Resultados e discussões

Na data de coleta desta pesquisa, 26 de fevereiro de 2025, o Diretório Miguilim contava com um total de 5.174 revistas, das quais 22,2% (1.147 revistas) já haviam sido atualizadas pelos editores. Até então, o único campo disponível para registro da indexação das revistas era denominado “Serviços de Informação”, em que os próprios editores autodeclaravam os serviços nos quais suas revistas estavam indexadas. À época, somente o Portal brasileiro de publicações e dados científicos em acesso aberto (Oasisbr)³ permitia, adicionalmente, a inserção do link de acesso direto ao registro da revista.

Todavia, a partir de análises realizadas pela equipe do Miguilim, constatou-se que os dados informados pelos editores apresentavam diversas inconsistências: em alguns casos, as revistas indicavam estar indexadas em serviços nos quais, na prática, não estavam; em outros, também se observou a ausência de informações sobre indexações efetivas, ou seja, revistas devidamente indexadas que não eram declaradas como tal. Além disso, o processo de curadoria dessas informações mostrava-se bastante moroso, dada a diversidade e a especificidade de cada serviço de informação.

Como forma de qualificar esse conjunto de dados, decidiu-se criar novos metadados que apresentam o link direto para os registros das revistas nos respectivos serviços de informação científica, ampliando a confiabilidade das informações disponibilizadas. A

³ Disponível em: <https://oasisbr.ibict.br/>

Tabela 1 apresenta os novos metadados criados e a quantidade de endereços de revistas brasileiras recuperados por meio do Projeto Laguna.

Tabela 1. Endereço “URL” Miguilim provenientes do Laguna

Base de dados	Passou a ter o endereço no Miguilim
Scopus [dc.relation.scopus]	371
Google Scholar [dc.relation.google scholar]	897
OpenAlex [dc.relation.openalex]	3432
SciELO [dc.relation.scielo]	285
Latindex [dc.relation.latindex]	2856
DOAJ [dc.relation.doaj]	1.421

Fonte: elaborada pelos autores, com base nos dados da pesquisa (2025).

Cada metadado apresentado na Tabela 1 foi estruturado de acordo com o padrão Dublin Core Qualificado, utilizando o elemento dc.relation com qualificadores específicos, como dc.relation.scopus ou dc.relation.openalex. A inclusão desses links no Miguilim representa um avanço importante na qualificação dos metadados disponíveis, facilitando a verificação da indexação das revistas por parte dos usuários e promovendo maior confiabilidade e transparência das informações apresentadas no Diretório.

Além das bases internacionais integradas por meio do Projeto Laguna, observou-se que serviços nacionais mantidos pelo próprio Ibict também não dispu

nham de links diretos em seus registros no Miguilim. Com o objetivo de suprir essa lacuna, foram realizados cruzamentos entre os identificadores únicos das revistas (ISSN e o ID do registro no Miguilim) e os registros disponíveis nesses serviços. Com essa estratégia, foi possível incluir no Miguilim os links de acesso direto a 1.366 revistas presentes no Portal Oasisbr e a 3.477 revistas registradas no Diretório de políticas editoriais das revistas científicas brasileiras (Diadorim).

A Tabela 2 apresenta a evolução do campo “Serviços de Informação” com base nos dados (URL) provenientes do Projeto Laguna e de outras fontes.

Tabela 2. Evolução do preenchimento do campo “Serviços de Informação”

Base de dados	Registros inicialmente preenchidos	Novos preenchimentos	Duplicatas removidas	Resultado final
Scopus	257	227	0	484
Google Scholar	393	836	1	1229
OpenAlex	0	3432	0	3432
SciELO	395	42	66	371
Latindex	3779	308	2	4087
DOAJ	896	892	3	1788
Oasisbr	1058	456	0	1514
Diadorim	4965	208	1	5172

Fonte: elaborada pelos autores, com base nos dados da pesquisa (Campos et al., 2025).

O cruzamento de dados permitiu a inserção de informações em registros que anteriormente não continham qualquer referência a esses serviços, evidenciando o potencial da interoperabilidade para qualificar e expandir a base de dados do Diretório. Além disso, o processo de curadoria possibilitou a identificação e remoção de duplicatas, contribuindo para a consistência e a confiabilidade dos registros. Como resultado, observa-se um crescimento expressivo no número total de revistas com informações vinculadas a serviços de indexação, o que amplia a representatividade e a visibilidade dessas revistas no Miguilim.

