*publications*

MDPI

*Article*

# The Coverage of Non-Traditional Research Outputs in Repositories and Current Research Information Systems: An Exploratory Study at the University of Bologna

**Alberto Ciarrocca, Ivan Heibi, Ahmadreza Nazari, Anna Nicoletti, Martina Pensalfini, Silvio Peroni *, Lucrezia Pograri, Pietro Tisci and Sergei Slinkin**

Department of Classical Philology and Italian Studies, University of Bologna, Via Zamboni 32, 40126 Bologna (BO), Italy; alberto.ciarrocca@studio.unibo.it (A.C.); ivan.heibi2@unibo.it (I.H.); ahmadreza.nazari@studio.unibo.it (A.N.); anna.nicoletti5@studio.unibo.it (A.N.); martina.pensalfini3@unibo.it (M.P.); lucrezia.pograri@studio.unibo.it (L.P.); pietro.tisci@studio.unibo.it (P.T.); sergei.slinkin@studio.unibo.it (S.S.)

* Correspondence: silvio.peroni@unibo.it

**Abstract**

The diversification of research outcomes produced within scholarly communication practices has led to a growing production of non-traditional research outputs (NTROs) such as datasets, software, databases, exhibitions, and multimedia materials, which are often poorly tracked by institutional systems. This study presents an exploratory study about knowledge production at the University of Bologna (UNIBO). This prominent national research institution offers a compelling case study to assess coverage, cross-repository overlap, and citation activity of NTROs across repositories. By harvesting and integrating metadata from the University of Bologna's institutional CRIS system, the institutional repository AMS Acta, the general-purpose repository Zenodo, the disciplinary archive Software Heritage, and OpenCitations to gather citation information, we analyse the availability of UNIBO-affiliated NTROs and show that, while the UNIBO CRIS platform (i.e., IRIS) remains the primary registry for information on NTROs, a substantial number of them are hosted exclusively in external repositories. These findings highlight structural gaps in tracking NTROs in UNIBO IRIS and underline the need for improved interoperability and coordinated Open Science strategies and policies, at least at the local level, to ensure recognition of diverse research outputs.

**Keywords:** current research information system; non-traditional publications; open science; open citations; scientometrics

Check for updates

Academic Editor: Andrew Kirby

Received: 10 January 2026
Revised: 7 April 2026
Accepted: 10 April 2026
Published: 15 April 2026

**Copyright:** © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

## 1. Introduction

The growing interdisciplinary nature of modern research, coupled with the diversified application of computational approaches to several research disciplines, including the Social Sciences and Humanities (SSH), has led to a significant increase in the creation of new and diverse research outputs, which include 3D models, virtual exhibitions, datasets, software, audiovisual materials, and grey literature (Okamura, 2019). Recently, pushing researchers to publish such non-traditional research outputs has been supported by several initiatives. Some of those, such as (UNESCO, 2021), support the view of research as an endeavour that involves several different expertises and, thus, competencies working collaboratively to reach a common research goal. As a direct consequence, the research

community and institutions advocating Open Science push for the adoption of inclusive and transparent processes that support the creation and, thus, publication of a wide range of research objects, from datasets to software, from documentation to other digital artefacts. Being a foundational part of the research, all these research objects should be treated as first-class publications in a broad sense.

Another aspect, for which the appropriate publication of all research outputs resulting from a research endeavour is crucial, concerns research assessment exercises. For instance, the agreement and related committments published by the Coalition for Advancing Research Assessment (2022), which have been signed by 953 institutions so far according to the CoARA website at https://www.coara.org/agreement/signatories/ (accessed on 26 March 2026), provides an important focus on the importance of including, while assessing research, diverse outputs beyond journal publications, thus including other kinds of published research object in the assessment process.

Given these premises, the tracking of such research outputs within research institutions is, year by year, becoming an important activity to address, due to the consequences it may entail for the management and administration of a research-performing organisation. However, for an institution, maintaining an overall view of all research objects created by its researchers can be challenging, since there is often no systematic approach to monitor all materials produced and published by its researchers.

Universities usually rely on Current Research Information Systems (CRIS) to track research outputs and manage associated metadata. CRIS platforms manage and exchange metadata for research activities, typically including information on researchers, institutions, projects, and outputs. Over time, CRIS platforms have evolved to support the entire research lifecycle, serving both as public showcases of institutional research and as internal tools for strategic planning (De Castro & Puuska, 2023). Nevertheless, CRIS platforms often fail to capture the full spectrum of research outputs published by researchers and made available in appropriate repositories (e.g., Zenodo), resulting in gaps in institutional repositories and undermining the visibility of certain studies (Azeroual & Schöpfel, 2019).

The objective of our research is to conduct an *exploratory study* about knowledge production at the University of Bologna (UNIBO). This prominent national research institution offers a compelling case study to assess coverage, cross-repository overlap, and citation activity of *non-traditional research products* (NTROs)—such as datasets, software, and other research materials—across repositories. In particular, within this context, the general research questions (RQs) we want to answer are:

1. What is the current coverage of NTROs created by researchers that are available in existing repositories?
    a. Is there any overlap among these repositories (i.e., research objects deposited in more than one)?
    b. How many citations (incoming and outgoing) are these research objects involved in?
2. To what extent are these research outputs mapped and recorded in CRIS platforms?

The choice of UNIBO is particularly appropriate for supporting an exploratory study since, while being the world's oldest university still in activity, it is one of the biggest universities in Italy and includes 31 distinct Departments distributed across five primary scientific areas: Humanities (Classical Philology and Italian Studies; Cultural Heritage; Education Studies; History and Cultures; Interpreting and Translation; Modern Languages, Literatures and Culture; Philosophy; Psychology; The Arts), Medicine (Biomedical and Neuromotor Sciences; Life Quality Studies; Medical and Surgical Sciences; Veterinary Medical Sciences), Science (Biological, Geological, and Environmental Sciences; Chemistry;

Industrial Chemistry; Mathematics; Pharmacy and Biotechnology; Physics and Astronomy), Social Studies (Economics; Legal Studies; Management; Political and Social Sciences; Sociology and Business Law; Statistical Sciences), and Technologies (Agricultural and Food Sciences; Architecture; Civil, Chemical, Environmental, and Materials Engineering; Computer Science and Engineering; Electrical, Electronic, and Information Engineering; Industrial Engineering).

In addition, in recent years, UNIBO has espoused Open Science as one of the guiding principles of its latest Strategic Plan 2022–2027 (Alma Mater Studiorum—Università di Bologna, 2022), also confirmed by the publication of important policies related to research publication and data (Alma Mater Studiorum—Università di Bologna, 2018) and research data management (Alma Mater Studiorum—Università di Bologna, 2023), the hiring of a specific unit of data stewards to support the application of Open Science principles, and the signing of relevant initiatives such as the Coalition for Advancing Research Assessment (2022), related to reforming research assessment practices, and the Barcelona Declaration on Open Research Information (2024), referred to the use and sharing of open metadata about research to assess researchers and institutions, to support strategic decision making, and to find relevant research outputs. All these activities have been key to disseminate correct practices of performing research, which have also pushed for considering NTROs as first-class publication entities that, as such, should be properly tracked within UNIBO's CRIS platform.

Within this broad disciplinary and policy-oriented scope, which enables us to have a balanced representation of all the primary scholarly disciplines, from STEM to SSH, and the Open Science policies that enforce a new vision of how to practice research within UNIBO, in this article we want to measure the publication of NTROs by UNIBO's researchers, comparing their availability in different repositories and their inclusion in UNIBO's CRIS platform, i.e., IRIS (Bollini et al., 2016), which is a CRIS solution used in the Italian context that enables researchers to contribute metadata about their scientific work directly. It is worth mentioning that populating IRIS is a manual activity performed by researchers, who often focus primarily on adding traditional publications (e.g., journal articles, conference papers, books) due to the incentives the Italian University system introduces for such kinds of publications (e.g., to be used in research assessment exercises), thereby avoiding often adding other research products such as datasets, databases, software which are not appropriately pushed by national policies.

To answer the RQs above within our exploratory study, we analyse UNIBO's contributions across three exemplars of different types of repositories: the *institutional* repository AMS Acta (https://amsacta.unibo.it/, accessed on 26 March 2026) (Vignocchi et al., 2006), the *disciplinary* repository Software Heritage (https://www.softwareheritage.org/, accessed on 26 March 2026) (Di Cosmo & Zacchiroli, 2023), and the *general-purpose* repository Zenodo (https://zenodo.org/, accessed on 26 March 2026) (European Organization For Nuclear Research & OpenAIRE, 2013). We introduce a further layer of analysis by examining the citation counts of these outputs using data from OpenCitations (https://opencitations.net, accessed on 26 March 2026) (Peroni & Shotton, 2020), a community-guided Open Science infrastructure dedicated to publishing open bibliographic metadata and citation data. Finally, we assess the extent to which the scholarly artefacts identified in these repositories are also included in UNIBO IRIS.

