



Eixo 6 – O mundo digital: apropriação e desafios

Preparação da Plataforma DSpace do Repositório da Universidade Estadual de Londrina - RIUEL para Coleta pelos sistemas GoPG/CAPES: relato de experiência da Universidade Estadual de Londrina

Preparation of the DSpace Platform for harvesting by the CAPES/GoPG systems: an experience report of State University of Londrina

Divino Ignácio Ribeiro Junior – Universidade do Estado de Santa Catarina (UDESC) – divino.ribeiro@udesc.br

Patrícia Ofélia Pereira de Almeida – Universidade Estadual de Londrina (UEL) – pereira@uel.br

Resumo: O artigo apresenta um relato de experiência sobre a preparação do Repositório Institucional da UEL para interoperabilidade com os sistemas da CAPES no contexto do GoPG. Objetiva compilar conhecimentos e experiências validadas para realização da preparação dos repositórios institucionais baseados na Plataforma DSpace. Fundamenta-se nos princípios de interoperabilidade, alinhamento semântico e padronização de metadados. Os procedimentos envolveram análise e tratamento de metadados, configuração do provedor OAI-PMH e customização de *crosswalks* para adequação ao vocabulário COAR. Conclui-se que a interoperabilidade demanda de ações contínuas de curadoria e padronização, contribuindo para a qualidade da informação e integração da produção científica institucional.

Palavras-chave: CAPES GoPG. RIUEL. Repositório Digital. Plataforma DSpace. Crosswalk de Metadados.

Abstract: This article presents an experience report on the preparation of the UEL Institutional Repository for interoperability with CAPES systems within the context of the GoPG program. It aims to compile validated knowledge and experiences related to the preparation of institutional repositories based on the DSpace platform. The study is grounded in the principles of interoperability, semantic alignment, and metadata standardization. The procedures included metadata analysis and processing, configuration of the OAI-PMH provider, and customization of crosswalks to comply with the COAR vocabulary. It concludes that interoperability requires continuous curation and standardization actions, contributing to information quality and the integration of institutional scientific production.

Keywords: CAPES GoPG. RIUEL. *Digital Repositories. DSpace Platform. Metadata Crosswalk.*



1 INTRODUÇÃO

1.1 O sistema GoPG/CAPES e os Repositórios Digitais

O Programa de Governança Colaborativa de Informações da Pós-Graduação (GoPG) é uma iniciativa da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) para viabilizar um ecossistema de serviços para integração entre sistemas acadêmicos utilizados no contexto da pós-graduação brasileira. Tem o intuito de estabelecer uma série de padrões e ferramentas para a interoperabilidade de dados entre diversos contextos e sistemas acadêmicos, fortalecendo a governança colaborativa da informação (CAPES, [2026]). Essa iniciativa visa promover a racionalização dos serviços informacionais, elevando a qualidade da informação dos sistemas envolvidos nos processos da pós-graduação brasileira.

É necessário considerar que as ações de preparação de um repositório de produção acadêmica incluem diversas providências de organização de dados e configurações dos sistemas, dos servidores das plataformas, além do exame detalhado da qualidade dos metadados do acervo, entre outras, que são peculiares à realidade de cada Instituição. Nesse contexto, é natural que as orientações da CAPES não sejam de natureza exaustiva, cabendo assim às instituições o papel de analisar suas especificidades de gestão de metadados e preparar seus ambientes de repositório para a coleta pelos sistemas GoPG.

A motivação para realização deste trabalho está pautada na contribuição com as organizações e outros atores envolvidos no ecossistema da pós-graduação brasileira, no sentido da formação de conhecimentos estratégicos e instrumentalização operacional objetiva, saneando problemas e preenchendo lacunas de informação existentes nas fontes sobre o assunto. Essas fontes, tal como a CAPES ([2026]), apresentam orientações fundamentais, princípios e padrões de forma detalhada, porém ainda não possuem a completude necessária para viabilizar as ações internas de preparação para os repositórios baseados na plataforma DSpace.

Diante disso, o presente artigo tem como objetivo de relatar a experiência de preparação do Repositório Institucional da Universidade Estadual de Londrina (RIUEL) para adequação às diretrizes de interoperabilidade do GoPG, compartilhando



conhecimentos e experiências validadas para realização da preparação dos repositórios institucionais baseados nessa plataforma.

