Repository Metadata Collection and Curation Practices Survey

A Barcelona Declaration / COAR joint survey

Available languages: English | Español | Português | 中文

Introduction:

Why this is important: Repositories play a key role in the scholarly communications landscape by preserving and providing access to a wide range of valuable research outputs. The metadata collected and made openly available by repositories supports discovery by search and indexing systems. It is used by researchers, institutions, funders, service providers, and policy makers for reporting, monitoring, research assessment, and the development of value-added services.

The Barcelona Declaration on Open Research Information and COAR have a shared interest in improving the quality of metadata in repositories. Accurate, granular metadata is essential for making repository content discoverable and reusable, and contributes to a trustworthy scholarly ecosystem.

Despite the important role repositories can play, metadata collection and curation practices vary significantly across the repository landscape, particularly in terms of metadata quality, completeness, consistency, and availability for reuse.

Purpose: This survey aims to gather comparable, high-level information about how repositories collect, curate, and make available research information (metadata), and to identify common practices, challenges and effective approaches to metadata curation for the broader community. The findings will help raise awareness of the value of high-quality metadata and inform ideas and strategies for repositories to strengthen their metadata practices.

About the repositories in scope in this survey: This survey focuses on institutional, national, disciplinary, federal, and other repositories that collect, curate, preserve, and make available scholarly resources and the metadata associated with those research outputs.

Who should respond: repository managers, repository librarians, technical leads, curators. The survey should ideally be completed by individuals with sufficient knowledge of repository workflows, metadata collection and curation practices, and decisions related to research information systems and infrastructure.

Note. If your organization operates more than one repository, please complete a **separate survey for each repository.**

Estimated time to complete: 20-30 minutes

The survey will be open from 3 November 2026 - 15 December 2026

Data use and privacy. *Your answers will be processed solely for the stated purpose of this survey, handled confidentially and stored securely. They may be included in aggregated/anonymized analyses.*

Select findings may be shared publicly (e.g. in reports or presentations), but no personal or identifiable information will be disclosed.

*They will **not** be used for marketing, advertising, or promotional outreach.*

Questions? Contact: survey@barcelona-declaration.org

Section 0. Respondent & Repository Context

Repository information
The following questions provide basic context about your repository

0.1 Institution name
0.2 Country
0.3 Repository public name
0.4 Repository home page URL (not the software/platform)

Respondent information
The following questions provide basic context about the respondent and will allow us to contact you with any questions about your repository.This information will not be shared publicly.

0.5 Contact person
0.6 Role of contact person in the organization
0.7 Contact email (optional)

Section 1. Repository Scope & Key Practices

This section provides essential context about your repository, including its scope and the types of research outputs it manages. It also focuses on key practices related to persistent identifiers and metadata.

1.1 Scope of the repository (check all that apply)
Please indicate which type(s) of research outputs does your repository contains. If something isn't listed, please choose "Other" and add a brief description.

- ☐ Publications (e.g. articles, books, conference outputs)
- ☐ Theses and dissertations
- ☐ Research data
- ☐ Software / code
- ☐ Other research outputs (e.g. reports, protocols, DMPs, non-textual outputs)
- ☐ Other (please specify): _____

1.2 Approximately how many records does your repository currently contain? (select one)

- Fewer than 10,000
- 10,000–99,999
- 100,000–999,999
- 1 million or more
- Not sure

1.3 Persistent identifiers at the record level
Does your repository assign or record persistent identifiers (PIDs) at the item level (e.g., DOI, Handle, ARK)? (check all that apply)

- ☐ The repository assigns new PIDs to deposited resources
- ☐ The repository records existing PIDs for deposited resources (for example, as assigned by the publisher of an article)
- ☐ No
- ☐ Not sure

1.3.1 If your repository assigns PIDs to records, which types does it assign? (check all that apply) [Conditional question, shown only if the repository assigns PIDs in Question 1.3)]

- ☐ DOI
- ☐ Handle
- ☐ ARK
- ☐ Other (please specify) : _____

1.3.2 If your repository does not assign PIDs, what is the reason? (check all that apply) [Conditional question, shown only if the response is "No" in Question 1.3)]

- ☐ PID assignment is not a priority for us
- ☐ Cost considerations
- ☐ Limited technical capacity
- ☐ Platform limitations
- ☐ Other (please specify) : _____

1.4 Metadata elements used in practice

To what extent are the following metadata elements populated in your repository? (Select one response for each metadata element)

Please focus on how commonly these elements are actually populated in practice across records, rather than on whether the repository system includes fields for them. You may also include less commonly used elements, such as contributor(s) / contributor statements (e.g. CRediT roles), data availability statements, retractions, or corrections, under "Other", if relevant.

Metadata element	Mandatory	Always, but not mandatory	Often	Sometimes	Rarely	Never
Title						
Author(s)						
ORCID(s)						
Author affiliation(s)						
ROR(s)						
Copyright holder						
License(s)						
COAR Resource Type Vocabulary						
Other resource type classification						
Subject heading(s) or keywords						

Project(s)						
Funding						
Date of Original Publication						
Abstract						
Other: _____						

1.4.1 Is there anything you would like to add to help contextualize your responses regarding metadata population practices? (optional)
For example, differences between collections, or metadata elements that are only used for specific types of outputs.

