THE ATLAS WHITEBOOK AND GUIDELINES:

FAIR RESEARCH ON ITALIAN DIGITAL CULTURAL HERITAGE

THE ATLAS WHITEBOOK AND GUIDELINES

FAIR RESEARCH ON ITALIAN DIGITAL CULTURAL HERITAGE

Authors

Alessia Bardi [iD] (alessia.bardi@isti.cnr.it)

Marina Buzzoni [iD] (mbuzzoni@unive.it)

Marilena Daquino [iD] (marilena.daquino2@unibo.it)

Riccardo Del Gratta [iD] (riccardo.delgratta@ilc.cnr.it)

Angelo Mario Del Grosso [iD] (angelo.delgrosso@ilc.cnr.it)

Roberto Rosselli Del Turco [iD] (roberto.rossellidelturco@unito.it)

Franz Fischer [iD] (franz.fischer@unive.it)

Sebastiano Giacomini [iD] (sebastiano.giacomin2@unibo.it)

Chiara Martignano [iD] (chiaramartignano@gmail.com)

Giorgia Rubin [iD] (giorgia.rubin@ilc.cnr.it)

Francesca Tomasi [iD] (francesca.tomasi@unibo.it)

Contributors

Laurent Fintoni

Copyright and License

Preface

Digital research in the humanities has flourished over the past two decades, producing a rich landscape of web-based projects across the Italian Cultural Heritage domain. Yet the long-term value of these efforts depends on how FAIR (Findable, Accessible, Interoperable, and Reusable) their outputs are, and on whether they are documented against shared and sustainable standards. This volume brings together the response of the ATLAS project (hereinafter: ATLAS) to that challenge in two complementary parts: a Whitebook that frames the context and infrastructure of the ATLAS initiative, and a practical set of Guidelines to help researchers design and produce FAIR research products in the Digital Humanities (DH).

Part 1, the Whitebook, surveys the state of the art and outlines the project's goals and methodology. It evaluates a set of pilot research products that guided the modeling process of the ATLAS knowledge graph, providing insights on how to effectively implement FAIR principles in the DH. It then presents the ATLAS model itself, with particular attention to the metadata required to catalogue heterogeneous research outputs. Finally, the Whitebook introduces the ATLAS web application and services, offering a roadmap for curators and users to enter, manage, and access data through a sustainable, standards-driven platform.

Part 2, the Guidelines, translates principles into practice. Building on the ALLEA (All European Academies) E-Humanities Working Group's recommendations, it provides targeted, phase-by-phase guidance for five common Digital Humanities output types: digital scholarly editions, text, software tools, ontologies, and linked open datasets. Rather than replacing disciplinary standards and best practices, the Guidelines complement them with FAIR-focused recommendations, including practical steps to further improve citation, documentation, and reuse.

Together, the two parts serve distinct yet complementary purposes. The Whitebook establishes a shared conceptual and technical foundation for the ATLAS catalogue, including its knowledge graph, search portal, and web application. The Guidelines support practitioners in their work by outlining concrete choices regarding standards, formats, identifiers, repositories and documentation practices that ensure the long-term robustness of research products across communities. Read in sequence, they move from vision to implementation; consulted

independently, each can support specific planning, development, or curation tasks.

This volume is intended for scholars, librarians, software developers, data stewards and cultural heritage professionals involved in the creation, maintenance or evaluation of digital research products. We also hope that those new to digital humanities will find it useful for gaining a clear overview of the field's outputs and the steps required to make them FAIR.

This work was conducted within the PRIN 2022 ATLAS project, funded by the Next Generation program of the European Commission for 24 months (October 2023 - October 2025), and is the result of a collaboration between the University of Bologna, Ca' Foscari University of Venice and two units of the Italian National Research Council—the Institute of Information Science and Technologies "Alessandro Faedo" (ISTI) and the Institute for Computational Linguistics "A. Zampolli" (ILC). Through the use of Semantic Web technologies to interlink scholarly outputs and by providing practical guidance on their creation and stewardship, ATLAS aims to strengthen the Italian Cultural Digital Heritage ecosystem, ensuring that today's research remains accessible, interoperable and reusable for future communities.

Author Responsibility Statement

Chiara Martignano is the corresponding author and is responsible for the following sections:

- Methodological Notes, Text Collections, and Digital Scholarly Editions in Chapter 1 of the Whitebook;

- Research Product, Text Collection, Digital Scholarly Edition, and 3D Digital Twin in Chapter 2 of the Whitebook;

- Introduction, Digital Scholarly Editions, Text Collections and From FAIR to top-notch of the Guidelines.

Sebastiano Giacomini is responsible for the following sections:

- Ontologies and Linked Open Data in Chapter 1 of the Whitebook;

- Methodological Notes, Research Product, Linked Open Data, and Ontology in Chapter 2 of the Whitebook;

- Chapter 3 of the Whitebook except section API and OpenAIRE Italian DH-CH Gateway;

- Linked Open Data, Ontology and From FAIR to top-notch of the Guidelines.

Giorgia Rubin is responsible for the following sections:

- Software in Chapter 1 of the Whitebook;

- Research Product, Software and Language Model in Chapter 2 of the Whitebook;
- API, OpenAIRE Italian DH-CH Gateway in Chapter 3 of the Whitebook;
- Software and From FAIR to top-notch of the Guidelines.

All authors contributed to all the remaining sections.

AI Statement

In the following publication, AI has been used to check grammar and spelling and to improve lexical forms.

Acknowledgments

Contents

Part 1
Whitebook

Abstract

In the last twenty years, numerous web-based Digital Humanities (DH) projects have emerged to collect data in the field of Italian Cultural Heritage. However, the research outputs produced by these projects risk obsolescence if not well-documented and based on shared standards and best practices. Moreover, these research products are often difficult to discover, and currently, no dedicated comprehensive catalogue exists. The ATLAS project—a joint effort of the University of Bologna, the Ca' Foscari University of Venice, and the Italian National Research Council Unit of Pisa (that includes the Institute of Information Science and Technologies "Alessandro Faedo"—ISTI—and the Institute for Computational Linguistics "A. Zampolli"—ILC)—aims to fill this gap. It seeks to create a catalogue that leverages Semantic Web technologies to interlink scholarly outputs and enhance the Italian Cultural Digital Heritage. The main outcomes of the ATLAS project are:

- a knowledge graph of DH research on Italian Cultural Heritage, accessible online via dedicated services and preserved in the ILC4CLARIN repository;

- a web application that allows users to explore the catalogued scholarly products and their contents, starting from the pool of selected research products used as pilots for the project;

- a search portal built on the OpenAIRE Connect Gateway that focuses on scholarly literature and data related to the research products included in the knowledge graph;

- a set of guidelines and best practices for producing FAIR (Findable, Accessible, Interoperable and Reusable) and high quality scholarly data in the Digital Humanities, which correspond to Part 2 of this volume;

- this whitebook for the curators and users of the catalogue.

References to the above results are available in the complete list of outputs produced by the project. This whitebook begins with a review of the state of the art and presents the ATLAS project's goals and methodologies. Chapter 1 evaluates the research products used as pilots for designing the ATLAS data model and provides guidelines for producing FAIR data. Chapter 2 presents the ATLAS data model, paying particular attention to metadata for cataloguing research products. Chapter 3 introduces the ATLAS web application and provides instructions for future curators and users on data entry and access.

Introduction

In recent years, the growing integration of the World Wide Web and its technologies has profoundly reshaped scholarly research, particularly within the Digital Humanities (DH) domain. These advancements have unlocked new avenues for preserving, sharing, and reusing research outputs, fostering unprecedented collaboration and dissemination [Tomasi 2022, Brügger and Finnemann 2013]. Yet, as vast amounts of scholarly data proliferate, the need for standardised models and guidelines to manage, aggregate, and explore this information effectively has become more pressing.

Several platforms in the broader scholarly landscape play a crucial role in providing persistent identification, long-term preservation, and enhanced findability of research data. Key services include Zenodo[1] and OpenAIRE.[2] The OpenAIRE network integrates various services, including community web portals such as the Digital Humanities and Cultural Heritage gateway,[3] which facilitate the discovery and sharing of research outcomes and Open Science practices.

Currently, multiple catalogues document DH research, including digital scholarly editions,[4] and project lists from national[5] and international associations,[6] and research centres. However, there remains a lack of comprehensive catalogues specifically focused on DH projects related to Italian Cultural Heritage, and no structured collections exist for DH research outputs using Semantic Web technologies.

In Italy, institutions have made significant progress in digitising and aggregating Cultural Heritage through Linked Open Data (LOD) collections. Notable initiatives include the *dati.culturaitalia*[7] platform by the Italian Ministry of Culture and the ArCO[8] project, which has created a Knowledge Graph based on the General Catalog of Italian Cultural Heritage. These efforts align with European digitisation projects like Europeana[9] and those carried out by research infrastructures such as ARIADNE.[10] While these initiatives provide interoperable LOD, a gap persists in the research framework for supporting best practices and improving the findability and reusability of Italian heritage-related DH data.

---

[1] https://zenodo.org/.

[2] https://www.openaire.eu/.

[3] https://dh-ch.openaire.eu/.

[4] Franzini, Greta (2012-) Catalogue of Digital Editions, https://doi.org/10.5281/zenodo.1161425.

Sahle, Patrick et al., a catalog of Digital Scholarly Editions, v.4.112 2020ff, last change 2024-06-06.

[5] For example AIUCD's list of DH projects: https://www.aiucd.it/progetti/.

[6] For example EADH's list of DH projects: https://eadh.org/projects.

[7] https://dati.cultura.gov.it/.

[8] https://dati.beniculturali.it/arco/index.php.

[9] https://www.europeana.eu/.

[10] https://www.ariadne-research-infrastructure.eu/.

The current models do not adequately address the complexities of today's DH landscape. DH projects produce very diverse outputs—including text collections, digital scholarly editions, linked open datasets, vocabularies, and software tools—each requiring specific descriptive approaches. Critical elements like textual typologies and edition criteria remain insufficiently addressed. Furthermore, current models lack effective solutions for connecting research activities to their corresponding Cultural Heritage objects, despite the opportunities offered by Linked Open Data.

The DH field also needs clear guidelines for creating high-quality scholarly data. While best practices and guidelines for the implementation of the FAIR (Findable, Accessible, Interoperable and Reusable) principles[11] already exist (e.g., The Recommendations of the ALLEA Working Group for E-Humanities "Sustainable and FAIR Data Sharing in the Humanities"[12] and PARTHENOS' "Guidelines to FAIRify data management and make data reusable"),[13] these resources need expansion to cover the full range of DH resources.

The ATLAS project (hereinafter: ATLAS) addresses these challenges by creating a comprehensive knowledge graph of DH research related to Italian Digital Cultural Heritage. By developing the ATLAS Ontology and its associated knowledge graph, the project seeks to establish a semantic framework that captures the diverse outputs of DH research, including digital editions, text collections, and datasets. ATLAS addresses the complexities of describing and linking scholarly data, ensuring that metadata is enriched and accessible, thus promoting the findability and reusability of these valuable cultural resources. Through detailed analysis and mapping of existing models and vocabularies, the project offers a new approach to integrating Italian DH resources into the global knowledge landscape, fostering greater collaboration and discovery across disciplines and institutions.

The ATLAS project has contributed to the Italian DH research landscape through several key outcomes, besides this whitebook. Firstly, there is the evaluation of pilot projects (i.e. model projects used as test cases) presented in Chapter 1. This evaluation was conducted to identify relevant metadata for different types of digital research outputs and to ensure data access and persistence. It also identified best practices for producing FAIR digital research products. Secondly, there is the ATLAS ontology, thoroughly described in Chapter 2, which is the model underlying the ATLAS knowledge graph. Based on international standards, this ontology aims to enrich the information available for different DH research outputs, while enhancing the interoperability of metadata. Thirdly, there is the knowledge graph on DH projects and scholarly data related to Italian Cultural Heritage, presented in Chapter 3, alongside the ATLAS web application, Application Programming Interfaces (API), and the OpenAIRE CONNECT

---

[11]https://www.go-fair.org/fair-principles/.

[12]https://doi.org/10.7486/DRI.TQ582C863.

[13]https://zenodo.org/records/2668479.

Gateway,[14] which enables access to the information included in it. Finally, there is a set of comprehensive guidelines for producing FAIR scholarly data in the Digital Humanities, which is presented in the second part of this volume.

For a full list of the ATLAS project outcomes and their respective links, please refer to the ATLAS Outputs section.

---

[14]https://connect.openaire.eu/.

Chapter 1
Pilot Analysis

Abstract

The ATLAS project aims to evaluate a pool of research products in order to identify descriptive metadata characterising research objects and projects via a bottom-up approach, as well as defining good practices for accessing and manipulating Cultural Heritage data. Metadata of selected scholarly products (also called pilots) were incorporated into the initial ATLAS knowledge graph (ATLAS-KG). The pilots fall into five main categories:

- collections of texts (ALIM, the Archive of the Italian Latinity of the Middle Ages; Biblioteca Italiana; Musisque Deoque; BUP - Digital Humanities);

- digital scholarly editions (VaSto - VArchi, STOria fiorentina; Codice Pelavicino Digitale; Digital Edition of Aldo Moro's works; Leges Langobardorum);

- software tools (EVT, Edition Visualization Technology; Voyant Tools);

- linked open datasets (Zeri & LODE; DanteSources; LiLa; Biflow);

- ontologies for the Cultural Heritage (CIDOC-CRM, Conceptual Reference Model; HiCO; SPAR Ontologies).

The pilots listed above were selected based on two main criteria. First, they represent advanced DH projects in the Italian Cultural Heritage domain, covering Italian and Latin sources and spanning from classical to contemporary periods. Second, for each pilot, at least one member of the ATLAS project team is directly involved in its development, providing access and in-depth knowledge to related resources.

The pilots were evaluated according to existing guidelines and best practices for implementing the FAIR principles [Wilkinson et al. 2016]. The results of this evaluation are presented in this first chapter, which is divided into five sections. Each section is dedicated to a category of digital research products, as listed above. The analysis focuses on the FAIRness of the research products, while we do not comment on their content and their effectiveness from a theoretical standpoint. We also summarise trends and practices that emerged from the pilots analysis and we suggest how to enhance the pilots' FAIRness.

Methodological notes

The initial phase of the ATLAS project focused on the analysis and evaluation of a curated selection of research products, which we refer to as 'pilots'. This analysis served two purposes: identifying cataloguing metadata to describe Digital Humanities research products and assessing the best practices for creating FAIR research outcomes. The results of this analysis helped us develop the ATLAS ontology (described in Chapter 2) and the ATLAS guidelines (outlined in Part 2 of this volume).

Selected pilots primarily consist of research products related to Italian historical and literary fields, mainly developed by Italian researchers. Notable exceptions include the Voyant Tools software and the CIDOC-CRM ontology, both widely adopted in the Italian scholarly environment, but developed abroad. The pilots span five categories: text collections, digital editions, ontologies, linked open data, and software tools. We have selected these categories because they represent five main ways in which data can be structured and organised in the field of DH. These categories allow us to represent research products created in disciplinary contexts that may be very distant from one another, regardless of the type of data or resource on which they are based. Pilots were chosen to represent ongoing Digital Humanities research on the Italian Cultural Heritage, covering periods from classical to 20th-century and encompassing both vernacular Italian and Latin. We prioritised research products that are considered valid examples in Italian Digital Humanities, wherein at least one ATLAS project team member is directly involved, ensuring high information quality and direct access to resources.

We evaluated the pilots using guidelines and best practices for producing FAIR research products that are shared within the broader scholarly community, focussing on the humanities' specificities, namely:

- 'The FAIR Guiding Principles for scientific data management and steward-ship' [Wilkinson et al. 2016];

- The GO-FAIR Initiative's practical guidelines for implementing FAIR principles;[15]

- PARTHENOS' Guidelines to FAIRify data management and make data reusable;[16]

- the Recommendations of the ALLEA Working Group for E-Humanities 'Sustainable and FAIR Data Sharing in the Humanities';[17]

- the Common Standards White Paper of OPERAS (the European Research Infrastructure for the development of open scholarly communication in the social sciences and humanities);[18]

---

[15]https://www.go-fair.org/fair-principles/.
[16]https://doi.org/10.5281/zenodo.2668479.
[17]https://doi.org/10.7486/DRI.TQ582C863.
[18]https://operas-eu.org/special-interest-group-living-book/operas-common-standards-white-paper-june-2021/.

- GoTriple Content Providers Handbook, which is a discovery platform for the social sciences and humanities.[19]

The analysis addresses four distinct yet interrelated dimensions:

- **metadata**, examining accessibility and adherence to standards;

- **data implementation**, evaluating licence clarity, use of standard formats, availability of persistent URLs or identifiers, and links to semantic web datasets;

- **data access and consultation methods**, assessing user interface and/or API accessibility, data download availability, and adoption of long-term preservation policies;

- **documentation**, reviewing website information clarity and completeness, availability of high-level and technical documentation, citation guidelines, and public detailed changelog and roadmap.

Following each pilot evaluation, a FAIRness evaluation summary is provided, listing the strengths of the pilot that make it findable, accessible, interoperable and reusable, alongside possible improvements in these areas.

It is important to note that our analysis does not assess the scientific merit or theoretical soundness of the research products. For instance, when examining a digital scholarly edition, we do not evaluate the effectiveness of the chosen editorial approach. We assume that any research product included in the ATLAS catalogue adheres to the theories and practices of its respective discipline. Instead, our analysis focuses on how the research products are created and managed from a FAIR perspective.

Text collections

A text collection is a type of digital archive that focuses on textual heritage, preserving texts digitally and providing web access. Beyond the texts, a text collection typically offers bibliographic metadata and editorial annotations.

Texts may be available in various formats, including digital facsimiles of documents (e.g., the 'Incunaboli in volgare' collection within Biblioteca Italiana) or critical editions (e.g., Musisque Deoque). For a text collection to be considered scientifically valid, its texts must be scholarly edited following shared methodologies, and its documentation must clearly explain the editorial criteria and workflow. For this reason, there is often a fine line between digital scholarly editions and digital text collections:

> "Editions widen their content. When they aim at including ever more documents and finally at completeness, and when the first level of representation may be just a digital facsimile with some metadata, then the edition looks more and more like an archive. [. . .] On the

---

[19]https://www.gotriple.eu/docs/gotriple-handbook-v2_0.pdf.

> other hand, digital archives are already critical on the bibliographic
> level and imply the possibility to incrementally add further critical
> information, other forms of representation (such as transcription)
> and may finally even present an edited text. [. . .] If we take the
> critical engagement and the application of scholarly knowledge as
> the defining characteristics of an edition, then we can say that from
> a certain point on, an archive starts to be an edition." [Sahle 2016]

Text collections can be created either as part of a service or institution, or as a
collection of contents that serve a specific community of users [Borgman 1999].
Research products typically fall into this second category, gathering texts focused
on a specific theme, topic, or area of study.

Text collections serve as essential tools for linguistic studies, providing a foun-
dation for computational analysis and queries. However, in computational
linguistics, corpora—structured sets of electronically stored and processed texts—
are typically not presented to users as complete text collections. In linguistic
corpora, users often cannot access individual texts in their entirety; instead,
they interact with the corpus through a graphical user interface. For our pilot
analysis, we examined text collections implemented as digital archives, though
it is worth noting that all of these also provide tools for conducting linguistic
queries and corpus analysis.