Os quantitativos apresentados na Tabela 1, quando comparados à coluna “Resultado Final” da Tabela 2, apontam para algumas discrepâncias nos dados, tendo em vista que a quantidade de URLs localizadas é menor do que a quantidade de respostas no campo “Serviços de Informação”. Assim, a Tabela 3 apresenta as taxas de divergência entre o preenchimento do campo “Serviços de Informação” e a presença de URL.

Tabela 3. Análise do campo “Serviços de Informação” em relação à presença de link (URL)

Base de dados	Scopus	Google Scholar	OpenAlex	Scielo	Latindex	DOAJ	Oasisbr	Diadorim
Revistas com o campo preenchido, mas sem URL	113	232	0	86	1231	367	149	1695
Revistas com o campo preenchido, com URL	371	897	3.432	285	2.856	1.421	1.365	3.477
Taxa de divergência	30%	37%	0%	31%	43%	26%	11%	49%

Fonte: elaborada pelos autores, com base nos dados da pesquisa (Campos et al., 2025a).

Observa-se que, com exceção do OpenAlex, que apresentava zero registros preenchidos inicialmente, todas as demais bases de dados demonstram discrepâncias entre o número de registros com link (URL) e aqueles sem correspondência. Foram observadas maiores divergências no Latindex e no Diadorim, enquanto o Oasisbr apresentou o menor índice de inconsistência. Após uma análise amostral aleatória com os registros sem URLs, identificaram-se três eventuais causas para essas divergências: I - em alguns casos, os registros das revistas não estavam efetivamente indexados nas bases informadas, o que indica inconsistências na autodeclaração feita por parte dos editores no Miguilim ou na importação original dos dados ao Diretório, ainda sem relação com o Projeto Laguna; II - o vínculo da revista na base de dados está ligado ao ISSN em outro formato além do eletrônico, o que inviabiliza a vinculação no Miguilim, que registra apenas revistas eletrônicas com registro de ISSN para tal formato; e III - limitações técnicas nos mecanismos de extração de dados utilizados, uma vez que cada base oferece diferentes formas de acesso às informações, o que pode ter impactado a completude dos dados recuperados pelo Projeto Laguna.

Adicionalmente, o Projeto Laguna também viabilizou a incorporação de dados oriundos de outros serviços que, embora não forneçam links diretos para as revistas, foram considerados relevantes para qualificar os registros do Miguilim. A Tabela 4 apresenta os dados complementares integrados a partir dessa iniciativa.

Tabela 4. Dados complementares incorporados a partir do Projeto Laguna

Base de dados	nº de inclusões
Índice H5 [dc.identifier.h5index]	1.985
Qualis periódicos 2017-2020 [dc.description.qualisclassification] e [dc.description.qualisarea]	5.174
Web of Science - Emerging Sources Citation Index (ESCI) [dc.relation.informationsservices]	229
Web of Science - Science Citation Index Expanded (SCIE) [dc.relation.informationsservices]	94
Web of Science - Social Sciences Citation Index (SSCI) [dc.relation.informationsservices]	16
Web of Science - Arts and Humanities Citation Index (AHCI) [dc.relation.informationsservices]	12

Fonte: elaborada pelos autores, com base nos dados da pesquisa (Campos et al., 2025a).

Por meio do Projeto Laguna, também foi possível incorporar informações quantitativas sobre as revistas, como o índice H5, que permite aferir o impacto das revistas com base em métricas utilizadas pelo Google Scholar Metrics. Por outro lado, a incorporação da Classificação Qualis Periódicos (2017-2020) foi incluída por se tratar de um dado relevante para os pesquisadores, especialmente no contexto da pós-graduação brasileira, embora esteja entrando em desuso (Schmidt, 2024). Por fim, o mapeamento nas diferentes coleções da Web of Science em que cada revista está indexada possibilitou identificar aquelas com maior inserção internacional, especialmente no que diz respeito a bases de dados com alto nível de exigência para indexação.

Os resultados apresentados evidenciam o potencial da interoperabilidade entre o Diretório Miguilim e o Projeto Laguna para qualificar as informações disponibilizadas pelo Diretório, especialmente no que tange à indexação em outras bases de dados. A utilização da ferramenta de importação em lote do Diretório permitiu ampliar a cobertura e a confiabilidade dos registros em um curto espaço de tempo, além de apontar fragilidades decorrentes da autodeclaração de algumas informações por parte dos editores científicos. Além disso, a incorporação de novos metadados e indicadores fortalece a transparência e a visibilidade das revistas científicas brasileiras, além de reforçar a importância na utilização de infraestruturas informacionais abertas e interoperáveis.

