SKYLINES AND GUIDELINES POLICIES AND PROCEDURE MANUALS FOR KNOWLEDGE ORGANIZATION IN INSTITUTIONAL REPOSITORIES

NKOS WORKSHOP 2022, 7 OCTOBER 2022

Bueno-de-la-Fuente, Gema Agustín-Lacruz, Carmen

Fujita, Mariângela S.L.

Terra, Ana Lúcia.







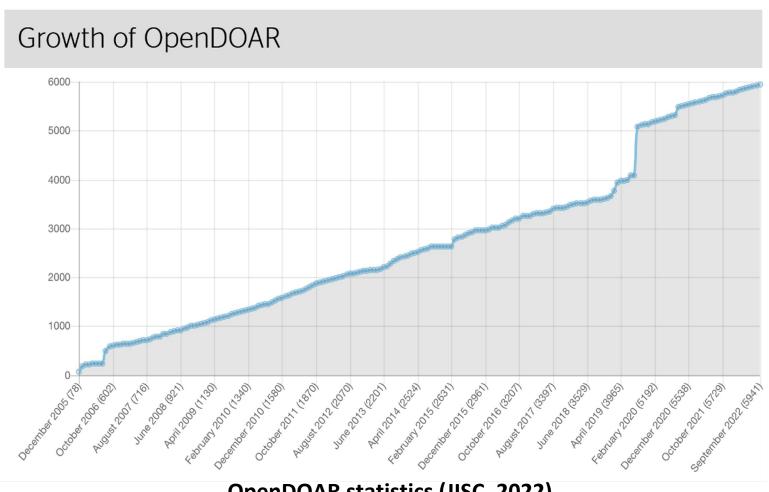
CONTENTS



- 1. Background
- 2. Aims
- 3. Methodology
- 4. Findings
- 5. Conclusions
- 6. Further Research

- Repositories are a strategy of academic institutions that emerged at the the late 20th century.
- Academic repositories began to develop the role of publishers by updating the process of scholarly communication.





OpenDOAR statistics (JISC, 2022)

https://v2.sherpa.ac.uk/view/repository_visualisations/1.html



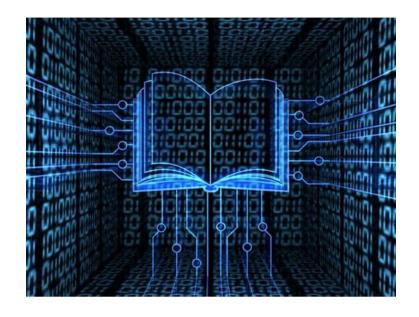
- Institutional repositories have specific approaches to information management and need to organize knowledge
- Academic libraries are much oriented to the collection management.
- Repositories are more focused on digital preservation, availability, and accessibility.
- Though responding to different needs, their close co-existence in administrative functions and structures, produces synergies and mutual improvements.

ACADEMIC LIBRARIES AND INSTITUTIONAL REPOSITORIES...



... share basic **knowledge organization** processes, such as
the attribution of metadata to
support access to collections.

...have common strategies on accessibility, interoperability, and standards adoption.



AIMS

The importance of developing professional documents (repository policies, guidelines, and procedural manuals) is highlighted.

These tools should include aspects such as: **metadata**, information processing, **controlled vocabularies**, auxiliary tables, and thesauri.



AIMS



To analyze:

- the recommendations about the knowledge organization and thematic characterization
- in the guidelines, policies, and procedure manuals
- of a sample of institutional repositories and networks.

METHODOLOGY



METHODOLOGY

At the **institutional repository level,** a data gathering **template** has been defined:

- institutional rules and procedure manuals
- authorship and traceability
- institutional organization
- thematic organization
- recommendations on subject metadata
- recommendations to use controlled vocabularies
- regularly update of recommendations.