The OAIS (Open Archival Information System) reference model[20] serves as the
conceptual foundation for most long-term digital preservation projects today.
Among the pilot text collections, only Biblioteca Italiana explicitly mentions
using the OAIS reference model in its implementation. The OAIS reference
model outlines key mechanisms for long-term information preservation and access.
Its functional model defines six core services: Ingest, Archival Storage, Data
Management, Preservation Planning, Access, and Administration. The model
centres on information packages containing both the preserved object and its
metadata. These come in three forms: Submission (SIP), Archival (AIP), and
Dissemination (DIP) Information Packages. Originally developed in 2003 and
later approved as ISO standard, OAIS has become the definitive reference for
long-term digital preservation at both national and international levels.

Another key international standard that guides text collection creation is the
National Information Standards Organization (NISO) Framework of Guidance for
Building Good Digital Collections.[21] This framework provides a comprehensive
set of principles organised around four core entities: collections, collected objects,
metadata, and initiatives, which are programmes or projects for creating and
managing collections.

Creating and maintaining a textual archive demands considerable time, funding,
and resources. It is no coincidence that three of the analysed pilots boast a thirty-

---

[20]Reference model for an Open Archival Information System (OAIS): https://public.ccsds.o
rg/pubs/650x0m2.pdf.
[21]https://www.niso.org/sites/default/files/2017-08/framework3.pdf.

year history, resulting from collaborations among numerous universities and support from various institutions. Simultaneously, we can perceive the diverse levels of completeness and technological sophistication in the databases published within the BUP "Digital Humanities" collection, which rely on time-limited funds, as they primarily stem from doctoral projects.

Over time, the Archive of the Italian Latinity of the Middle Ages (ALIM), Musisque Deoque (MQDQ), and Biblioteca Italiana have undergone extensive restructuring to adapt to technological advancements. All three databases have adopted XML/TEI [22] for text encoding. TEI (Text Encoding Initiative) is the standard format for producing electronic texts and is widely used for the preparation of text collections that serve both as digital libraries and archives, and linguistic corpora. Other common formats that were built on the framework and experience of the TEI, are dedicated to more specific types of texts, such as MEI (Music Encoding Initiative)[23] for music notation, Epidoc (Epigraphic documents in TEI XML)[24] for inscriptions, and CEI (Charters Encoding Initiative) [25] for charters.

ALIM and Biblioteca Italiana are based on Muruca,[26] a framework for creating and publishing digital libraries. Using existing software is preferable to developing new software to ensure easier maintenance of the text collection. However, long-term access to resources still depends on the maintenance of the publishing platform. MQDQ's solution of publishing its texts on an external repository (ILC4CLARIN)[27] appears to be the only zero-cost option currently available to guarantee long-term data access. MQDQ also sets a commendable example by implementing RESTful (Representational State Transfer) APIs to query the database, thereby promoting interoperability and data reuse.

All pilots lack clear indications of database status and detailed documentation of changes and updates. These shortcomings stem from two main challenges in the field of research and Cultural Heritage. First, it is difficult to predict funding availability, which affects planning for database implementation and maintenance. Second, tracking editorial work is complex when it is carried out collaboratively by large groups. The latter issue can be addressed by adopting tools that facilitate collaborative editing on shared resources (e.g., GitHub and other versioning platforms, such as Perseus Digital Library'sGitHub repositories)[28] and by publishing resources on external certified repositories. Planning regular publications—ideally annually or biannually—or updates (e.g., Corpus Corporum's[29] 'What's new?' section) after substantial modifications makes it easier to document editorial work and indicate progress status.

---

[22]https://www.tei-c.org/.
[23]https://music-encoding.org/.
[24]https://epidoc.stoa.org/.
[25]https://www.cei.lmu.de/.
[26]https://www.muruca.org/.
[27]https://ilc4clarin.ilc.cnr.it/.
[28]https://github.com/PerseusDL.
[29]https://mlat.uzh.ch/home.

Today, most text collections are encoded in XML/TEI , but research projects often lack the time and resources to produce detailed encodings that capture all philological and material phenomena in texts. Various projects, including ALIM and Biblioteca Italiana, have implemented a multi-level encoding system comprising: a "light" or "base" level mainly focused on structural aspects, such as page divisions, and one or two more advanced levels. The basic level is applied to all texts, while the other levels are progressively implemented in the collection. To encourage the adoption of the XML/TEI standard, despite limited resources, we recommend considering starting the editorial workflow with the automatic transcription of the reference texts and then using automatic tools for conversion into the XML/TEI format.

The text collections we selected as pilots are:

- **ALIM** (Archive of the Italian Latinity of the Middle Ages), an ongoing project whose main goal is to collect and publish all Latin texts produced in Italy during the Middle Ages. ALIM offers reliable TEI-based editions of both literary and documentary texts, making ALIM an invaluable resource for philologists, historians and literary scholars.

- **Biblioteca Italiana**, a digital library of more than 3000 texts representative of Italian heritage spanning between the Middle Ages and the 20th century. All texts are associated with detailed bibliographic metadata, while texts within the BibIt corpus are also encoded in XML/TEI .

- **Musisque Deoque** (MQDQ), a digital archive of scholarly edited poetic texts in Latin. The project's main corpus includes 642 works, spanning between the origins and the 7th century AD and amounting to a total of 343.709 verses and about 2.300.000 tokens. The 'Poeti d'Italia in lingua latina' corpus comprises Latin poetic texts written in Italy during the Middle Ages and is composed of 3.200.000 tokens. The scholarly editions of the texts are encoded in XML/TEI and are characterised by rich critical apparatuses and detailed metrical analysis.

- **BUP - Digital Humanities**, an editorial collection of digital scholarly editions and databases. All resources are described with detailed bibliographic metadata. The digital scholarly editions are encoded in XML/TEI with a very rich markup and are published using the open source visualisation software EVT.

ALIM

The ALIM database (Archivio della Latinità Italiana del Medioevo, website: https://alim.unisi.it/) aims to collect all texts written in Latin in Italy during the Middle Ages. Initiated in the mid-1990s[30] and continuing to this day, ALIM is the product of a collaborative effort among six Italian universities: Verona (coordinated by Prof. Antonio De Prisco), Suor Orsola Benincasa Naples (Prof. Edoardo D'Angelo, who is also the national coordinator of the entire project), Palermo (Prof. Giorgio Di Maria), Ca' Foscari Venice (Prof. Marina Buzzoni), Siena-Arezzo (Prof. Francesco Stella), and Basilicata (Prof. Fulvio Delle Donne).

The primary objective of the database is to provide scholarly reliable texts, mainly targeting academics such as philologists, literary historians, and historians specialising in medieval culture, science, and institutions. Additionally, the text collection serves as a valuable teaching resource.

In addition to its primary objective, ALIM contributes to the European Dictionary of Medieval Latin, sponsored by the Union Académique Internationale in Brussels. To support this effort, ALIM's digital library includes both public and private documentary sources from the 8th to the 15th centuries. The ALIM website also features "Lexicon," a tool for lexical analysis that generates indexes, frequency diagrams, and comparisons of linguistic forms across texts.

The current version of ALIM's digital library (ALIM 2.0) debuted in 2016, following a comprehensive overhaul of the original version's architecture and editorial workflow.[31] ALIM 2.0 employs XML/TEI (P5) encoding for both documentary and literary sources and is exclusively accessible via its website. This website is constructed using Muruca, a modular framework designed for creating and publishing digital libraries. All texts are open access, allowing users to freely download them in HTML, XML, TXT, or PDF formats.

The website's homepage highlights two key components: diverse access points to the texts and various collections. The access points include the authors' index, works' index, and searchable lists for literary and documentary sources. The collections function as thematic sections within the ALIM database, showcasing unique and rare texts. However, the homepage lacks crucial information such as citation guidelines and the date of the last update.

ALIM contains 781 literary sources and 6,654 documentary sources. While these numbers are not explicitly stated in the database's presentation, they are evident from the respective lists. Both literary and documentary source lists offer full-text and proximity search functions. Users can filter the literary texts by author, work, type (prose or verse), historical period, genre, and collection. The documentary sources can be filtered by collection, historical period, place,

---

[30]The ALIM project has been sustained by PRIN funds (granted by the Italian government) from 1996 to 2017, with additional support from the National Research Council (CNR), the National Academic Union (UAN), and the participating universities.

[31]ALIM 1.0 remains accessible at http://www.alim.dfll.univr.it/.

and "corpus" (referring to the codex or other document where the sources are preserved).

Each text is accompanied by a set of "bio-bibliographical" metadata. These include: work title, author name, historical period, text type (e.g., letter), record entry date (often empty), style (prose or verse), genre (e.g., comedy, lyric poetry), source type (documentary or literary), text dimensions (total words and characters), encoding information, notes (editorial and similar), and source. The bibliographic reference of the text source is provided in full. While the source type (e.g., "critical edition") is not explicitly listed among the metadata, the database's general presentation states that authoritative and scholarly editions were used as text sources.[32] For documentary sources, the style and genre are not specified.

Figure 1: the text of Dante Alighieri's De vulgari eloquentia, as presented in the ALIM database.

The database presentation is concise yet effectively illustrates its contents and guides users through the website's features. The documentation provides detailed information on the website's architecture and, most importantly, the data model for TEI text encoding. This includes the schema in DTD and RNG formats, a teiHeader template, and a comprehensive handbook. ALIM's TEI encoding procedure comprises three distinct levels:

- a "base-level" that marks up the text structure;

- a "medium-level" that encodes semantic features such as quotations, names, places, and works;

---

[32]Metadata attached to documentary sources are slightly different. However, it was not possible to analyse them thoroughly, as the consultation of documentary sources is currently malfunctioning.

- an "advanced-level" that adds an "editorial" layer, addressing abbreviations, corrections, and critical notes.

The base-level encoding has been applied to all texts, while the medium-level has been implemented for only some. The advanced-level remains experimental, having been applied to just a limited selection of texts. The documentation fails to specify whether an existing model or standard was used to define the metadata. Additionally, it lacks a clear indication of the ALIM database's completion level and future plans for long-term preservation. Moreover, when switching to the English version of the website, some content remains in Italian, such as the metadata labels.

FAIRness evaluation summary

Table 1: Analisi FAIR della collezione

	Strengths	Improvements

F		• Assign a persistent identifier (e.g., DOI,[33] Handle)[34] to the text collection.
		 • Register and index the text collection in a searchable resource.
		 • The website's homepage lacks essential information:
		   – citation guidelines;
		   – date of last update.

A	• The text collection is open access. • The data dump is deposited in a certified repository.	• Consider implementing a RESTful API.
I	• Users can download all texts.	
R	• Documentation is concise and effective.	• Consider changing license to CC BY SA.

---

[33]https://www.doi.org/the-identifier/what-is-a-doi/.
[34]https://www.handle.net/index.html.

Biblioteca Italiana

Biblioteca Italiana (website: http://www.bibliotecaitaliana.it/) is a digital library that collects representative texts of Italian tradition and literature from the Middle Ages to the 20th century. Originating in 1996 from the work of a consortium of sixteen Italian universities, Centro Interuniversitario Biblioteca Italiana Telematica (CIBIT), the digital library is now directed by Beatrice Alfonzetti and Stefano Asperti, with Amedeo Quondam as its founder and president. The current website, launched in 2016, is hosted by the Department of Literature and Modern Cultures at La Sapienza University of Rome. This version of the digital library was developed in collaboration with the Italian Ministry of Culture,[35] the Central Institute for the Union Catalogue of Italian Libraries and Bibliographic Information (ICCU),[36] and BEIC - Biblioteca Europea di Informazione e Cultura.[37]

While it primarily targets scholars, Biblioteca Italiana's text collection is also a valuable teaching resource. It offers three distinct sections or collections. 'BibIt', the core component of Biblioteca Italiana, contains 1,632 complete works in text format. These are based on authoritative scholarly editions, encoded in XML/TEI , and are all freely accessible, downloadable, and searchable. BibIt provides tools for contextual searches, proximity searches, full-text searches, and the creation of dynamic concordances in KWIC (Keywords in Context) format. 'Scrittori d'Italia' is the digital reproduction of the eponymous book series, founded in 1910 by the Laterza publishing house. It comprises 179 works (in 287 volumes), totalling 125,171 text-images. Lastly, 'Incunaboli in volgare' features 1,604 Italian incunabula, amounting to more than 200,000 images. These incunabula are freely available for online consultation and are accompanied by technical and management metadata.

Biblioteca Italiana's text collections are accessible exclusively through their website, which is based on Muruca. Users can access all three sections via a unified list and search form. The list offers filtering options by collection, genre, historical period, author, and title. While metadata searches (e.g., title, author, publisher, etc.) are available across all collections, full-text search functionality is limited to the 'BibIt' collection.

As previously mentioned, only works from the 'BibIt' collection are available as XML/TEI files. The website offers a handbook for the XML/TEI (P4) encoding, along with the corresponding DTD schema and its extensions. The incunabula and volumes from 'Scrittori d'Italia' are accessible solely as images, which users can navigate using interactive indexes, as shown in Figure 2.

---

[35]http://librari.beniculturali.it/it/.

[36]https://www.iccu.sbn.it/it/.

[37]https://www.beic.it/.

The images from both 'Incunaboli in volgare' and 'Scrittori d'Italia' collections do not implement the International Image Interoperability Framework (IIIF)[38], and their licencing terms and reuse permissions are not clearly specified.

Figure 2: the incunabulum of the Convivio by Dante Alighieri, published within Biblioteca Italiana.
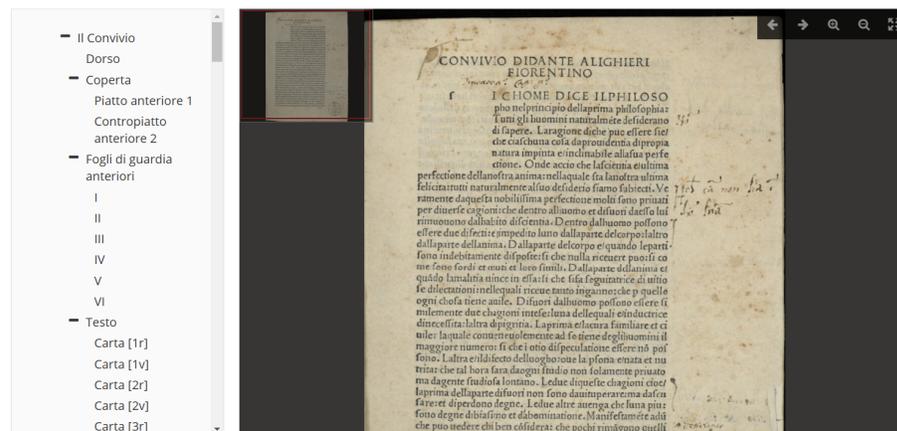
Each text is associated with a concise yet comprehensive set of metadata. Some metadata fields are common across all three collections (such as author, title, genre, and historical period), while others are specific to the collection to which the text belongs. For the complete description of the print sources of the published texts, Biblioteca Italiana relies on integration with SBN for the 'Bibit' and 'Scrittori d'Italia' sections; for the incunabula section, it refers to the descriptive record of the Incunabula Short Title Catalogue (ISTC) of the British Library and to the General Index of Incunabula in Italian Libraries (IGI). The print sources used in the BibIt collection are typically critical editions, though this is not explicitly stated in the metadata. Biblioteca Italiana's editorial work is limited to XML/TEI -encoding of the published texts from these print sources, without additional editing.

The architecture of Biblioteca Italiana is based on the logical model OAIS (Open Archival Information System).[39] The metadata management system is based on the METS[40] framework, integrated with a series of auxiliary sub-schemas, including the MODS[41] schema for the bibliographic description of digital documents and their original sources, and the MIX[42] schema for the

---

[38]https://iiif.io/.
[39]See references in the text collections' standards and guidelines section.
[40]http://www.loc.gov/standards/mets/.
[41]http://www.loc.gov/standards/mods/.
[42]http://www.loc.gov/standards/mix/.

Figure 3: the metadata associated with the text of Vittorio Alfieri's Antigone, published in the BibIt collection.

Titolo
Antigone

Autore:	Alfieri, Vittorio	Pubblicazione:	Roma: Biblioteca Italiana, 2003
Genere:	Letteratura teatrale	Periodo:	700

Descrizione fonte cartacea

Autore:	Alfieri, Vittorio	Titolo:	Tragedie
Pubblicazione:	Firenze: Sansoni, 1985	Altra Responsabilità:	Toschi, Luca
Record SBN:	IT\ICCU\CFI\0021280		

Descrizione versione digitale

Dimensione:	91390 bytes

Links

File XML   File METS   File MAG   Vai al testo

description of technical metadata for digital images. From the metadata stored in the METS records, metadata in TEI Header format are generated and inserted into the XML/TEI files of the texts, as well as metadata in MAG[43] format—a schema developed at the ICCU—for interoperability with other national projects. The metadata for each text is available for download in various formats. Notably, the text and its metadata are presented on separate web pages. While the metadata page includes a link to the text page, the text page lacks a direct link back to the metadata page.

Each text appears to have a unique identifier for internal use. This identifier is present in the URL of the webpage displaying the text. However, the URL is not provided as a permalink in the metadata associated with the text.

---

[43]http://www.iccu.sbn.it/opencms/opencms/it/.

FAIRness evaluation summary

	Strengths	Improvements
F		• Assign a persistent identifier (e.g., DOI,[44] Handle)[45] to the text collection. • Register and index the text collection in a searchable resource (e.g., OpenAIRE Explore).[46] • Assign each text a persistent identifier. • The website's homepage lacks essential information about the database: – citation guidelines for the database; – licence terms for the available texts; – clear indication of the database's status, completion level, and future plans.
A	• The text collection is open access. • The data dump of the text collection is deposited in a certified repository in a standard format. • The texts in the collection are accessible via an intuitive and user-friendly search form.	• Deposit the data dump in a certified repository such as Zenodo, to ensure its long-term preservation and access. • A RESTful API would be very valuable for retrieving text metadata and performing textual queries.
I	• Metadata are encoded in standard formats. • Texts are encoded in the standard XML/TEI format. • Users can download the metadata of all texts in the collection and the TEI-encoded texts from the BibIt collection.	• To enhance interoperability and long-term durability, update the texts' encodings from version P4 to version P5 of the TEI schema. • The 'Incunaboli in volgare' and the 'Scrittori d'Italia' subcollections would benefit from the IIIF framework, allowing for image reuse.

R

- The documentation concisely yet effectively presents the text sub-collections and their contents.
- The technical documentation details the encoding format and schemas used.

- Indicate licence terms for the images.
- Maintain a proper changelog to document all updates and changes.
- In the documentation, provide information about future plans and long-term maintenance.

Musisque Deoque

'Musisque Deoque: Un archivio digitale di poesia latina, dalle origini al Rinascimento italiano' (MQDQ, website: https://www.mqdq.it/) is a comprehensive database of Latin poetry. Established in 2005,[47] this digital archive was designed to enable researchers to explore texts not only in their authoritative versions but also to examine textual variations found in critical apparatuses, as stated on the database's homepage.

Over the years, the initial database expanded with new textual archives: 'Carmina Latina Epigraphica' for Latin inscriptions, 'Poeti d'Italia in lingua latina' for Italian poets who composed in Latin, and 'Hellenica' for ancient Greek poetry.

In 2018, the Venice Centre for Digital and Public Humanities relaunched the project as 'MQDQ Galaxy' to ensure long-term sustainability of, and open access to, the MQDQ archives. This recent funding enabled API access to the MQDQ database. The project now involves a large group of researchers coordinated by the following Italian universities: Ca' Foscari University of Venice (Prof. Paolo Mastandrea), University of Calabria (Prof. Raffaele Perrelli), University of Parma (Prof. Gilberto Biondi), University of Perugia (Prof. Loriano Zurli), and University of Naples Federico II (Prof. Valeria Viparelli).