The rest of the article is organised as follows. Section 2 details the methodology used to gather and analyse the data necessary to answer the RQs within the context of the University of Bologna, with particular attention to the tools and approaches employed to ensure transparency and reproducibility of the workflow, which can also be potentially customised and adopted in future studies with data from different universities. Section 3

presents the results of the data analysis, supported by infographics and visualisations to provide a clear overview of the outcomes. Lastly, Section 4 discusses the interpretation of the findings, including their limitations and impact, and concludes the article by outlining future work.

## 2. Materials and Methods

This section outlines the methodology adopted for our exploratory study, with particular attention to the tools and approaches used to ensure transparency and reproducibility. All datasets generated are openly accessible on Zenodo (Nazari et al., 2025), together with the research protocol (Ciarrocca et al., 2025a), which provides a comprehensive overview of the entire workflow summarised in Figure 1, and the data management plan (Pograri & Tisci, 2025). The software developed for data extraction, processing, and analysis is available in the project's GitHub repository (https://github.com/open-sci/2024-2025/, accessed on 26 March 2026) and mirrored on Zenodo (Ciarrocca et al., 2025b).

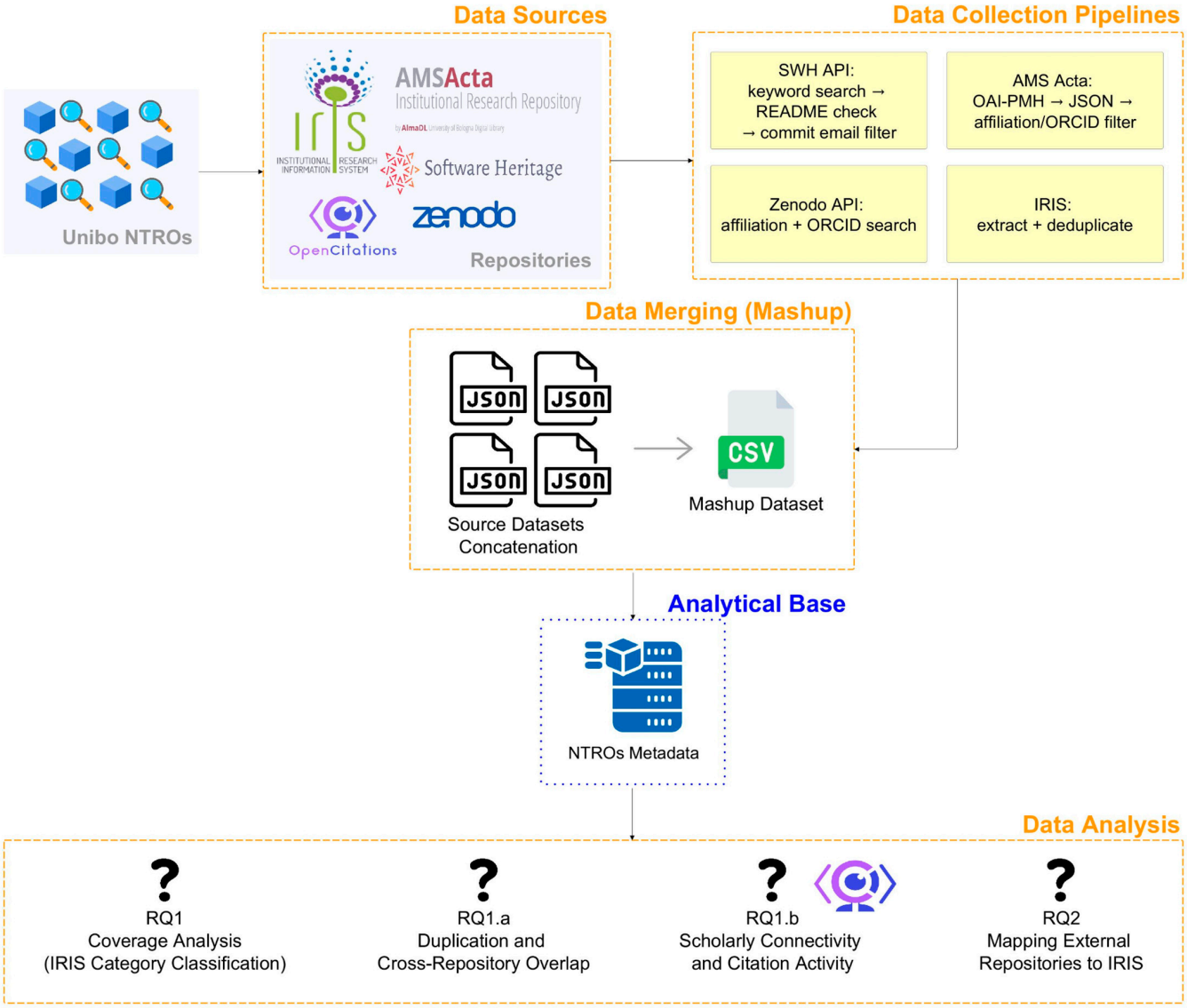



**Figure 1.** Methodological workflow for harvesting, integrating, and analysing UNIBO non-traditional research outputs (NTROs).

### *2.1. Data Sources Selection*

The subjects of our exploratory study are non-traditional research outputs produced by personnel affiliated with UNIBO, which include, among others, software, databases, exhibitions, patents, multimedia content, prototypes, datasets, and educational materials. To investigate where such outputs are stored, described, and disseminated, we employed a methodology to collect and analyse associated metadata from a selection of repository types, which may be used by UNIBO's researchers depending on the intended audience (i.e., general or specialist). The repository types considered are those described in the UNIBO's guidelines for research data management (Alma Mater Studiorum—Università di Bologna & ARIC—Research Division, Research Services and Division Projects Coordination Unit, Data Stewards, 2024), which classifies them into three distinct categories:

- *Disciplinary* repositories, which use metadata schemas specifically designed for either a certain academic discipline or, for the scope of the present study, a specific kind of research object used consistently across multiple disciplines/domains. This kind of repository can offer greater visibility and make it easier to share data within the relevant scientific community.
- *Institutional* repositories, which are made available to the members of an academic or research institution and usually provide validation and support services to ensure the quality of the datasets deposited.
- *General-purpose* repositories, which gather different kinds of data and materials from different disciplines and research contexts. They provide a solid platform for data preservation, visibility and accessibility.

In the context of this exploratory study, we have considered:

(a) An institutional repository, i.e., AMS Acta—the institutional repository of UNIBO designed to ensure long-term open access to scientific outputs.
(b) A disciplinary repository, i.e., Software Heritage—the world's largest public archive of software and associated development history, preserving (as of 26 March 2026) over 28 billion unique source files from more than 430 million projects.
(c) A general-purpose repository, i.e., Zenodo—an interdisciplinary platform developed by CERN that provides secure long-term preservation and open metadata access (European Organization For Nuclear Research & OpenAIRE, 2013).

The information from the repositories mentioned above is compared with the data in the UNIBO installation of IRIS (https://cris.unibo.it/, accessed on 26 March 2026), the UNIBO's CRIS platform used by the university to track research production, and OpenCitations, Open Science and community-guided open scholarly infrastructure dedicated to the publication of open bibliographic metadata and citation data, which is used to measure the number of citation links involving all the NTROs identified in the present study. Two distinct aspects justify the choice of OpenCitations. On the one hand, OpenCitations is an Open Scholarly Infrastructure that provides free, reusable citation data and related services for any purpose, enabling the reproducibility of all analyses conducted on this data. On the other hand, recently Andreose et al. (2026) have shown how the coverage OpenCitations provides of publication records and related citation links is comparable, at least in the context of UNIBO, to what is currently observed in well-known proprietary systems such as Web of Science and Scopus. In addition, the OpenCitations Index (Heibi et al., 2024), i.e., the OpenCitations collection storing citation links between bibliographic resources, provides citation data by mashing up several different sources (as of 26 March 2026, they include Crossref, DataCite, OpenAIRE, the NIH Open Citation Collection, and the Japan Link Centre), positioning it as one of the most complete open sources available for citation data.

The data collection procedure relied on data extraction via publicly documented, accessible APIs provided by the services/sources introduced in the previous step of the methodology. This approach facilitated the systematic harvesting of relevant information. In particular, dedicated API endpoints were used to access Zenodo, Software Heritage, and OpenCitations. In contrast, AMS Acta was queried via its OAI-PMH interface, and at the time of the work, IRIS data were downloaded from the most recent available dump (Amurri et al., 2025). A detailed description of the data collection processes for each data source is presented in the following subsections.

