



A INTEGRAÇÃO DOS REPOSITÓRIOS INSTITUCIONAL E DE DADOS DE PESQUISA DA UNICAMP E A OTIMIZAÇÃO DA PRODUÇÃO CIENTÍFICA EM ACESSO ABERTO

THE INTEGRATION OF UNICAMP'S INSTITUTIONAL AND RESEARCH DATA REPOSITORIES AND THE OPTIMIZATION OF SCIENTIFIC PRODUCTION IN OPEN ACCESS

Alessandra Karyne Neves, Universidade Estadual de Campinas - akaryne@unicamp.br

Erica Cristina de Carvalho Mansur, Universidade Estadual de Campinas - ericacc@unicamp.br

Eixo Temático 6: O mundo digital: apropriação e desafios

INTRODUÇÃO

Hoje as pesquisas científicas e intelectuais no meio acadêmico tem cada vez mais a necessidade de serem disponibilizadas, seja para divulgação das produções, seja para o compartilhamento e reuso dos dados utilizados nas pesquisas ou até mesmo para prestar contas aos órgãos de fomento. O acesso aberto é cada vez mais o caminho para o conhecimento sem fronteiras e sem elitização. Os repositórios institucionais nas universidades visam dar transparência, resguardar e preservar toda a informação produzida na instituição.

Qualquer dado é suscetível a perdas ou alterações e, muitas vezes, fazer uma cópia ou realizar um backup não é suficiente para garantir a existência, integridade e perenidade do documento. Cada mídia tem um determinado período de duração ou logo se torna obsoleta, portanto, também não oferece garantias para o acesso a esses dados no futuro, Lima; Pinto; Farias (p.1, 2020). Pensando na divulgação e neste resguardo dos dados e das produções das instituições é que os repositórios tomaram vulto na área de informação atualmente. Visto que estas duas ferramentas Repositórios Institucionais e Repositórios de Dados tem se mostrado tão necessários



e importantes nas instituições, a interoperabilidade entre eles se faz necessária para ligar todas as partes de uma pesquisa. Contudo esse relato de experiência tem como objetivo mostrar a integração entre essas duas ferramentas.

FUNDAMENTAÇÃO TEÓRICA

O Sistema de Bibliotecas da Unicamp (SBU) tem como missão “prover informação, por meio de produtos e serviços de excelência, para as atividades de ensino, pesquisa e extensão, garantindo um ambiente de respeito à diversidade e à socialização” (SBU-UNICAMP, 2019). Partindo deste princípio, de prover informação, em 2002 o SBU iniciou um projeto de digitalização de todas as teses e dissertações da Unicamp, com o propósito de disponibilizar na íntegra todos estes trabalhos produzidos na universidade. Surgiu assim a Biblioteca Digital da Unicamp (BDU) que em 2010 atingiu a meta de disponibilizar 100% de todos os trabalhos na internet e obteve resultados em nível mundial, com acessos e downloads tanto do Brasil quanto do exterior.

Em 2013 surgiu uma demanda da Fundação de Amparo à Pesquisa (FAPESP), para que as universidades estaduais do Estado de São Paulo tivessem seus Repositórios Institucionais. Com isso, o SBU começou a trabalhar com essa demanda, que em 2013 ficou conhecida como Biblioteca Digital da Produção Intelectual e Científica, sendo apresentada no lançamento do Repositório da Produção Científica do Conselho de Reitores das Universidades Estaduais Paulistas (CRUESP), que

atua como um mega buscador contendo os metadados dos Repositórios das 3 universidades estaduais.

Após esse lançamento houve vários estudos no SBU sobre a ferramenta repositório, para entender primeiramente a diferença entre Bibliotecas Digitais e Repositórios, partindo do princípio de que o SBU estava trabalhando com as produções técnico-científicas na BDU. Porém, neste sistema são indexados outros materiais: obras raras, hemerotecas, periódicos, anais de congresso que não necessariamente são materiais produzidos na Unicamp, segundo Santos (2022, p.36):

Possivelmente a maior diferença entre biblioteca digital e repositório está se firmando em relação ao tipo de documentação. Enquanto os



repositórios têm características científicas, mesmo quando disponibiliza para acesso online literatura cinzenta, bibliotecas digitais, por seu turno, dissemina em grande parte a memória técnica, sendo muito utilizado por órgãos de governo para dar maior visibilidade a sua documentação digital, promovendo a transparência, de forma a atender a Lei de Acesso à Informação (LAI).