The MQDQ website is available in Italian and English. However, its landing page lacks some crucial information: citation guidelines for the digital archives, a standard licence for the texts, and an external identifier such as a ISSN code. The footer provides an official email address and a brief update date. A more detailed changelog is available on the Ca' Foscari University website,[48] though it is not formatted as a proper changelog.

MQDQ's primary aim is to provide an extensive corpus of Latin poetic texts for advanced lexical searches, including both edited texts and critical apparatuses. While primarily targeting scholars, MQDQ's authoritative and comprehensive text collection also serves as a valuable teaching resource.

---

[44]https://www.doi.org/the-identifier/what-is-a-doi/.

[45]https://www.handle.net/index.html.

[46]https://explore.openaire.eu/.

[47]Paolo Mastandrea, a professor at the University Ca' Foscari of Venice, initiated the project with four PRIN grants from the Italian Government (1999, 2001, 2005, and 2007).

[48]https://pric.unive.it/projects/mqdq-galaxy/home#c11784.

Users can access the digital archive through various means: a general search interface focusing on lexical and metrical features, an alphabetic index, a chronological index, and specialised search features for lexical and metrical co-occurrences.

From a technical perspective, MQDQ consists of digital editions encoded in XML/TEI . However, the standard and format used for the texts' metadata are not clearly specified. The digital editions are based on previous printed editions. MQDQ editors revise, transcribe, and mark up these printed editions in XML/TEI .

Each text includes information about the base text, the digital edition's editor(s), the data curator(s), and a permalink. Furthermore, each text is accompanied by a critical apparatus, a list of witnesses, the meter, and a metrical scan. Within the text body, portions subject to variation in the tradition are highlighted (Figure 4). When users click on these highlighted portions, the corresponding entry from the critical apparatus appears on the side.

The metrical scans of the texts were developed as part of the Pedecerto[49] project, which created a tool for automatic verse meter analysis and a search function that queries the entire MQDQ database. These metrical scans are available for download from the MQDQ website as XML/TEI files.[50]

Figure 4: the metadata associated with the text of Vittorio Alfieri's Antigone, published in the BibIt collection. https://www.mqdq.it/texts/ALBINOV%7Cfr ag%7C001).

Neither the bibliographic references for the base texts nor those for the witnesses are provided in full or linked to external web resources. Additionally, the website lacks a comprehensive bibliography. The data modelling process is undocumented, and the database's information structure is implicit.

As mentioned earlier, the MQDQ database comprises three distinct archives: Carmina Latina Epigraphica, Poeti d'Italia in lingua latina, and Hellenica. These archives vary in terms of completeness and functionality. Hellenica offers a limited

---

[49]https://www.pedecerto.eu/public/.

[50]https://www.pedecerto.eu/public/pagine/autori.

range of works, accessible through the MQDQ portal via a dedicated alphabetic index[51] and search feature.[52] Poeti d'Italia in lingua latina presents a broader range of works, explorable through chronological and alphabetical indexes, a search feature, and an index organised by metrical scheme.[53] Carmina Latina Epigraphica is a more specialised collection, accessible through its own, as well as the platform's shared indexes and search feature. This sub-collection is uniquely presented in detail from a theoretical perspective on a dedicated web page.[54]

The overall extent of the MQDQ text compilation is not explicitly specified. The homepage does not indicate the number of collected texts. However, the Pedecerto website reports that the MQDQ database contains 345,996 lines.[55] It is difficult to assess the database's completeness relative to the research project's goals. While one might assume the project is ongoing and that the database owners intend to preserve and maintain it long-term, the website does not explicitly state these intentions or future plans.

The complete MQDQ database can be downloaded as a zipped folder containing XML/TEI files from the ILC4CLARIN repository.[56] These files are available under the CC BY-SA 4.0 licence. The downloadable version on the CLARIN platform dates from 2021, and there is no indication of a more recent version available for download. However, the website states that the database was last updated on 19 September 2025.

The MQDQ database can be queried through RESTful APIs developed in recent years as part of the MQDQ Galaxy project. Access to these APIs is restricted to authenticated users. The RESTful API provides access to data such as the total number of authors, verses by meter, and multiword occurrences.

---

[51] https://mizar.unive.it/hellenica/public/indici/autori/idautori/1.

[52] https://mizar.unive.it/hellenica/public/ricerca/avanzata.

[53] https://www.poetiditalia.it/public/indici/metri.

[54] https://www.mqdq.it/public/ce/presentazione.

[55] Information taken from https://www.pedecerto.eu/public/pagine/arte.

[56] https://dspace-clarin-it.ilc.cnr.it/repository/xmlui/handle/20.500.11752/OPEN-555.

FAIRness evaluation summary

	Strengths	Improvements
F	• Each text is assigned a unique URL, facilitating precise referencing. • The text collection is indexed in OpenAIRE Explore. • Texts metadata include essential scientific information (author, title, editors, base text reference, witnesses, and previous editions).	• Assign a persistent identifier (e.g., DOI) to the database. • The website's homepage lacks essential information about the database: − citation guidelines for the database; − date of the last update; − clear indication of the database's status, completion level, and number of collected texts; − link to the downloadable version of the text collection; − link to the API.
A	• The text collection is open access. • The database offers multiple access points: a graphical user interface (MQDQ website) and RESTful APIs. • The data dump of the text collection is deposited in a certified repository in a standard format. • Both simple and sophisticated search queries are supported. • The website features comprehensive indexes for easy navigation.	• Consider implementing a RESTful API to enable easier text retrieval. • Regular updates to the downloadable version of the database should be implemented, ideally every six months to a year.
I	• Texts are encoded in the standard XML/TEI format. • Users can download all texts of the collection and the metrical scans.	
R	• The licence under which the downloadable data dump is available is stated clearly.	• Clearly state in the website's home page the licences under which the texts are available. • The scientific documentation needs expansion, detailing the editorial criteria, data model, and metadata standards used. It should also include a bibliography of the base texts and witnesses.

BUP - Digital Humanities

The editorial collection Digital Humanities, published by BUP (Basilicata University Press) and directed by Prof. Fulvio Delle Donne, focuses on digital scholarly editions and databases. The publications included in the collection are developed within the Department of Human Sciences, primarily by PhD candidates on the 'Culture e Saperi dell'Europa mediterranea dall'Antichità all'Età contemporanea' doctoral program. The editorial process is guided by a board of scholars coordinated by Prof. Alessandro Di Muro and comprising Dr Cristiano Amendola, Prof. Teofilo De Angelis and Dr Martina Pavoni.

Although it is not a traditional collection of texts, the Digital Humanities project and its digital scholarly editions are an interesting pilot for the ATLAS project for several reasons. Firstly, it demonstrates how digital products can be enhanced and promoted by treating them with the same care as printed editions. For example, each digital scholarly edition is assigned an ISBN code. If an edition is available in multiple formats, such as XML/TEI and PDF, then each format receives its own code. Additionally, the editorial collection itself has an ISSN code. Secondly, the collection has developed a publishing framework and editorial workflow based on XML/TEI encoding and EVT visualisation software, enabling the production of high-quality digital scholarly editions at low cost.

The collection's website provides comprehensive metadata for each digital scholarly edition, including the author and title of the edited work, the editor's name, ORCID[57] and brief biography, a short author biography, keywords, an abstract, page count, languages used, ISBN codes, links to the PDF and XML/TEI editions (visualised with EVT), and the publication date.

The Digital Humanities collection also includes eight databases. Although it is not stated explicitly, most of these databases are still under development or incomplete, while others are temporarily inaccessible. For this reason, they will not be analysed individually. Each database is presented on the editorial collection's website with metadata similar to that of digital scholarly editions. Some databases collect texts encoded in XML/TEI format. Additionally, the databases seem to have been implemented using a custom, ready-to-use publishing framework developed by the University of Basilicata's technical support team.

---

[57]ORCID (Open Researcher and Contributor ID) is a unique identifier that individuals can use when engaging in research and scholarly activities. Provided by the eponymous global, not-for-profit organisation, these identifiers enable transparent and trustworthy connections between researchers, their contributions and their affiliations. https://orcid.org/.

FAIRness evaluation summary

	Strengths	Improvements
F	• The digital scholarly editions in the collection are assigned a persistent identifier. • All publications are introduced by a well-structured set of comprehensive metadata. • The digital scholarly editions are indexed in a searchable resource.	
A	• All publications in the collection are open access.	• Deposit all publications in a certified repository for long-term preservation.
I	• The digital scholarly editions in the collection are encoded in the standard XML/TEI format and can be downloaded in PDF.	
R	• The licences under which the publications are available are stated clearly.	• Indicate, especially for the databases, the status and level of completeness.

Digital scholarly editions

To introduce digital scholarly editions we rely on a well-known definition [Sahle 2016]:

> "Scholarly digital editions are scholarly editions that are guided by a digital paradigm in their theory, method and practice".

According to Sahle, one of the major implications of the above mentioned "digital paradigm" is that a digital edition "as a publication is a process rather than a product. It grows incrementally not only before its final release, but also during its availability to the public".

The digital paradigm has not changed the editorial workflow completely [Mancinelli and Pierazzo 2020], but in order to produce a digital edition new activities are necessary alongside the ones usually applied in the printed-oriented workflow (e.g., transcription of primary sources, collation, etc.).[58] These new activities, namely digitisation, data modelling, and encoding, are part of the "source-output conceptual and technological model" (*ibid.*), where the source is

---

[58]See also: Leonardi, Lino. 2021. "Filologia digitale del Medioevo italiano." Griseldaonline 20 (2): 77–89. https://doi.org/10.6092/issn.1721-4777/12817.

an annotated text, usually encoded in XML/TEI , from which various outputs in multiple formats are derived.

Documents' digitisation is a necessary step when they are not already available in electronic form. Digitised documents serve both as a starting point of the editorial workflow to (automatically) transcribe the texts and as part of the final product, in order to provide readers with the possibility of verifying editors' readings first-hand. Thanks to IIIF and its compliant viewers (such as OpenSeadragon,[59] Mirador,[60] and Universal Viewer),[61] it is now easier to integrate the original documents' and witnesses' facsimiles in a digital edition.

The encoding step aims at producing one or multiple files containing the data of the edition, i.e., the text(s) and the annotations that document editorial interventions and allow readers to interpret the text(s). The TEI guidelines have gradually emerged as the "de facto standard" [Mancinelli and Pierazzo 2020]. However, choosing XML/TEI as encoding format is not always straightforward. Editors often desire to adhere to the TEI standard but lack the technical skills for implementation. When funds and resources are available, this issue is frequently addressed by either entrusting the encoding to researchers with digital philology expertise or by developing bespoke editing tools that facilitate encoding, also known as "haute couture editing" [Pierazzo 2019]. The following graph [Kurzmeier et al. 2024] compares the number of digital scholarly editions, catalogued in Franzini's Catalogue of Digital Editions, that are based on XML/TEI against the number of those that are not TEI-compliant or not even in XML format divided by their initiation year. In the first years shown in the chart, corresponding to the period before and immediately after TEI's inception, the numbers are understandably low. The graph's trend likely reflects biases in the catalogue's compilation timing and methods.

Considering the trend depicted in the graph and the fact that the TEI guidelines have been continuously supported and developed by a large international scholarly community, it is logical to strongly recommend XML/TEI as the standard text encoding format. In case of peculiar types of texts, other TEI -compatible and XML-based models may be used, namely: MEI for music notation, Epidoc for inscriptions, and CEI for medieval charters.

Questions may arise about the actual FAIRness of a digital edition using XML/TEI , particularly regarding interoperability and reuse [Szleszyński et al. 2025]. The TEI guidelines offer flexibility in element interpretation and usage, potentially leading to diverse encodings of identical textual phenomena across editions. However, this issue can be readily addressed by providing users with the applied schema and clearly describing editorial criteria in the documentation.

---

[59]https://openseadragon.github.io/.
[60]https://projectmirador.org/.
[61]https://universalviewer.io/.

Figure 5: a graph comparing the number of digital scholarly editions based on XML/TEI (1) with those that are neither TEI -compliant nor in XML (0.5 and 0, respectively), organised by their initiation year.
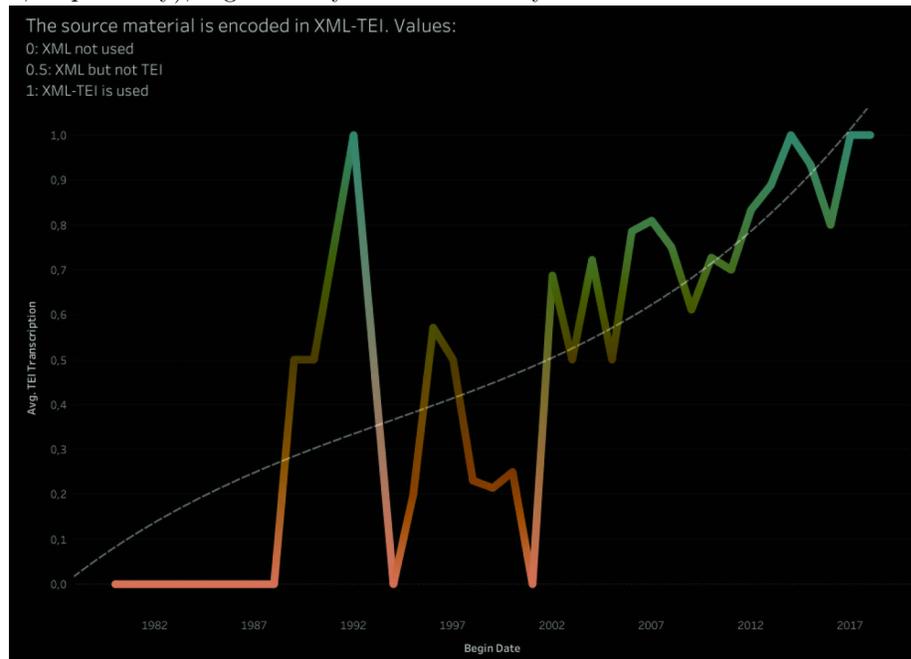
Addressing the initial concern of editors struggling with TEI , two strategies are now easily adoptable. First, use editing software that automatically exports data in XML/TEI , favouring generic tools over custom-developed ones. Numerous free tools are available for various editorial tasks like collation, transcription, and lemmatisation. The second approach involves converting data from other formats to TEI . TEI has long offered the conversion tool OxGarage, recently relaunched as TEIGarage,[62] for this purpose.

Finally, the last step of the "source-output" model is the presentation (or data visualisation). Various outputs can be derived from the edition's source file(s), e.g., PDF, e-book formats. However, most of the time the output is an HTML-based website, often generated by software tools such as EVT and TEIPublisher.[63] To foster reuse and enhance interoperability, a digital edition should provide users with downloadable outputs in multiple standard formats [Spadini and Losada Palenzuela 2025] and also strive to connect with and build upon existing scholarly work.

The digital editions we selected as pilots are:

- **VaSto - VArchi, STOria fiorentina. Edizione digitale**, an international project that aims at producing the annotated digital edition of the *Storia fiorentina* by Benedetto Varchi (1503-1565). The edition is TEI compliant and visualised through the open source software EVT, which has been customised to provide support for genetic editions.

- **Codice Pelavicino Digitale** is a project providing a digital scholarly edition of a historical document of crucial importance for the Italian Cultural Heritage with regard to historical research of the XII-XIII century. The edition is TEI compliant and visualised through the open source software EVT.

- **Digital Edition of Aldo Moro's works**, a critical edition of Aldo Moro's published and unpublished texts, with their historical introduction. The edition reused well-known ontologies in the CH domain and is available in RDFa and XML/TEI format.

- **Leges Langobardorum**, the TEI -based diplomatic-interpretative editions of two witnesses of the Edict of Rothari, produced within a wider editorial project that stems from the ALIM digital library and aims at creating a complete critical edition of all Langobardic laws' main witnesses.

---

[62]https://teigarage.tei-c.org/.

[63]Created in 2015 by the company eXist Solutions in the context of the Swiss e-codices association, TEIPublisher (https://teipublisher.com/index.html) allows you to publish XML files on the web (mainly according to the TEI standard), generating a website, or convert documents to other formats, such as PDF and ePUB.

VaSto - VArchi, STOria fiorentina. Edizione digitale

The VaSto project (VArchi, STOria fiorentina. Edizione digitale, website: ht tps://dharc-org.github.io/progetto-vasto/) aims to publish the uncensored version of Benedetto Varchi's *Storia fiorentina*, as attested in the manuscript Corsiniano 1532 (Biblioteca Nazionale dei Lincei e Corsiniana, Rome). VaSto is a collaborative effort between /DH.arc[64] (University of Bologna) and Concordia University, with support from the CarisBo foundation, led by professors Dario Brancato and Paola Italia.

The project's primary output is an XML/TEI -encoded digital edition of the *Storia fiorentina*. To date, only the work's preface has been published as a pilot version. The initial release in May 2020 used the first beta version of EVT 2.0 software, while the current version, released a month later, employs the second beta. The edition is both diplomatic—presenting the text as it appears in the Corsiniano 1532 manuscript alongside digital images—and critical-genetic, incorporating editorial interventions that restore the author's intended meaning and enhance readability. EVT's interface allows users to toggle between these two edition levels or compare them side-by-side, as shown below.

Figure 6: side-by-side comparison of the diplomatic and interpretative transcriptions within the VaSto edition GUI (graphical user interface).
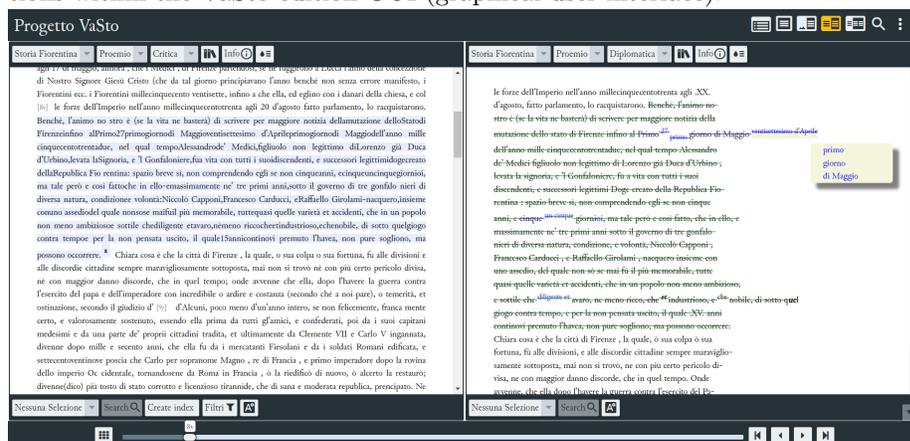
The edition is also "annotated", as stated in the documentation, because names of people and places mentioned in the edited work are marked up and linked to their respective (Virtual International Authority File (VIAF)[65] entries and/or descriptions on external web sources when available.

VaSto is also a collaborative edition, as it was implemented with the participation of students from the *Scholarly Editions* lab held at the University of Bologna in

---

[64]https://centri.unibo.it/dharc/en.

[65]https://viaf.org/en.