#### 2.1.1. Software Heritage

To identify software repositories affiliated with the University of Bologna within the SWH archive, we implemented a data extraction pipeline in Python (Ciarrocca et al., 2025b), designed to programmatically search, retrieve, and filter relevant entries from software artefacts. The pipeline uses the public API of Software Heritage (https://archive.softwareheritage.org/api/1, accessed on 26 March 2026) and relies on open-source libraries for file and data handling, as well as for managing API throttling and request delays. Furthermore, failed requests and errors are automatically detected and handled during execution.

The primary goal of the extraction was to isolate the source code repositories likely associated with UNIBO. The current identification approach follows these steps. First, it performs a keyword-based full-text search via Software Heritage API using a predefined set of keywords (i.e., "unibo", "alma mater", "alma mater studiorum", "university of bologna", "almamaterstudiorum", "università di bologna", "universita di bologna"). Then, for each candidate source code repository returned, we check its most recent snapshot in Software Heritage to inspect its README file (if present), convert it to lowercase, and look for the presence of one or more of the keywords used in the initial search. In addition, we use the repository's full commit history to identify the presence of UNIBO email addresses (domains "@unibo.it" or "@studio.unibo.it"). We list all repositories that satisfy at least one of the UNIBO affiliation criteria.

To assess the quality of the extraction performed using the aforementioned heuristics, we selected a sample of 100 random records returned by the process. From a manual check of this sample, all these records were indeed managed by UNIBO personnel, with equal distribution between UNIBO researchers and UNIBO students.

#### 2.1.2. Zenodo

The data extraction from Zenodo was performed through its REST API service, documented at https://developers.zenodo.org/#rest-api (accessed on 26 March 2026), in particular using the '*records*' operation. The process was implemented using a two-step approach to capture UNIBO's research outputs.

The first step involved querying the records using the specified affiliation strings in the "creators" and "contributors" fields in the Zenodo metadata, by searching for the keywords: "University of Bologna", "Università di Bologna", "UNIBO", and "Alma Mater Studiorum". These data were then complemented with Zenodo records that included ORCIDs in the "creators" and "contributors" fields of researchers affiliated with UNIBO extracted from UNIBO IRIS. The metadata for all retrieved Zenodo data were stored in appropriate JSON files.

To assess the quality of the extraction performed using the aforementioned heuristics, we selected a sample of 100 random records returned by the process. Of these, 75 were matched through explicit affiliation with the University of Bologna and 25 through the ORCIDs of UNIBO researchers (that are specified in the UNIBO IRIS dump). While the

records selected via ORCIDs are considered correct, we have manually checked the string match in the affiliations and found no errors.

#### 2.1.3. AMS Acta

For AMS Acta, bibliographic metadata was retrieved via its OAI-PMH API endpoint at https://amsacta.unibo.it/cgi/oai2 (accessed on 26 March 2026), using the Python library Sickle (Loesch, 2020), a lightweight client designed to interact with OAI-PMH services and facilitate harvesting of metadata records by handling the OAI protocol natively and supporting automatic pagination. Our process began by querying the AMS Acta endpoint to retrieve the ePrint IDs of AMS Acta records. Then, for each harvested ePrint ID, a direct call was made to the JSON export URL provided by the AMS Acta OAI-PMH API endpoint, which returned structured bibliographic metadata in JSON format, including affiliation data and ORCID identifiers when available.

It is worth noting that, while serving as UNIBO's institutional repository for data, AMS Acta also hosts data not provided by UNIBO researchers but by associate partners from external institutions. Thus, to isolate records associated with UNIBO, we first checked whether the selected records included UNIBO information in the affiliation metadata and, if so, kept the record. Otherwise, we checked each author's ORCID to determine whether it was included in the ORCID list extracted from UNIBO IRIS; if so, the record was retained; otherwise, it was discarded. For multi-authored publications, having at least one matching author (by affiliation or ORCID) was sufficient for inclusion in the final filtered dataset, which is stored in JSON format.

To assess the quality of the extraction performed using the aforementioned heuristics, we selected a sample of 100 random records returned by the process. Of these, 75 records were correctly matched using the ORCIDs of UNIBO researchers (as specified in the UNIBO IRIS dump). When an ORCID was not available, we relied on the affiliation field and checked whether it contained any reference to the University of Bologna. In particular, we have checked the remaining 25 records, and all of them belonged correctly to the University of Bologna.

### *2.2. Data Merging*

The merging process consolidated data from different sources into a single, homogeneous file to simplify analysis and answer the RQs. In addition to data from Zenodo, AMS Acta, and Software Heritage, the UNIBO IRIS dump has been considered in this phase to consolidate all data in a single location for comparison.

The final dataset, organised in tabular format, was built by concatenating the records from the four source datasets, extending the columns of the final file according to the fields included in the information downloaded from the sources, where some of the columns were aligned based on their shared semantics in case the same concept was reported with different naming conventions. In addition, for each record in the final merged file, we added information on the source from which it was extracted. The final table containing all the metadata had the following columns: "title", "id", "doi", "creators", "orcid", "date", "description", "resource_type", "url", "type", "rights", "publisher", "relation", "communities", "swh_id", "keywords", "src_repo", "issn", "pmid".

### *2.3. Data Analysis*

The data analysis is organised in four steps. All Python scripts developed to implement this methodology, along with the associated data, are publicly available online—as indicated in the Data Availability Statement at the end.

#### 2.3.1. Coverage Analysis

We have considered the 15 publication types included in UNIBO IRIS, which are defined as referring to non-traditional research outputs. They are presented in Table 1, along with their English translations.

**Table 1.** Mapping of IRIS publication types for non-traditional research outputs, in Italian and English.

| IRIS Publication Type (Original—Italian) | IRIS Publication Type (English Translation) |
|---|---|
| 7.01 Carta tematica e geografica | 7.01 Thematic and geographic map |
| 7.02 Carta geologica | 7.02 Geological map |
| 7.03 Prodotto dell'ingegneria civile e dell'architettura | 7.03 Civil engineering and architecture product |
| 7.04 Software | 7.04 Software |
| 7.05 Banche dati | 7.05 Databases |
| 7.06 Composizione musicale | 7.06 Musical composition |
| 7.07 Disegno | 7.07 Drawing |
| 7.08 Design | 7.08 Design |
| 7.09 Performance | 7.09 Performance |
| 7.10 Prodotto artistico e spettacolare: Manufatto | 7.10 Artistic and performance product: Artefact |
| 7.11 Prodotto artistico e spettacolare: Prototipo d'arte e relativi progetti | 7.11 Artistic and performance product: Art prototype and related projects |
| 7.12 Attività espositiva: Mostra o Esposizione | 7.12 Exhibition activity: Exhibition or Show |
| 7.13 Rapporto tecnico | 7.13 Technical report |
| 7.14 Audiovisivi | 7.14 Audiovisual products |
| 7.15 Test psicologici | 7.15 Psychological tests |

A script iterates over each row of the mashup dataset described in Section 2.2 and, for each row, it considers the *type* and *resource_type* fields and tries to match them (via text normalisation and regular expressions) with at least one of the UNIBO IRIS publication types considered. If the script finds a match, a new column, *iris_cat*, is added to the row; otherwise, the row is discarded. For instance, all Zenodo resources labelled as "publication", "presentation", "poster", "lesson", and "other" are excluded because they do not match any of the types listed in Table 1. Similarly, a similar assumption has also been made with AMS Acta records: those labelled as "monograph", "book section" (e.g., book chapters), "preprint", "article", and "conference item" (e.g., proceeding papers and posters) are excluded from the analysis. Thus, the resulting dataset, named IRIS-MATCH, retains only the labelled rows and comprises 20 columns: the original metadata introduced in Section 2.2, plus the additional *iris_cat* field.

#### 2.3.2. Duplication and Cross-Repository Overlap

Starting from the IRIS-MATCH dataset, we apply two different strategies to detect duplicates among the list of entities:

- DOI match: As the DOI is the most authoritative, cross-platform identifier for scholarly objects, we have adopted it as our primary strategy for detecting duplicates. In particular, if two rows expose the same DOI, we mark them as representing the same entity;
- Title match: Since many UNIBO IRIS and AMS Acta records lack a DOI and Zenodo versions may carry a different DOI (preprint vs. published record), an exact title match, with pre-normalisation of the text (e.g., removing punctuation and converting to lowercase), can help capture additional duplicate entities.

Other identifiers, such as PMID, were initially considered for deduplication purposes. However, since none appeared in more than one of the sources considered, they did not yield any additional match procedure.

#### 2.3.3. Citation Coverage

We retrieved citation data and analysed these relationships using the OpenCitations Index API (https://api.opencitations.net/index/v2/, accessed on 26 March 2026). In particular, starting from the IRIS-MATCH table, we collected citation data for each entity associated with a DOI. For each DOI, two API operations were used:

- `/references/{doi}`—to retrieve outgoing citations (other works cited by the object);
- `/citations/{doi}`—to retrieve incoming citations (other works citing the object).

