
Importação da produção intelectual da Universidade Federal de Minas Gerais a partir dos currículos da Plataforma Lattes: um desafio para ampliação do acesso aberto

Izabel Antonina de Araujo
Universidade Federal de Minas Gerais
antoninaizabel@gmail.com

Luciana Gonçalves Silva Souza
Universidade Federal de Minas Gerais
lucianags9@gmail.com

Resumo

O artigo aborda as importantes iniciativas da Universidade Federal de Minas Gerais na promoção do acesso aberto a sua produção intelectual. A instituição vem trabalhando intensivamente para implementar o seu Repositório Institucional visando promover a acessibilidade e a difusão dos resultados de suas pesquisas que se aproxima de 20 mil itens documentais produzidos anualmente. Trata-se de uma iniciativa para criar processos facilitadores para importação dos dados referenciais da produção da Universidade da Plataforma Lattes para o Repositório Institucional. Utiliza-se como critério para padronização a Rede de Repositórios da América Latina, LA Referencia. Descreve a contextualização acerca do Repositório Institucional da UFMG, com a apresentação dos procedimentos utilizados para a coleta e o aproveitamento dos dados da Plataforma Lattes e a síntese dos resultados alcançados até o momento. Conclui-se que a importação dos dados de forma padronizada para o repositório facilita a gestão e agiliza a disponibilização da informação. Ressalta-se a importância da integração e interoperabilidade entre a Plataforma Lattes e os repositórios e a necessidade de inserir metadados no formulário de preenchimento do Currículo Lattes que são essenciais para facilitar o trabalho de tratamentos dos dados e recuperação da informação.

Palavras-chave: Acesso aberto. Produção intelectual. Plataforma Lattes. Importação de metadados. Repositórios institucionais

Importation of the intellectual production of the Federal University of Minas Gerais from the curriculum of the Lattes Platform: a challenge to expand open access

Abstract

The article addresses the initiatives of the Federal University of Minas Gerais in promoting open access to their intellectual production. The institution has been working intensively to implement its Institutional Repository aiming to promote the accessibility and dissemination of the results of its researches, which approximates 20 thousand documentary items produced annually. It is an initiative to create processes to facilitate the importation of reference data from the of the Plataforma Lattes to the Institutional Repository. The Latin American Repository Network, LA Reference, is used as the criterion for standardization. Describes the contextualisation about the Institutional Repository of UFMG, presenting the procedures used for the collection and use of data from the Lattes Platform and the synthesis of the results achieved so far. It is concluded that the import of the data in a standardized way to the repository facilitates the management and expedites the availability of the information. Evidence the importance of integration and interoperability between the Lattes Platform and the repositories and the need to insert correctly the metadata in the Lattes Curriculum because they are essential to facilitate the work of data processing and retrieval of information.

Key words: Open access, Intellectual production, Platform Lattes, Import metadata, Institutional repositories.

Introdução

Nos últimos anos, o acesso aberto à informação é tema de destaque mundial. Vários países desenvolvem ações que priorizam a divulgação de suas pesquisas. Nesta perspectiva a Universidade Federal de Minas Gerais (UFMG), produtora de conhecimento de interesse público em todas as áreas do conhecimento, nos últimos dois anos vem trabalhando intensivamente para implementar o seu Repositório Institucional (RI). Muitas ações têm sido desenvolvidas no sentido de facilitar o povoamento do repositório, tendo em vista que a Instituição contabiliza aproximadamente 20 mil itens documentais anuais, que incluem livros, capítulos de livros, artigos, trabalhos apresentados em eventos, relatórios de pesquisas, softwares, projetos de pesquisa, material de ensino e aprendizagem, memoriais, entre outros. Todavia, a maioria dessa produção encontra-se dispersa em suas Unidades e Departamentos, em publicações periódicas ou outros meios, sem receber um tratamento analítico e, às vezes, com acesso limitado, havendo dificuldade para a recuperação e preservação dessa informação.

Com o objetivo de reunir e disponibilizar no Repositório a produção intelectual (científica, técnica, artística e cultural) produzida e, devido ao volume da produção, foi necessário pensar em formas de povoamento mediado, realizando a importação da produção de uma base de dados confiável. Com base em experiências de outras instituições de ensino, optou-se por utilizar a Plataforma Lattes, que integra a base de dados de currículos, denominada Currículo Lattes, que hoje é adotada por instituições de fomento, pesquisa e ensino de todo o país, sendo também reconhecida internacionalmente por sua consistência, alcançada pela qualidade do sistema e das informações nele contidas (BRASIL, [2018]). A Plataforma Lattes permite a alguns órgãos, definidos em sua política, a extração dos dados públicos da base de Currículo Lattes utilizando um sistema web próprio, denominado Extrator Lattes (BRASIL, [201-]).

Encontrar alternativas para o povoamento do repositório foi um desafio, tendo em vista que a UFMG se propõe a trabalhar a política de autoarquivamento gradativamente. Este trabalho descreve a experiência nos processos de importação dos dados e as alternativas encontradas para melhoria da qualidade dos dados obtidos.