Figure 7: the named entities lists within the VaSto edition GUI.
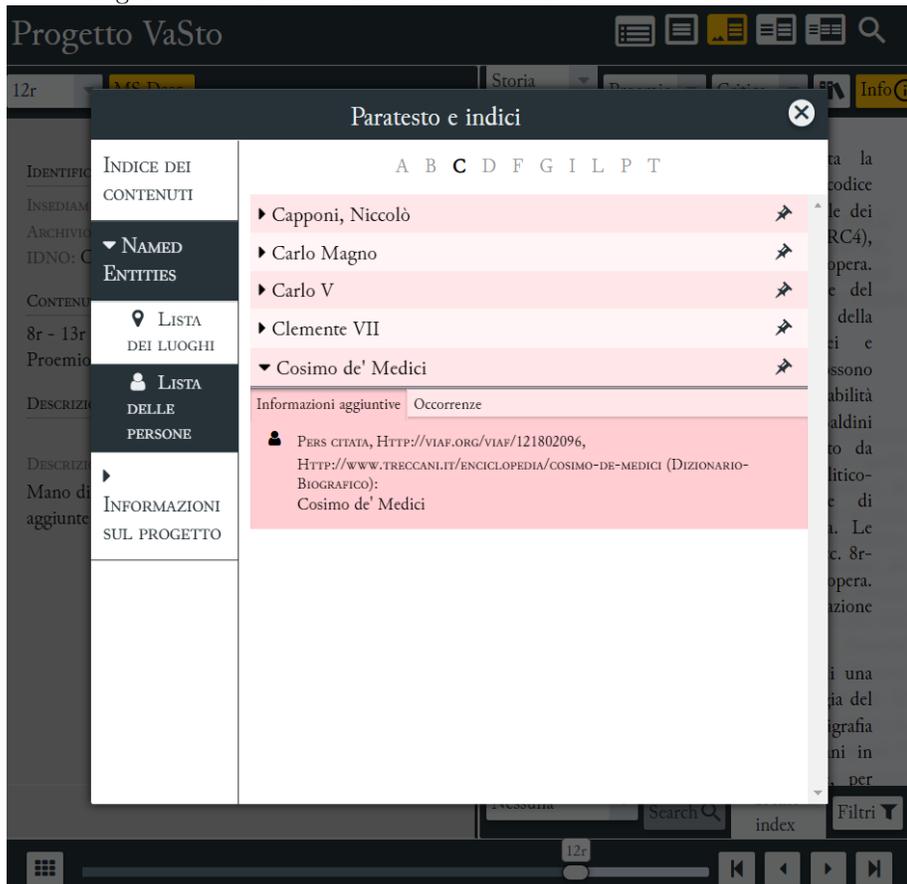
2020. The edition's XML/TEI document is freely accessible and downloadable from the project's GitHub repository.[66] This single document contains the diplomatic edition, the critical edition, and the annotated named entities.

The editors aim to complete the edition of the *Storia fiorentina*, presenting both the manuscript version of the author's last will and the censored version that has become the work's *vulgata*. However, the documentation lacks information about the full edition's publication timeline and the current status of the editorial work.

In addition to the digital edition, the VaSto project has produced three secondary outcomes:

- a timeline developed with TimelineJS,[67] depicting the writing process of the *Storia fiorentina* and key historical events described in the work;

- a map created with Leaflet,[68] illustrating the places mentioned in the *Storia fiorentina*;

- VastoCollection, which showcases digital images of the RC4 manuscript and portraits of people mentioned in the *Storia fiorentina*. This collection is built using Omeka[69] and IIIF, with catalographic records structured according to the Dublin Core standard.

The website provides comprehensive documentation from both philological and technical perspectives. The philological section[70] offers a description of the edited text and its history, the author's biography, a bibliography, and a list of witnesses—though these are not linked to external web sources. The technical section[71] includes a detailed guide for interacting with the edited text. That section also explains how the EVT 2.0 software was customised to meet the project's requirements, which is particularly valuable for EVT's developers as it fosters synergy with the user community. While the documentation does not provide the encoding schema, it lists the TEI elements used to represent editorial (corrections, critical notes, etc.) and material (page and line breaks, deletions, etc.) phenomena. The website's copy requires revision, as typos are present— for instance, "IIF" instead of "IIIF". The credits page is well-structured and comprehensive, presenting not only the editorial team and collaborators but also the official releases, attribution, and licence under which the edition is available (Creative Commons Attribution-NonCommercial-NoDerivs 3.0 Unported).

---

[66] https://github.com/ValentinaPasqual/ProgettoVasto/blob/master/evt2beta2/data/text/pilot_proemio.xml.

[67] https://timeline.knightlab.com/.

[68] https://leafletjs.com/.

[69] https://www.omeka.net/.

[70] https://dharc-org.github.io/progetto-vasto/Progetto.html.

[71] https://dharc-org.github.io/progetto-vasto/Funzionalit%C3%A0.html.

FAIRness evaluation summary

	Strengths	Improvements
F	• Comprehensive documentation includes attribution and version history.	• Register the edition in a searchable resource to increase its findability.
A	• The edition is open access. • Users can access the TEI-encoding of the edition.	• Deposit the edition in a certified repository for long-term preservation.
I	• The edition is encoded in the standard XML/TEI format. • Names mentioned in the edition are linked to authority records.	• Publish the edition's schema as a downloadable document.
R	• The licence under which the edition is available is stated clearly.	• Provide a clear indication of the edition project's status and timeline. • Polish the website's copy for clarity.

Codice Pelavicino Digitale

Codice Pelavicino Digitale (website: https://pelavicino.labcd.unipi.it/) is the digital documentary edition of the eponymous codex preserved at the Archivio Capitolare Lunense of Sarzana. Professor Enrica Salvatori (University of Pisa) initiated the project in 2014. Over the years, numerous collaborators have joined the extensive editing team, contributing to the codex transcription, encoding, and development of consultation tools. The edition is now complete, as clearly stated on the website's homepage, featuring 529 documents from the codex and a comprehensive set of accompanying materials. Currently, the edition undergoes periodic corrections, revisions, and progressive feature improvements.

The project's primary outcome is an open-access, TEI -encoded digital edition, visualised using EVT 1.0 software (version 1.3.2). The graphical user interface allows users to read diplomatic transcriptions of the documents alongside digital images of the respective folios. Each document is prefaced by a detailed, toggleable description. Within the transcribed text, users can interact with editorial notes (indicated by dots) and named entities.

Named entities mentioned in the text, specifically places, people, families, and organisations, can be explored via lists. Each entry in these lists includes a description and links to occurrences in the documents. In addition to the named entities' lists, users can access a glossary and a chronological index presenting the historical events mentioned in the documents, ordered by date or document. A full-text search function is also available.

Figure 8: side-by-side comparison of the text and the manuscript image within the Codice Pelavicino GUI.
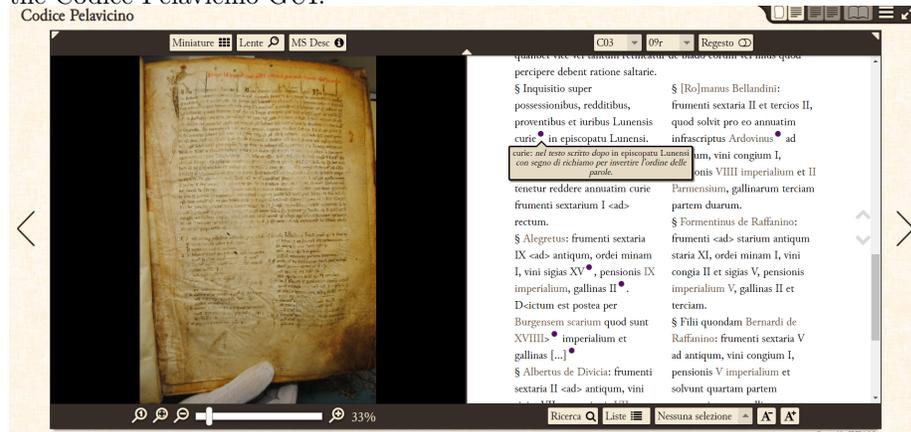
To facilitate data reuse, the edition's XML/TEI files are available for download under the Creative Commons CC BY 4.0 licence from the ILC4CLARIN repository.[72] The named entities' lists can also be downloaded from the same platform.[73] However, the manuscript's digital images are not available due to copyright restrictions. Implementing the manuscript's facsimiles with IIIF and the named entities as linked open data would significantly enhance this already commendable work.

The edition's website offers a rich set of materials:

- interactive map of places mentioned in the codex;

- named entities' search form;

- list of all charters contained in the codex;

- notaries' list;

- comparative table of current and previous document numbering.

- timeline, developed with the TimelineJS tool, illustrating historical events mentioned in the codex.

- A TEI -based glossary of terms that were more challenging to interpret.

The documentation is comprehensive from both philological and technical perspectives, including attribution, licences, bibliography, and publications. The philological section clearly outlines the editorial criteria and standards applied to implement the digital edition, namely the 'Guidelines for Editors of Scholarly

---

[72]http://hdl.handle.net/20.500.11752/OPEN-1012.

[73]https://dspace-clarin-it.ilc.cnr.it/repository/xmlui/handle/20.500.11752/OPEN-1011.

Editions'[74] by the Modern Language Association and the 'Criteria for Reviewing Scholarly Digital Editions', version 1.1[75] published by the Institut für Dokumentologie und Editorik. While the technical section does not provide a downloadable TEI schema, it offers a detailed explanation of the elements used, illustrating how peculiar aspects of the texts, such as coins, professions, and notaries' signs, were encoded. This information can be invaluable for other edition projects of similar documents. Lastly, the documentation also illustrates how the visualisation software tool, EVT 1.0, was customised to meet the edition's requirements.

FAIRness evaluation summary

	Strengths	Improvements
F	• The edition is indexed in a searchable resource.	
A	• The edition is open access. • Users can access the TEI-encoding of the edition. • The data dump of the edition is deposited in a certified repository. • The edition's website provides indexes, search features, and various visualisation tools that enhance the discoverability of the contents.	
I	• The edition is encoded in the standard XML/TEI format. • Names mentioned in the edition are linked to authority records.	• Publish the manuscript's digital images using IIIF to enhance accessibility and interoperability. • Implement named entities as linked open data to enhance the already high-quality scientific work.
R	• The licences under which the edition and all its related materials are available are stated clearly.	

---

[74]http://www.mla.org/cse_guidelines.
[75]http://www.i-d-e.de/aktivitaeten/reviews/criteria-for-reviewing-scholarly-digital-editions-version-1-1.

Digital Edition of Aldo Moro's works

The National Edition of Aldo Moro's Works (website: https://aldomorodigi
tale.unibo.it/) is a collaborative effort involving numerous Italian universities
and partner institutions. While scholars from various Italian universities curate
the edition, a team of researchers from the University of Bologna manages
its technical-scientific implementation. Initiated in 2021 and still ongoing, the
project aims to publish all of Aldo Moro's works, both published and unpublished.
To date, 806 works have been released across eight volumes, organised into two
main sections: the first covering Moro's religious, journalistic, and political
writings, speeches, and interviews; the second dedicated to his academic works.
Both sections follow a chronological arrangement.

The published works are freely accessible on the edition's website. A banner
displayed on the landing page informs users that the edition is under development,
with new content and features forthcoming. Notably, the edition's bibliographic
reference, complete with ISBN code and DOI, is prominently displayed in the
page footer.

Users can access works through a search function, analytical indexes for people,
places, organisations, and bibliographic references mentioned by Moro, or a
step-by-step interface guiding them through the edition's volumes. Navigation
proceeds from section to volume, tome, and finally to individual works. Each
tome is preceded by its bibliographic reference and links to an introduction and
historical-critical note, which are displayed in different web pages, each with its
own DOI.

Figure 9: presentation of a tome of Aldo Moro's works within the digital edition
GUI.

Alongside the title and the edited text, each work's edition comprises:

- an introductory section with an abstract, the full bibliographic reference of the edited work and of the witnesses;

- interactive lists of people, places, organisations, citations, and bibliographic references mentioned in the edited text, along with their occurrence counts;

- a metadata sheet detailing themes, data curator, author's roles, typologies, document status, date, event location, identifier, licence, and additional notes.

The texts are available for download in XML/TEI, HTML-RDFa, and PDF formats. The interface facilitates sharing via email and social media, and provides a copy function for the work's URL. Users can search works by title and keywords, and filter them based on document type, themes, Moro's roles, date, and publication status.

Beyond the corpus edition, the project has produced an RDF dataset of "structural, intertextual, and contextual data related to Aldo Moro's works". This knowledge base was used to design the edition's consultation interface and its search functionalities. The Turtle-encoded dataset is freely accessible and downloadable from the Zenodo repository.[76] Controlled vocabularies for roles, themes, and document types are also available on Zenodo in Turtle format. This data is integrated into the edition's documents using RDFa (Resource Description Framework in Attributes),[77] which embeds structured data within web pages. The added information includes:

- licence, persistent identifier (DOI), and bibliographic citation of the document;

- references to people, places, organisations, works cited by Moro or researchers, and quotations;

- mentioned entities, their attested forms, authority control based on Wikidata records, and controlled forms of personal names;

- notes comprising commentary by researchers and/or Moro himself.

The data modelling draws on several existing ontologies, including: Bibliographic Reference Ontology (BiRO),[78] Discourse Elements Ontology (DEO),[79] Dublin Core Metadata Terms (DCTerms),[80] FRBR-aligned Bibliographic Ontology (FaBiO),[81] Friend Of A Friend vocabulary (FOAF),[82] and Expression of Core FRBR Concepts in RDF (FRBRcore).[83]

---

[76]http://doi.org/10.5281/zenodo.5592157.
[77]Resource Description Framework in Attributes. https://www.w3.org/TR/rdfa-primer/.
[78]http://purl.org/spar/biro.
[79]http://purl.org/spar/deo.
[80]http://purl.org/dc/terms/.
[81]http://purl.org/spar/fabio.
[82]http://xmlns.com/foaf/0.1/.
[83]http://purl.org/vocab/frbr/core.

The project's workflow is noteworthy for addressing a common challenge in digital edition projects. Works were initially transcribed in Word, then digitised, indexed, enriched with information, and published through a complex process involving multiple contributors. A custom software tool, KWICKWOCKWAC,[84] facilitated this process by converting Word documents to Web pages, enabling text markup, metadata insertion, and linking entities to existing Web authority records. This approach bypasses the frequent issue of editors lacking XML/TEI skills, which often leads to not adopting TEI for text encoding or, as in this case, developing custom editing tools.

The project's website offers comprehensive documentation on the edition's preparation, detailing the workflow, editorial criteria, data modelling, website development, and scientific outcomes. These outcomes encompass the corpus of edited works, the software code, the RDF dataset, controlled vocabularies, technical documentation on data modelling, and other materials produced during the edition's creation. To ensure long-term access and citability, all scientific outputs except the corpus are published on Zenodo under CC BY 4.0 and CC BY-NC 4.0 licences. The website's terms and conditions page provides citation guidelines for the various outputs.

FAIRness evaluation summary

	Strengths	Improvements
F	- The edition is indexed in a searchable resource. - All parts of the edition are assigned a persistent identifier.	
A	- The edition is open access. - All parts of the edition are deposited individually in a certified repository. - The edition's website provides indexes, a search feature, and a user-friendly interface that enhance the discoverability of the contents.	If possible, consider publishing the whole set of the edition files in various formats (XML/TEI, PDF and RDF-a) as a data dump in an external repository in order to facilitate the reuse of the textual corpus as a whole.

---

[84]https://aldomorodigitale.unibo.it/markup/.

	Strengths	Improvements
I	- The edition files are encoded in the standard XML/TEI format. - Users can download the edition files in various standard formats. - The edition files are also accessible as an RDF dataset, employing semantic web technologies to facilitate data reuse and further studies.	
R	- The licences under which the edition and all its related materials are available are stated clearly. - Comprehensive documentation details scientific aspects and technical implementation.	

Leges Langobardorum

The digital scholarly edition of the Leges Langobardorum, edited by Marina Buzzoni and Roberto Rosselli Del Turco, was launched in 2014 within ALIM and developed over the years through various collaborators and funding from the PRIN project "A Digital Scholarly Edition of the 'Leges Langobardorum': Social Memory, Historical Legacy and Valorisation of Cultural Heritage."

In 2024, diplomatic-interpretative editions of two Italian manuscripts containing the Edict of Rothari (the renowned collection of Langobardic laws of high historical interest) were published: Vercelli, Biblioteca Capitolare Eusebiana, CLXXXVIII (ms 2) and Ivrea, Biblioteca Capitolare, XXXIV (5) (ms 3).

The project's ultimate goal is to publish a critical edition of the *Edict* and all the *Leges Langobardorum*, based on complete collation and recensio of extant witnesses. This edition will include a digital glossary for easier consultation, digital facsimiles, and diplomatic editions. Upon completion, the critically established texts of all selected witnesses will be incorporated into the ALIM digital library.

For our purposes, we will analyze the two diplomatic-interpretative editions of the *Codex Vercellensis*[85] and *Codex Eporedianus*[86] (presented above) published on the Edizioni Ca' Foscari platform using the open source visualisation software EVT. These editions demonstrate how by-products produced during the progressive development of an ambitious long-term editorial project can be valued as research results in their own right.

---

[85] http://doi.org/10.30687/978-88-6969-888-0/001.
[86] http://doi.org/10.30687/978-88-6969-889-7/001.

Both editions were implemented in XML/TEI. The TEI document headers describe the chronology of editorial interventions and name the collaborators who contributed to creating the editions. Ideally, these collaborators' names should also appear in the external metadata to increase their visibility and recognize their work. The files underlying the editions, including encoding schemas and those related to the *Leges* project, have been published in a GitHub repository.[87] Publishing the editions in a certified repository would ensure their long-term preservation.

The decision to publish the editions in open access with Edizioni Ca' Foscari is both valid and sustainable. This publisher has implemented a platform that effectively presents all editorial product information—from DOIs and authors' ORCIDs to licenses and ISBNs—while offering downloadable metadata in standard Dublin Core format. Additionally, the EVT software for visualizing digital critical editions is integrated into the platform, with editions downloadable in XML/TEI format. The editions have been registered and indexed in the University of Turin's IRIS AperTO system and in OpenAIRE Explore.

In both editions, the EVT software allows users to consult and compare the diplomatically and interpretatively transcribed texts alongside photographic reproductions of the witness. The visualisation tool also offers a full-text search feature. In the *Codex Eporedianus* text, names of places, people, and Langobardic terms appear highlighted, but clicking these elements reveals no additional information in the dedicated panel, and the named entities list in the table of contents appears incomplete. The *Codex Vercellensi*s edition interface offers the option to highlight these same elements, but their apparent absence in the text may mislead users.

FAIRness evaluation summary

	Strengths	Improvements
F	• Both editions are introduced by a rich set of metadata. • Both editions are indexed in a searchable resource.	• Include contributors' names in the metadata.
A	• Both editions are open access. • Users can access the TEI-encodings of the editions. • A search function and a user-friendly interface enhance the editions' readability.	• Deposit the edition in a certified repository for long-term preservation.

---

[87]https://github.com/RobertoRDT/LegesLangobardorum.

	Strengths	Improvements
I	• Both editions are encoded in the standard XML/TEI format. • The metadata associated with both editions are compliant with an international standard (Dublin Core) and available for download.	• Publish the edition's schema as a downloadable document.
R	• The licence under which the edition is available is stated clearly.	• Allow users to download the editions in multiple standard non-proprietary formats, such as JSON, to foster reuse.

Software tools

In the field of digital humanities, as well as in other academic areas, software tools are progressively gaining recognition as research products and their development is considered by many scholars as a vital research activity [Schreibman and Hanlon 2010, Ramsay and Rockwell 2012]. Software is also considered an integral part of Cultural Heritage, as denoted by initiatives such as Software Heritage,[88] which aims to collect and preserve software in source code form, "because software embodies our technical and scientific knowledge and humanity cannot afford the risk of losing it".

Despite the digital turn of the humanities and the significant progress achieved in encoding, analysing, and visualising cultural and textual resources, the design of software tools often lacks shared, formal, and reusable models [Almas and Beaulieu 2013, Shillingsburg 2015, Robinson and Bordalejo 2016]. This shortcoming frequently results in the development of ad hoc, project-specific applications that are difficult to extend, reuse, or maintain beyond their original context [Ciotti 2014, Schmidt2014]. As the need for extensible, reusable, and modular applications continues to grow within digital humanities research [Pierazzo 2014, Driscoll and Pierazzo 2016], a more critical reflection on software design and architecture becomes increasingly necessary.