Considerações finais

O processo de interoperabilidade entre o Diretório Miguilim e o Projeto Laguna se mostrou efetivo, ao contribuir para o preenchimento e a verificação de informações acerca do ecossistema das revistas científicas brasileiras. Os dados integrados permitem traçar um panorama mais preciso sobre os serviços de informação em que essas revistas estão indexadas, favorecendo a análise crítica de sua visibilidade, alcance e inserção em outras bases de dados, nacionais e internacionais.

A integração desses dados traz benefícios em diversos aspectos e para diferentes atores desse ecossistema. Para pesquisadores, explicita informações sobre cada

revista, além de fornecer um instrumento valioso para estudos científicos sobre a editoração científica brasileira. Para editores e administradores de portais de revistas, a interoperabilidade entre os sistemas reduz o retrabalho no preenchimento de informações em múltiplas plataformas, ao mesmo tempo em que promove a padronização e a consistência dos metadados.

Eventuais desafios para a consolidação da interoperabilidade – como a atualização contínua das informações, a heterogeneidade das fontes e a adequação dos padrões de metadados adotados – exigem atenção constante das equipes técnicas envolvidas. No entanto, os avanços demonstrados neste trabalho indicam que o esforço conjunto entre iniciativas como o Miguilim e o Laguna fortalece a infraestrutura nacional de informação científica.

Conflito de Interesses

Os autores declaram não haver conflitos de interesses.

Disponibilização dos Dados de Investigação

Os dados da pesquisa foram disponibilizados no Repositório de dados de pesquisa do Ibict (Aleia), podendo ser acessados em: <https://doi.org/10.48472/aleia/Y9K90W>

CRedit – Contribuições dos Autores

Phillipe de Freitas Campos | Concetualização; Escrita – redação original, Curadoria dos dados

Denise Aparecida Freitas de Andrade | Concetualização, Escrita – redação original, Curadoria dos dados

Fabio Lorensi do Canto | Escrita – revisão e edição, Recolha de dados

Francisco da Costa Carvalho | Desenvolvimento de software

Bianca Amaro | Escrita – revisão e edição, Supervisão

Referências

Campos, F. F., Andrade, D. A. F. de., Canto, F. L. do., Carvalho, F. C., & Amaro, B. (2025a). Conjunto de dados “Levantamento dos indexadores das revistas científicas eletrônicas brasileiras no Diretório Miguilim por meio do projeto Laguna”, <https://doi.org/10.48472/aleia/Y9K90W>, Aleia, VI.

Campos, F. F., Andrade, D. A. F., Amaro, B., Canto, F. L., & Carvalho, F. C. (2025b). Directorio de las revistas científicas electrónicas brasileñas (Miguilim): Desarrollo, curaduría, funcionalidades y perspectivas futuras. *Revista Interamericana De Bibliotecología*, 48(2), 1–15. <https://doi.org/10.17533/udea.rib.v48n2e357709>

Campos, F. F., Sena, P. M. B., Carvalho Segundo, W. L. R., Amaro, B., & Vilas Boas, R. F. (2023). Contribución del Instituto Brasileño de Información en Ciencia y Tecnología a la promoción del acceso abierto y la ciencia abierta: Análisis de sus infraestructuras.

Revista Científica, 48(3), 56–66. <https://doi.org/10.14483/23448350.20924>

Canto, F. L., Pinto, A. L., Dias, T. M. R., Gavron, E. M., & Talau, M. Gsm_hdata: a bibliometric tool to analyze data from google scholar metrics. *Mobile Networks Applications* 29, 754–761. <https://doi.org/10.1007/s11036-023-02258-9>

Carvalho Segundo, W. L. R. de, Pinto, A. L., Canto, F. L. do, & Neubert, P. (2023). Projeto Laguna: infraestrutura de um lago de dados científicos em acesso aberto. *BiblioCanto*, 3(2), 133–138. <https://periodicos.ufrn.br/bibliocanto/article/view/33825>

Marcondes, C. H., & Sayão, L. F. (2001). Integração e interoperabilidade no acesso a recursos informacionais eletrônicos em C&T: a proposta da Biblioteca Digital Brasileira. *Ciência da Informação*, 30(3). <https://doi.org/10.1590/S0100-19652001000300004>

Schmidt, Sarah (2024, 26 de outubro). Qualis-periódicos será substituído por classificação com foco nos artigos - Capes definiu três novos procedimentos de avaliação que poderão ser combinados de acordo com o que cada área do conhecimento definir. *Revista Pesquisa Fapesp*. Disponível em: <https://revistapesquisa.fapesp.br/qualis-periodicos-sera-substituido-por-classificacao-com-foco-nos-artigos/>. Acesso em: 9 abr. 2025.