

Skylines and guidelines: policies and procedure manuals for knowledge organization in institutional repositories

Bueno-de-la-Fuente, Gema (University of Zaragoza, Spain)

Agustín-Lacruz, Carmen (University of Zaragoza, Spain)

Fujita, Mariângela S.L. (Universidade Estadual Paulista, Brasil)

Terra, Ana Lúcia (Universidade de Coimbra, Portugal)

Repositories emerged in the late 20th century as a new strategy of academic institutions with the aim of storing, managing, and preserving huge collections together with the metadata describing them. Academic repositories began to develop the role of publishers by updating the process of scholarly communication (Villalobos and Gomes, 2018) and improving the dissemination of their intellectual production, through open access to knowledge. Nowadays, the number of repositories and the consequent availability of information have increased enormously (e.g. OpenDOAR statistics. JISC, 2022).

Like digital libraries, academic repositories have specific approaches to information management and need to organize knowledge through representation techniques, which are important for information retrieval (Hjørland, 2021). Today, academic libraries coexist with institutional repositories and often share administrative functions and structures. Although originally responding to different needs, this close co-existence in administrative structures produces synergies and mutual improvements. While libraries are very much oriented to the technical process of collection management (bibliographic selection to support reference and cataloging services), repositories focus on digital preservation, availability, and accessibility.

Both information systems have common strategies on accessibility, interoperability, and standards adoption. Therefore, they share basic knowledge organization processes, such as the attribution of metadata to support access to collections. Villalobos and Gomes (2018) highlight the lack of efforts to standardize metadata in repositories. Knowledge Representation can be a mean to enhance the value of digital repositories (Araújo and Silva, 2021).

The importance of developing professional documents (repository policies, guidelines, and procedural manuals) is highlighted (Villalobos and Gomes, 2018). These tools should include aspects such as: metadata, information processing, controlled vocabularies, auxiliary tables, and thesauri (Sousa, 2012). The use of natural language and keywords produces fewer effective results when compared to controlled vocabulary search. The larger the collection of a repository, the more sophisticated the subject-controlled vocabulary should be (Phillips, Taver and Zavalina, 2019).

The authors of this paper continue previous research on the organization of knowledge in institutional repositories (Fujita and Tolare, 2019; Terra, Agustín-Lacruz, Bernardes, Fujita and Bueno-de-la-Fuente 2021) and its treatment in the scientific literature (Fujita, Agustín-Lacruz,

Tolare, Terra and Bueno-de-la-Fuente, 2021 and 2022). As a result, the need to also analyze repositories' professional documents was identified.

The aim of this work is to analyze the recommendations of the guidelines, policies, and procedure manuals of a sample of institutional repositories and networks, in terms of knowledge organization and thematic characterization of their contents. We present an exploratory study on these repositories' professional documents.

A convenience sample has been selected, including representative sources at 3 different levels: institutional repositories: Repositório Institucional UNESP (Univ. Estadual Paulista, <https://repositorio.unesp.br>), Estudo Geral (Univ. de Coimbra, <https://estudogeral.sib.uc.pt>), and Repositorio Zagan (Univ. Zaragoza, <https://zagan.unizar.es>); national aggregators: Oasisbr (Brazilian Open Access Publications and Scientific Data Portal, <https://oasisbr.ibict.br>), RCAAP (Repositórios Científicos de Acesso Aberto de Portugal, <https://www.rcaap.pt>), and RECOLECTA (Recolector de Ciencia Abierta, <https://recolecta.fecyt.es>); together with international aggregators and networks: OpenAIRE (<https://www.openaire.eu>), COAR (Confederation of Open Access Repositories, <https://www.coar-repositories.org>), and LARReferencia (Red de repositorios de acceso abierto a la ciencia, <http://www.lareferencia.info>).

At the institutional repository level, a template has been defined to gather information from the sample, such as: institutional rules and procedure manuals openly available; authorship and traceability; institutional organization and organic responsibility; thematic organization of the contents; recommendations on subject metadata elements; recommendations to use controlled vocabularies (thesaurus, taxonomies, classification systems), and if they are regularly updated.

At the international level, it is observed that guidelines and policies that include recommendations on subject metadata encoding have been adopted, always recognising the autonomy of repositories to implement them according to their needs. OpenAIRE guidelines (2022a) just suggest the use of classification schemes or controlled vocabularies, and if no specific classification scheme is used, they recommend the Dewey Decimal Classification (DDC). OpenAIRE maintains a list of supported subject classification schemes, to code the qualifier of dc:subject element (OpenAIRE, 2022b). COAR (Confederation of OpenAccess Repositories) have an editorial board on controlled vocabularies, and have developed vocabularies on resources types, access rights, and version types, but not for subject elements (COAR, 2021). For its part, LARReferencia, whose metadata collection policies were initially founded on the Driver guidelines, is currently in the process of adapting to the OpenAIRE 4.0 (OpenAIRE 2022a).