Software tools may differ a lot from an architectural point of view.[89] However, research tools are often developed in similar ways within academia by individual researchers or project teams relying on time-limited funds. This context typically favours rapid development aimed at immediate research goals, often at the expense of long-term sustainability and architectural soundness. As a consequence, development tends to be fragmented and discontinuous, making it

---

[88]https://www.softwareheritage.org/.

[89]For a detailed categorisation of software tools, please refer to ISO's "Systems and software engineering — Framework for categorization of IT systems and software, and guide for applying it". https://www.iso.org/standard/63611.html.

challenging to maintain tools in the long-term and update them to new standards and technologies. For example, the software EVT has been reengineered twice to accommodate technological changes. Other tools, like Collate[90] (relaunched as CollateX),[91] required a complete rewrite to remain accessible and usable.

Three key factors that can render a software tool inaccessible are the lack of documentation, the impermanence of research projects' web domains, and the absence of user technical support [Martignano 2023]. If a tool's development is unstable, it is unlikely to provide users with adequate technical support. Commercial software and services typically offer both documentation and technical support, but may be inaccessible due to costs. Some applications offer trial periods or free plans suitable for small to medium-sized datasets. For instance, Transkribus[92] provides free transcription for 3,000 printed pages and 500 handwritten pages. For open-source software, inaccessibility often stems from a lack of technical documentation and poor source code readability.

Free software tools are more accessible and likely to be adopted by a larger audience, but only if they are user-friendly, customisable, and meet users' needs. Researchers often prefer developing new tools from scratch rather than adopting existing ones [Pierazzo 2019], to maintain full control over the tool's functionality and user interface. To encourage reusability, tools should be highly customisable and open-source. EVT is a successful example of an open-source tool that has been reused and further developed through collaboration with various research projects [Rosselli Del Turco et al. 2019].

A key design approach for producing effective and reusable software is called "Domain Driven Design (DDD)" [Evans 2004]. According to DDD, developers should involve domain experts since the early stages of the software's design process, in order to effectively capture the software's application domain. In digital humanities, this implies modelling entities such as texts, documents, editions, witnesses, or collections in ways that reflect scholarly practices and research questions. The lack of formal domain specifications and shared abstractions has been identified as a major limitation in many digital humanities tools, particularly in the field of scholarly editing [Almas and Beaulieu 2013, Robinson and Bordalejo 2016].

Another crucial factor in developing robust and reliable software is applying best practices and methodologies from software engineering, such as object-oriented programming and SOLID principles, and the use of design patterns [Silén 2024, Ackerman and Gonzalez 2011]. However, in the field of Digital Humanities, not all software developers have a strong background in software engineering, which can lead to poorly implemented tools:

> "The development of applications in the field of Digital Humanities (DH) does not adequately take into account domain modelling, software design principles and software engineering methodologies. In

---

[90]https://digitalmedievalist.org/2012/04/01/collate-text-editing-software/.

[91]https://collatex.net/.

[92]https://www.transkribus.org/it.

> fact, many systems developed in the context of DH-related projects have not been conceived to be modular, extensible, and scalable: they only tend to solve specific problems such as data-driven and project-oriented tools. In addition, most projects focus on the requirements of humanists (as end users), but leave out the needs of software developers" [Del Grosso et al. 2017].

Software tools should leverage current technologies, such as the World Wide Web Consortium (W3C) standards[93] for web development, and adhere to digital humanities standards (e.g., XML/TEI). This approach ensures interoperability among various tools, fostering a more cohesive and efficient digital humanities ecosystem.

Finally, there currently are guidelines that help developers make their software tools and their associated metadata FAIR, in particular the Research Software MetaData Guidelines[94] and the FAIR Principles for Research Software [Chue Hong et al. 2021].

The software tools we selected as pilots are:

- **EVT - Edition Visualization Technology**, is an open source software to visualise digital scholarly editions on the basis of TEI/XML-encoded documents. It is easy to configure and deploy on the Web (using HTML, CSS and Javascript for long-term support), it is fully customisable, and it includes several useful research tools.

- **Voyant Tools**, is an open-source, web-based text reading and computer-assisted analysis environment for scholars. It allows scholars to explore and query texts in several linguistic tasks and analyses such as word frequency, keyword analysis, and topic modelling.

EVT

EVT (Edition Visualization Technology, http://evt.labcd.unipi.it/) is an open-source, client-only software for visualising digital scholarly editions in XML/TEI format. Conceived by Roberto Rosselli Del Turco and developed by students of the Master's degree programme in Digital Humanities of the University of Pisa, the tool was initially created for the digital edition of the Vercelli Book.[95] Through collaboration with other projects, notably the Digital Codex Pelavicino (see Codice Pelavicino Digitale), it later evolved to accommodate various editions. It is now developed by DH graduate students, researchers, and professional software developers.

EVT 1.0, released in 2013, supports diplomatic-interpretative editions and "digital documentary editions" [Pierazzo 2014], using XSLT transformations to generate HTML from XML/TEI. EVT 2.0, launched in 2016, was built with

---

[93]https://www.w3.org/TR/?status%5B0%5D=standard.

[94]https://fair-impact.github.io/RSMD-guidelines/.

[95]http://vbd.humnet.unipi.it/beta2/.

the JavaScript framework AngularJS[96] to support critical editions. It later incorporated all functionalities from the first version. A notable example of EVT 2.0 usage is the digital edition of Benedetto Varchi's *Storia fiorentina* (see VaSto). In October 2024, EVT 3.0 beta was released, essentially rebooting EVT 2.0 with Angular,[97] a more modern JavaScript framework. All EVT versions are free and open-source. This pilot analysis focuses on version 2.0, given its stability and widespread adoption.

EVT 2.0[98] is available under the AGPL-3.0 licence. It uses standard web technologies (JavaScript, CSS, HTML) and integrates existing tools, particularly OpenSeadragon[99] for image viewing and VisColl[100] for visualising manuscript structure.

To use EVT, users insert their materials (XML/TEI files, images, and VisColl files) into the software folder and specify paths in the configuration file. This file also allows interface customisation, such as selecting edition levels and visualisation modes. The EVT2-Config-Generator[101] tool simplifies this process. Users can customise TEI element presentation via a dedicated CSS stylesheet. Comprehensive user information is available in the repository's README file[102] and in the manual provided in each release archive.

EVT 2.0's modular architecture facilitates collaborative development. The repository's README provides developer instructions, and the code is well-commented to encourage third-party contributions. However, EVT 2.0 lacks a detailed, structured presentation of data extraction and modelling from source materials. EVT 3.0 addresses this by modelling data through TypeScript's class system and providing software contribution guidelines in the repository wiki.[103]

A complete history of EVT 2.0 releases is not available on GitHub, which only shows recent releases. Previously, the EVT 2.0 GitHub repository was for private development use, with releases published on the platform SourceForge.

---

[96]https://angularjs.org/.

[97]https://angular.dev/.

[98]EVT 2.0's source code GitHub repository: https://github.com/evt-project/evt-viewer.

[99]https://openseadragon.github.io/.

[100]https://viscoll.org/.

[101]http://evt.labcd.unipi.it/evt2-config/.

[102]https://github.com/evt-project/evt-viewer/blob/master/USER_README_EN.md).

[103]https://github.com/evt-project/evt-viewer-angular/wiki.

FAIRness evaluation summary

	Strengths	Improvements
F	• The software is indexed in a searchable resource.	
A	• The software is freely available and open-source. • The software is deposited in a certified repository.	
I	• The software is based on standard web technologies.	
R	• The licence under which the software is available is stated clearly. • The modular and well-documented source code of the software facilitates reuse and customisation.	• Include a detailed description of data extraction and modelling in the documentation. • Provide external developers with a complete release and change history. • Provide software contribution guidelines (addressed in EVT 3.0).

Voyant Tools

Voyant Tools (https://voyant-tools.org/) is a web-based environment for text reading and analysis. First released in 2016, the software was designed and developed by Stéfan Sinclair, building on existing text analysis tools (HyperPo[104] and Taporware).[105] Following Sinclair's passing in 2020, Geoffrey Rockwell of the University of Alberta now leads the project, supported by Andrew MacDonald as principal programmer and Cecily Raynor at McGill University.

Voyant Tools aims to facilitate reading and interpretive practices for both the general public and digital humanities scholars. The software's interface is available in English and thirteen other languages. Its open-source code is available on GitHub[106] under a GPL3 licence, with the latest version (2.6.17) released in September 2024.

To use Voyant Tools, users can paste or upload texts or collections in various formats (plain text, HTML, XML, PDF, RTF, MS Word, and Pages) on the software's landing page. Multiple URLs to textual resources can also be inserted.

After tokenising and analysing the textual materials, the application displays results in a graphical interface with a "default skin" — a set of default tools

---

[104]https://web.archive.org/web/20121110191405/http://hyperpo.org/.

[105]https://edutechwiki.unige.ch/en/Taporware.

[106]Voyant Tools' source code GitHub repository: https://github.com/voyanttools.

including word distribution graphs, a word cloud visualising the most frequent words, and a summary providing information such as word count and distinctive words. Users can choose from 29 different tools to display in the graphical user interface (GUI). These tools interact: clicking an element in one tool updates the information in others. Users can search the corpus using various tools, and export the tools and displayed data as HTML snippets for embedding in external web pages.

Uploaded texts and collections are cached on the software's servers for about a month. This allows users to bookmark and share URLs referring to a text collection, enabling multiple users to work on the same texts across different sessions. Voyant provides basic access management features to control who can access a given text collection. For those preferring local use, VoyantServer — a standalone version — allows running Voyant Tools without storing documents on the software's server.

Voyant Tools provides demonstrative corpora for users to explore its features. The software's comprehensive documentation details each tool individually and offers tutorials. It also includes citation guidelines. While the documentation doesn't specify the programming languages used (Java and XSLT), this information can be inferred from the GitHub repository. A notable section of the documentation outlines the design principles guiding the software's development, such as "domain sensitivity" and "skinnability", offering valuable insights into its conceptual framework.

FAIRness evaluation summary

	Strengths	Improvements
F	• The software is indexed in a searchable resource.	• Enrich the metadata available in the searchable resource.
A	• The software is freely available and open-source.	• Deposit the software source code in a certified repository, to guarantee long-term accessibility and preservation.
I		

	Strengths	Improvements
R	- The licence under which the software is available is stated clearly in the documentation. - The modular organisation of the software and the comprehensive documentation of the software facilitate reuse and customisation. - The interface is highly customisable, allowing users to tailor the environment to their specific needs.	Consider displaying the version number and the licence terms of the software more prominently.

Linked open data

The expression Linked Data can be used to define "a set of best practices for publishing and connecting structured data on the Web" [Bizer et al. 2009]. Tim Berners-Lee's guidelines for five-star linked open data have become widely recognised. These guidelines outline a progressive approach to creating linked data:

1. "Available on the web (whatever format) but with an open licence, to be Open Data.

2. Available as machine-readable structured data (e.g., Excel instead of an image scan of a table).

3. As (2) plus non-proprietary format (e.g., CSV instead of Excel).

4. All the above plus, Use open standards from W3C (RDF and SPARQL) to identify things, so that people can point at your stuff.

5. All the above, plus: Link your data to other people's data to provide context." [Berners Lee 2009]

Over the last years, many GLAM institutions have been leveraging the potential of Linked Data to open previously siloed collections and regain their centuries-old recognition as intermediaries between users and Cultural Heritage through high-quality data publication [Marden et al. 2013, Daquino 2021]. Notable examples include the Library of Congress Linked Data Service, whose first dataset dates back to 2009. Since then, many cultural institutions have embraced the same challenge, converting their holdings into collections of RDF triples and making them available through dedicated query services and APIs.

The construction and exploration of linked datasets is made possible by some fundamental Web technologies, namely URIs (Uniform Resource Identifiers), HTTP (HyperText Transfer Protocol), and W3C-recommended standard formats

such as RDF (Resource Description Framework),[107] SPARQL (SPARQL Query Language for RDF),[108] SKOS (Simple Knowledge Organization System),[109] OWL (Ontology Web Language).[110] On a conceptual level, their development is ensured by ontologies, such as the popular CIDOC's Conceptual Reference Model (CIDOC-CRM) and the Friend of a Friend (FOAF) vocabulary, which provide domain-specific vocabularies and define the meaning of shared terms. However, the production of LOD collections in a native way has received little attention so far and many Cultural Heritage datasets are still dependent on some intermediate technologies which store their content in a traditional format. Overcoming such a limitation would make data consistency easier to maintain while maximising knowledge reusability and entity reconciliation [García et al. 2017].

To make linked open datasets truly FAIR, additional measures are necessary. Essential information such as licence, version, and version history must be included. Data provenance management requires careful attention, as it is paramount also for project management purposes, e.g., to monitor the editorial process and to keep track of data versions. Models and technologies used should be thoroughly documented. Adopting shared models, vocabularies, standards, and protocols at national and international levels is vital, as exemplified by Zeri & LODE's use of Italian government standards for photos and artworks and LiLa's integration of Latin into the Linguistic Linked Open Data (LLOD) Cloud.[111]

The FAIRness of data also hinges on usability. While open and free access is fundamental, accessibility depends on the ease of use of services like LodView[112] and the quality of their documentation. Long-term access must be ensured, preferably through external open access repositories like Zenodo, CLARIN repository or other robust, actively maintained tools. URIs require careful consideration—they should be uniform across different data access modes, provide stable access to web resources, and ideally support dereferencing, enabling both humans and machines to obtain meaningful descriptions of the identified entities. This allows both humans and machines to resolve identifiers to authoritative representations, served via content negotiation and accessed through web browsers, interfaces, or automated agents, thus enabling the consumption, generation, verification, and reuse of data within distributed and continuously evolving knowledge infrastructures (Pareti 2016). Lastly, for a dataset to be truly FAIR, it must be comprehensively documented, especially regarding its data model, used resources (standards, vocabularies, ontologies), workflow, and data preparation tools.

The linked open datasets we selected as pilots are:

---

[107] https://www.w3.org/2001/sw/wiki/RDF.

[108] https://www.w3.org/2001/sw/wiki/SPARQL.

[109] https://www.w3.org/2001/sw/wiki/SKOS.

[110] https://www.w3.org/2001/sw/wiki/OWL.

[111] https://linguistic-lod.org/

[112] https://lodview.it/.

- **Zeri & LODE**, the linked open dataset of the Federico Zeri photo archive, which describes the photographic documentation of 30000 pieces of modern artworks, collected by Zeri, a notable historian. The descriptions comply with the cataloguing standards of the Italian Ministry of Culture.

- **DanteSources**, a digital encyclopaedia of Dante Alighieri's work, including bibliographic references of primary sources cited or quoted in Dante's works.;

- **LiLa - Linking Latin**, a knowledge base of linguistic resources and natural language processing tools for Latin, implemented according to the Linked Data paradigm.

- **Biflow - Toscana Bilingue Catalogue**, a highly structured LOD-based online resource which represents the social and cultural history of translation in the Tuscan Middle Ages. It contains detailed descriptions of about 2.000 MSS and of their respective texts, translations and scribes.

Zeri & LODE

Zeri&LODE (website: https://data.fondazionezeri.unibo.it/) is a project of the University of Bologna aimed at enhancing Federico Zeri's repository catalogue. This repository includes an art library (46,000 volumes, 37,000 auction catalogues, 60 periodicals) and a photo archive (290,000 photographs of monuments and artworks). The project's goal is to create an RDF dataset of Zeri's photo archive, building on the cataloguing work conducted by the Federico Zeri Foundation and the University of Bologna. This initiative is part of a larger endeavour led by PHAROS, an International Consortium of Photo Archives that aims to create an open and freely accessible digital research platform allowing for comprehensive consolidated access to photo archive images and their associated scholarly documentation.

The catalogue implementation employed two Italian metadata content standards issued by the ICCD (Central Institute for Catalogue and Documentation) of the Italian Ministry of Cultural Heritage: Scheda F[113] for photographs (Scheda di fotografia) and Scheda OA[114] for artworks (Scheda Opera d'Arte). These standards served as the foundation for the dataset modelling. The initial release of the Zeri Photo Archive RDF dataset (April 2016) represents a significant subset of data already available on the Zeri Catalog website[115] and discoverable through the Europeana Portal.[116] The dataset primarily covers Modern Art (15th-16th centuries), describing about 19,000 artworks and over 30,000 photographs through approximately 11 million RDF statements.

---

[113]http://www.iccd.beniculturali.it/index.php?it/473/standard-catalografici/Standard/10.
[114]http://www.iccd.beniculturali.it/index.php?it/473/standard-catalografici/Standard/29.
[115]http://catalogo.fondazionezeri.unibo.it/.
[116]http://www.europeana.eu/portal/it/search?q=PROVIDER%3A%22Federico+Zeri+Fo undation%22.

Developed using W3C standard technologies, RDF and SPARQL, the linked open dataset is accessible via a dedicated SPARQL endpoint, a web user interface, and LodView, a linked open data browser. The documentation provides quick-access links to facilitate navigation in this browser. A downloadable version[117] of the dataset is available in the AMSActa repository of the University of Bologna.

The concise documentation includes licence information, attribution guidelines, references to models used in building the data model, and the URI creation schema to facilitate data reuse. However, some aspects of the dataset creation and workflow remain unclear, such as the use of external tools and development duration. While seemingly secondary, this information could benefit researchers undertaking similar projects. Additionally, the latest version date appears only in the attribution, and the dataset's status is ambiguous, as it represents only a subset of the materials in the Zeri Archive.

The data model incorporates existing ontologies: CIDOC-CRM, the SPAR Ontologies, and the HiCO Ontology. To enhance interoperability, the Zeri&LODE project developed two custom ontologies: FEntry Ontology[118] and OAEntry Ontology,[119] which represent the aforementioned Scheda OA and Scheda F standards. The documentation also provides links to the RDF mappings of these standards. Technical terms for artwork and photograph descriptions were sourced from the AAT Getty Thesaurus. GeoNames[120] was used for place names, while VIAF, Getty ULAN,[121] DBpedia,[122] and Wikidata were used for people's names.

FAIRness evaluation summary

	Strengths	Improvements
F	• The dataset has been assigned persistent identifiers.	• Register the dataset in a searchable resource. • Prominently display the dataset's version number, as it is currently only mentioned in the attribution.

---

[117]http://amsacta.unibo.it/5157/.

[118]http://www.essepuntato.it/2014/03/fentry.

[119]http://purl.org/emmedi/oaentry.

[120]https://www.geonames.org/.

[121]http://www.getty.edu/research/tools/vocabularies/ulan/.

[122]http://wiki.dbpedia.org/services-resources/datasets/dbpedia-datasets.

	Strengths	**Improvements**
A	- The dataset is open access. - The complete dataset is available as a data dump on an external open access repository (AMSActa) that ensures its long-term preservation. - There are multiple access points to the dataset.	
I	- The dataset is based on standard W3C technologies. - It is possible to download the dataset. - Builds the data model by leveraging existing models and standards. - Publishes native ontologies in open access, making them freely available.	
R	- The licences under which the dataset and the metadata are available are stated clearly. - The dataset comes with comprehensive guidelines for data reuse.	Include a brief description of the dataset's creation process and any tools used in the documentation.

DanteSources

DanteSources (https://dantesources.dantenetwork.it/) is a web tool for retrieving and visualising data about the sources of Dante Alighieri's works. Developed between 2013 and 2016, it is a joint effort of the Institute of Information Science and Technologies 'Alessandro Faedo' (ISTI) of the Italian National Research Council and the Department of Philology, Literature, and Linguistics at the University of Pisa.