Um repositório institucional é um ecossistema de serviços de informação voltado à preservação e disseminação da produção intelectual de uma organização, organizada em coleções digitais. Ele garante a memória e a visibilidade de ativos como artigos, teses e documentos técnicos. Sua origem remonta ao final dos anos 90, surgindo como resposta à crise na comunicação científica. Na época, o alto custo das assinaturas de periódicos e o controle proprietário das editoras restringiam o acesso ao conhecimento, motivando a busca por alternativas digitais abertas (CAFÉ *et. al.*, 2003).

Movimentos como o Open Archives Initiative (OAI) e marcos como a Convenção de Santa Fé (1999) consolidaram o modelo. Iniciativas como o Arxiv.org foram pioneiras ao promover o autoarquivamento e a interoperabilidade técnica entre plataformas globais. Atualmente, os repositórios digitais são pilares da Ciência Aberta, conforme se observa nas recomendações da UNESCO (2022), prova da consolidação de um canal da comunicação científica essencial para as instituições que atuam na produção do conhecimento científico.

1.2 Os processos do GoPG em Repositórios Digitais

As etapas de implementação do GoPG incluem a *interoperabilidade organizacional* (gestão institucional), a *interoperabilidade semântica* (padronização dos dados), e a *interoperabilidade técnica* (integração entre sistemas e plataformas) (Plataforma Sucupira, [2026]), responsável pela preservação da qualidade da informação durante a transferência dos dados entre os sistemas.

Nesse cenário, o *crosswalk* é um recurso que faz o alinhamento semântico e o mapeamento de metadados (alinhamento sintático), com vistas a facilitar o agrupamento de informações armazenadas em diferentes fontes (Senso; Hidalgo-Delgado; Hipola, 2016, tradução nossa). Em outras palavras, criam-se instrumentos de equivalência entre padrões de metadados distintos, para que o sistema faça a leitura do formato na fonte e a conversão para o formato no sistema de destino.

No contexto do GoPG, esse processo envolve a utilização de vocabulários controlados, como os recomendados pela *Confederaton of Open Access Repositories* (COAR) (Plataforma Sucupira, [2026]), empregados na padronização de tipos



documentais e demais elementos descritivos para fins de interoperabilidade entre repositórios e sistemas de coleta de dados científicos.

2 PROCEDIMENTOS METODOLÓGICOS

Trata-se de um relato de experiência sobre a customização e aplicação de recursos disponibilizados pela Plataforma Sucupira para a interoperabilidade de dados da produção acadêmica do RIUEL com os sistemas da CAPES. A coleta desses dados possui relevância para a consolidação de indicadores de produção científica e para os processos de gestão e acompanhamento da pós-graduação institucional.

As atividades foram desenvolvidas a partir da análise dos metadados existentes no RIUEL, exportação e tratamento dos dados, configuração do provedor OAI-PMH da plataforma DSpace, customização do arquivo de *crosswalk* *oai_capes.xls* (Plataforma Sucupira, [2026]), e validação dos formatos disponibilizados para coleta pelos sistemas da CAPES.

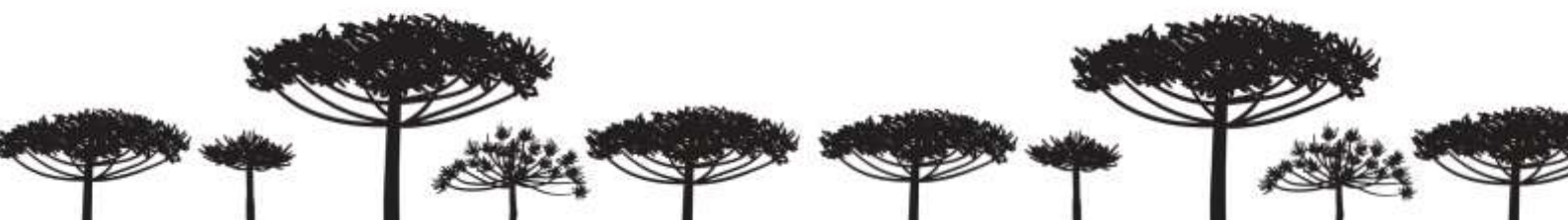
3 RESULTADOS E DISCUSSÃO

A partir da contextualização da pesquisa e dos procedimentos realizados no RIUEL, apresentaremos de forma descritiva os processos aplicados.