Both incoming and outgoing citations are accompanied by the corresponding Open Citations Identifiers (OCIs) (Peroni & Shotton, 2019), unique identifiers for citations issued by OpenCitations.

#### 2.3.4. Mapping External Repositories to IRIS

To quantify the degree of overlap with external repositories, we calculated the percentage of Zenodo and AMS Acta objects already present in UNIBO IRIS. To do so, we built three separate datasets, one for each repository: IRIS, Zenodo, and AMS Acta. In the first step of our data analysis, Software Heritage was excluded because it had no overlapping entries. The matching strategy followed the previously described approach, prioritising DOI-based matching and falling back to title-based matching for records without a DOI.

## 3. Results

By querying the related APIs (between 25 and 26 June 2025) introduced in Section 2, we obtained several matches from all three sources considered, i.e., Software Heritage, Zenodo, and AMS Acta. In Software Heritage, the initial keyword-based extraction yielded approximately 4000 candidate entries, which were subsequently filtered to approximately 1000 repositories by verifying institutional markers in README.md files and contributors' email addresses.

Data extraction from Zenodo yielded 5434 records: 3949 from the initial filtering by affiliation and 1485 from queries created using ORCIDs extracted from IRIS. In particular, we collected records with different types specified. Half of the Zenodo records gathered were labelled as *publication*, while the second most common entity type was *dataset* with 1273 entries. Software was a relatively rare research product, accounting for only 352 records out of the 5434. The remaining results are classified into the following categories: physical object, workflow, video, lesson, event, model, image, presentation, poster, and other.

Regarding AMS Acta, the first extraction using the JSON exporter, after obtaining the ePrint IDs, yielded approximately 5000 records, which were filtered, as explained in Section 2.3, to approximately 1000 records from researchers affiliated with the University of Bologna. In Figure 2, we show the number of distinct publication types in AMS Acta, totalling 383 records across datasets and software.

Using these data, we analysed the coverage of UNIBO personnel's research outcomes across three sources (RQ1). In this analysis, we included UNIBO IRIS data to understand the percentage of research objects that were not already mapped in the UNIBO CRIS system. The numbers of this analysis are listed in Table 2.

We have also measured the counts for each publication type considered in the study to assess their representativeness across all sources analysed (including UNIBO IRIS), as summarised in Table 3. IRIS UNIBO is the largest host in terms of coverage, particularly for artistic/performance artefacts. At the same time, Zenodo extends the UNIBO CRIS with additional records for datasets (loosely mapped to category *7.05 Databases*) and software. Software Heritage contributes only to software, as expected, but at a scale comparable to Zenodo.

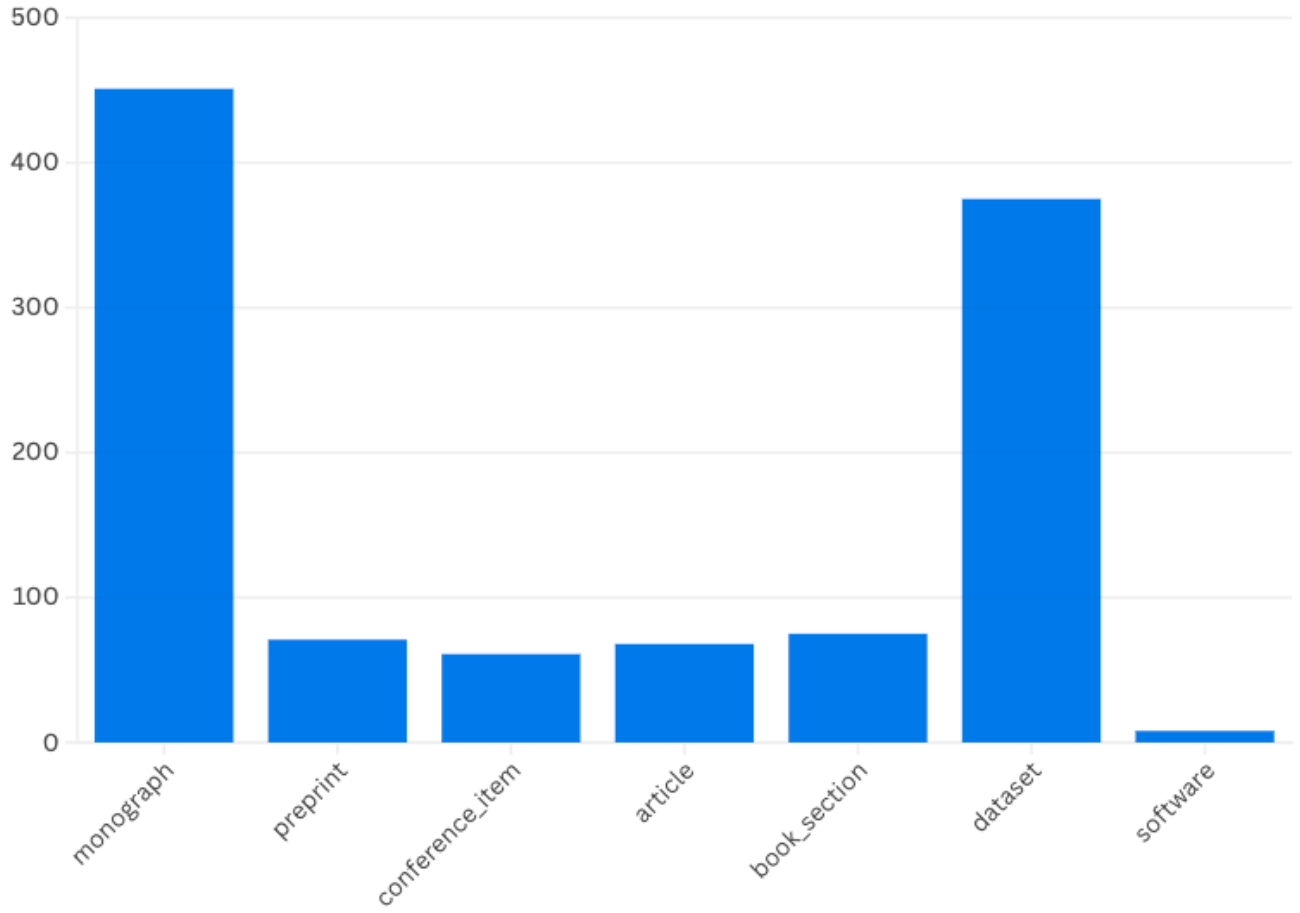



**Figure 2.** Number of records in AMS Acta by research object type.

**Table 2.** The number of non-traditional research outputs authored by UNIBO personnel found in UNIBO IRIS, Zenodo, Software Heritage, and AMS Acta. The last row, in italics, details the overall amount of publications from all the sources.

| Source | Non-Traditional Publications by UNIBO Personnel | Percentage of the Total |
|---|---|---|
| UNIBO IRIS | 4711 | 54% |
| Zenodo | 2522 | 29% |
| Software Heritage | 1014 | 12% |
| AMS Acta | 452 | 5% |
| *Total* | *8699* | *100%* |

In terms of cross-repository overlap, we applied two complementary duplication detection strategies on the IRIS-MATCH subset: (i) DOI matching (case-insensitive) and (ii) exact title matching after text normalisation (lowercasing and punctuation removal). At the record level, these strategies yielded 285 rows (3.3% of the 8699 subset) whose DOI or normalised title occurs in at least two repositories. However, since multiple rows can represent the same underlying research object, we additionally computed an object-level overlap by collapsing duplicate rows into unique objects using a deterministic key (DOI when available, otherwise normalised title). Under these assumptions, 89 unique objects are shared between at least two repositories, as shown in Table 4. In particular, we observe overlaps in AMS Acta, Zenodo and UNIBO IRIS, while no overlaps involving Software Heritage have been encountered.

Using the OpenCitations Index API, we computed the number of citations that involve any of the non-traditional research outputs we identified in our study (RQ1.b). The identified citations included all sources except Software Heritage (since we did not identify

any DOI in it, which is required to obtain citations from OpenCitations), as shown in Figure 3. Although Zenodo listed around 29% of the total number of identified non-traditional research outputs, it accounted for only a portion of the total citations. At the same time, the majority came from UNIBO IRIS records. In addition, UNIBO IRIS showed a predominance of outgoing citations due to the presence of more textual publications (i.e., 7.13 Technical report) that may contain several references to other works.