Sendo assim, houve uma mudança no nome do Repositório que a partir de 2014 passou a ser conhecido como Repositório da Produção Científica e Intelectual da Unicamp (RI), dado que o RI iria abranger apenas as produções técnico-científicas e intelectuais da universidade. Um segundo estudo realizado foi com vistas à qualidade do Repositório da Unicamp em relação a: metadados, arquivos, curadoria, políticas, direitos autorais, entre outros; pois é sabido que o SBU tem como valores a “excelência, cooperação, acessibilidade, comprometimento, inovação, satisfação do usuário, sustentabilidade, sensibilidade para com o outro e competência em informação” (UNICAMP, 2019).

MÉTODO DA PESQUISA

Para o RI atender os valores relatados, foi realizado um estudo com os usuários para verificar qual a necessidade e expectativa para o uso do RI, e a Divisão de Tratamento da Informação (DTRI) do SBU teve o cuidado de ouvir e adequar o sistema para atender a todos. Sendo assim, seguem alguns destaques no quadro 1:

Quadro 1 - Dados prioritário para construir o RI

Metadados	Além dos básicos temos: URL do artigo, DOI, unidades, departamentos, agradecimentos
Curadoria	Padronização segundo AACR2
Agências de Fomento	Padronização das agências brasileiras com o número de processos e as estrangeiras estão registradas nos agradecimentos
Assuntos	Vocabulário controlado tanto em português quanto em inglês e termos livres
Documentos	Todos as publicações têm o arquivo em PDF-A
Identificadores	Handle, DOI e Orcid

Fonte: Própria, 2022

Com essa qualidade de indexação, o RI cresceu com excelência e trouxe com ele materiais tais como: teses, dissertações, trabalhos de conclusão de curso, trabalhos de especialização, relatórios de pós-doutorado, e-books, capítulos de livros, artigos, e ainda com a expectativa de colocar outras produções como as artísticas,



por exemplo. O resultado desse trabalho no RI foi a disponibilização de indicadores precisos e geração de relatórios, atendendo às necessidades dos usuários.

Consolidado o repositório institucional, surgiu na academia a ideia de disponibilizar os dados de pesquisa, ou seja, informações geradas ou obtidas na investigação científica, independente do seu formato. “Os dados de pesquisa fornecem a informação necessária para reproduzir uma pesquisa, se complementados com a documentação associada (Unicamp, 2019)”. A Unicamp começou a estudar a possibilidade de ter um Repositório de Dados de Pesquisa e criou a Comissão de Gestão de Dados de Pesquisa (CGPD), que tem a “incumbência de sustentar e promover a política institucional de dados de pesquisa da Unicamp propondo ações neste campo segundo as melhores práticas em âmbito nacional e internacional para apreciação e deliberação pela Comissão Central de Pesquisa” (Unicamp/PRP, 2022). Como resultado destes estudos tivemos o lançamento em 2019 do Repositório de Dados de Pesquisa da Unicamp (REDU), que é qualificado como instrumento oficial incumbido de armazenar conteúdos digitais na forma de software, dados brutos de pesquisa, gravações de áudio e vídeo, questionários, códigos computacionais, fotografias e imagens, planilhas, entre outros.

O REDU está dividido em grandes áreas (artes e humanidades, biomédicas, ciências aplicadas, exatas e tecnológicas) apenas docentes, pesquisadores e discentes (autorizados pelos docentes) podem realizar depósitos de metadados - que passam por curadoria e o depositante não é obrigado a dar acesso aos dados de pesquisa. Segue no quadro abaixo alguns destaques:

Quadro 2 - Dados prioritário para construir o REDU

Metadados	Além dos básicos: filiação dos depositantes, resumo sobre os dados de pesquisa e metadados específicos
Curadoria	Padronização segundo AACR2
Agências de Fomento	Padronização das agências brasileiras com o número de processo
Documentos	São três opções: depositar no REDU, colocar o endereço da URL de onde está depositado ou ainda colocar uma nota com os dados do responsável para contato
Identificadores	DOI e Orcid

Fonte: Própria, 2022



Conforme o relatado, o SBU disponibiliza todas essas ferramentas com vistas ao acesso à informação e divulgação das produções técnico-científicas. No entanto, percebendo o crescimento significativo tanto de documentos quanto da qualidade dos mesmos nas ferramentas digitais, veio a indagação: como fazer a integração entre o RI e o REDU, partindo do princípio de que os dois repositórios são de produções da Unicamp?

