May 12, 2025 Version 2

Open Science 24/25 - Protocol V.2

DOI
dx.doi.org/10.17504/protocols.io.4r3l264ppv1y/v2

Alberto Ciarrocca[1], Anna Nicoletti[2], Ahmadreza Nazari[3]
[1]University of Bologna; [2]Unibo; [3]Student

Open Science

Silvio Peroni
University of Bologna

DOI: https://dx.doi.org/10.17504/protocols.io.4r3l264ppv1y/v2

Protocol Citation: Alberto Ciarrocca, Anna Nicoletti, Ahmadreza Nazari 2025. Open Science 24/25 - Protocol. **protocols.io** **https://dx.doi.org/10.17504/protocols.io.4r3l264ppv1y/v2**Version created by **Anna Nicoletti**

Protocol status: In development
We are still developing and optimizing this protocol

Created: May 07, 2025

Last Modified: May 12, 2025

Protocol Integer ID: 217826

Keywords: researchers across various repository, citation impact of unibo, integration of these research object, research object, overlap among these repository, such research object, citation analysis tool, current research information system, various repository, much of such research object, analyzed metadata, repository, research output, open science practices at the university, existing repository, citation dynamic, citation pattern, dissemination of diverse research output, many citation, researcher, coverage of these research output, diverse research output, strengthening open science practice, citation impact, generalist repository, repository usage, open science, protocol purpose this study, research question, unibo personnel, metadata, methodology the project team, university of bologna, dissemination, providing insight, dissemination pattern, unibo, database, project team, integration within unibo, opencitation, research

Abstract

Purpose
This study aims to evaluate the representation and dissemination of diverse research outputs - such as software, databases, exhibitions, audio-visual materials, and others - produced by the University of Bologna's (UNIBO) researchers across various repositories. It seeks to assess the extent of overlap among these repositories, analyze citation dynamics involving these research objects, and determine their integration within UNIBO's [Current Research Information System (IRIS)](https://cris.unibo.it).
The Research Questions (RQs) are formulated as follows:
1. What is the current coverage of these kinds of research objects created by UNIBO personnel in existing repositories?
2. Is there any overlap among these repositories (i.e., research objects deposited in more than one)?
3. How many citations (incoming and outgoing), as in OpenCitations, are these research objects involved in?
4. How much of such research objects are actually mapped in IRIS?

Methodology
The project team has systematically collected and analyzed metadata from selected institutional, disciplinary, and generalist repositories, including [AMS Acta](https://amsacta.unibo.it/), [Software Heritage](https://www.softwareheritage.org/), and [Zenodo](https://zenodo.org/). Relevant data and metadata were extracted using APIs and web scraping techniques. The analysis identified cross-repository depositions to assess overlaps and employ citation analysis tools, such as [OpenCitations](https://opencitations.net/), to examine citation patterns. Additionally, the collected data were cross-referenced with IRIS to evaluate its coverage of these research outputs.

Value
This study shed light on the dissemination patterns and citation impact of UNIBO's diverse research outputs, providing insights to enhance their visibility and accessibility. The findings can inform strategies to optimize repository usage and improve the integration of these research objects within IRIS, ultimately strengthening open science practices at the University of Bologna and promoting a more cohesive approach for maximizing the impact of its research outputs.

Troubleshooting

Data Gathering

1 From Zenodo use the OAI-PMH protocol to harvest metadata in JSON format.

Dataset

Zenodo NAME

https://developers.zenodo.org/#rest-api LINK

Expected result

A JSON file providing the following properties:
'creator', 'date', 'description', 'identifier', 'publisher', 'relation', 'rights', 'title', 'type'

2 From AMSActa use the OAI-PMH protocol to harvest metadata in CSV format.

Dataset

AMSActa NAME

https://doi.org/10.6092/unibo/amsacta/7736 LINK

Expected result

	A	B	C	D	E	F	G
	ITEM_ID	RM_PERSON_ID	PID	ORCID	FIRST_NAME	LAST_NAME	PLACE

CSV

3 From Software Heritage use API to identify and extract metadata related to code/software authored by UNIBO researchers.

Dataset

Software Heritage NAME

https://archive.softwareheritage.org/api/1/ LINK

Expected result

JSON

4 Convert and normalize data

4.1 Write scripts to convert JSON to tabular format (CSV) for SH and Zenodo repository

4.2 Normalize metadata fields across all sources, including: title, authors, affiliations, publication year, DOI, held in (Datasets where the item was retrievedrepository using Pandas or OpenRefine

Expected result

	A	B	C	D	E	F	G
	PROJECT_ID	TITLE	AUTHOR	AFFILIATION	PUBL_DATE	DOI	RETRIEVED FROM

CSV

5 Filter data relevant to the project

5.1 by author affiliation (e.g., “University of Bologna”) where possible

5.2 identify research object types relevant to the project (e.g., software, audio-visual materials).

6 Integrate datasets

6.1 Create a single CSV file combining all normalized and filtered metadata

6.2 Include metadata flags indicating source repository

7 one CSV file with unified metadata from the three repositories as previously normalised.

Expected result

	A	B	C	D	E	F	G
	PROJECT_ID	TITLE	AUTHOR	AFFILIATION	PUBL_YEAR	DOI	RETRIEVED FROM

Data Analysis

8 Assess the repository coverage using simple descriptive statistics (Answers Research Question 1)

8.1 Count the number of research objects per repository

8.2 Determine distribution by object type and publication year

9 Detect overlap across repositories (Answers Research Question 2)

9.1 Identify duplicates or matches using identifiers (e.g., DOI if present, title similarity)

9.2 Double-check matches found using other relevant data (e.g. author, publication date) to reasonably confirm them

9.3 Tag entries present in multiple repositories

10 Compare the results with IRIS metadata (Answers Research Question 4)

10.1 Cross-check the unified dataset against the IRIS bibliographic dump (https://doi.org/10.6092/unibo/amsacta/7736)

10.2 Mark objects that are also registered in IRIS

11 Collect citation data (Answers Research Question 3)

11.1 Use OpenCitations to gather incoming and outgoing citations for entries with a DOI

11.2 Record citation counts for each object

12 A Jupyter-Notebook file with the related CSV files

Expected result

Jupyter-Notebook, CSV

Data Visualisation

13 Visualise key findings

13.1 Bar chart showing number of objects per repository (Research Question 1)

13.2 Venn diagram showing repository overlap (Research Question 2)

13.3 graph showing the citation distribution chart (Research Question 3)

13.4 Bar chart showing the percentage of mapped entries in IRIS (Research Question 4)

14 A Jupyter-Notebook representing the results with visualisations

Expected result

Jupyter-Notebook

Data Publication

15 The software will be deposited in an appropriate data repository and will be accompanied by proper documentation