Built on semantic web technologies, DanteSources' RDF dataset encompasses 714 cited works, 273 cited authors, and 45 distinct thematic areas. Users can search the data by Dante's work, primary source, cited author, thematic area, and type of reference in both Dante's works and primary sources. The tool presents the distribution of primary sources, authors, and thematic areas cited by Dante through tables and charts, with the underlying data available for download as CSV files. The referenced publications are often not openly accessible or readily

available. The dataset is also accessible via a SPARQL endpoint.[123] The URIs extracted from the SPARQL endpoint do not identify an existing resource.

While the documentation effectively guides data exploration, it falls short in detailing the data model and dataset. The website provides the project's RDF schema (which references the CIDOC-CRM and FOAF ontologies) and full bibliographic references for the analysed editions of Dante's works. However, it omits crucial information such as licensing and citation guidelines. The dataset's status remains ambiguous, though it appears comprehensive, covering all of Dante Alighieri's works except the Comedy.

FAIRness evaluation summary

	Strengths	Improvements
F	• The dataset is indexed in a searchable resource.	• Register the dataset in a searchable resource. • Prominently display the dataset's version number, as it is currently only mentioned in the attribution. • Fix the URIs extracted from the SPARQL endpoint so that they identify existing resources.
A	• The dataset is open access. • There are multiple access points to the dataset. • The dataset can be accessed via an intuitive graphical interface.	• Deposit the complete dataset available as a data dump in a certified repository.
I	• The dataset is based on standard W3C technologies. • The RDF schema is included in the documentation.	• Provide a detailed description of the data model, including the models and standards used.
R		• Indicate the licence clearly. • Describe in the documentation the website's development process, including tools and technologies employed. Additionally, provide users with a guide. • Include demo queries in the documentation.

---

[123]https://dantesources.dantenetwork.it/sparql.

LiLa

The LiLa (Linking Latin, website: https://lila-erc.eu/) project, led by Professor Marco Passarotti, is funded by the European Research Council and based at the Cattolica University's CIRCSE (Centro Interdisciplinare di Ricerche per la Computerizzazione dei Segni dell'Espressione) research centre. Running from 2018 to 2023, LiLa aimed to build a Linked Data-based Knowledge Base of Linguistic Resources and Natural Language Processing (NLP) tools for Latin. Its primary goal is to connect and leverage existing resources, enhancing the study and analysis of Latin texts, as stated in the project's website:

> "LiLa integrates existing and new linguistic data for Latin. It combines resources like corpora, lexica, and NLP tools from various providers. LiLa also generates new data by enhancing existing resources. This includes:
>
> - Adding PoS-tagging and lemmatisation to Latin texts
>
> - Standardising annotations across Latin treebanks
>
> - Expanding Latin WordNet and Latin-Vallex
>
> - Increasing coverage of the Index Thomisticus Treebank
>
> Additionally, LiLa develops new models for PoS-tagging and lemmatisation, aiming to create an optimal NLP pipeline for Latin."

The LiLa Knowledge Base comprises a lemma bank, lexical resources (e.g., Word Formation Latin, Latin Vallex 2.0, Latin WordNet), and textual resources (e.g., Index Thomisticus Treebank (ITTB), UDante, Lucani Pharsalia). LiLa's webpage presents each data component with three links: a Zenodo publication for attribution, a GitHub repository for download, and a link to the LodView browser for data access. LiLa offers three exploration methods for its knowledge base: a triplestore, a user-friendly query interface, and an interactive search platform. The SPARQL query interface includes useful sample queries to familiarise users with the Knowledge Base's content and data model. Additionally, LiLa developed the "TextLinker" web service to populate its knowledge base by lemmatising and Part-of-Speech (PoS) tagging raw Latin text and linking tokens to the LiLa Lemma Bank.

While the TextLinker and interactive search platform have intuitive interfaces, they lack user examples and instructions. The documentation also omits information on tool development and open-source status.

The project aimed to create a knowledge base adhering to FAIR principles through the following procedures:

- Assigning unique URIs to all (meta)data, lemmas, word types, and metadata tags, enabling precise identification and linking of linguistic elements across resources.

- Using HTTP for data retrieval and SPARQL for querying, facilitating data reuse and citation tracking.

- Archiving all resources on Zenodo for long-term preservation.

- Employing standard semantic web vocabularies and ontologies to describe object relations, though specific ontologies and vocabularies are not mentioned in the documentation.

- Providing open access to all resources.

- Using standardised URIs and offering detailed provenance information.

- Releasing new data under the CC BY-SA licence, software under GNU LGPL3, and Zenodo metadata under CC0.

- Integrating Latin into the multilingual Linguistic Linked Open Data (LLOD) cloud.[124]

The documentation lacks crucial information about the LiLa Knowledge Base's underlying data model, newly created ontologies, and the application of existing RDF vocabularies.

Having seemingly achieved all its goals, the project can be considered complete.

FAIRness evaluation summary

	Strengths	Improvements
F	- The datasets are indexed in a searchable resource.	- Prominently display the datasets in the website, alongside their version numbers and citation guidelines.
A	- The datasets are open access. - The datasets are deposited in certified repositories. - The data can be accessed via user-friendly tools with intuitive interfaces that facilitate exploration.	
I	- The dataset is based on standard W3C technologies. - The knowledge graph was built by reusing and leveraging existing resources.	

---

[124]http://linguistic-lod.org/llod-cloud.

Strengths	Improvements
R	- Prominently display the licences under which the datasets and the tools are available. - Provide a detailed description of the knowledge graph's data model and newly created ontologies, specifying how existing RDF vocabularies were utilised. - Include examples and user guides for all tools. - Describe the implementation process of the tools, considering publishing them as open-source to facilitate reuse. - In the documentation, specify the tools used for creating the knowledge base.

Biflow

BIFLOW (Bilingualism in Florentine and Tuscan Works, website: https://ca talogobiflow.vedph.it/) is a research project funded by the European Research Council and hosted by the Ca' Foscari University of Venice and EHESS (École des hautes études en sciences sociales), Paris. Directed by Prof. Antonio Montefusco, the project investigates literary documents that circulated simultaneously in multiple languages in medieval Tuscany, particularly in Florence, from the late 13th to the early 15th century.

The project's primary outcome is an RDF-based catalogue, accessible through a SPARQL endpoint[125] implemented with the tool RDF store-js[126] (which was last updated in 2016, raising concerns about its long-term viability), or via a more user-friendly website interface. However, it is worth noting that the data provided through this service doesn't fully align with the catalogue's RDF serialisations, particularly regarding the base URI of entities and properties.

The catalogue is organised into entries collecting bilingual textual dossiers, which include the source text and its various translations. Each dossier is identified by a code derived from the author's name and the work's title. Every dossier provides a content summary, an essential bibliography, and an interactive graph showing relationships between versions.

Within each dossier, descriptive sub-entries for each text version offer varying degrees of detail on textual history, manuscript tradition, and editorial history.

---

[125]https://catalogobiflow.vedph.it/sparql/.

[126]https://github.com/antoniogarrote/rdfstore-js.

Sub-entries are denoted by a letter added to the dossier code: uppercase for source texts, and lowercase for translations. All texts include a census and description of the codices comprising their textual traditions. The manuscripts are described with essential information and links to other cataloguing sites. However, links to authority records are missing for works and authors. Users can download the entire dossier—including entries for each version and the manuscript census—in PDF and RDF formats. The RDF documents lack information about data provenance.

The catalogue supports querying via full-text search or by author/translator, manuscripts, title, incipit/explicit, language, genre, and textual typology.

While the catalogue has an associated ISSN code and each dossier includes the editor's name and attribution, the website lacks permalinks to dossiers and other catalogue objects.

The catalogue's data model is formalised as an OWL ontology—the Biflow-Toscana Bilingue ontology—visualisable with the WebVOWL[127] tool and documented with Ontospy.[128] This ontology expands on existing models, particularly eFRBRoo, CIDOC-CRM, and the Biblissima ontology.[129]

The catalogue's completion status remains unclear. The dataset isn't available for download and appears accessible only through the Biflow website, potentially jeopardising its long-term preservation. The website requires refinement, notably in its publications section. Moreover, the English version is incomplete and only partially translated.

FAIRness evaluation summary

	Strengths	Improvements
F		• Register the dataset in a searchable resource. • Clearly indicate the project's current status. • Prominently display the dataset's version number, as it is currently only mentioned in the attribution.
A	• The dataset is open access. • There are multiple access points to the dataset. • The dataset can be accessed via an intuitive graphical interface.	• Deposit the complete dataset as a data dump in a certified repository. • Use a more up-to-date and stable tool for the implementation of the SPARQL endpoint.

---

[127]https://service.tib.eu/webvowl/.
[128]https://lambdamusic.github.io/Ontospy/.
[129]https://doc.biblissima.fr/ontologie/bibma/.

	Strengths	**Improvements**
I	- The dataset is based on standard W3C technologies. - The documentation includes the dataset's ontology.	- Link data, such as work titles, author names, and manuscripts, to external authority systems (e.g., GeoNames, VIAF). - Add data provenance in the RDF documents. - Homogenise the base URIs.
R		- Indicate the licence clearly. - Refine the website by adding missing content and completing translations.

Ontologies

Ontologies are typically designed to provide "a vocabulary describing a domain of interest and a specification of the meaning of terms in that vocabulary" [Euzenar and Shvaiko 2007]. While ontologies often target specialists in specific fields, they also serve as valuable models for broader disciplinary areas, and practical applications. Through their properties and structure, they aim to provide a detailed vocabulary, explicitly defining some terms and the relationships existing between them [Guarino et al. 2009]. CIDOC exemplifies this approach, having emerged from the need to standardise diverse models and practices in the field of Cultural Heritage. Its comprehensive scope enables various applications, encourages reuse, and allows for extensions into more specific models. To effectively reach a general audience, providing detailed documentation with numerous concrete examples is crucial.

A crucial aspect in the design of a new ontology concerns the systematic examination of existing models that address the same or closely related domains. Such analysis is indispensable for identifying conceptual overlaps, assessing established modelling choices, and determining the extent to which existing vocabularies may be reused. This process should lead to the definition of a native alignment, formulated during the early stages of ontology development. Ontology alignment—namely, "the process of establishing correspondences or mappings between concepts, relationships, or entities in different ontologies" [Naveen et al. 2024]—supports both the reuse of well-established terms and the achievement of interoperability between datasets and systems. Explicitly addressing these activities helps clarify the conceptual and technical effort required to develop a new ontology.

Formalising an ontology requires specific markup languages. Among notable examples, such as RDF and RDF schema, the Web Ontology Language OWL has been standing as the most popular semantic technology for the definition

of vocabularies since its release in 2004 [Matentzoglu et al. 2013]. Its current version (OWL 2) has offered a Description Logic (DL) for various well-known vocabularies in the DH domain, including HiCO and the set of SPAR ontologies [Peroni and Shotton 2018].

The effectiveness of a semantic vocabulary hinges on its modelling quality. Therefore, following shared guidelines and standardised best-practices is essential. For instance, the Simplified Agile Methodology for Ontology Development (SAMOD)[130] has been designed to guide ontology engineers through an iterative workflow. Additionally, involving the user community is paramount, as demonstrated by the SPAR ontologies. It is essential to test the efficacy of an ontology through real-world applications, as seen with HiCO's use in various research projects. A notable difference in the frequency of updates and maintenance often emerges when an ontology is developed by a large community rather than individual scholars. Finally, maintaining the functionality of IRIs is a critical aspect of ontology upkeep.

A successful ontology must adhere to FAIR principles. This can be supported by several tools, including catalogues (e.g., LOV,[131] ODP),[132] documentation services (e.g., LODE,[133] WIDOCO,[134] WebVOWL),[135] and ontology repositories.

The ontologies we selected as pilots are:

- **CIDOC-CRM**, the *de facto* golden standard for museum and cultural objects description, which was expanded through the years with many new compatible models.

- **HiCO**, an extension of CIDOC-CRM which describes the interpretive process underlying uncertain or questionable information and may be used for describing competing artwork attributions.

- **SPAR Ontologies**, a suite of ontologies for describing the publishing domain, which includes the standard ontologies for the bibliographic domain.

---

[130]https://essepuntato.it/papers/samod-owled2016.html.

[131]https://lov.linkeddata.es/dataset/lov.

[132]ODP (Ontology Design Patterns): http://ontologydesignpatterns.org/wiki/Main_Page.

[133]Live OWL Documentation Environment: https://essepuntato.it/lode/.

[134]WIzard for DOCumenting Ontologies (WIDOCO): https://zenodo.org/badge/latestdoi/11427075.

[135]https://github.com/VisualDataWeb/WebVOWL.

CIDOC-CRM

The CIDOC Conceptual Reference Model (CRM, website: https://www.cidoc-crm.org/) is a tool for integrating Cultural Heritage information across diverse datasets. It provides a formal structure for describing concepts and relationships in Cultural Heritage documentation, enabling data integration from multiple sources.

CIDOC-CRM aims to create a shared understanding of Cultural Heritage information through a common semantic framework. It serves as a language for domain experts and implementers to define information system requirements and guide conceptual modelling, acting as a "semantic glue" between different Cultural Heritage information sources, as stated in the ontology's documentation.

The CRM consists of a standard base (CRMbase) and modular extensions. These extensions, developed with or by research communities, support specialised research questions while remaining compatible with the base ontology. This approach ensures high information integrity and integration. The extensions are:

- LRM_{OO} — Library Reference Model;

- $PRESS_{OO}$ — Model for publishing periodicals;

- CRMact — Model for activity plan;

- CRMarchaeo — Excavation model;

- CRMba — Model for archaeological buildings;

- CRMdig — Model for provenance metadata;

- CRMgeo — Spatiotemporal model;

- CRMinf — Argumentation model;

- CRMsci — Scientific observation model;

- CRMsoc — Model for social phenomena;

- CRMtex — Model for the study of ancient texts.

CIDOC-CRM is developed by a volunteer community, the CIDOC-CRM Special Interest Group, under ICOM's International Council for Documentation. Members include institutions involved in researching and documenting human history.

The development of the CRM began in 1996. In 2006, CRM gained recognition as an official ISO standard.[136] While the latest version of the model is 7.3,[137] the most recent stable and official version—which also serves as the current ISO standard—is 7.1.3,[138] released in February 2024. All previous versions are

---

[136]https://www.iso.org/standard/85100.html.
[137]https://www.cidoc-crm.org/Version/version-7.3.
[138]https://www.cidoc-crm.org/Version/version-7.1.3.

presented on the website in a chronologically organised table. Each stable version can be downloaded in various standard formats (XML, JSON-LD, RDF) and is accompanied by comprehensive documentation in PDF and DOCX formats. The website offers users learning materials, use cases, and best practices, along with mappings and information to facilitate the model's reuse.

CRMbase can be visualised online via an intuitive and friendly interface. Its documentation is available in English, German, Greek, French, Portuguese, Russian and Chinese. All classes and properties are assigned a URI.

FAIRness evaluation summary

	Strengths	**Improvements**
F	• Detailed presentation of current and previous versions, clearly differentiated by stability status.	• Register the model and its compatible models in a searchable resource.
A	• The model is open access.	• Deposit the stable versions of the ontology in a certified repository.
I	• The model is an official ISO standard. • All stable versions are available for download in various standard formats. • The model is formalised as an ontology using standard and non-proprietary formats. • The model supports interoperability with other domain-relevant standards.	
R	• Users can report issues, fostering continuous improvement.	

SPAR Ontologies

The Semantic Publishing and Referencing Ontologies (SPAR Ontologies, website: http://www.sparontologies.net/) are a suite of OWL 2 DL[139] ontology modules for creating machine-readable RDF metadata for semantic publishing and referencing. David Shotton created the first ontology, CiTO (Citation Typing Ontology), in 2009. Since 2010, Shotton and Silvio Peroni have led the development of the SPAR suite. With contributions from numerous collaborators, the SPAR ontologies have evolved into interoperable ontological modules that

---

[139]http://www.w3.org/TR/owl2-syntax/.

reuse existing vocabularies. The SPAR suite now encompasses sixteen distinct ontologies:

- FRBR-aligned Bibliographic Ontology (FaBiO): Describes publishable entities and bibliographic references.

- Citation Typing Ontology (CiTO): Characterises citation types and rhetoric.

- Bibliographic Reference Ontology (BiRO): Defines bibliographic records, references, and collections.

- Citation Counting and Context Characterisation Ontology (C4O): Records in-text citations, contexts, and global citation counts.

- Document Components Ontology (DoCO): Provides vocabulary for document components.

- Publishing Status Ontology (PSO): Characterises publication status throughout the publishing process.

- Publishing Roles Ontology (PRO): Describes roles in the publication process.

- Publishing Workflow Ontology (PWO): Describes publication workflow steps.

- Essential FRBR in OWL2 DL Ontology (FRBR): Expresses IFLA's Functional Requirements for Bibliographic Records.

- Discourse Elements Ontology (DEO): Provides vocabulary for rhetorical document elements.

- Scholarly Contributions and Roles Ontology (SCoRO): Describes scholarly contributions and roles.

- Funding, Research Administration and Projects Ontology (FRAPO): Describes research project administrative information.

- DataCite Ontology (DataCite): Enables description of DataCite metadata properties

- Bibliometric Data Ontology (BiDO): Describes bibliometric data.

- Five Stars of Online Research Articles Ontology (FiveStars): Characterises online journal article attributes.

- FAIR* Reviews Ontology (FR): Describes reviews of scholarly resources.

On the suite's website, each ontology has its own page featuring comprehensive documentation, usage examples, graphical representations, references, and a link to the GitHub repository where the ontology's source code can be found. Users can access and download the ontology document in various standard non-proprietary formats: RDF/XML, Turtle, N-Triples, and JSON-LD. Each

Figure 10: structure of the SPAR ontologies.
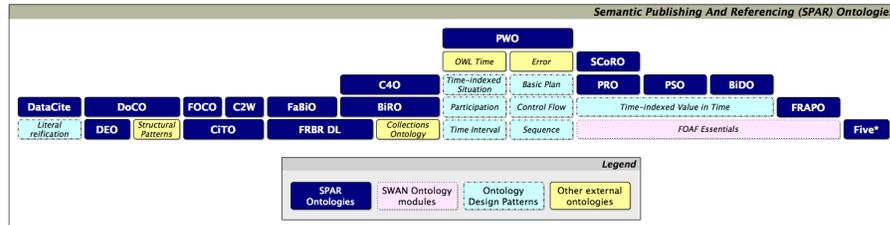
ontology can be visualised online using the LODE tool, which provides a user-friendly interface for browsing classes and properties. All ontologies are available under the CC BY 4.0 licence and are assigned a DOI and IRIs, created using the PURL and W3id.org[140] systems. However, the PURL-generated links do not work properly.

The SPAR ontologies' website offers extensive and detailed documentation, including examples of ontology usage and references to external projects that have implemented the suite. Additionally, the documentation provides contribution guidelines for those interested in proposing a new ontology for inclusion in the SPAR suite.

FAIRness evaluation summary

	Strengths	Improvements
F	• Each ontology is assigned a unique DOI.	• Index all the ontologies in a searchable resource.
A	• All ontologies are open access.	• Deposit all ontologies in a certified repository.
I	• All ontologies in the suite are available for download in various standard formats. • The suite is built by reusing existing vocabularies.	
R	• The modular structure of the suite facilitates the reuse of individual or multiple ontologies and the integration of new ones. • Each ontology is well-documented.	

---

[140]http://w3id.org.