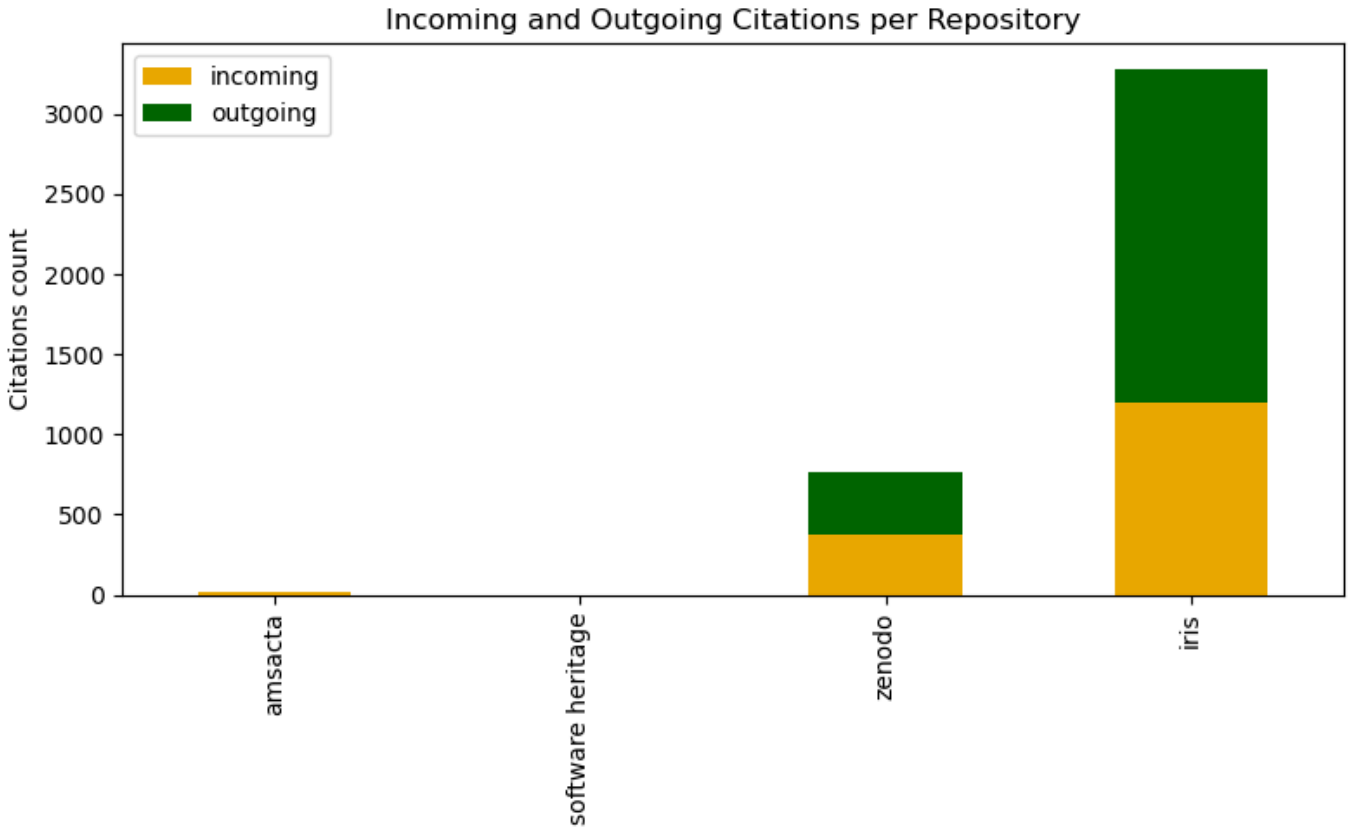



**Figure 3.** The incoming and outgoing citations of the non-traditional publications from each source.

**Table 3.** The number of items found in each of the four sources considered—i.e., IRIS, AMS Acta, Zenodo, and Software Heritage—and the specific Population of subsets, according to IRIS categories (shown by their code as defined in Table 1).

| Code | IRIS | AMS Acta | Zenodo | Software Heritage |
|---|---|---|---|---|
| 7.01 | 120 (2.55%) | 0 (0.00%) | 0 (0.00%) | 0 (0.00%) |
| 7.02 | 75 (1.59%) | 0 (0.00%) | 0 (0.00%) | 0 (0.00%) |
| 7.03 | 225 (4.78%) | 0 (0.00%) | 27 (1.07%) | 0 (0.00%) |
| 7.04 | 589 (12.50%) | 8 (1.77%) | 375 (14.87%) | 1014 (100.00%) |
| 7.05 | 257 (5.46%) | 375 (82.96%) | 1275 (50.56%) | 0 (0.00%) |
| 7.06 | 148 (3.14%) | 0 (0.00%) | 21 (0.83%) | 0 (0.00%) |
| 7.07 | 16 (0.34%) | 0 (0.00%) | 371 (14.71%) | 0 (0.00%) |
| 7.08 | 10 (0.21%) | 0 (0.00%) | 0 (0.00%) | 0 (0.00%) |
| 7.09 | 185 (3.93%) | 0 (0.00%) | 4 (0.16%) | 0 (0.00%) |
| 7.10 | 156 (3.31%) | 0 (0.00%) | 0 (0.00%) | 0 (0.00%) |
| 7.11 | 9 (0.19%) | 0 (0.00%) | 0 (0.00%) | 0 (0.00%) |
| 7.12 | 1996 (42.37%) | 0 (0.00%) | 0 (0.00%) | 0 (0.00%) |
| 7.13 | 838 (17.79%) | 69 (15.27%) | 449 (17.80%) | 0 (0.00%) |
| 7.14 | 0 (0.00%) | 0 (0.00%) | 0 (0.00%) | 0 (0.00%) |
| 7.15 | 87 (1.85%) | 0 (0.00%) | 0 (0.00%) | 0 (0.00%) |

**Table 4.** The number of objects shared between each repository.

| Source Pair | Overlapping Objects | 7.04 Software | 7.05 Databases | 7.13 Technical Report |
|---|---|---|---|---|
| AMS Acta—UNIBO IRIS | 45 | 0 | 19 | 26 |
| UNIBO IRIS—Zenodo | 40 | 1 | 20 | 19 |
| AMS Acta—Zenodo | 4 | 0 | 4 | 0 |

## 4. Discussions

This study highlights the current landscape of dissemination for NTROs at the University of Bologna. Our findings demonstrate that while UNIBO IRIS remains the primary system listing the majority of these research objects (54% of the overall entries found), even if several NTROs that are not listed in UNIBO IRIS have been uploaded to generalist platforms such as Zenodo (29%) and disciplinary archives such as Software Heritage (12%), thus significantly extending the coverage, particularly for datasets and software. AMS Acta, though smaller in scale, still contributes to the preservation of institutional knowledge, particularly technical reports and data records. UNIBO IRIS is also the main responsible for tracking a particular NTRO type, i.e., *7.12 Exhibition activity*, which mirrors a specific UNIBO's behaviour held within various (predominantly Humanities) departments.

The overlap analysis we conducted shows that only 3.3% of NTROs are duplicated across multiple repositories, suggesting that many remain siloed in specific repositories. This observation may reflect either platform specialisation or a broader lack of coordinated deposition strategies that may be due to a lack of standardised interoperability procedures implemented across different repositories. Notably, Software Heritage appears entirely non-overlapping with other repositories, underscoring both its specificity and the challenges of interoperability. However, they have recently partnered with Zenodo to automatically upload open-source software published there directly to the Software Heritage database (see https://www.softwareheritage.org/2024/11/13/software-heritage-zenodo-integration/, accessed on 26 March 2026). Thus, the situation may change in the near future.

Citation data from OpenCitations further confirms disparities in visibility and scholarly engagement compared with traditional publications, such as journal articles. IRIS records dominate citation activity with over 1200 incoming and 2000 outgoing citations, followed by Zenodo. In contrast, Software Heritage entries, despite their volume and potential value, recorded no citation activity in OpenCitations, probably due to the marginal role that software still plays in formal citation practices (i.e., they are not often formally cited by including them within the bibliographic reference list of research papers), and for the fact that the primary persistent identifier used by Software Heritage for software is the SoftWare Hash IDentifiers (SWHID, https://www.swhid.org/, accessed on 26 March 2026), which is currently (as of 26 March 2026) not tracked in OpenCitations, making the retrieving of citations to a given software not possible using its SWHID. Overall, the citation analysis we conducted is constrained by the fact that a large portion of the entities in OpenCitations are identified primarily by their DOIs, and that OpenCitations lists entities only if they are involved in at least one citation. This situation excludes many NTROs and, as a result, the low citation counts retrieved may reflect infrastructural limitations, as well as the omission of specific scholarly citation practices for certain kinds of NTROs.