RESULTADOS: INTEGRAÇÃO ENTRE OS REPOSITÓRIOS DA UNICAMP

Como foi explanado, o Repositório Institucional é, sumariamente, o destino final de uma pesquisa, ou seja, a disponibilização do trabalho. Por outro lado, o Repositório de Dados de Pesquisa disponibiliza os materiais brutos e sem análise, que são importantes por abrirem possibilidades para surgimento de vários produtos.

O SBU fez um estudo paralelo dos dois repositórios com o propósito de integração dos mesmos. Primeiramente levantou-se os dados de pesquisa depositados no REDU por um docente específico e quais foram os produtos resultantes destes dados. Verificou-se que deles foram produzidos um artigo e uma tese. Uma segunda etapa do estudo foi averiguar se estes produtos, ou seja, o artigo e a tese também foram depositados do RI.

Realizado esse levantamento, a próxima etapa consistiu no trabalho entre a DTRI e a Diretoria de Tecnologia da Informação (DTI) do SBU, para verificar quais metadados e informações seriam adequados para fazer a integração entre o RI e o REDU, com vistas a possibilidade de que o usuário que estiver no RI fosse direcionado ao REDU e vice-versa.

O RI está no Software Sophia, que trabalha com o gerenciador em MARC 21 e o REDU está no Software Datavers que trabalha apenas com Dublin Core – o que impossibilitou a pretendida integração no mesmo metadado. Porém, o importante neste momento para a Unicamp é que os dois repositórios estivessem com hiperlink que possibilitasse o acesso a um documento de um repositório para o outro.



Nesta tentativa, para o RI foi escolhido o metadado localização e acesso eletrônico, ou seja, a tag 856\u, pois todos os acessos ficam nesse campo, conforme a figura abaixo:

Figura 1 - Tabela Marc tag 856

[-]	856	4	0	Localização e acesso eletrônico		
	u			URI		
	y			Ligação textual	DOI	
[-]	856	4	0	Localização e acesso eletrônico		
	u			URI		
	y			Ligação textual	Texto completo	
[-]	856	4	0	Localização e acesso eletrônico		
	u			URI		
	y			Ligação textual	Dados de pesquisa	

Fonte: gerenciador Sophia, 2022

Já para o REDU foi escolhido o metadado Related Publication (figura 2), escolhido porque permite colocar a referência completa que está no RI e por ter um campo a parte URL para colocar o Handle.

Figura 2 - Metadado Dublin Core

Related Publication

Citation

+

ID Type **ID Number**

Select...

URL

Fonte: Software Datavers, 2022

Após todos os estudos anteriormente citados, partiu-se para a execução destes procedimentos e verificação se os metadados escolhidos seriam adequados. Para isso, foi selecionado um depósito no REDU com o título “Workflow research object for a molecular dynamics simulation”. Conforme pode ser observado na figura abaixo, foi referenciado mais de um produto final, sendo eles uma tese e um artigo. Isto é



permitted because the Dataverse system allows the repetition of the Related Publication metadata. As can be observed, the information of the handle must be placed in the URL box to be a hyperlink:

Figura 3 - Exemplo dos dados do REDU

Related Publication

Citation

CARVALHO, Lucas Augusto Montalvão Costa. Reproducibility and reuse of experiments in eScience: workflows, ontologies and scripts = Reprodutibilidade e reuso de experimentos em eScience: workflows, ontologias e scripts. 2018. 1 recurso online (111 p.) Tese (doutorado) - Universidade Estadual de Campinas, Instituto de Computação, Campinas, SP.