HiCO

The Historical Context Ontology (HiCO, website: https://marilenadaquino.gi thub.io/hico/) is an OWL 2 DL ontology designed to represent the context of scholarly claims about Cultural Heritage objects. Developed by Marilena Daquino (University of Bologna), with contributions from Silvio Peroni and Francesca Tomasi, HiCO wasn't born from a research project but has found application in various research initiatives. It is used to describe art historians' attributions in the Zeri dataset and in digital editions of Paolo Bufalini's notebook[141] and Vespasiano da Bisticci's letters.[142]

HiCO extends PROV-O,[143] a W3C-recommended ontology for data provenance description, adding terms to describe aspects of hermeneutical activity. It also incorporates the CiTO Ontology (from the SPAR Ontologies) to link attributions to related sources. HiCO has been developed according to the SAMOD methodology. The ontology is accessible online via the LODE tool and downloadable as an OWL document under the Creative Commons 4.0 licence. The current version, 2.0, was released in 2020. HiCO is also hosted in a public GitHub repository,[144] although the version available appears to be older than the one published via LODE. The documentation provides detailed component descriptions, prefaced by the ontology's scope, a visual representation of its components, and a real-world usage example. The IRIs are built using PURL, but they currently do not function correctly. Consequently, properties used in Linked Open Data collections (e.g., Zeri archive) are identified by URIs that do not lead to existing web pages.

FAIRness evaluation summary

	Strengths	Improvements
F	• The ontology has been assigned a permanent URI (https://w3id.org/hico/). • HiCO's current version is assigned a specific IRI, while previous versions remain accessible online.	• Register the ontology in a searchable resource. • Fix the PURL IRIs to ensure proper resolution of ontology identifiers.
A	• The ontology is open access.	• Deposit all versions of the ontology in a certified repository.

---

[141]https://projects.dharc.unibo.it/bufalini-notebook/.

[142]https://vespasianodabisticciletters.unibo.it/.

[143]https://www.w3.org/TR/prov-o/.

[144]https://github.com/marilenadaquino/hico.

	Strengths	Improvements
I	- The ontology is based on standard and non-proprietary formats. - The ontology reuses existing ontologies.	
R	The documentation is comprehensive.	

Chapter 2
Data Model

Abstract

In this chapter we present the data model of the ATLAS knowledge graph, focusing on the outcomes from the project's first phase, namely the cataloguing metadata for describing various types of research products (identified through the pilots' analysis), and the framework serving as the foundation for the data model.

The framework was constructed by comparing and mapping existing models, particularly Schema.org,[145] RO-crate,[146] DCAT,[147] and the OpenAIRE knowledge graph.[148] It is based on two main concepts: Research Project and Research Product. The latter represents all types of digital objects produced by scholarly research projects. The Research Product class offers a suite of properties that help identify the research product from both cataloguing (e.g., title, description, release date, version, licence) and technical (e.g., format, access points) perspectives.

To distinguish different types of research products, from the Research Product have been derived seven sub-classes: Text Collection, Digital Scholarly Edition, Ontology, Linked Open Data, Software, Language Model, and 3D Digital Twin. This modelling approach highlights the unique characteristics of each type of digital object produced in scholarly research. Each subclass expands on the Research Product class, allowing for the description of specific types of research products with additional properties. After some methodological notes, the second chapter presents these classes and their related properties, illustrated with examples from the pilots and accompanied by recommendations for data entry.

Methodological notes

Two key research questions guided our data modelling process:

1. How can we categorise and represent different types of research products?

2. How can we represent research projects and their relationship to the products they develop?

To address these questions, we began by comparing existing models with similar aims and scope. We focused initially on the Data Catalog Vocabulary

---

[145] http://schema.org.

[146] Research Object Crate: https://www.researchobject.org/ro-crate/.

[147] Data Catalog Vocabulary: https://www.w3.org/TR/vocab-dcat-2/.

[148] Data model documentation of OpenAIRE's knowledge graph: https://graph.openaire.eu/docs/data-model/.

(DCAT), Schema.org, DC Terms,[149] and the FRBR-aligned Bibliographic Ontology (FaBiO).[150]

DCAT is a W3C recommended RDF vocabulary designed to represent data catalogues published on the Web. It is based on seven main classes, namely:

- `dcat:Catalog`, a collection of metadata about resources like datasets or services.

- `dcat:Resource` which can be a dataset, a data service, or any other type of resource described in a catalog. It is not used independently but serves as a base for more specific types like `dcat:Dataset`, `dcat:DataService`, and `dcat:Catalog`.

- `dcat:Dataset`, a collection of data published or managed by a single person, group, or organisation. It can include various types of information like numbers, text, images, or sounds.

- `dcat:Distribution`, a specific dataset form that people can access, such as a downloadable file.

- `dcat:DataService`, a set of operations (like an API) that allows access to one or more datasets.

- `dcat:DatasetSeries`, a group of related datasets that are published separately but share some common characteristics.

- `dcat:CatalogRecord`, which contains information about a catalogue entry, such as who added it and when.

Some of DCAT's classes and properties, particularly Dataset, were incorporated into Schema.org. Using Schema.org instead of DCAT allows access to a wider pool of classes and properties while still utilising the main classes defined in DCAT. Among these, the class Dataset seemed to be a valuable starting point for describing research products in our model.

Dublin Core is a crucial international standard for cataloguing metadata. Initiated in the mid-1990s, the DCMI Metadata Terms ontology now comprises fifteen main elements applicable to the metadata description of all resource types, including digital scholarly objects. These elements are contributor, coverage, creator, date, description, format, identifier, language, publisher, relation, rights, source, subject, title, and type. The ontology also provides a set of classes (the DCMI Type Vocabulary) to categorise the nature or genre of the resource. Some of these classes were viable options to represent different types of research products in our data model, namely: Collection, Dataset, InteractiveResource, Service, Software, and Text.

---

[149]DCMI Metadata Terms: https://www.dublincore.org/specifications/dublin-core/dcmi-terms/.

[150]https://sparontologies.github.io/fabio/current/fabio.html.

Lastly, FaBiO, one of the SPAR Ontologies, describes publications and other publishable entities (e.g., journal articles, conference papers, books) from a bibliographic perspective. While this ontology does not offer classes that perfectly fit the different types of pilots we analysed, it may prove useful in connecting the concept of research product to IFLA's WEMI (Work, Expression, Manifestation, Item) framework. This connection may facilitate the integration of our data model with relevant models in the Cultural Heritage field, such as CIDOC-CRM.

In the subsequent phase of the modelling process, we examined the data models of systems and knowledge graphs that like ATLAS aim at representing and cataloguing scholarly outputs, in particular, the models of: the OpenAIRE scholarly knowledge graph, SKG-IF (Scholarly Knowledge Graphs Interoperability Framework),[151] RO-Crate, IRIS (Institutional Research Information System)[152] and the KNOT catalogue.[153] The comparison between these models enabled us to identify common properties and select those most appropriate for our objectives. We formalised the comparison between these models and our own data model as a comprehensive mapping.[154]

The data model of the OpenAIRE scholarly knowledge graph comprises different classes to represent organisations, projects, communities, individual people, and, more importantly, research products and data sources, i.e., the sources from which the metadata of graph objects are collected. In this data model, research products may be categorised as publications, data, software, or other products with additional properties.

The Data source and Research product classes, with the same categorisation for the latter, are also proposed in the SKG-IF, a model developed by an Interest Group of the RDA (Research Data Alliance) on Open Science Graphs for FAIR Data. In this model, the concept of "research product" is categorised as:

- "**Literature:** Intended for reading by humans (article, thesis, peer-review, blog posts, books, reports, patents, etc.)

- **Research data:** Self-contained, persistently identified digital assets intended for processing (e.g. files containing: tables, metadata collections, dumps; persistent dynamic queries to scientific databases)

- **Research software:** (definition from RDA WG) Research Software includes source code files, algorithms, scripts, computational workflows, and executables that were created during the research process or for a research purpose. [...]

- **Other products**: any digital asset, uniquely identified, whose nature does not fall in the first three types" [Mannocci and Baglioni 2023]

---

[151]https://skg-if.readthedocs.io/en/v1.0/.

[152]https://wiki.u-gov.it/confluence/display/public/UGOVHELP/IRIS+-+Institutional+Research+Information+System.

[153]https://projects.dharc.unibo.it/knot/.

[154]The mapping is deposited in Zenodo and can be consulted freely: https://doi.org/10.5281/zenodo.13993057.

RO-Crate is a community-driven initiative aimed at developing a lightweight approach for packaging research data along with their associated metadata. The research object (RO) is fundamentally a collection of data or a "crate" (e.g., papers, data files, software, references to other research). In order to make it easier to track, archive, and attribute, the crate is accompanied by a plain text file named the *RO-Crate Metadata Document,* which includes metadata for each item within the collection. The RO-crate specification is based mainly on Schema.org and other existing standards, such as the W3C Web Annotation Data Model,[155] W3C PROV, Dublin Core Terms, and ORCID.

IRIS[156] is a Java-based platform for managing and enhancing research outputs adopted by numerous Italian universities. While originally focussing on publications, IRIS's modular structure was progressively expanded to describe research and public engagement activities and projects as well. Two modules are particularly relevant to ATLAS:

- Activities & Projects (AP): Gathers information on research projects, contracts, and initiatives. Allows data entry to highlight scientific value and collaborations.

- Institutional Repository / Open Archive (IR/OA): Stores and enhances publication outputs. Provides an interoperable system for managing and disseminating publications, compliant with MIUR and the OpenAIRE knowledge graph requirements [Bollini et al. 2016].

IRIS's data model is mainly based on DC Terms.

The data model developed for the KNOT catalogue[157] proved particularly useful as a reference for our model's general framework. The KNOT data model (KNOT-DM)[158] is based on DCAT, CIDOC-CRM, and PROV-O, which are used to describe published data, Cultural Heritage information, and academic provenance, respectively. The KNOT data model (KNOT-DM) distinguishes between research projects and their products (termed "digital scholarly objects" in KNOT), representing products as `dcat:Catalog` and `prov:Entity`, and projects as prov:Activity. Following DCAT, it further differentiates between data (`dcat:Dataset`), data access services (`dcat:Dataservice`), and data publications (`dcat:Distribution`).

---

[155]https://www.w3.org/TR/annotation-model/.

[156]IRIS is developed by Cineca—an Italian consortium comprising universities, research centers, and the Italian Ministry of Education (MIUR).

[157]The University of Bologna and the Central Institute for the Digitisation of Cultural Heritage (ICDP) developed the KNOT catalogue to explore ways of integrating digital Cultural Heritage from Italian universities, with a focus on Digital Humanities, into the ICDP Digital Library. The Digital Humanities Advanced Research Center (/DH.ARC) created the catalogue.

[158]KNOT data model documentation: https://icdp-digital-library.github.io/KNOT/website /ENG/data_model.html.

To differentiate types of research products, the KNOT-DM uses a type attribute and a native controlled vocabulary, the KNOT Taxonomy,[159] as values. These include corpus, database, dataset, digital archive, digital catalogue, digital edition, digital library, digital platform, digital repository, knowledge base, knowledge graph, ontology, and software.

Going back to the above mentioned research questions, we decided to maintain KNOT-DM's conceptual distinctions between research products and projects, and between research products and their access services and publications. However, we took a different approach to representing research product types. We modelled various research products, starting with those from our pilots, as subclasses of a generic 'research product' concept, formalised as `schema:Dataset`. These subclasses ('Ontology', 'Software', 'Linked Open data', 'Digital Scholarly Edition', and 'Text Collection') are more thorough and specific to the domain of the ATLAS catalogue than the ones offered in the models presented above, usually limited to 'Publication', 'Software' and 'Other'.

After addressing the research questions and establishing our general approach, we continued data modelling by analysing the pilots. The pilots' analysis was crucial for identifying common properties across different types of research products and those specific to particular product types. This analysis, extensively described in the previous chapter, yielded valuable insights and led to the selection of domain-specific metadata.[160]

The data model was then formalised as an OWL 2 DL ontology. The resulting ATLAS Ontology 1.0[161] is based on the classes and properties of Schema.org (version 28.0), a vocabulary which has already been shown to be suitable for describing and aggregating metadata for Cultural Heritage objects [Freire et al. 2018]. The complexity of the digital Cultural Heritage research domain required the integration of other, more detailed models concerning the DH domain, particularly the FaBiO (FRBR-aligned bibliographic ontology) and DC Terms. The first version of the ATLAS ontology introduces new classes and properties to enhance granularity and specificity, facilitating coherent connections between classes across different vocabularies. It also describes the five common research product types identified in the pilot analysis: text collections; digital scholarly editions; software tools; linked open datasets; and ontologies.

Initially, we tested and validated the newly created model by reusing metadata from the preliminary analysis of pilot resources to develop our first Knowledge Graph.[162] Subsequently, we conducted a further application-based[163] evaluation through a hands-on session, held on 26 March 2025 during the ATLAS Workshop

---

[159]KNOT Taxonomy: https://github.com/icdp-digital-library/KNOT/blob/main/data_model/controlled_vocabularies/1.5/ktx.ttl.

[160]A comprehensive list of metadata and examples illustrating how they were used to describe the pilots is freely available on Zenodo: https://doi.org/10.5281/zenodo.13993057.

[161]https://doi.org/10.5281/zenodo.14058231.

[162]https://doi.org/10.5281/zenodo.14058143.

[163]For a complete description of the ATLAS web application, refer to Chapter 3.

in Bologna.[164] This second validation stage involved scholars, researchers and students in the cataloguing process, enabling them to provide valuable feedback and insights.[165]

Much of the feedback we received requested the inclusion of additional research product types in the ontology. Although the ATLAS ontology was designed with a modular approach from the outset that allows research product types to be easily included as subclasses of `atlas:ResearchProduct`, we have decided to expand our model. To allow users to describe a wider range of research products, we modified the ontology as follows:

- We developed two new classes for two other common categories of DH research outcomes: language models (`atlas:LanguageModel`) and 3D digital twins of cultural objects (`atlas:3DDigitalTwin`).

- We added a new property to `schema:Dataset` that enables users to categorise a wider range of research products that are not included in the Atlas Ontology as a class by associating them with a term of the KNOT taxonomy .

These novelties led to the development of the second version of the ontology: ATLAS Ontology 2.0,[166] which is described in detail in this chapter. To test the final version of the ontology, we held a hands-on session on 22 September 2025 during the final ATLAS workshop in Venice.[167] This second application-based evaluation involved PhD and Master's students, who were asked to browse and search the catalogue for data, providing valuable feedback.

In order to identify the metadata necessary for effectively describing language models and 3D digital twins in our catalogue, we selected a few pilot research products (listed below) and consulted researchers from outside the ATLAS project team, particularly those specialising in 3D digital twins. Our team lacked experience in developing this type of research product. For this reason, we have chosen not to present these products in detail alongside those that were used as pilots from the beginning.

Unlike the other pilots, no formal FAIRness evaluation was conducted for these research products. Our work with these products was limited to exploring metadata requirements and understanding how they could be represented in the catalogue. Consequently, no FAIRness evaluation summaries are available and the process followed here slightly differs from that applied to the other pilots.

- **Minerva** (website: https://nlp.uniroma1.it/minerva/) is a family of large language models (LLMs) developed by Sapienza NLP[168] in collaboration with the Future Artificial Intelligence Research (FAIR) foundation[169] and

---

[164]https://dh-atlas.github.io/workshop.html.

[165]https://doi.org/10.5281/zenodo.15173473.

[166]https://dh-atlas.github.io/deliverables/ontology/2.0/index-en.

[167]https://dh-atlas.github.io/workshop.html.

[168]https://nlp.uniroma1.it/.

[169]https://fondazione-fair.it/en/.

CINECA, an Italian IT company. The models and their data are open, with approximately half of the pretraining data composed of Italian text.

- **OLMo 2** (website: https://allenai.org/olmo) is a family of language models developed from start to finish with open and accessible training data, open-source training code and reproducible training recipes. It was developed by Ai2,[170] a non-profit AI research institute based in Seattle.

- **BrancacciPOV** (website: http://brancaccipov.cnr.it/) is a project developed by the Institute of Cultural Heritage Science[171] of the National Research Council in collaboration with the Municipality of Florence. The project aims to improve access to the Brancacci Chapel and preserve its cycle of frescoes, which were created by Masaccio and Masolino da Panicale. The result is an immersive virtual reality environment that allows users to explore the frescoes and is freely accessible online.

- **DiSpLay** (Digital Spolia Layering, website: https://pric.unive.it/projects /display/home) is a georeferenced, interactive map of reused sculptures (called *spolia*) preserved in Venice and its lagoon, in both architectural contexts and as freestanding structural elements. It was developed within the Venice Centre for Digital and Public Humanities.[172] Users can browse the map online and access 3D models of the *spolia*.

Description

At the core of the ATLAS model lies the `schema:Dataset` class, which defines research products as structured sets of information focused on specific topics of interest. We chose this class for three main reasons. Firstly, as mentioned in the previous section, this class seemed like a good starting point for our modelling process as it can cover different types of research products. As it is defined as "a body of structured information describing one or more topics of interest", we concluded that the defining feature of a research product is that it is accompanied by relevant, structured metadata and documentation, making it a scholarly output. While this choice may not be ideal for representing all types of products included in our model, particularly software tools, we wanted to emphasise the importance of the 'research' aspect. Secondly, we wanted to use a Schema.org class, as this model has proven to be more interoperable and easier to integrate than the other models we analysed. Lastly, this class provides many properties that align perfectly with the metadata we wanted to include in our data model to thoroughly describe research products.

The `schema:Dataset` class branches into seven specialised subclasses (`atlas:LinkedOpenData`; `atlas:DigitalScholarlyEdition`; `atlas:Ontology`; `atlas:Software`; `atlas:TextCollection`; `atlas:LanguageModel`; `atlas:3D-DigitalTwin`), each designed to provide a more detailed definition and

---

[170]https://allenai.org/.

[171]https://www.ispc.cnr.it/.

[172]https://www.unive.it/pag/39287.

categorisation of various types of Research Products. Consequently, each
subclass is characterised by distinct properties, and for enhanced clarity, each is
also aligned as a subclass of a corresponding class from the FaBiO ontology.
This alignment also establishes a meaningful connection with the FRBR model,
positioning ATLAS' Research Products as Expressions within that framework.

Figure 11: overview of the ATLAS ontology classes.

The relationship between a Research Product and the Research Project respon-
sible for its creation is expressed through the `schema:producer` property, which
links a `schema:Dataset` instance to a `schema:Organization`, more specifically
to a `schema:ResearchProject`.

The ATLAS ontology also aims to provide detailed information on the ac-
cess methods for Research Products. To support this, two key classes from
Schema.org—`schema:SoftwareApplication` and `schema:WebSite`—have been
integrated. The former encompasses any type of computer programme, including
SPARQL endpoints or API services, which facilitates access to the content of a
`schema:Dataset` by offering specific functionalities. In other cases, a Research
Product may be made accessible through a dedicated website, as seen with digital
edition visualisation platforms. These websites typically use advanced software
to process the content of the Research Product and generate an interactive
browsing interface.

The illustration below presents a simplified use case drawn from the pilots. A
Research Product, specifically an `atlas:LinkedOpenData` (Zeri Photo Archive
RDF Dataset) is connected to its Research Project (`schema:ResearchProject`),
Zeri & LODE, and SPARQL endpoint (`schema:SoftwareApplication`).

In the following sections, we provide a detailed presentation of the Research
Product class and its subclasses. For information on other classes and prop-

Figure 12: example of how the ATLAS ontology can be used to represent the Zeri and LODE research project, as well as its outputs.

erties included in the ATLAS ontology, please refer to the ATLAS ontology documentation.[173]

Research Product

In the ATLAS ontology research products are modelled as instances of `schema:Dataset`,[174] which is defined as "a body of structured information describing some topic(s) of interest".