METHODOLOGY

1. DESCRIPTION	2. COLLECTION MANAGEMENT	3. METADATA	
Repository name	Deposit mode (self-archiving, delegated deposit, etc.)	Metadata curation (authority control)	
URL	Assistance instruments: FAQ; tutorials; procedure manuals; support mail or chat.	Recommendations on subject metadata elements	
Regulatory framework	Assistance instruments regularly updated	Recommendations to use controlled vocabularies	
Institutional organization and organic responsibility	Authorship of instruments (clearly stated)	Indexing policies known by authors	
Management units	Digital preservation policies	Use of standardized classification system (e.g. UDC)	
Nº of records	Use of persistent identifiers (DOI, Handle, URN, ORCID, etc.)	Use of controlled vocabularies or ontologies with concepts endowed with persistent identifiers (e.g. LCSH LD, UNESKOS)	
Institutional rules openly available	Collections	Number of topic / subject	
Institutional rules updated	Organizing systems Percentage of subjects /		

FINDINGS AT LOCAL LEVEL

SPANISH REPOSITORY

1. DESCRIPTION

Repository name

URL

Regulatory framework

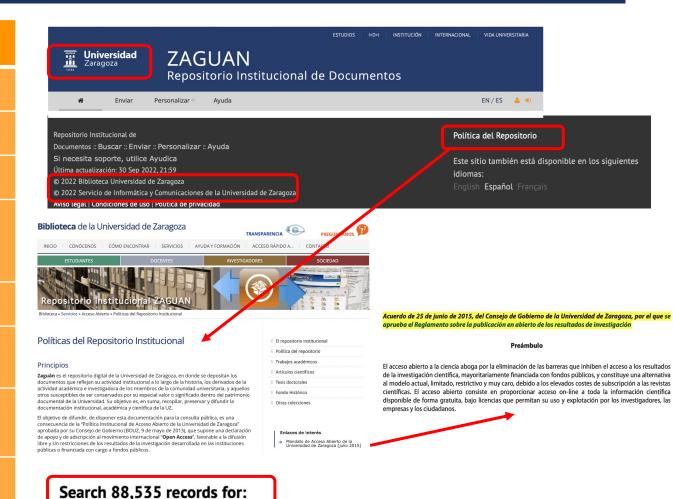
Institutional organization and organic responsibility

Management units*

Nº of records

Institutional rules openly available

Updated data (Institutional rules)



FINDINGS AT LOCAL LEVEL

BRAZILIAN REPOSITORY

2. COLLECTION MANAGEMENT

Deposit mode (self-archiving, delegated deposit, etc.)

Assistance instruments: FAQ; tutorials; procedure manuals; support mail or chat

Assistance instruments regularly updated

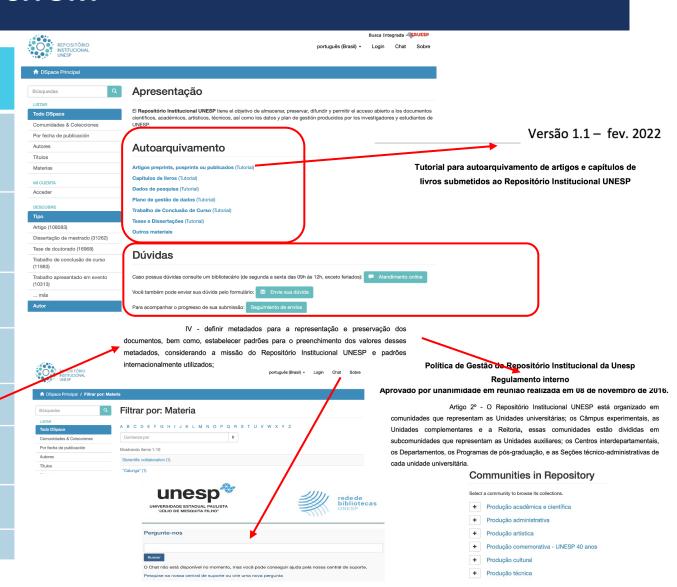
Authorship of instruments (clearly stated)

Digital preservation policies

Use of persistent identifiers (DOI, Handle, URN, ORCID...)