3.1 Análise dos metadados do acervo

A tarefa inicial, ponto de partida do processo de implantação das diretrizes do GoPG no RIUEL, foi a análise detalhada de todo o conjunto de metadados do Repositório, totalizando 10987 registros de publicações, entre os quais 8394 dissertações e 2475 teses (os demais são trabalhos de graduação ou especialização), contados na ocasião da escrita deste trabalho. Os dados foram exportados do sistema e tratados por processos que tradicionalmente se utilizam em atividades de *Data Warehouse*, campo de tecnologias originárias da área de Ciência da Computação.

A partir de um painel de análise, foram realizadas diversas tarefas, entre as quais destacam-se:



- Tratamento do metadado de tipo de documento, para preparação da conversão para as recomendações da Open AIRE/Vocabulário COAR;
- Verificação da nomenclatura dos nomes dos PPG existentes no acervo, a partir das informações da CAPES;
- Normalização de identificadores Lattes e ORCID;
- Identificação de informações sobre o tipo de acesso pois as informações sobre embargo não fazem parte dos metadados, e isso exigiu uma análise específica nos dados da plataforma;
- Mapeamento de identificadores da produção local;
- Alinhamento de vocabulários de assunto e de Alinhamento de vocabulários de Área de Conhecimento com as áreas da CAPES;
- Padronização de formatos de data, idiomas e outros tipos de campo.

Vale ressaltar que esta lista não é exaustiva, porém ela sintetiza as principais atividades realizadas. A maior parte dessas atividades foi baseada nas orientações técnicas divulgadas no portal do GoPG.

3.2 Preparação do provedor de dados (OAI) do DSpace

Uma etapa fundamental da preparação para Coleta do GoPG é a configuração do ambiente do Provedor de Dados OAI do DSpace. Em linhas gerais, o Provedor de Dados OAI é parte da arquitetura de compartilhamento de recursos armazenados no DSpace, existente na plataforma há mais de 20 anos, baseada nas especificações da Open Archives Initiative (2002). Ele é um módulo do sistema que constrói uma espécie de *cache* de dados, estruturados em formato XML com a finalidade específica de estarem disponíveis para coleta por *harvesters* (podemos traduzir como ‘coletadores’, mesmo que a tradução literal seja ‘colheita’) que são outros sistemas que consomem esses dados, e toda comunicação é feita através de protocolo HTTP via WEB, o que simplifica muito a implantação, evitando as complexidades que implementações de protocolos de troca de registros apresentavam, a exemplo do Z39.50, originária nos anos 1970 (LIBRARY OF CONGRESS, 2026) , bem antes das tecnologias WEB.

Essa etapa envolve uma sequência de operações de configuração no DSpace, a fim de habilitar a visualização do formato da CAPES no *end-point* OAI do Repositório. De forma resumida as operações são as seguintes:

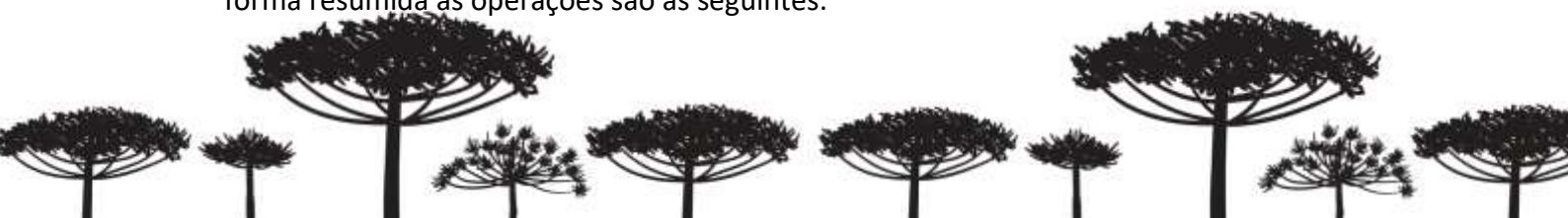
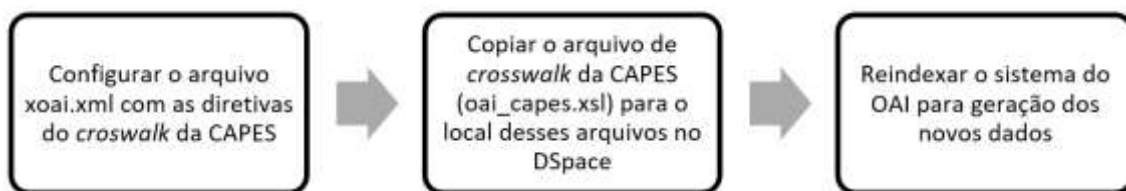


