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 late 20th century.

Academic repositories began to develop the role of publishers by updating the process of scholarly communication.
BACKGROUND

Growth of OpenDOAR

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

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

**Local**
- Institutional repositories
  - Repositório Institucional UNESP (Univ. Estadual Paulista)
  - Estudo Geral (Univ. de Coimbra)
  - Repositorio Zaguan (Univ. Zaragoza)

**National**
- National aggregators and networks
  - Oasisbr
  - RCAAP
  - Recolecta

**International**
- International aggregators and networks
  - OpenAire
  - LAReferencia
  - COAR
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

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

## 1. DESCRIPTION

<table>
<thead>
<tr>
<th>Repository name</th>
<th>ZAGUAN Repositorio Institucional de Documentos</th>
</tr>
</thead>
<tbody>
<tr>
<td>URL</td>
<td><a href="http://zaguan.unizar.es/">http://zaguan.unizar.es/</a></td>
</tr>
<tr>
<td>Regulatory framework</td>
<td></td>
</tr>
<tr>
<td>Institutional organization and organic responsibility</td>
<td></td>
</tr>
<tr>
<td>Management units*</td>
<td></td>
</tr>
<tr>
<td>Nº of records</td>
<td></td>
</tr>
<tr>
<td>Institutional rules openly available</td>
<td></td>
</tr>
<tr>
<td>Updated data (Institutional rules)</td>
<td></td>
</tr>
</tbody>
</table>

### Repository name

**ZAGUAN Repositorio Institucional de Documentos**

**Repository name:** Universidad de Zaragoza

**URL:** http://zaguan.unizar.es/

**Regulatory framework:**

**Institutional organization and organic responsibility:**

**Management units:**

**Nº of records:**

**Institutional rules openly available:**

**Updated data (Institutional rules):**

**Search 88,535 records for:**
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
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

3. Bibliotecas da UC

Compete às Bibliotecas da UC:

Zelar pela qualidade dos metadados relativos às publicações da comunidade académica de cada Unidade Organizativa, Departamento ou unidade de 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 descriptive categorised in archives.

Accurately categorising material using a controlled vocabulary increases the likelihood of finding the material.

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 searching by increasing the likelihood that the relevant materials will be returned by the user's search.

---

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? The process that RECOLECTA carries out

**Identify**
- You must have a resource in one of the following categories:
  - Institutional repository
  - Researcher profile
  - Researcher portal

**Registry**
- Register the validation and validation code
- Register in the RECOLECTA platform
- Register or access with your login credentials if you are already registered

**Validation**
- RECOLECTA Criteria
- Validation process
- Report

**Harvesting**
- If your resource is eligible, you can request the harvesting of your resource
- RECOLECTA will require the validation code

**Search Engine**
- http://recolecta.redreference.com

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 weekly basis.
- **Search engine**: Allows free locating and access to all scientific publications deposited in Spanish repositories through a single interface.
- **User support**: recolecta@fecyt.es

---

**WHAT IS IT?**
- SERVICES
- OPEN ACCESS
- FECYT ALLIANCES
- DOCUMENTS
- NEWS & EVENTS

**La Referencia**
- Red de repositorios de acceso abierto a la ciencia

The **Federated Institutional Repositories of Scientific Publications Network** is a Latin American network of open access repositories. Through its services, it supports national Open Access strategies in Latin America through a platform with interoperability standards, sharing and giving visibility to the scientific production generated in institutions of higher education and scientific research.

In 2020, the **FECYT** has signed its adherence to the **La Referencia**. This alliance will contribute to increasing the visibility for Spanish 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.