Collections

Organizing systems



FINDINGS AT LOCAL LEVEL

PORTUGUESE REPOSITORY

3. METADATA

Metadata curation (authority control)

Recommendations on subject metadata elements

Recommendations to use controlled vocabularies

Indexing policies known by authors

Use of standardized classification system (e.g. UDC)

Use of controlled vocabularies or ontologies with concepts endowed with persistent identifiers (e.g. LCSH LD, UNESKOS)

Number of topic / subject

Percentage of subjects / records

https://estudogeral.sib.uc.pt/politica.jsp?locale=en

ESTUDO GERAL Repositório científico da UC

idicio comunidades publicações investigadores unidades de IRO pi 3. **Bibliotecas da UC**a coleções

Compete às Bibliotecas da UC:

Política de Acesso Livre da UC

(aprovado a 23 de Setembro de 2010)

a Zelar pela qualidade dos metadados relativos às publicações da comunidade académica de cada Unid Organica, μεραπαπεπτο ου υπίσασε σε I&D.

https://estudogeral.sib.uc.pt/help/index.html#controlledvocabulary

SUBMIT: Controlled Vocabulary top of submit top

A controlled vocabulary is a set of terms which form a dictionary of descriptio categorised in archives.

Accurately categorising material using a controlled vocabulary increases the li

To enter a controlled vocabulary term in the form, select "Subject Categories"

https://estudogeral.sib.uc.pt/help/index.html#subjectsearch

SUBJECT CATEGORY SEARCH top

A controlled vocabulary is a set of terms which form a dictionary of descriptions of particular types of content or subject matter. These are maintained by standards bodies in order to standardise the way that similar materials are categorised in archives. This aids searchine by increasing the likelihood that the relevant materials will be returned by the user's search.

Percorrer por assunto





FLUC - Departamento de Filosofia, Comunicação e Informação : [710]

Subject Categories

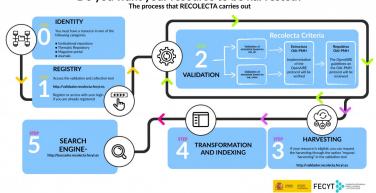
- Research Subject Categories
 - HUMANITIES and RELIGION
 - LAW/JURISPRUDENCE
 - SOCIAL SCIENCES
 - Social sciences
 - Education
 - Subject didactics
 - Nursing education
 - Music education
 International education
 - Psychology
 - Applied psychology
 - Environmental psychology
 - Cognitive science

https://estudogeral.sib.uc.pt/controlledvocabulary/controlledvocabulary. jsp?ID=dc_subject&vocabulary=srsc

FINDINGS AT NATIONAL LEVEL

SPANISH NETWORK

Do you want your resource to be harvested?



SERVICES

There is an important infrastructure of resources in Spain, made up of institutional repositories, thematic repositories and open access journals where research publications can be published or deposited openly.

> RECOLECTA aggregates all these resources in a single platform that guarantees its interoperability and allows open access to all national scientific publications.

The services of RECOLECTA to open access repositories are:

- · Validator: Allows repositories to self-assess, as many times as they want, their degree of compliance with international interoperability guidelines and identify erroneous records.
- Collector: Aggregates the metadata of the repositories that are part of the RECOLECTA community on a
- · Search engine: Allows free locating and access to all scientific publications deposited in Spanish repositories through a single interface.
- User support: recolecta@fecyt.es



In 2020, the FECYT has signed its adherence to the LA Reference. This alliance will contribute to increasing the visibility for Spain's scientific production collected in RECOLECTA by including it in an aggregator and an international search engine that counts on the presence of 10 countries, more than 100 institutions and gives access to more than 2 million scientific documents in open access.