Figura 1 - Etapas das operações no DSpace



Fonte: os autores

Descrição: fluxograma com 03 etapas sobre a preparação do *crosswalk* no DSpace

A validação do êxito desta configuração é obtida pelo acesso ao provedor de dados do DSpace. O processo foi testado em instalações 7.6, 8.3 e 9.2 do DSpace. No caso do RIUEL, o *end-point* do provedor OAI está disponível em <https://repositorio.uel.br/srv-c0003-s01/oai/request?verb=ListMetadataFormats>. Se houver algum erro nas configurações, esse provedor de dados poderá ficar indisponível, ou até mesmo todo o *backend* do DSpace poderá não funcionar. Uma vez confirmada a disponibilidade do provedor de dados, é necessário verificar se o formato de metadado recém adicionado aparece na lista de formatos de metadados disponíveis.

Numa instalação do *backend* padrão, o provedor de dados é encontrado completando a URL do *backend* com a adição do caminho */oai/request?verb=Identify*. A Figura 1 apresenta a tela de identificação do provedor de dados do Repositório da UEL.

Figura 2 - Interface de identificação do provedor OAI do RIUEL

DSpace OAI-PMH Data Provider	
Response Date	2024-09-07 21:28:31
Repository Information	
Repository Name	Repositório Institucional da UEL
E-Mail Contact	repositorio@digital@unesp.com.br
Repository Identifier	repositorio.uel.br
Sample Identifier	oai:repositorio.uel.br:523456789/1234
Protocol Version	2.0
Earliest Registered Date	2024-09-12 13:30:27
Date Granularity	YYYY-MM-DD [timestamps]
Deletion Mode	timestamp

Fonte: UEL ([2026]).

Descrição: Captura de tela com informações do provedor de dados do RIUEL.

Para listar os formatos de metadados ativos, deve-se clicar no link *'Metadata Formats'*, na parte superior da página, e rolar a tela para visualizar o formato da CAPES recém adicionado (Figura 2).



Figura 3 - Lista de formatos de metadados registrados no Repositório



Fonte: UEL [2026].

Descrição: Captura de tela do provedor de dados do RIUEL, destacando o formato de metadado OAI CAPES configurado na plataforma DSpace.

3.3 Configuração e customização do arquivo de *crosswalk oai_capes.xml*

Uma importante etapa é a customização do arquivo *oai_capes.xml*. Esse arquivo é gerado com regras de conversão dos metadados mais comuns em repositórios brasileiros. Por exemplo, repositórios mais antigos, especialmente aqueles que tem origem na distribuição do DSpace adaptado para o mtd_br versão 2 (distribuição do TEDE 2 mencionada anteriormente), também estão parcialmente contemplados neste modelo. É importante compreender de que se trata de um modelo genérico, e precisa ser modificado para incluir ou ajustar o mapeamento de metadados que o repositório da instituição utiliza.

Para consecução desse mapeamento, foi realizado um estudo da estrutura metadados atual, com a finalidade de demarcar os metadados que necessitam de algum ajuste ou inclusão no mapeamento na configuração do *crosswalk*.

A Figura 3 ilustra um dos ajustes realizados: a conversão do campo '*dc.type*', original do RIUEL, para o campo de destino '*dcterms.type*', e fazendo a conversão automática dos valores para o padrão do vocabulário COAR.

A estratégia deste exemplo é preservar a nomenclatura de tipos usada no sistema de busca do repositório, já conhecida pela comunidade acadêmica da Universidade, e expor o valor do metadado convertido para o formato requerido pela coleta GoPG. Dessa maneira, não é necessário fazer a alteração de todos os metadados do sistema, ação que teria grande impacto de manutenção, além de impor uma adaptação dos usuários ao novo vocabulário.

A validação desta conversão foi realizada pela comparação automatizada dos registros XML expostos pelo provedor de dados e os metadados exportados pelo DSpace, com a utilização de um script em Python codificado para este fim. Esse tipo de



processo de comunicação, no qual a linguagem do sistema deve ser conhecida pelo usuário, determinando a possibilidade de interação pela busca (KOBASHI, 1994, p.31).