These findings carry important implications for institutional research evaluation and, more generally, Open Science policies. First, the results show a gap between the NTROs listed in UNIBO IRIS and those that are retrieved from the other three repositories. Indeed, while the overlap between UNIBO IRIS and the repositories is low, a relevant part of the NTROs are not mapped in IRIS. Such a missing compromises discoverability, limits the recognition of diverse scholarly contributions, and potentially may distort research

assessment frameworks when only NTROs described in UNIBO IRIS are considered, which is usually the case in local and national research assessment exercises. However, the UNIBO IRIS installation has recently been extended to track data links to existing traditional publications. It means that, even if specific kinds of research products (such as datasets) are not currently tracked in UNIBO IRIS, it is possible to interlink them with other traditional publications authored by UNIBO personnel using datasets' persistent identifiers, such as DOIs, even if such datasets have been deposited in external repositories (e.g., Zenodo). However, this possibility does not allow a researcher to create a primary record about a dataset in UNIBO IRIS if it is not associated with a traditional publication.

A similar analysis has been conducted by several other institutions in the past, which have increasingly focused on how research outputs are published and stored, particularly within the framework of Open Science. For example, Hurrell (2023) led a study at the University of Calgary examining the digital institutional repositories of 77 universities that are signatories of the Declaration on Research Assessment (DORA) (Cagan, 2013). The study found that even institutions committed to reforming research assessment struggle to prioritise the collection of non-traditional outputs such as datasets, software, and creative works in their institutional repositories. In particular, while 96% of repositories surveyed in the UK included some non-traditional content, these outputs represented only a small fraction of the total, with institutional focus remaining heavily skewed toward traditional publications. This situation suggests that, despite technical readiness and policy frameworks promoting inclusivity, a more active commitment is needed to elevate the visibility and integration of diverse scholarly outputs within institutional ecosystems.

In parallel with these policy-oriented studies, a significant body of work has also emerged around methodologies for cross-repository analysis, which is essential for understanding the broader landscape of research dissemination. Cross-repository matching typically involves specifying attributes (such as metadata fields like titles, authors, identifiers, or content descriptors) to identify duplicate records and overlapping content across repositories. Recent advances include the use of binary classification models based on repository-level embeddings to predict conceptual similarity (Rokon et al., 2021), as well as multi-level analysis techniques that incorporate aspects such as source code, documentation, licensing, and dependency management (Zhang et al., 2024).

### *4.1. Temporal Analysis*

To understand the variability of the publication of NTROs in time in the three different repositories considered in the study, and also their presence in UNIBO IRIS, we have run a temporal analysis from 2004, the first year observed in IRIS that includes a relevant number of publication from UNIBO researchers, to 2025, considering that this last year is only partially covered considering when we have run the ingestion process of the data considered in our study (end of June 2025). The outcomes of this analysis are summarised in Figure 4.

From a purely institutional perspective, UNIBO researchers initially recorded a limited number of NTRO types in IRIS from 2004 to 2010, whereas, from 2011 onward, the variability of these types increased. There is a clear difference in the number of items between two specific time windows, i.e., 2004–2013 and 2014–2025, with a significant reduction in the number of recorded *exhibitions* (7.12) in IRIS. Analysing the data, it is not clear what has led to this decrease in time, and further investigation is needed to explore the rationale for this trend.

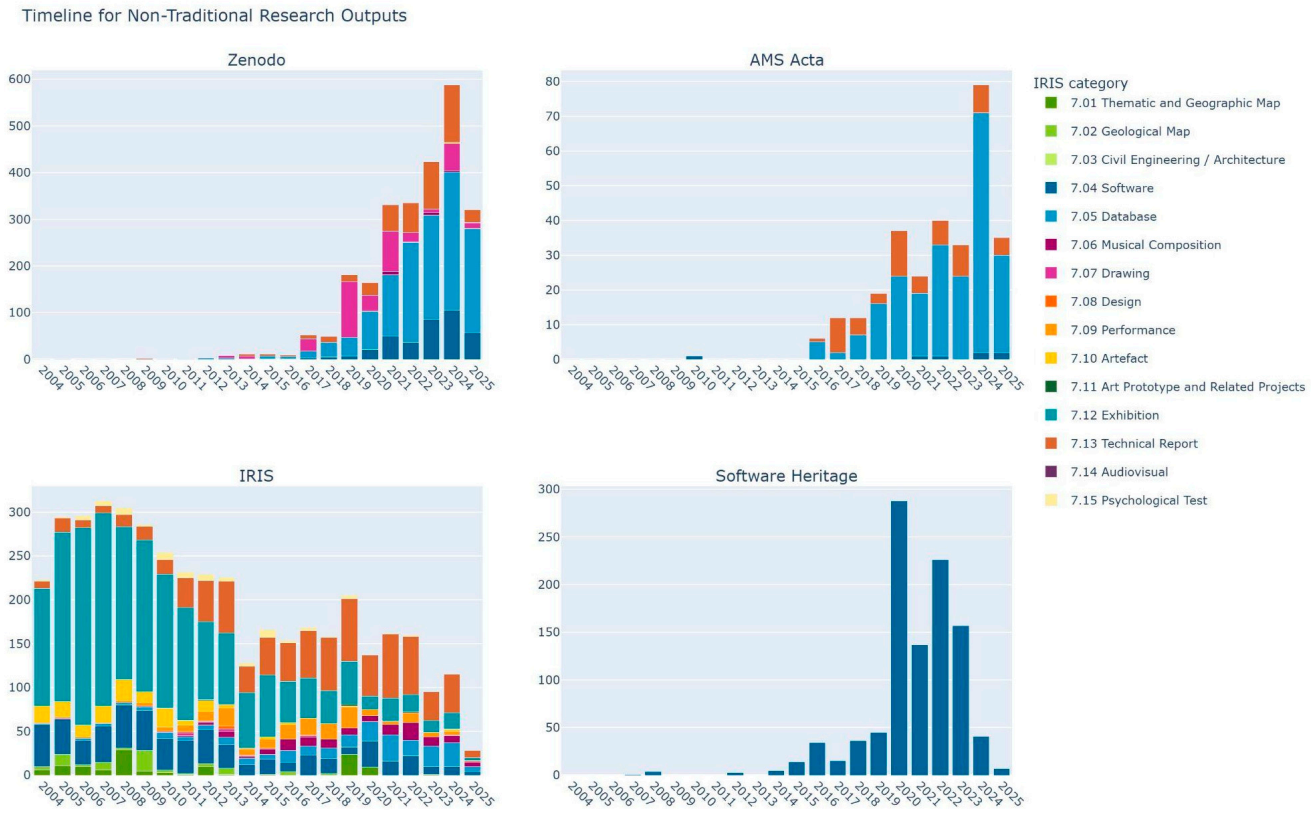



**Figure 4.** A temporal analysis of the NTROs retrieved from the various sources.

A different argument can, instead, explain the clear increase in the *databases* (7.05) from 2020 onward. This increment is probably due to the latest national Research Quality Evaluation exercise (in Italian: *Valutazione della Qualità della Ricerca*, or *VQR*, https://www.anvur.it/en/research/evaluation-research-quality, accessed on 26 March 2026), which is a nationwide analysis, held by the National Agency for the Evaluation of Universities and Research (ANVUR, https://www.anvur.it/en, accessed on 26 March 2026), that measures the quality of scientific output produced by universities and research institutions, considering research products published between 2020 and 2024. In this latest exercise, some NTROs have also been considered as possible research products to present, which justifies the increase in the number of databases being among the most used NTROs across different disciplines.

Similar considerations can be drawn from analysing the trends in Zenodo and AMS Acta. In addition, we can also notice another, more recent, behaviour that involves these repositories. However, first, it is necessary to clarify one aspect concerning the approach for aligning specific Zenodo and AMS Acta product types with the classification provided by UNIBO IRIS, as shown in Table 1. UNIBO IRIS does not currently have any particular category for representing *datasets*. However, this category is among the primary ones used by several repositories, including Zenodo and AMS Acta, to represent generic data collections. Thus, for this analysis, we have decided, as anticipated in the previous sections, to align the Zenodo and AMS Acta datasets with the IRIS category *7.05 Database*, even though it is not a perfect match.

Having clarified this aspect, we can observe a clear increase in the number of NTROs published on both Zenodo and AMS Acta in 2024, which seems to be confirmed also in the data related to 2025, which reached half of the number of the 2024 NTRO publications in half a year, considering that the data we have available for 2025 as of the end of June. This trend could be the result of a series of activities that UNIBO has formally pushed,

starting from the release of the official policy dedicated to research data management (Alma Mater Studiorum—Università di Bologna, 2023) and the related seminars and conferences organised, for UNIBO researchers, by the UNIBO data stewards to share good practices to manage and publish research data in relevant repositories, also following the CoARA commitments (Coalition for Advancing Research Assessment, 2022) that UNIBO signed with several other Italian universities. During these sharing events, which UNIBO researchers well attended, the two primary repositories mentioned as possible venues to use for publishing and sharing data were, indeed, Zenodo and AMS Acta, introduced as general-purpose and institutional repositories. Thus, these activities can be a plausible justification for the rise in their use during the past two years.