1.5 Primary metadata schema(s) in use
Please indicate which metadata schema(s) are primarily used to describe items in the repository (check all that apply)

- ☐ Dublin Core
- ☐ Local schema: (please specify the standard or metadata schema it is based on, if any) _____
- ☐ DataCite
- ☐ Not sure
- ☐ None
- ☐ Other (please specify): _____

1.5.1 Please specify the local schema and, if applicable, the standard or metadata schema on which it is based (short text) [Conditional question (shown only if the response is “Local schema” in Question 1.5)]

Section 2. Metadata Collection and Curation Practices

In this section, please indicate how metadata is collected and managed in your repository, including the main metadata curation and enrichment practices applied and how consistently they are used across repository records.

2.1 How is metadata collected in your repository? (check all that apply)

- ☐ API ingestion (pull or push from another system)
- ☐ Data download / import (e.g., bulk files, exports)
- ☐ Manual entry by library / repository staff
- ☐ Manual entry by author or author's proxy (e.g. student or admin staff)
- ☐ Other (please specify): _____

2.2 Curation & enrichment activities

Which of the following curation or enrichment metadata practices are applied in your repository?

Please indicate how commonly each practice is applied across repository records by selecting one response for each activity.

Activity	Always	Often	Sometimes	Rarely	Never
Completeness checks					
Schema / profile validation (checking compliance with the metadata schema)					
PID standardization (e.g. checking, cleaning, or standardizing identifiers such as DOIs or ORCID iDs)					
Metadata harmonization (e.g. standardizing names, formats, or text fields)					
Deduplication					
Error correction					
Rights / licensing & access checks					
Metadata enrichment (e.g. adding subject classification)					
Other (please specify)					

2.3. Do the curation / enrichment activities described above apply to all new submissions to your repository? (select one)

- Yes, all activities apply to all submissions
- No, these activities are applied selectively
- Not sure

2.3.1 If activities are applied selectively, please briefly explain (short answer)

2.4 Is a manual approval step by the repository required before records become publicly available? (choose one)

- Yes
- No

2.5 On average, how much time are repository staff spending on curating each individual record? (short answer)
For example: 10 minutes per record

Section 3. Metadata Sources and Population Practices

In this section, please indicate whether your repository obtains metadata from local/internal and/or external sources and, if so, how metadata from these sources is added to repository records. Direct metadata entry by authors or author proxies is covered in Question 2.1 and is therefore not included here.

3.1 Source category (check all that apply)

- ☐ Imported from local, internal sources (e.g. CRIS)
- ☐ Imported from external sources (e.g. OpenAlex, OpenAIRE)

3.2 Local metadata sources used to populate or enhance your metadata records

For each local (internal) source used by your repository, select the method used most often to ingest or add metadata to repository records. Select one option per row. If a source is not used, leave the row blank.

Local source / Collection options	Manual entry	Batch upload (e.g. CSV / XML)	API	Harvesting (e.g. OAI-PMH, OpenAIRE)	Identifier-based imports	Other
Institutional CRIS / researcher information system						
Library catalog						
Graduate school (for theses / dissertations)						
Institutional Publisher						
Other internal source						

If you selected "Other internal source" please indicate which other internal sources are used to populate or enhance your metadata records.

3.2.1 Is there anything you would like to add about the local metadata sources or ingest methods used by your repository? (optional)

For example, if more than one ingest method is used for a source, please identify the source and list any additional methods beyond the one selected in the matrix.

3.3 External metadata sources used to populate or enhance your metadata records
For each external source used by your repository, select the method used most often to ingest or add metadata to repository records. Select one option per row. If a source is not used, leave the row blank.

External source/Ingest options	Manual entry	Batch upload (CSV/XML)	API	Harvesting (OAI-PMH, OpenAIRE)	Identifier based imports	Other
OpenAlex						
Crossref						
DataCite						
ORCID						
National or International authority files						
National open science aggregators / platforms (including national CRIS)						
Publisher platforms						
Scopus						
Web of Science						
Dimensions						
Google Scholar						
Other external sources						

If you selected "Other external sources" please indicate which other external sources are used to populate or enhance your metadata records.

3.3.1 Is there anything you would like to add about the external metadata sources or ingest methods used by your repository? (optional)
For example, if your repository uses more than one ingest method for a source, please identify the source and list any additional methods not selected in the matrix.

3.4 Which metadata source(s) typically require the most curation effort in your repository? (free text)
Please indicate the main source(s) and, if possible, briefly explain why.

Section 4. Metadata Availability and Reuse

In this section, please indicate whether the metadata records in your repository are reusable, and how they can be accessed by others (e.g. through interfaces or downloads).

4.1 Is your repository metadata publicly available? (choose one)

- Yes
- Partially
- No
- Not sure

4.2 Through which interfaces can your metadata be accessed? (check all that apply)

- ☐ OAI-PMH interface
- ☐ Other API (documented endpoint)
- ☐ Manual / bulk downloads (e.g., CSV/XML exports)
- ☐ Public web interface (UI)
- ☐ Other: _____
- ☐ None

4.3 Is your repository metadata made available under a specific license? (choose one)

- Yes
- No
- Not sure

4.3.1 If yes, which license is used? (check all that apply)

- ☐ CC0 / public domain
- ☐ CC BY
- ☐ Other CC license
- ☐ Other: _____

4.3.2 If yes, how are users made aware of the license? (brief answer)

_____

Section 5. Additional input

In this section, we invite you to share additional context on how metadata is used in the repository, as well as any challenges or opportunities related to your ingest and curation practices.

5.1 Primary use case(s) and purposes (free text)
Describe the main ways your institution uses repository metadata and how they support your institution's goals (e.g., discovery, reporting, open access monitoring, compliance, interoperability).

5.2 Is there anything else you would like to add regarding metadata ingest and curation practices? (free text)