---

**Table of Scientific Publications**

<table>
<thead>
<tr>
<th>Year</th>
<th>Author/Author(s)</th>
<th>Title/Title</th>
<th>Language</th>
<th>Category</th>
</tr>
</thead>
<tbody>
<tr>
<td>2021</td>
<td>UNESCO</td>
<td>Recomendaciones de Open Science de UNESCO</td>
<td>English</td>
<td>Otros</td>
</tr>
<tr>
<td>2021</td>
<td>FECYT, CBIC, ISCIII y el INIA</td>
<td>Resultados del proyecto Infraestructuras y Estandares para la Ciencia Abierta (INESIS) y manual de buenas practicas para la gestión de repositorios institucionales, sistemas de gestión curricular y revistas científicas</td>
<td>Español</td>
<td>Documentos FECYT</td>
</tr>
<tr>
<td>2021</td>
<td>Grupo de trabajo de evaluación de repositorios</td>
<td>Guía para la evaluación de repositorios institucionales de investigación</td>
<td>Español</td>
<td>Documentos FECYT</td>
</tr>
<tr>
<td>2021</td>
<td>FECYT</td>
<td>API RECOLECTA de enriquecimiento del metadato fuente de financiaciación, MANUAL</td>
<td>Español</td>
<td>Documentos FECYT</td>
</tr>
<tr>
<td>2021</td>
<td>FECYT</td>
<td>Guía de ayuda del buscador de publicaciones y proyectos RECOLECTA</td>
<td>Español</td>
<td>Documentos FECYT</td>
</tr>
<tr>
<td>2019</td>
<td>Press</td>
<td>Declaration to the Framework on open science, Spain</td>
<td>Español</td>
<td>Declaraciones</td>
</tr>
<tr>
<td>2018</td>
<td>Paolo Budroni, University of Vienna and Stefan Hanslik, Austrian Federal Ministry for Education in close cooperation with the European Commission</td>
<td>The Vienna Declaration on the European Open Science Cloud</td>
<td>English</td>
<td>Declaraciones</td>
</tr>
<tr>
<td>2017</td>
<td>Comisión de Seguimiento sobre el Grado de Cumplimiento del Artículo 37 de la Ley de la Ciencia</td>
<td>Hacia un acceso abierto por defecto</td>
<td>Español / English</td>
<td>Documentos FECYT</td>
</tr>
</tbody>
</table>
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 Repositórios de Acesso Aberto 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.
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 directory 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 Directory.

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 La Referência.
17. Subject (MA)

datacite:subject

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-fed" 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.


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

<table>
<thead>
<tr>
<th>Documento</th>
<th>Descargas</th>
<th>Fecha</th>
<th>Descargar</th>
</tr>
</thead>
<tbody>
<tr>
<td>Criterios de Calidad de LA Referencia - Julio 2015</td>
<td>1763</td>
<td>29 Nov 2016</td>
<td>Descargar (pdf, 204 KB)</td>
</tr>
<tr>
<td>Directrices de LA Referencia - Versión extendida (ppt.)</td>
<td>1735</td>
<td>05 Abr 2018</td>
<td>Descargar (pptx, 9.89 MB)</td>
</tr>
<tr>
<td>Directrices de LA Referencia - Versión abreviada (ppt.)</td>
<td>1335</td>
<td>05 Abr 2018</td>
<td>Descargar (pptx, 7.06 MB)</td>
</tr>
<tr>
<td>Directrices de LA Referencia - Versión basada en la infografía (ppt.)</td>
<td>1095</td>
<td>05 Abr 2018</td>
<td>Descargar (pptx, 1.47 MB)</td>
</tr>
<tr>
<td>Directrices DRIVER 2.0 (2129 descargas)</td>
<td>2129</td>
<td>29 Nov 2016</td>
<td>Descargar (pdf, 920 KB)</td>
</tr>
</tbody>
</table>

https://www.lareferencia.info/es/recursos/directrices-metadatos
# FINDINGS AT INTERNATIONAL LEVEL

**COAR**


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

## 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

---

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.
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

- OpenAIRE (2022a). OpenAIRE Guidelines for Literature Repositories v. 4.0. https://openaire-guidelines-for-literature-repository-managers.readthedocs.io/en/v4.0.0/
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