ESPAÑOL | ENGLISH

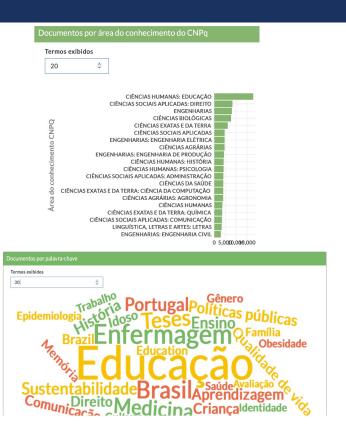
Año/ Year	Autor/ Author	Titulo/ Title	Idioma/ Language	Categoría/ Category
2021	UNESCO	Recomendaciones de Open Science de UNESCO	English	Otros
2021	FECYT, CSIC, ISCIII y el INIA	Resultados del proyecto Infraestructuras y Estándares para la Ciencia en Abierto (INEOS) y manual de buenas prácticas para la gestión de repositorios institucionales, sistemas de gestión curricular y revistas científicas	Español	Documentos FECYT
2021	Grupo de trabajo de evaluación de repositorios	Guía para la evaluación de repositorios institucionales de Investigación	Español	Documentos FECYT
2021	FECYT	API RECOLECTA de enriquecimiento del metadato fuente de financiación. MANUAL	Español	Documentos FECYT
2021	FECYT	Guía de ayuda del buscador de publicaciones y proyectos RECOLECTA	Español	Documentos FECYT
2020	COAR Working Group	COAR Community Framework for Good Practices in Repositories	English	Otros
2020	Rebiun	Declaración a favor del conocimiento abierte y costenible	Español	Doclaraciones
2018	Paolo Budroni, University of Vienna and Stefan Hanslik, Austrian Federal Ministry for Education in close cooperation with the European Commission	The Vienna Declaration on the European Open Science Cloud	English	Declaraciones
2017	Comisión de Seguimiento sobre el Grado de Cumplimiento del Artículo 37 de la Ley de la Ciencia	Hacia un acceso abierto por defecto	Español / English	Documentos FECYT



FINDINGS AT NATIONAL LEVEL

BRAZILIAN NETWORK





What are the partners linked to Oasisbr?

Oasisbr collects, disseminates and gives access to all content available in the Brazilian Digital Library of Theses and Dissertations (BDTD). In addition, Oasisbr gives access to the scientific content present in the Scientific Repository of Open Access of Portugal (RCAAP), and the content made available by RCAAP is collected by Oasisbr and presented to the Brazilian scientific community. The contents aggregated by Oasisbr are also collected by the Rede Repositorios de Acceso Abierto a la Ciencia (LA Referencia), which in turn is collected by the European aggregator OpenAIRE. This information network enables Oasisbr to fulfill its primary mission, to increase the impact of Brazilian Science around the world.

FINDINGS AT NATIONAL LEVEL PORTUGUESE NETWORK

About RCAAP content

The RCAAP portal is an aggregator (meta-repository) that collects the description (metadata) of documents deposited in various institutional repositories, research data repositories and scientific journals in Portugal. The portal saves the full text of these documents to improve the search results but does not save any document.

When selecting the title of a document at search results in the RCAAP portal, you access, when possible, to their full content in the repository where the document is. So that when difficulties arise in view or download the document that it should be the repository in question and not the portal RCAAP. In addition to being able to search the Portuguese scientific production, you can also choose to search the Brazilian scientific production which currently is comprised of several repositories and journals aggregated by OASISbr project.

The portal RCAAP tries to collect only the scientific literature and academic full text of which is in free access. However, there may be some documents that do not meet these requirements, or may be restricted due to embargoes or may not contain the full text of the document, only metadata records.

For more information, see the list of repositories in the <u>directory</u> of the current aggregate RCAAP.





About RCAAP

RCAAP portal aims to collect, aggregate and index Open Access scientific contents from Portuguese institutional repositories.

RCAAP constitutes a single entry point for searching, discovery and recall of thousands of scientific and scholarly publications, namely journal articles, conference papers, thesis and dissertations, distributed by several Portuguese repositories. A list of the repositories aggregated in the portal is available in the <u>Directory</u>.

RCAAP portal is one of the main components from the project Repositórios Científicos de Acesso Aberto de Portugal. RCAAP project is an initiative from UMIC Knowledge Society. Agency, developed by FCCN Fundação para a Computação Científica Nacional, with the technical and scientific collaboration from Minho University.