### *4.2. Limitations*

From a methodological perspective, the Software Heritage extraction process described in Section 2.1.1, although effective, is time-consuming, with a runtime of approximately 40 h. This extended duration reflects not only the complexity of the heuristic filtering process but also the large scale of the Software Heritage archive, which contains, as of 26 March 2026, over 28 billion source files from more than 430 million projects[1].

The approach used to identify matches with Software Heritage listed repositories introduces the risk of missing relevant repositories due to naming inconsistencies and incomplete metadata, and could be significantly improved. In addition, it was not possible to cross-check coverage between Software Heritage and UNIBO IRIS due to the absence of shared persistent identifiers, such as DOIs, and the limited metadata exposed by the SWH API. Indeed, the heuristic adopted for affiliation matching in Software Heritage is a best-effort approach, and the fact that the software it exposes is not associated with prominent persistent identifier schemes used in CRIS platforms, such as DOIs, prevents us from achieving robust matching at scale in UNIBO IRIS.

The approach described in Section 2.3.1, used to infer institutional affiliation in Software Heritage by exploiting commit email domains and README.md content, did not allow us to reconcile Software Heritage records easily with the structured, more publication-oriented metadata available in IRIS. As a result, direct linkage or overlap analysis between the two sources was infeasible. Indeed, the Software Heritage API provided a limited set of metadata elements, and each record had minimal descriptive data. The fields included only the repository URLs, the related revision and directory identifiers, and the list of authors. From the URLs, we attempted to extract software names. Still, this method was imprecise and led to inconsistencies because users often used different names for their software and the repositories that hosted it. Some of the repositories were also inaccessible due to moving or deletion.

The fact that Software Heritage uses SWHIDs to identify the software it preserves adds complexity to deduplication and cross-repository overlap identification. Indeed, while the SWHID scheme is one of the primary standard identifier schemes for software, it has not been systematically tracked by several repositories and CRIS platforms, thus making it difficult to use it for the aforementioned purposes.

Finally, a limitation of the study, particularly related to RQ1.b, is the use of only one source, i.e., OpenCitations, to extract citation links involving the retrieved NTROs. Indeed, even if OpenCitations is one of the primary open sources providing this kind of information, it is not the only one available in the scholarly domain. In principle, citation data from OpenCitations could be complemented with other data from different providers, such as ScholeXplorer (https://scholexplorer.openaire.eu, accessed on 26 March 2026) (Baglioni et al., 2020) and the Data Citation Corpus (https://makedatacount.org/find-a-tool/, accessed on 26 March 2026) (DataCite & Make Data Count, 2025), to measure if

the combined citation data (and other potential relational links that are not formal citations) involving NTROs may show a more comprehensive coverage.

## 5. Future Developments and Perspectives

Future improvements on the Software Heritage data collection pipeline described in Section 2.1.1 could involve refining the filtering criteria by expanding the keyword set for the initial query to include UNIBO's department acronyms—e.g., *Dipartimento*, *DISI*, *FICLIT*, etc.—and UNIBO's campus names—e.g., *Campus di Cesena*, *Campus di Ravenna*, *Campus di Rimini*. In addition, a confidence-based scoring system could replace the current binary checks. This system would assign weights to multiple affiliation signals, such as matches in repository URLs, author email domains, and README.md content, and evaluate repositories based on a cumulative score, to improve recall without compromising precision. Furthermore, automating integration with UNIBO IRIS could streamline researcher workflows and improve metadata coverage across institutional systems, making it more comprehensive and consistent.

Another possible future development is the reuse of the proposed methodology, including the software, to address the same RQs introduced here, involving different Italian institutions that adopt IRIS as a platform for implementing their CRIS systems, aiming to expand the study and add additional evidence. For instance, in principle, comparisons against Zenodo and Software Heritage can be run if input data from another IRIS platform are available and similar string queries (using the related university's names and acronyms) are used to filter data via the APIs. In this way, it would be possible also to create appropriate benchmarks for comparison, enabling others to repeat similar studies and to monitor and re-run the same analysis over time.

Of course, in this context, the involvement of AMS Acta would not be appropriate, as it primarily contains data from UNIBO's researchers and serves as the university's institutional repository. However, it could be replaced in the pipeline by another repository implementing a specific software extension for the data collection step that returns the same output format used by our software, thus enabling the reuse of the full pipeline. Similarly, we could draw on additional data from other existing open scholarly infrastructures to gather contextual data alongside OpenCitations' citation data and enrich the study, such as institutional affiliations and funding providers from OpenAIRE (https://openaire.eu, accessed on 26 March 2026).

More broadly, this work highlights the importance of coordinated efforts to align institutional repositories with Open Science infrastructures—something that has been currently pushed at UNIBO as a consequence of this study and other investigations performed in the context of UNIBO's participation in CoARA (https://www.coara.org, accessed on 26 March 2026) and the Barcelona Declaration on Open Research Information (DORI, https://barcelona-declaration.org/, accessed on 26 March 2026). Achieving genuine interoperability requires technical harmonisation—for instance, through shared metadata schemas—alongside procedural and cultural shifts in the ways researchers publish and register their outputs. Repositories should be conceived with interoperability as a core principle, enabling cross-platform citation, discovery, and usage analytics. The Open Science community has addressed this aspect in recent years. It has proposed tools to enable technical and semantic interoperability, such as the Scientific Knowledge Graph-Interoperability Framework (SKG-IF, https://skg-if.github.io/, accessed on 26 March 2026), which was recently published as an RDA Recommendation (Mannocci et al., 2025). In the text of this recommendation, there are already mentions of existing infrastructures, including the CRIS of Novi-Sad University (https://cris.uns.ac.rs/sr, accessed on 26 March 2026), which is implementing SKG-IF to facilitate better exchange and mashups of research metadata. Along

the same lines, the European Open Science Cloud (EOSC) Association (https://eosc.eu/, accessed on 26 March 2026) is actively working to implement an interoperability framework (Corcho et al., 2021) to support the creation and setting up of the EOSC (Burgelman, 2021), i.e., a European research data ecosystem offering a trusted, secure and sovereign common space of high-quality, FAIR (Wilkinson et al., 2016) research data and related services. These and other initiatives are fundamental steps toward better alignment and interoperability among institutional services provided by research-performing organisations.

Finally, some of the outcomes of this analysis have already been shared with various groups and task forces at the University of Bologna responsible for aligning current international practices with the expectations of local and global policies, for instance, concerning the identification and tracking of specific types of NTROs. Indeed, this and other studies conducted in the past months, e.g., by Andreose et al. (2026), will inform decisions on possible policy revisions and infrastructure implementations.

**Author Contributions:** Conceptualization: S.P.; Data curation: A.C., A.N. (Ahmadreza Nazari), M.P., P.T. and L.P.; Formal analysis: A.N. (Anna Nicoletti) and S.S.; Funding acquisition: S.P.; Investigation: A.C., I.H., A.N. (Ahmadreza Nazari), A.N. (Anna Nicoletti), M.P., S.P., P.T., L.P. and S.S.; Methodology: A.C., I.H., A.N. (Ahmadreza Nazari), A.N. (Anna Nicoletti), M.P., S.P., P.T., L.P. and S.S.; Project administration: S.P.; Resources: I.H.; Software: A.C., A.N. (Ahmadreza Nazari), A.N. (Anna Nicoletti), M.P., P.T., L.P. and S.S.; Supervision: S.P.; Validation: I.H. and S.P.; Visualization: A.N. (Anna Nicoletti) and S.S.; Writing—original draft preparation: A.C., A.N. (Ahmadreza Nazari), A.N. (Anna Nicoletti), M.P., P.T., L.P. and S.S.; Writing—review and editing: I.H. and S.P. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the European Union Grant Agreements No 101188018 (GRAPHIA Project) and No 101187940 (LUMEN Project). The APC was funded by the European Union.

**Data Availability Statement:** The original metadata used in this study were retrieved from publicly available repositories: Zenodo (European Organization For Nuclear Research & OpenAIRE, 2013), Software Heritage (Di Cosmo & Zacchiroli, 2023), AMS Acta (Vignocchi et al., 2006), IRIS—the University of Bologna's CRIS system (Bollini et al., 2016) and OpenCitations (Peroni & Shotton, 2020). All processed datasets generated during the project are openly accessible on Zenodo (Nazari et al., 2025), as are the research Protocol (Ciarrocca et al., 2025a) and the Data Management Plan (Pograri & Tisci, 2025). All scripts used for data extraction, processing, and analysis are available on the project's GitHub repository (https://github.com/open-sci/2024-2025/, accessed on 26 March 2026) and mirrored on Zenodo (Ciarrocca et al., 2025b).