At the national level, it is noted that the aggregators OASISBr, RCAAP and Recolecta, adhere to these international policy frameworks for the metadata harvesting and validation.

The preliminary results have shown that the three national aggregators share a common framework of policies and guidelines on metadata aggregation. In addition, the study of local repositories, raised the following limitations: a) Manuals and guides are difficult to locate and access, identify their authorship, and trace their update; b) Lack of tools for vocabulary control; c)

Low level of adoption of content schemes for subject metadata, and d) Very high number of keywords without vocabulary control.

Finally, further research is needed on the effective use of specific content schemes for the coding of controlled vocabularies and classification schemes, for which a selective harvesting of the sample repositories would be needed.

References

- Araújo, D. O. & Silva, M. B. (2021). Repositório digital: a delimitação de um conceito por meio de mapa conceitual. *Informação@Profissões*, 10(3), 18-33. <https://www.uel.br/revistas/uel/index.php/infoprof/article/view/44878>
- COAR (2021). *COAR Repository toolkit*. <https://www.coar-repositories.org/news-updates/coar-repository-toolkit>
- Fujita, M. S. L. & Tolare, J. B. (2019). Vocabulários controlados na representação e recuperação da informação em repositórios brasileiros. *Informação & Informação*, 24(2), 93-125. <http://dx.doi.org/10.5433/1981-8920.2019v24n2p93>
- Fujita, M. S. L., Agustín-Lacruz, M. C., Tolare, J. B., Terra, A. L. & Bueno-de-la-Fuente, G. (2021). A organização do conhecimento em repositórios institucionais: Uma análise da literatura recente publicada em periódicos de biblioteconomia e ciência da informação. In Silva, C. G. D., Revez, J. & Corujo, L. (eds.). *Organização do Conhecimento no Horizonte 2030: Desenvolvimento Sustentável e Saúde*, 703-716. <https://doi.org/10.51427/10451/50067>
- Fujita, M. S. L., Agustín-Lacruz, M. C., Tolare, J. B., Terra, A. L. & Bueno-de-la-Fuente, G. (2022). Institutional repositories and knowledge organization: A bibliographic study from Library and Information Science. *Education for Information*, Pre-press, 1-16. <http://dx.doi.org/10.3233/EFI-220015>
- Hjørland, B. (2021). Information Retrieval and Knowledge Organization: A Perspective from the Philosophy of Science. *Information*, 12(3), 135. <https://doi.org/10.3390/info12030135>
- JISC (2022). *OpenDOAR: OpenDOAR Statistics. Growth of OpenDOAR*. https://v2.sherpa.ac.uk/view/repository_visualisations/1.html
- OpenAIRE (2022a). *OpenAIRE Guidelines for Literature Repositories v. 4.0*. <https://openaire-guidelines-for-literature-repository-managers.readthedocs.io/en/v4.0.0/>
- OpenAIRE (2022b). *OpenAIRE API. Vocabularies: dnet:subject_classification_typologies*: http://api.openaire.eu/vocabularies/dnet:subject_classification_typologies
- Phillips, M. E., Taver, H. & Zavalina, O. L. (2019). Using metadata record graphs to understand controlled vocabulary and keyword usage for subject representation in the UNT theses and dissertations collection. *Cadernos BAD*, 1, 61-76. <https://dcpapers.dublincore.org/pubs/article/download/4237/2431.pdf>
- Sousa, B. A. D. (2012). Proposta de Criação de um Repositório Institucional para o IFPB. *Revista Brasileira de Biblioteconomia e Documentação*, 8(1), 66-84. <http://rbbd.febab.org.br/rbbd/article/view/196/228>
- Terra, A. L., Agustín-Lacruz, C. Bernardes, O., Fujita, M. S. L. & Bueno-de-la-Fuente, G. (2021). Subject-access metadata on ETD supplied by authors: A case study about keywords, titles and abstracts in a Brazilian academic repository. *Journal of Academic Librarianship*, 47(1). <https://doi.org/10.1016/j.acalib.2020.102268>
- Villalobos, A. P. de O. & Gomes, F. A. (2018). Análise dos repositórios das universidades federais brasileiras. *PontodeAcesso*, 12(3), 126-144. <https://periodicos.ufba.br/index.php/revistaici/article/view/27929>