The current version of RCAAP portal was developed based on the platform from the network $\underline{\mathsf{La}}$ Referência.

Back to Homepage Top

About RCAAP Help

All rights reserved.











FINDINGS AT INTERNATIONAL LEVEL OPENAIRE

17. Subject (MA)

datacite:subject



♠ OpenAIRE Guidelines for institutional and thematic Repository Managers

17.1. Cardinality

Mandatory if applicable

Occurrence: 0-n

17.2. Definition and Usage Instruction

Subject, keyword, classification code, or key phrase describing the resource (occurrences: 0-n).

Usage Instruction

In the subject property two kinds of values are possible: encode either a keyword or a classification.

In general, choose the most significant and unique words for keywords, avoiding those too general to describe a particular resource.

For keywords/keyphrases that are not controlled by a vocabulary or thesaurus either encode multiple terms with a semi-colon separating each keyword/keyphrase; or repeat the element for each term. There are no requirements regarding the capitalization of keywords though internal (within archive) consistency is recommended.

Where terms are taken from a standard classification schema: encode each term using the additional attributes of the subject property. Encode the complete subject descriptor according to the relevant scheme. Use the capitalisation and punctuation used in the original scheme.

It is recommended to use an URI when using classification schemes or controlled vocabularies especially when codified schemes are used DDC or UDC. Service providers can recognise encoding schemas more easy when the schema is "URI-fied" by an authority namespace.

If no specific classification scheme is used we recommend the Dewey Decimal Classification (DDC). More information about the DDC and the DDC Summaries can be found at https://www.oclc.org/en/dewey/resources.html. Please note that OCLC owns all copyright rights in the Dewey Decimal Classification system. Dewey, Dewey Decimal Classification, DDC, OCLC and WebDewey are registered trademarks of OCLC.

FINDINGS AT INTERNATIONAL LEVEL

LAREFERENCIA



LA Referencia forma parte de la Confederación de Repositorios de Acceso Abierto (COAR) y a través de RedCLARA colaboran con el Proyecto OpenAIRE.

2 - Directrices Metadatos https://www.lareferencia.info/es/recursos/directrices-metadatos				
Documentos				
Criterios de Calidad de LA Referencia - Julio 2015 (1763 descarga	as) Popular	29 Nov 2016	Descargar (pdf, 204 KB)	
Directrices de LA Referencia - Versión extendida (ppt.) (1735 des	scargas)	05 Abr 2018	Descargar (pptx, 9.89 MB)	
Directrices de LA Referencia - Versión abreviada (ppt.) (1335 des	scargas)	05 Abr 2018	Descargar (pptx, 7.06 MB)	
Directrices de LA Referencia - Versión basada en la infografía (p	pt.) (1095 descargas)	05 Abr 2018	Descargar (pptx, 1.47 MB)	
Directrices DRIVER 2.0 (2129 descargas)		29 Nov 2016	Descargar (pdf, 920 KB)	

FINDINGS AT INTERNATIONAL LEVEL

COAR







What is a controlled vocabulary?

A controlled vocabulary is an organized arrangement of words and phrases used to index content and/or to retrieve content through browsing or searching. It typically includes preferred and variant terms and has a defined scope or describes a specific domain.

COAR Community Framework for Good Practices in Repositories Version 2 – updated on August 29, 2022