4 CONSIDERAÇÕES FINAIS

A aplicação de tecnologias para preparação de dados para interoperabilidade sempre foi um desafio no campo das plataformas para Repositórios Digitais. A promoção de recomendações para boas práticas para Ciência Aberta tem sido amplamente disseminada, debates e eventos tem sido cada vez mais frequentes, mas a concretização dessas recomendações ainda é algo que depende da capacidade da instituição em absorver o conhecimento tecnológico envolvido. Sem esse tipo de conhecimento, não se implementam recursos para interoperabilidade, não há compartilhamento de produções, e conseqüentemente, essas recomendações para Ciência Aberta não se concretizam.

Ainda no escopo dos desafios, mesmo com a documentação da CAPES para o GoPG, há diversos aspectos técnicos a serem considerados no contexto da instituição, para então se realizar a preparação do Repositório.

Outro aspecto desse processo é a melhoria da qualidade dos metadados do repositório expostos para coleta. A padronização e alinhamento de informações do repositório melhora sensivelmente, qualificando ainda mais o trabalho realizado pela instituição, e isso deve ser considerado como um impacto indireto relevante da criação do GoPG/CAPES.

Espera-se que esta experiência contribua para que outras instituições enfrentem desafios semelhantes com maior segurança, reconhecendo que a jornada rumo à interoperabilidade é tanto um exercício técnico quanto de governança de dados. O investimento em padronização de metadados não apenas facilita a integração com sistemas externos, mas também fortalece a qualidade e a descoberta dos acervos digitais institucionais.



REFERÊNCIAS

CAFÉ, Lígia; MELO, Bianca Amaro de; BARBOZA, Elza Maria Ferraz. **Repositórios institucionais**: nova estratégia para publicação científica na Rede. In: CONGRESSO BRASILEIRO DE CIÊNCIAS DA COMUNICAÇÃO, 26., 2003, Belo Horizonte. Anais... São Paulo: Intercom, 2003. Disponível em: <http://www.portcom.intercom.org.br/pdfs/133031748743561923574505019461348858722.pdf>. Acesso em 14 de maio de 2026.

CAPES. **Programa de Governança Colaborativa de Informações da Pós-Graduação (GoPG)**. [2026]. Disponível em: <https://www.gov.br/capes/pt-br/aceso-a-informacao/acoes-e-programas/avaliacao/programa-de-governanca-colaborativa-de-informacoes-da-pos-graduacao-gopg>. Acesso em: 13 maio 2026.

PLATAFORMA SUCUPIRA. **Guia de arquitetura e interoperabilidade de dados da pós-graduação**. [2026]. Disponível em: <https://sucupira.capes.gov.br/guia-de-interoperabilidade>. Acesso em: 13 maio 2026.

KOBASHI, N. Y. **A elaboração de informações documentárias**: em busca de uma metodologia. São Paulo: USP. Tese (Doutorado em Ciências da Comunicação) - Escola de Comunicações e Artes, Universidade de São Paulo, 1994

LEIVA-MEDEROS, A.; SENSO, J. A.; HIDALGO-DELGADO, Y.; HIPOLA, P. Working framework of semantic interoperability for CRIS with heterogeneous data sources. **Journal of Documentation**, v. 73, n. 3, p. 481-499, 2017. DOI: <https://doi.org/10.1108/JD-07-2016-0091>. Disponível em: <https://www.emerald.com/jd/article-pdf/73/3/481/1529333/jd-07-2016-0091.pdf>. Acesso em: 14 maio 2026.

LIBRARY OF CONGRESS. **Z39.50 Gateway and Resources**. Disponível em: <https://www.loc.gov/standards/z3950/agency/resources/>. Acesso em: 14 maio 2026.

OPEN ARCHIVES INITIATIVE. **The Open Archives Initiative Protocol for Metadata Harvesting version 2.0**. 14 jun. 2002. Document Version 2015-01-08. Disponível em: <https://www.openarchives.org/OAI/openarchivesprotocol.html>. Acesso em: 14 maio 2026.

UNESCO. **Recomendação da Unesco sobre Ciência Aberta**. Paris: UNESCO, 2022. Disponível em: https://unesdoc.unesco.org/ark:/48223/pf0000379949_por. Acesso em: 14 de maio. 2026.

UNIVERSIDADE ESTADUAL DE LONDRINA – UEL. **Repositório Institucional da Universidade Estadual de Londrina – RIUEL**. [2026]. Disponível em: <https://repositorio.uel.br/home>. Acesso em: 14 maio 2026.

