

Apr 28, 2025

Version 1

Open Science 24/25 V.1

DOI

<https://dx.doi.org/10.17504/protocols.io.4r3l264ppv1y/v1>

Alberto Ciarrocca¹, Anna Nicoletti², Ahmadreza Nazari³

¹University of Bologna; ²Unibo; ³Student

Open Science



Silvio Peroni

University of Bologna

Create & collaborate more with a free account

Edit and publish protocols, collaborate in communities, share insights through comments, and track progress with run records.

Create free account

OPEN  ACCESS



DOI: <https://dx.doi.org/10.17504/protocols.io.4r3l264ppv1y/v1>

Protocol Citation: Alberto Ciarrocca, Anna Nicoletti, Ahmadreza Nazari 2025. Open Science 24/25. **protocols.io**
<https://dx.doi.org/10.17504/protocols.io.4r3l264ppv1y/v1>

License: This is an open access protocol distributed under the terms of the **Creative Commons Attribution License**, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited

Protocol status: In development

We are still developing and optimizing this protocol

Created: April 10, 2025

Last Modified: April 28, 2025

Protocol Integer ID: 126519

Keywords: researchers across various repository, notable overlap among these repository, citation analysis tool, citation impact of unibo, citation analysis, various repository, analyzed metadata, repository, interoperability across repository, citation pattern, citation dynamic, citation impact, current research information system, integration of these research object, research object, substantial portion of these research object, certain identical research object, interdisciplinary research object, software heritage, broader range of interdisciplinary research object, research output, citation network, diverse research output, multiple repository, generalist repository, selected repository, dissemination of diverse research output, open science practices at the university, citation, repository usage, outgoing citation, coverage of the diverse research output, research, strengthening open science practice, researcher, metadata, coverage of these research output, open science, gaps in institutional tracking, frequent r

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/>).

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.

Findings

The analysis reveals that the coverage of the diverse research outputs produced by the personnel of the University of Bologna varies significantly across the selected repositories. AMS Acta primarily hosts institutional outputs, while Software Heritage focuses on software-related contributions, and Zenodo encompasses a broader range of interdisciplinary research objects. There is a notable overlap among these repositories with certain identical research objects being deposited across multiple repositories. Citation analysis indicates that while these research objects generate both incoming and outgoing citations, their citation networks are unevenly distributed, with software and databases receiving more frequent references. Moreover, a substantial portion of these research objects remains unmapped in IRIS, highlighting gaps in institutional tracking and integration. These findings emphasize the need for a more cohesive strategy to enhance visibility, discoverability, and interoperability across repositories.

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.

Data Gathering

- 1 From Zenodo use the OAI-PMH protocol to harvest metadata in XML format.
- 2 From AMSActa use the OAI-PMH protocol to harvest metadata in XML format.
- 3 From Software Heritage use API to identify and extract metadata related to code/software authored by UNIBO researchers.
- 4 Convert and normalize data
 - 4.1 Write scripts to convert XML to tabular format (CSV) for each repository
 - 4.2 Normalize metadata fields across all sources, including: title, authors, affiliations, publication year, DOI, repository using Pandas or OpenRefine
- 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

**Expected result**

CSV

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, title similarity)
 - 9.2 Double-check matches found by title similarity and decide
 - 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