ID Type Select... **URL** <https://hdl.handle.net/20.500.12733/163>

ID Number

Citation

Carvalho, Lucas A. M. C. C.; Silveira, Rodrigo L.; Pereira, Caroline S.; Skaf, Munir S.; Medeiros, Claudia Bauzer. Provenance-Based Retrieval: Fostering Reuse and Reproducibility Across Scientific Disciplines. In Proceedings IPAW (Proceedings of the 6th International Workshop on Provenance and Annotation of Data and Processes), Springer International Publishing, 183-193, 2016

ID Type doi **ID Number** 10.1007/978-3-319-40593-3_17

URL <https://hdl.handle.net/20.500.12733/166>

Fonte: Software Datavers, 2022

As previously reported, the deposit in REDU passes through curation and publication, after these actions automatically a DOI is generated, which will be used to refer to the research data in the RI, as it is a persistent identifier.

Figura 4 - Exemplo da tese: título Reproducibility and reuse of experiments in eScience

856	4	0	Localização e acesso eletrônico	
			URI	https://doi.org/10.47749/T/UNICAMP.2018.1079656
856	4	0	Localização e acesso eletrônico	
			URI	https://doi.org/10.25824/redu/KX3EYF
856	4	0	Localização e acesso eletrônico	
			URI	https://hdl.handle.net/20.500.12733/1635322

Fonte: gerenciador Sophia, 2022



Figura 5 - Exemplo do artigo: título Provenance-based retrieval

856	4	0	Localização e acesso eletrônico	
u			URI	https://link.springer.com/chapter/10.1007/978-3-319-40593-3_17
y			Ligação textual	Texto completo
856	4	0	Localização e acesso eletrônico	
u			URI	https://doi.org/10.25824/redu/KX3EYF
y			Ligação textual	Dados de Pesquisa
856	4	0	Localização e acesso eletrônico	
u			URI	https://hdl.handle.net/20.500.12733/1661817
y			Ligação textual	Texto completo

Fonte: gerenciador Sophia, 2022

Todos os estudos e experimentos demonstraram êxito na busca pelo objetivo pretendido de se realizar a integração entre o RI e o REDU. Na sequência deste trabalho apresentamos como ficou a visualização para o usuário no RI e no REDU.

No RI o usuário tem apenas o link com o DOI informando que se trata de um dado de pesquisa. Clicando nessa opção o mesmo é direcionado ao REDU para o depósito específico, ficando da seguinte forma:

Figura 6 - Exemplo dos documentos no RI

Sites
 DOI: https://doi.org/10.1007/978-3-319-40593-3_17
 Texto completo: https://link.springer.com/chapter/10.1007/978-3-319-40593-3_17
 Dados de Pesquisa: <https://doi.org/10.25824/redu/KX3EYF>

Arquivos
 000389496000017 PDF

Sites
 DOI: <https://doi.org/10.47749/T/UNICAMP.2018.1079656>
 Dados de Pesquisa: <https://doi.org/10.25824/redu/KX3EYF>

Fonte: Sophia WEB, 2022



No REDU, como mostrado anteriormente o usuário irá visualizar a referência e logo abaixo o Handle que está com hiperlink direcionado para o RI nas produções relacionadas:

Figura 7 - Exemplo do documento no REDU

Citation Metadata ▲	
Dataset Persistent ID	doi:10.25824/redu/KX3EYF
Publication Date	2019-11-28
Title	Workflow research object for a molecular dynamics simulation
Author	Carvalho, Lucas Augusto Montalvão Costa (Universidade Estadual de Campinas. Instituto de Computação) - ORCID: 0000-0002-2412-7183
Contact	<p>✉ Use email button above to contact.</p> <p>Carvalho, Lucas Augusto Montalvão Costa (Universidade Estadual de Campinas. Instituto de Computação) Medeiros, Claudia Maria Bauzer (Universidade Estadual de Campinas. Instituto de Computação)</p>
Description	<p>This is a workflow research object (WRO) created using the Taverna system robundle API. This WRO encapsulates all artifacts needed to reproduce the molecular dynamics experiment published in Silveira, R.L. and Skaf, M. S. Molecular Dynamics Simulations of Family 7 Cellobiohydrolase Mutants Aimed at Reducing Product Inhibition. J. Phys. Chem. B 119, 9295-9303 (2015). DOI: https://doi.org/10.1021/jp509911m. The WRO contains the input data for that experiment, and the encapsulated workflow runs in the Taverna workflow system. The structure and contents of this WRO are documented in https://lucasaugustomcc.github.io/s2rwro/ and further explained in Carvalho's PhD thesis (Reproducibility and reuse of experiments in eScience) in http://repositorio.unicamp.br/jspui/handle/REPOSIP/333317 (2016-07-10)</p>
Subject	Chemistry; Computer and Information Science
Keyword	Scientific workflows Reproducibility in eScience Molecular dynamics Chemistry
Related Publication	<p>CARVALHO, Lucas Augusto Montalvão Costa. Reproducibility and reuse of experiments in eScience: workflows, ontologies and scripts = Reprodutibilidade e reuso de experimentos em eScience: workflows, ontologias e scripts. 2018. 1 recurso online (111 p.) Tese (doutorado) - Universidade Estadual de Campinas, Instituto de Computação, Campinas, SP. https://hdl.handle.net/20.500.12733/1635322</p> <p>Carvalho, Lucas A. M. C. C.; Silveira, Rodrigo L.; Pereira, Caroline S.; Skaf, Munir S.; Medeiros, Claudia Bauzer. Provenance-Based Retrieval: Fostering Reuse and Reproducibility Across Scientific Disciplines. In Proceedings IPAW (Proceedings of the 6th International Workshop on Provenance and Annotation of Data and Processes), Springer International Publishing, 183-193, 2016 doi: 10.1007/978-3-319-40593-3_17 https://hdl.handle.net/20.500.12733/1661817</p>
Notes	There are several other publications associated with this WRO
Language	English
Producer	Carvalho, Lucas Augusto Montalvão Costa (Universidade Estadual de Campinas. Instituto de Computação) (Carvalho, Lucas A. M. C.)
Contributor	Supervisor : Medeiros, Claudia Maria Bauzer
Grant Information	Fundação de Amparo à Pesquisa do Estado de São Paulo: FAPESP: 2017/03570-3 Fundação de Amparo à Pesquisa do Estado de São Paulo: FAPESP: 2014/23861-4 Fundação de Amparo à Pesquisa do Estado de São Paulo: FAPESP: 2013/08293-7

Fonte: Software Datavers, 2022

CONSIDERAÇÕES FINAIS

O projeto foi aprovado por todas as comissões envolvidas nos repositórios. Como é um trabalho que acabou de ser desenvolvido, o SBU está na fase de



divulgação, até o momento é possível relatar que a comunidade científica da Unicamp tem recebido essa integração com boa expectativa de maior divulgação dos trabalhos científicos.

REFERÊNCIAS

LIMA, J. S.; PINTO, V. B.; FARIAS, M. G. G. O bibliotecário na gestão de dados de pesquisa: uma revisão sistemática. **Em Questão**, Porto Alegre, v. 26, n. 3, p. 43–69, 2020. DOI: 10.19132/1808-5245263.43-69. Disponível em: <https://seer.ufrgs.br/index.php/EmQuestao/article/view/90551>. Acesso em: 23 jul. 2022.

SANTOS, Gildenir Carolino; SHINTAKU, Milton (Org.). **Ecosistemas e inovações tecnológicas: da construção às boas práticas**. Campinas: Unicamp/BCCL; Brasília: Ibict, 2022. p. 31-52 DOI: 10.22477/ISBN9786588816363. Disponível em: <https://econtents.bc.unicamp.br/omp/index.php/ebooks/catalog/view/157/190/677>. Acesso em: 20 julho 2022.

UNIVERSIDADE ESTADUAL DE CAMPINAS. Pró-Reitoria de Pesquisa. Gestão de dados de pesquisa. 2021. Disponível em: <https://www.prp.unicamp.br/pt-br/cgdp>. Acesso em: 20 julho 2022.

UNIVERSIDADE ESTADUAL DE CAMPINAS. Sistema de Bibliotecas. Missão, valores e princípios. 2019. Disponível em: <http://www.sbu.unicamp.br/sbu/missao-sbu/>. Acesso em: 20 julho 2022.