**Acknowledgments:** This study was based on the outcomes of the analysis held during the Open Science 2024/2025 course (https://www.unibo.it/en/teaching/course-unit-catalogue/course-unit/2024/443753, accessed on 26 March 2026) of the Second Cycle International Degree in Digital Humanities and Digital Knowledge (https://corsi.unibo.it/2cycle/DigitalHumanitiesKnowledge, accessed on 26 March 2026) of the University of Bologna, taught by S.P. We thank all the people who participated in the workshop organised in the context of the Open Science course who provided insightful feedback and, thus, implicitly contributed to some of the aspects of the present study.

**Conflicts of Interest:** S.P. is the Director of OpenCitations. I.H. is the Chief Technology Officer of OpenCitations. The authors declare no conflicts of interest.

## Notes

[1] For further details on the Software Heritage archive, including its dimensions and internal storage architecture, see the official documentation at https://docs.softwareheritage.org/devel/swh-storage/archive-copies.html (accessed on 26 March 2026).

## References


Alma Mater Studiorum—Università di Bologna. (2018). *University policy on open access to research publication and data [policy]*. Alma Mater Studiorum—Università di Bologna. Available online: https://www.unibo.it/en/attachments/the-university-policy-on-open-access-to-research-publications-and-data-1/@@download/file/PolicyAteneoAccessoApertoPubblicazioniDatiRicerca_EN_checked.pdf (accessed on 26 March 2026).

Alma Mater Studiorum—Università di Bologna. (2022). *Strategic plan 2022–2027 [strategic plan]*. Alma Mater Studiorum—Università di Bologna. Available online: https://pianostrategico.unibo.it/en (accessed on 26 March 2026).

Alma Mater Studiorum—Università di Bologna. (2023). *The university policy on research data management [policy]*. Alma Mater Studiorum—Università di Bologna. Available online: https://www.unibo.it/en/attachments/university-policy-on-research-data-management/@@download/file/The-University-Policy-on-research-data-management.pdf (accessed on 26 March 2026).

Alma Mater Studiorum—Università di Bologna & ARIC—Research Division, Research Services and Division Projects Coordination Unit, Data Stewards. (2024). *University guidelines for research data management [guidelines]*. Alma Mater Studiorum—Università di Bologna. Available online: https://www.unibo.it/en/research/open-science/university-guidelines-for-research-data-management (accessed on 26 March 2026).

Amurri, A., Giachino, E., Misuraca, E., & Peroni, S. (2025). *UNIBO IRIS bibliographic data dump, dated 30 May 2025, updated on 3 July 2025 [dataset; CSV]*. AMS Acta. [CrossRef]

Andreose, E., Di Marzo, S., Heibi, I., Peroni, S., & Zilli, L. (2026). Analysing the coverage of the University of Bologna's bibliographic and citation metadata in OpenCitations collections. *Scientometrics*, *131*(2), 845–871. [CrossRef]

Azeroual, O., & Schöpfel, J. (2019). Quality issues of CRIS data: An exploratory investigation with universities from twelve countries. *Publications*, *7*(1), 14. [CrossRef]

Baglioni, M., Manghi, P., & Mannocci, A. (2020). Context-driven discoverability of research data. In M. Hall, T. Merčun, T. Risse, & F. Duchateau (Eds.), *Digital libraries for open knowledge—24th international conference on theory and practice of digital libraries, TPDL 2020, Lyon, France, August 25–27, 2020, Proceedings* (Vol. 12246, pp. 197–211). Springer. [CrossRef]

Barcelona Declaration on Open Research Information. (2024). *Barcelona declaration on open research information: Commitments*. Available online: https://barcelona-declaration.org/commitments/ (accessed on 26 March 2026).

Bollini, A., Mennielli, M., Mornati, S., & Palmer, D. T. (2016). IRIS: Supporting & managing the research life-cycle. *Universal Journal of Educational Research*, *4*(4), 738–743. [CrossRef]

Burgelman, J.-C. (2021). Politics and open science: How the european open science cloud became reality (the untold story). *Data Intelligence*, *3*(1), 5–19. [CrossRef]

Cagan, R. (2013). San Francisco declaration on research assessment. *Disease Models & Mechanisms*, *6*(4), 869–870. [CrossRef] [PubMed]

Ciarrocca, A., Nazari, A., Nicoletti, A., Pensalfini, M., Peroni, S., Pograri, L., Slinkin, S., & Tisci, P. (2025a). *Open Science 24/25—Protocol V.2*. Zenodo. [CrossRef]

Ciarrocca, A., Nazari, A., Nicoletti, A., Pensalfini, M., Peroni, S., Pograri, L., Slinkin, S., & Tisci, P. (2025b). *Software for the final project of the open science course a.y. 2024/2025*. Zenodo. [CrossRef]

Coalition for Advancing Research Assessment. (2022). *Agreement on reforming research assessment [policy]*. European Science Foundation. Available online: https://coara.eu/agreement/the-agreement-full-text/ (accessed on 26 March 2026).

Corcho, O., Kurowski, K., Ojsteršek, M., Choirat, C., van de Sanden, M., & Coppens, F. (2021). *EOSC interoperability framework*. Publications Office of the European Union. [CrossRef]

DataCite & Make Data Count. (2025). *Data citation corpus data file (version v4.1) [data set]*. DataCite. [CrossRef]

De Castro, P., & Puuska, H.-M. (2023). *Research information management systems: Covering the whole research lifecycle*. Available online: https://dspacecris.eurocris.org/handle/11366/2471 (accessed on 26 March 2026).

Di Cosmo, R., & Zacchiroli, S. (2023). The software heritage open science ecosystem. In T. Mens, C. De Roover, & A. Cleve (Eds.), *Software ecosystems* (pp. 33–61). Springer International Publishing. [CrossRef]

European Organization For Nuclear Research & OpenAIRE. (2013). *Zenodo: Research. Shared*. CERN. [CrossRef]

Heibi, I., Moretti, A., Peroni, S., & Soricetti, M. (2024). The OpenCitations index: Description of a database providing open citation data. *Scientometrics*, *129*(12), 7923–7942. [CrossRef]

Hurrell, C. (2023). Research assessment reform, non-traditional research outputs, and digital repositories: An analysis of the Declaration on Research Assessment (DORA) signatories in the United Kingdom. *Evidence Based Library and Information Practice*, *18*(4), 2–20. [CrossRef]

Loesch, M. (2020). *Sickle: OAI-PMH for humans (version 0.7.0) [Python]*. Available online: https://archive.softwareheritage.org/swh:1:dir:417f7113d3eaffc808d80848d003ffa9418a92a6;origin=https://github.com/mloesch/sickle;visit=swh:1:snp:c541c472ed3bbd6f193199d2629a6a129df5164d;anchor=swh:1:rev:14ef88e1aecce1b767ef7a3824e90c009899e1f7 (accessed on 26 March 2026).

Mannocci, A., Peroni, S., & Scientific Knowledge Graphs Interoperability Framework (SKG-IF) WG. (2025). *Scientific Knowledge Graphs Interoperability Framework (SKG-IF): Enabling data exchange for SKGs (Working group recommendation 1.0.0)*. Research Data Alliance. [CrossRef]

Nazari, A., Ciarrocca, A., Nicoletti, A., Pensalfini, M., Peroni, S., Tisci, P., Pograri, L., & Slinkin, S. (2025). *Datasets for the final project of the Open Science course a.y. 2024/2025*. Zenodo. [CrossRef]

Okamura, K. (2019). Interdisciplinarity revisited: Evidence for research impact and dynamism. *Palgrave Communications*, *5*(1), 141. [CrossRef]

Peroni, S., & Shotton, D. (2019). *Open citation identifier: Definition*. Figshare. [CrossRef]

Peroni, S., & Shotton, D. (2020). OpenCitations, an infrastructure organization for open scholarship. *Quantitative Science Studies*, *1*(1), 428–444. [CrossRef]

Pograri, L., & Tisci, P. (2025). *Open Science end-of-the-course project DMP—2024/2025*. Zenodo. [CrossRef]

Rokon, M. O. F., Yan, P., Islam, R., & Faloutsos, M. (2021). Repo2Vec: A comprehensive embedding approach for determining Repository similarity. *arXiv*, arXiv:2107.05112. [CrossRef]

UNESCO. (2021). *UNESCO recommendation on open science* (p. 36). Programme and Meeting Document No. SC-PCB-SPP/2021/OS/UROS. UNESCO. [CrossRef]

Vignocchi, M., Sacchi, S., & Morroia, F. (2006). Alma-DL, the digital library of the University of Bologna: A case study. *Archiving Conference*, *3*(1), 31–36. [CrossRef]

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., . . . Mons, B. (2016). The FAIR guiding principles for scientific data management and stewardship. *Scientific Data*, *3*, 160018. [CrossRef] [PubMed]

Zhang, H., Zhang, L., Fang, L., & Filgueira, R. (2024). Multi-level AI-driven analysis of software repository similarities. In *2024 IEEE 20th international conference on e-science (e-Science)* (pp. 1–10). IEEE. [CrossRef]

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.