Essential Characteristic Desired Characteristic 1. Discovery .9 The repository facilitates linking in the metadata record between 1.1 The repository enables users to apply basic Dublin Core elated contents such as preprints, published articles, data, and metadata to its records, as well as more granular elements ftware (e.g. including PIDs for related resources held elsewhere) (e.g. to support multilingualism, FAIR-compliance, disciplinebased, and regional metadata schemas) 10 The repository supports PIDs for authors, funders, institutions, funding programmes and grants, and other relevant entities 1.2 The repository supports harvesting of metadata using 1.11 The metadata in the repository are made available under a Creative Commons public domain dedication / waiver (CC0) 1.3 In cases where the resource has been withdrawn, the repository provides a tombstone page and the metadata 1.12 In the case of research data, the repository supports identifiers record remains publicly available for data at multiple levels of granularity, where appropriate (for example, if there there is research using a subset of the full dataset 1.4 The repository assigns persistent identifiers (PIDs) that and a citation of the data subset is needed) point to the landing page of the resource 1.13 The repository facilitates the use of controlled vocabularies in 1.5 The repository offers a search facility its metadata records 1.6 The metadata in the repository are indexed by external 1.14 The metadata in the repository are available for download in a academic discovery services and aggregators standard bibliographic format at no cost to the user 1.7 The repository is included in one or more disciplinary or general registry of repositories 1.8 The metadata in the repository are available in humanreadable and machine-readable formats

5. Quality assurance

- **5.1** The repository undertakes lightweight review (and enhancement if needed) of basic metadata upon submission of the resource
- **5.2** The repository provides documentation or has a policy outlining what curation processes are applied to the resources and the metadata

COAR Confederation of Open Access Repositories Controlled Vocabularies

The use of controlled vocabularies for bibliographic metadata "ensures that everyone is using the same word to mean the same thing". The continuous revision, update and maintenance of the COAR Controlled Vocabularies and its adoption by the most commonly used open repository software is a way to enhance the interoperability across repositories and with other related systems such as harvesters, CRIS systems, data repositories and publishers.

News & Undates

What We Do

The COAR Controlled Vocabularies are governed and maintained by an Editorial Board. In order to define the controlled vocabularies, the Editorial Board analyzes existing vocabularies and dictionaries and will adopt the most appropriate existing terms and definitions whenever possible. In cases where there are gaps identified by the community, new terms are defined by the group. The COAR Controlled Vocabulary Editorial Board also translates vocabulary terms into numerous languages.

COAR Controlled Vocabularies

Resource Type Vocabulary

Access Rights Vocabulary

Version Type Vocabulary

http://bit.ly/coar_vocabularies

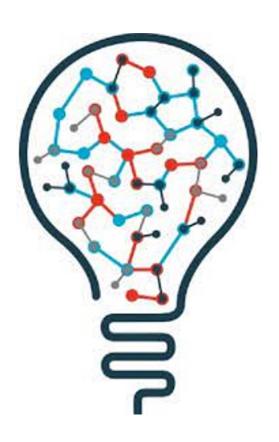
CONCLUSIONS LOCAL LEVEL



At the local level (institutional repositories) the study raised the following barriers:

- a) Manuals and guides are difficult to locate and access, identify their authorship, and trace their update.
- a) Lack of tools for vocabulary control.
- a) Low level of adoption of content schemes for subject metadata.
- a) Very high number of keywords without vocabulary control.

CONCLUSIONS NATIONAL LEVEL



At the national level (aggregators and networks of repositories):

- a) Delegate the decision of adopting controlled vocabularies to each repository.
- a) Refer to international policies and guidelines regarding subject control.

CONCLUSIONS INTERNATIONAL LEVEL



At the international level (aggregators and networks of repositories):

- a) Recommendations on subject metadata encoding are included in their policies and guidelines
- a) The autonomy of repositories to implement the policies according to their needs is always recognised

FURTHER RESEARCH



- to expand the sample under study, with a larger number of repositories and networks.
- to perform a selective harvesting of the sample repositories, checking the effective use of concept schemes for specific controlled vocabularies
- to draw up questionnaires addressed to those in charge of repositories and networks at the national and international levels.

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). <u>Institutional repositories and knowledge organization</u>: A bibliographic study from <u>Library and Information Science</u>. 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

SKYLINES AND GUIDELINES POLICIES AND PROCEDURE MANUALS FOR KNOWLEDGE ORGANIZATION IN INSTITUTIONAL REPOSITORIES

NKOS WORKSHOP 2022, 7 OCTOBER 2022



THANK YOU

Gema Bueno-de-la-Fuente; Carmen Agustín-Lacruz, Mariângela S. L. Fujita, Ana Lúcia Terra