Integralização entre sistemas de informação: novos desafios para as instituições

Disponibilizar informação qualificada requer padrões para descrever essas informações de forma estruturada para facilitar sua organização e recuperação. Para Souza e Alves (2009), a organização das informações de forma consistente, ou seja, por meio de estruturas de metadados, é de grande relevância quando se busca a interoperabilidade entre aplicações sustentadas por plataformas de hardware e software diferentes. Nesta direção é fundamental estabelecer a comunicação com sistemas internos e externos de forma a partilhar e integrar informação e estes sistemas devem ter padronizados e normalizados.

Para Campos (2007) a padronização é a base para o entendimento dos metadados, sendo este um compromisso que deve ser realizado respeitando a estrutura do padrão escolhido. Esta padronização é responsável por permitir a reutilização e compartilhamento dos metadados, ou seja, é a base para permitir a interoperabilidade. Nesta perspectiva, a padronização permite a transposição de metadados entre diferentes bases de dados, reduzindo esforços na inclusão repetitiva de conteúdos.

Assim, a construção de padrões entre as bases de dados científicas brasileiras auxiliaria os repositórios institucionais em suas estratégias de povoamento, vistas como um dos desafios para os gestores de repositórios, conforme aponta Sarvo et al. (2016). Para eles, o povoamento “apresenta-se como uma atividade complexa que exige grande demanda de recursos para sua realização”.

Como alternativa para racionalização de processos e para simplificar o auto arquivamento da produção, ou até mesmo para contribuir para o seu incentivo, instituições de pesquisas investem na recolha automática de dados de bases científicas. Nesta perspectiva, algumas instituições de pesquisas brasileiras como a Universidade Estadual Paulista “Júlio de Mesquita Filho” (UNESP), a Universidade de São Paulo (USP) e a Universidade Estadual de Campinas (UNICAMP) estão implementando iniciativas de povoamento automatizado dos repositórios.

Com base em experiências relatadas por Matias (2015), Vidotti et al. (2015) e Sarvo et al. (2016) sobre a importação de dados a partir da Plataforma Lattes, a UFMG optou por iniciar o povoamento do seu Repositório Institucional seguindo o mesmo procedimento. Dessa forma, estabeleceu como meta inicial importar os dados referenciais da produção dos seus pesquisadores dos últimos quatro anos registrada na Plataforma Lattes (cerca de 80 mil documentos). Para isso foi necessário estabelecer procedimentos de coleta, definição de regras e algoritmos para padronizar os metadados correspondentes entre a Plataforma Lattes e o Repositório UFMG. Este processo se deu em seis etapas conforme descrito na Figura 1.

Processo de importação para povoamento do Repositório Institucional da UFMG

O trabalho foi realizado pela equipe de bibliotecários envolvidos no processo de implementação do Repositório UFMG em parceria com a equipe de tecnologia de informação da Instituição. A Figura 1 ilustra o processo de povoamento realizado, que será descrito a seguir.

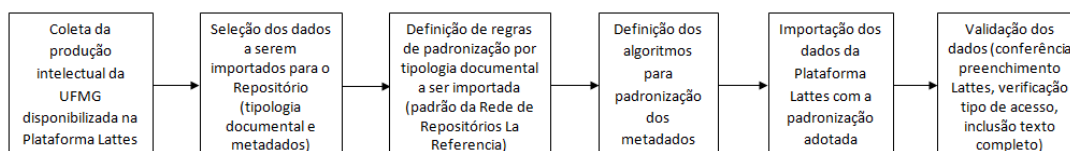


Figura 1: Fluxo de povoamento do Repositório da UFMG com publicações científicas

Inicialmente, a coleta dos dados referentes à produção intelectual da instituição se deu por meio da busca dos registros disponibilizados pelo Currículo Lattes, que foram recuperados e dispostos em planilhas em formato “.xlsx” para seleção dos dados que seriam passíveis de importação no Repositório, considerando as tipologias documentais definidas previamente pelo Comitê Gestor do RI UFMG, a saber: artigos, livros, capítulos de livros, apresentação de trabalhos e trabalhos publicados em eventos, e a definição dos metadados necessários à representação dos documentos importados.

A partir dos registros selecionados, foram identificadas inconsistências de padronização, como as seguintes:

- a) Títulos de livros, capítulos, eventos e trabalhos incompletos;
- b) Múltiplas entradas para o mesmo autor;
- c) Ausência de preenchimento de metadados principais;
- d) Preenchimento incorreto de metadados;
- e) Uso composto de caracteres maiúsculos e minúsculos no preenchimento dos metadados;
- f) Repetição de registros em produções de múltiplas autorias, já que a importação considera os dados inseridos por cada autor do documento.

Assim, foram definidos filtros automáticos para padronização mínima dos dados importados, considerando que o volume da produção a ser incluída no Repositório é considerável e o ajuste manual dos dados seria bastante dispendioso. Tendo em vista que a Plataforma Lattes disponibiliza as informações em acesso livre, mas não possui um protocolo padrão de metadados interoperável com o Repositório, o processo demandou o estabelecimento e refinamento de regras. Além disso, poucas publicações contam com o número DOI (Digital Object Identifier), padrão para identificação de documentos em ambientes digitais, o que poderia facilitar a recuperação dos metadados para inclusão no Repositório.

Diante disso, buscou-se automatizar a padronização da entrada dos autores e dos caracteres quanto ao uso de maiúsculas e minúsculas, seguindo a formatação da rede latinoamericana LA Referencia – Rede Federada de Repositórios Institucionais de Publicação Científica, padrão indicado pelo Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT), órgão nacional encarregado de gerenciar a informação científica e tecnológica do país.

Com a definição dos algoritmos para tratamento dos dados obteve-se uma melhor uniformização do conteúdo que, então, foi importado para o Repositório. A partir desses dados foi realizado o processo de validação dos registros, com a conferência do conteúdo importado a partir do preenchimento dos autores no Currículo Lattes, visando alterar erros de grafia, títulos incompletos e inclusão de metadados não preenchidos; verificação do tipo de acesso dos documentos, para exibição de dados completos apenas dos registros disponíveis em acesso aberto, e inclusão do texto completo desses documentos no RI.

Considerações

O processo desenvolvido para importação dos dados referenciais da Plataforma Lattes utilizando as regras de padronização dos metadados da Rede de Repositórios LA Referencia contribui para agilizar o povoamento do repositório da UFMG, facilita o trabalho de gestão e recuperação da informação, tendo em vista que recuperou o registro de 80 mil

documentos. Além disso, contribui para maior visibilidade das pesquisas desenvolvidas na instituição e a dos seus pesquisadores.

Considera-se que a importação da produção foi bem sucedida, os documentos foram recuperados com os metadados padronizados conforme as regras estabelecidas. Todavia, alguns metadados do Repositório ficaram sem preenchimento pelas seguintes razões: falta de preenchimento dos dados por parte dos pesquisadores e a ausência do metadado “tipo de acesso”, considerado essencial para fazer a gestão da informação no Repositório e que não está disponível no formulário padrão do Currículo Lattes. Desta forma, demanda ações de preenchimento, validação e conferência dos metadados pela equipe técnica do repositório.

Ressalta-se a importância da integração e interoperabilidade entre a Plataforma Lattes e os repositórios institucionais e da necessidade de estabelecer um formulário padrão para preenchimento de metadados de forma a facilitar o tratamento dos dados importados destas bases. Acredita-se que a inclusão de critérios para o preenchimento do Currículo Lattes, utilizando campos controlados, e a conscientização dos autores para a inclusão correta dos conteúdos pode resultar em maior qualidade das informações disponibilizadas na Plataforma, que alimentarão os repositórios. Estas ações facilitarão a gestão e agilidade na disponibilização da informação com ganhos qualitativos para toda a comunidade acadêmica e a sociedade em geral.

Referências

BRASIL. Conselho Nacional de Desenvolvimento Científico e Tecnológico. Extração de dados da Plataforma Lattes – PoSIC. [Em linha]. Brasília : CNPQ. [Consult. 20 Abr. 2018]. Disponível na Internet: <<http://memoria.cnpq.br/documents/313759/dad0813b-6eb1-45b5-9084-fdbd3f6c78cc>>.

BRASIL. Conselho Nacional de Desenvolvimento Científico e Tecnológico. Sobre a Plataforma Lattes. [Em linha]. Brasília: CNPQ. [Consult. 19 Abr. 2018]. Disponível na Internet: <<http://lattes.cnpq.br/>>.

CAMPOS, Luiz Fernando de Barros (2007) – Metadados digitais: revisão bibliográfica da evolução e tendências por meio de categorias funcionais. *Encontros Bibli: Revista Eletrônica de Biblioteconomia*. 1518–2924. Vol. 12., No 23 (2007). [Consult. 04 Set. 2018]. Disponível na Internet: <[URL:https://periodicos.ufsc.br/index.php/eb/article/view/1518-2924.2007v12n23p16](https://periodicos.ufsc.br/index.php/eb/article/view/1518-2924.2007v12n23p16)>.

MATIAS, M. (2015) – Base referencial para o povoamento de repositórios institucionais: coleta automatizada de metadados da Plataforma Lattes. São Carlos : Universidade Federal de São Carlos. 86 p. Dissertação de mestrado.

SARVO, Denilson de Oliveira et al. (2016) – Ineficácia do povoamento automatizado de repositórios institucionais a partir da Web of Science. ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 17, Salvador, 2016.

VALMORBIDA, Willian (2011) – Análise e implementação de um sistema integrado de busca baseado nos padrões de metadados e protocolos de interoperabilidade utilizados por catálogos on-line de bibliotecas e repositórios digitais. Lisboa : Centro Universitário de Lisboa, UNIVATES, 2011. 115 p. Monografia de graduação.

VIDOTTI, Silvana Aparecida Borsetti Gregorio et al. (2015) – Coleta de dados a partir dos currículos da Plataforma Lattes: procedimentos utilizados no Repositório Institucional UNESP. PontodeAcesso. ISSN 1981-6766. Vol. 9, No 3 (2015), p. 117-132.