As illustrated in Schema.org's data and datasets overview,[175] "[w]hen describing collections of packaged data, for example as published in scientific, scholarly or governmental 'open data' repositories, the Dataset type can be used, alongside DataCatalog to indicate the overall collection, and DataDownload for specific representations of a dataset." These three classes are designed for applications that publish or integrate different kinds of data, such as the ATLAS catalogue. In the ATLAS ontology, we implement this approach by associating each Research Product with a Research Project (where applicable) and providing a downloads property.

---

[173]https://dh-atlas.github.io/deliverables/ontology/index-en.html.

[174]https://schema.org/Dataset.

[175]https://schema.org/docs/data-and-datasets.html.

The Research Product class includes multiple properties, some mandatory to ensure essential identifying information. The recommended properties are designed to thoroughly describe research products following FAIR principles.

Prefixes used in RDF examples

>
@prefix **atlas**: <https://w3id.org/dh-atlas/>
@prefix **bibframe**: <http://id.loc.gov/ontologies/bibframe/>;
@prefix **dcterms**: <http://purl.org/dc/terms/>
@prefix **db**: <http://dbpedia.org/ontology/>
@prefix **etv**: <https://w3id.org/dh-atlas-vocabularies/etv/>
@prefix **osw**: <http://ontosoft.org/software>
@prefix **prov**: <http://www.w3.org/ns/prov#>
@prefix **rdfs**: <http://www.w3.org/2000/01/rdf-schema#>
@prefix **schema**: <https://schema.org/>
@prefix **skos**: <http://www.w3.org/2004/02/skos/core#>
@prefix **vann**: <http://purl.org/vocab/vann/>
@prefix **xsd**: <http://www.w3.org/2001/XMLSchema#>

Title   *Mandatory*

Value type: **String**	*Domain:* **Dataset**, **Research Project**, **Web Site**
Cardinality: **MANY**	
RDF property: **schema:name**	

The title of the research product and its abbreviations or aliases. If multiple titles in different languages are available, it is possible to indicate them all.

```
<https://w3id.org/dh-atlas/1729030147-4807222> a schema:Dataset ,
        <http://purl.org/spar/fabio/Anthology> ,
        atlas:TextCollection ;
    ...
    schema:name "Archivio della Latinità Italiana del Medioevo
    ↪  (ALIM) Digital Library"@en ;
```

Description *Mandatory*

Value type: **String**	*Domain:* **Dataset**, **Research Project**, **Web Site**
Cardinality: **MANY**	
RDF property: **schema:description**	

Brief text that describes the research product, its content and main features.

The description should highlight the research product's innovative features.

```
<https://w3id.org/dh-atlas/1729156722-0141878> a schema:Dataset ,
        <http://purl.org/spar/fabio/Anthology> ,
        atlas:TextCollection ;
    rdfs:label "Biblioteca Italiana"@en ;
...
    schema:description "Biblioteca Italiana is a digital library
    ↪  of texts representative of the Italian cultural and
    ↪  literary tradition from the Middle Ages to the 20th
    ↪  century, with more than 3,500 titles in its catalog. It
    ↪  is divided into three main collections: BibIt,
    ↪  Incunaboli, Scrittori d'Italia."@en ;
```

Creator *Mandatory*

Value type: **URI**	*Domain:* **Dataset**
Cardinality: **MANY**	*Range:* **schema:Person**, **schema:Organization**
RDF property: **schema:creator**	

The research product's creator(s). It may be a single person, a group of people or an organisation.

Use this field to indicate the editors of **digital scholarly editions** and **text collections**.

We recommend using ORCID identifiers for people and the Research Organization Registry for organizations[176].

```
<https://w3id.org/dh-atlas/1729160920-6521974> a schema:Dataset ,
        <http://purl.org/spar/fabio/Anthology> ,
        atlas:TextCollection ;
    rdfs:label "Musisque Deoque (MQDQ)"@en ;
```

---

[176]https://ror.org/.

```
    schema:creator
    ↪   <https://w3id.org/dh-atlas/1729169549-7938185>,
  ... > .
<https://w3id.org/dh-atlas/1729169549-7938185> rdfs:label "Paolo
↪   Mastandrea"^^xsd:string .
```

Contributor

Value type: **URI**	*Domain:* **Dataset**
Cardinality: **MANY**	*Range:* **schema:Person**, **schema:Organization**
RDF property: **schema:contributor**	*Subproperty:* **atlas:referencedAuthor**

A person, group, or organisation that has contributed to the creation or development of the research product.

We recommend using ORCID identifiers for people and the Research Organization Registry for organizations.

```
<https://w3id.org/dh-atlas/1729175944-0618887> a schema:Dataset ,
        <http://purl.org/spar/fabio/Anthology> ,
        atlas:TextCollection ;
    schema:contributor
    ↪   <https://w3id.org/dh-atlas/1729175545-28137> ,
... > .
<https://w3id.org/dh-atlas/1729175545-28137> rdfs:label "Martina
↪   Pavoni"^^xsd:string .
```

Publisher   *Mandatory*

Value type: **URI**	*Domain:* **Dataset**
Cardinality: **MANY**	*Range:* **schema:Person**, **schema:Organization**
RDF property: **schema:publisher**	

The publisher of the research product. It may also be a university or other institution hosting the research product.

We recommend using ORCID identifiers for people and the Research Organization Registry for organizations.

```
<https://w3id.org/dh-atlas/1729185052-855823> a schema:Dataset ,
        <http://purl.org/spar/fabio/MetadataDocument> ,
        atlas:DigitalScholarlyEdition ;
    rdfs:label "VaSto - Varchi, Storia fiorentina Digital
    ↪  Edition"@en ;
    schema:publisher
    ↪  <https://w3id.org/dh-atlas/1729185609-482096> ;
... > .
<https://w3id.org/dh-atlas/1729185609-482096> rdfs:label "/DH.ARC
↪  - Digital Humanities Advanced Research Centre"^^xsd:string .
```

Landing Page   *Mandatory*

Value type: **URL**	*Domain:* **Dataset, Computer Program, Web Site, Organization**
Cardinality: **ONE**	*Range:* **xsd:anyURI**
RDF property: **schema:url**	

The URL of the web page where the research product is presented, and is possible to access its distributions.

A landing page can be either a website built from scratch or a README file in a repository.

```
<https://w3id.org/dh-atlas/1729019160-8370216> a schema:Dataset ,
        <http://purl.org/spar/fabio/MetadataDocument> ,
        atlas:DigitalScholarlyEdition ;
    rdfs:label "Codice Pelavicino Digital Edition"@en ;
    schema:url <https://pelavicino.labcd.unipi.it/evt/> ;
... > .
```

Identifier

Value type: **String**	*Domain:* **Dataset, Agent**
Cardinality: **MANY**	
RDF property: **schema:identifier**	

A unique identifier of the research product.

The same research product can be assigned multiple identifiers of different kinds.

We recommend using international standard codes such as ISBN and ISSN, and persistent identifiers such as DOI, W3ID and Handle.

```
<https://w3id.org/dh-atlas/1729190822-8668764> a schema:Dataset ,
        <http://purl.org/spar/fabio/MetadataDocument> ,
        atlas:DigitalScholarlyEdition ;
    rdfs:label "National Edition of Aldo Moro's Works"@en ;
    schema:identifier "https://doi.org/10.6092/unibo/aldomoro"@en
    ↪   ;
... > .
```

Release Date   *Mandatory*

Value type: **Date**	*Domain:* **Dataset**
Cardinality: **ONE**	*Range:* **xsd:date**
RDF property:	
schema:datePublished	

The date the current version of the research product was released.

If this information is not available, enter the release date of a previous version of the research product.

If the full date can not be applied, indicate the year.

```
 <https://w3id.org/dh-atlas/1729010538-925614> a schema:Dataset ,
        <http://purl.org/spar/fabio/ComputerProgram> ,
        atlas:Software ;
    rdfs:label "Edition Visualization Technology (EVT)"@en ;
    schema:datePublished "2020-01-01"^^xsd:date ;
... > .
```

Current Version

Value type: **String**	*Domain:* **Dataset**
Cardinality: **ONE**	
RDF property: **schema:version**	

The number of the research product's current version.

We recommend using shared numbering schemes, such as:

- SemVer (Semantic Versioning),[177] e.g., 1.0.0

- CalVer (Calendar Versioning)[178]

---

[177]https://semver.org/.
[178]https://calver.org/.

If this information is not available, you may enter "Unknown".

```
<https://w3id.org/dh-atlas/1729187122-1272922> a schema:Dataset ,
        <http://purl.org/spar/fabio/ComputerProgram> ,
        atlas:Software ;
    rdfs:label "Voyant tools"@en ;
    schema:version "2.06.14"@en ;
... > .
```

Access Rights   *Mandatory*

Value type: **URI**	*Domain:* **Dataset**
Cardinality: **ONE**	
RDF property:	
dcterms:accessRights	

This field indicates how the research product may be accessed. If multiple access rights exist, use this field to specify access for the main part of the product. Use Has Part to detail access rights for any subsets with different permissions.

The expected values are defined in the COAR Access Rights vocabulary[179] and namely are:

- embargoed access, when the publication of the research product in open access is delayed due to copyright constraints;

- metadata only access;

- open access;

- restricted access, when the research product can be accessed only by authorised users.

```
<https://w3id.org/dh-atlas/1729073242-5875242> a schema:Dataset ,
        <http://purl.org/spar/fabio/DataFile> ,
        atlas:LinkedOpenData ;
    rdfs:label "Biflow-Toscana Bilingue RDF Dataset"@en ;
    dcterms:accessRights
    ↪   <http://purl.org/coar/access_right/c_abf2> ;
... > .
```

---

[179]http://vocabularies.coar-repositories.org/documentation/access_rights/.

Access Point

Value type: **URI**	*Domain:* **Dataset**
Cardinality: **MANY**	*Range:* **Computer Program, Web Site**
RDF property: **atlas:servedBy**	

URL of the website or data service (e.g., API, SPARQL endpoint) that grants access to the research product.

```
<https://w3id.org/dh-atlas/1728943937-3007112> a schema:Dataset ,
        <http://purl.org/spar/fabio/DataFile> ,
        atlas:LinkedOpenData ;
    rdfs:label "Zeri Photo Archive RDF Dataset"@en ;
    ...
    atlas:servedBy <https://w3id.org/dh-atlas/1728987090-372862>,
        <https://w3id.org/dh-atlas/1728988048-7197032>,
        <https://w3id.org/dh-atlas/1728993673-7262628> .
<https://w3id.org/dh-atlas/1728987090-372862> rdfs:label "Zeri
↪   Photo Archive RDF Dataset - SPARQL endpoint"^^xsd:string .
<https://w3id.org/dh-atlas/1728988048-7197032> rdfs:label "Zeri
↪   Photo Archive RDF Dataset - SPARQL query
↪   interface"^^xsd:string .
<https://w3id.org/dh-atlas/1728993673-7262628> rdfs:label "Zeri
↪   Photo Archive RDF Dataset - LodView RDF Browser"^^xsd:string
↪   .
```

License   *Mandatory*

Value type: **URI**	*Domain:* **Dataset**
Cardinality: **ONE**	
RDF property: **schema:license**	

The licence under which the research product is available to the public.

Expected values are theURIs to the terms defined in the Licences vocabulary[180] by the Italian Government and the licences listed on the Open Source Initiative's website.[181]

If this information is not available, select "Unknown License."[182]

---

[180]https://schema.gov.it/lodview/controlled-vocabulary/licences.

[181]https://opensource.org/licenses.

[182]https://schema.gov.it/lodview/controlled-vocabulary/licences/C_Unknown.

```
<https://w3id.org/dh-atlas/1729034329-548886> a schema:Dataset ,
        <http://purl.org/spar/fabio/DataFile> ,
        atlas:LinkedOpenData ;
    rdfs:label "LiLa Knowledge Base"@en ;
    schema:license
    ↪   <https://w3id.org/italia/controlled-vocabulary/licences/
    A31_CCBYSA40> ;
... > .
```

Downloads *Mandatory*

Value type: **URL**	*Domain:* **Dataset**
Cardinality: **MANY**	*Range:* **xsd:anyURI**
RDF property: **schema:distribution**	

The URL of the download or download page of the research product's files.

```
<https://w3id.org/dh-atlas/1728948776-872414> a schema:Dataset,
        <http://purl.org/spar/fabio/OntologyDocument> ,
        atlas:Ontology ;
    rdfs:label "Historical Context Ontology (HiCO)"@en ;
    schema:distribution
    ↪   <http://marilenadaquino.github.io/hico/current/hico.owl>
    ↪   ;
... > .
```

Status

Value type: **URI**	*Domain:* **Dataset**
Cardinality: **ONE**	
RDF property:	
schema:creativeWorkStatus	

The research product's status in its lifecycle.

The expected values are defined in the EU Dataset Status vocabulary[183] and namely are:

- Completed.

- Deprecated, when it is recommended that the contents of this dataset be no longer used.

---

[183]http://publications.europa.eu/resource/dataset/dataset-status.

- Under development, the dataset may be in an incomplete or faulty state.

- Discontinued, if the dataset is no longer produced or updated.

- Withdrawn, when the dataset is no longer meant to be published.

```
<https://w3id.org/dh-atlas/1728995301-3097954> a schema:Dataset ,
        <http://purl.org/spar/fabio/DataFile> ,
        atlas:LinkedOpenData ;
    rdfs:label "DanteSources RDF Dataset"@en
    schema:creativeWorkStatus
    ↪   <http://publications.europa.eu/resource/authority/dataset
    -status/COMPLETED> ;
... > .
```

Format

Value type: **URI**	*Domain:* **Dataset**
Cardinality: **MANY**	
RDF property:	
schema:encodingFormat	

The file format(s) in which the research product's data are stored.

Expected values areURIs to the terms of the EU File Type vocabulary[184] or of the IANA Media Types[185] list.

```
<https://w3id.org/dh-atlas/1728948776-872414> a schema:Dataset,
        <http://purl.org/spar/fabio/OntologyDocument> ,
        atlas:Ontology ;
    rdfs:label "Historical Context Ontology (HiCO)"@en ;
 schema:encodingFormat
 ↪   <http://publications.europa.eu/resource/authority/
 file-type/JSON_LD>,
<http://publications.europa.eu/resource/authority/file-type/OWL>,
<http://publications.europa.eu/resource/authority/file-type/RDF_
XML> ;
... > .
```

---

[184]http://publications.europa.eu/resource/dataset/file-type.

[185]http://www.iana.org/assignments/media-types/media-types.xhtml.

Metadata Standards

Value type: **URI**   *Domain:* **Dataset**
Cardinality: **MANY**
RDF property: **purl:conformsTo**

Standards to which the research product's metadata conform.

Recommended values can be found in:

Riley, Jenn, 2018, "Seeing Standards: A Visualization of the Metadata Universe", https://doi.org/10.5683/SP2/UOHPVH, Borealis, V3, UNF:6:gWl/jicj8wtJm4Grmph7TQ== [fileUNF]

```
<https://w3id.org/dh-atlas/1729010538-925614> a schema:Dataset ,
        <http://purl.org/spar/fabio/ComputerProgram> ,
        atlas:Software ;
    rdfs:label "Edition Visualization Technology (EVT)"@en ;
    dcterms:conformsTo <http://www.tei-c.org/> ;
... > .
```

Language

Value type: **URI**   *Domain:* **Dataset**
Cardinality: **MANY**
RDF property: **schema:inLanguage**

The language(s) in which the research product is expressed.

If multiple languages are used—for example, in labels of an ontology or various texts in a collection—we recommend indicating all of them.

Expected values are URIs to the terms of the EU Language vocabulary.[186]

```
<https://w3id.org/dh-atlas/1728948052-580608> a schema:Dataset,
        <http://purl.org/spar/fabio/OntologyDocument> ,
        atlas:Ontology ;
    rdfs:label "Semantic Publishing and Referencing Ontologies
    ↪  (SPAR) "@en ;
    schema:inLanguage
    ↪  <http://publications.europa.eu/resource/authority/
    language/ENG> ;
... > .
```

---
[186]http://publications.europa.eu/resource/authority/language.

Type   *Mandatory*

Value type: **URI**	*Domain:* **Dataset**
Cardinality: **ONE**	
RDF property:	
schema:additionalType	

The type of research product.

Expected values are: Digital Scholarly Edition; Text Collection; Linked Open Data; Ontology; Software; Language Model; 3D Digital Twin; Other.

To use further values, select 'Other' and then fill out the field Custom Type.

```
<https://w3id.org/dh-atlas/1728948736-9982028> a schema:Dataset,
        <http://purl.org/spar/fabio/OntologyDocument> ,
        atlas:Ontology ;
    rdfs:label "CIDOC Conceptual Reference Model (CRM)"@en ;
    schema:additionalType
    ↪   <http://purl.org/spar/fabio/OntologyDocument> ;
... > .
```

Custom Type

Value type: **URI**	*Domain:* **Dataset**
Cardinality: **MANY**	
RDF property: **dcterms:type**	

The type of research product, specified using a term from a controlled vocabulary.

This field only becomes available once you have selected 'Other' in the Type field.

Expected values are terms of the KNOT Taxonomy. It is possible to use terms from other relevant taxonomies.

```
<https://w3id.org/dh-atlas/1728948736-9982028> a schema:Dataset,
    rdfs:label "The Medieval Kingdom of Sicily Image Database"@en
    ↪   ;
    dcterms:type <http://w3id.org/knot/taxonomy/image_database> ;
... > .
```

Size

Value type: **String**	*Domain:* **Dataset**
Cardinality: **ONE**	*Range:* **xsd:String**
RDF property: **schema:size**	

The standardised size of a research product, expressed as the number of individual items. For clarity, the value must include the unit of measurement; for example, '500 documents', '230,000 tokens' or '1,000 triples'. We particularly recommend filling out this field for **text collections,** to give users a more detailed overview of the level of completion of the collection.

For **language models**, use this field to indicate the number of parameters in the model.

```
<https://w3id.org/dh-atlas/1757973188-9464774 a schema:Dataset,
        atlas:LanguageModel ;
    rdfs:label "Minerva 7B"@en ;
    schema:size "7 billion parameters"@en ;
... > .
```

```
<https://w3id.org/dh-atlas/1758628052-4316523 a schema:Dataset,
        atlas:TextCollection ;
    rdfs:label "Digital library of Late-Antique Latin Texts
    ↪  (digilibLT)"@en ;
    schema:size "Near 400 texts"@en ;
... > .
```

Research Project

Value type: **URI**	*Domain:* **Dataset**
Cardinality: **ONE**	*Range:* **schema:ResearchProject**
RDF property: **schema:producer**	

The research project within which the research product was created.

Not all research products originate from a research project.

Indicating the research project provides context, helping to situate the research product within its timeline, development stage, and the broader research environment in which it was conceived and created.

```
ALIM
<https://w3id.org/dh-atlas/1729030147-4807222> a schema:Dataset ,
        <http://purl.org/spar/fabio/Anthology> ,
```

```
        atlas:TextCollection ;
    rdfs:label "Archivio della Latinità Italiana del Medioevo
    ↪  (ALIM) Digital Library"@en ;
    schema:producer
    ↪  <https://w3id.org/dh-atlas/1729017470-5332708> ;
... > .
```

```
ALIM's Research Project
<https://w3id.org/dh-atlas/1729017470-5332708> a
↪  schema:Organization,
        schema:ResearchProject ;
    rdfs:label "Archivio della Latinità Italiana del Medioevo
    ↪  (ALIM) Project"@en ;
    schema:description "ALIM (Archivio della Latinità Italiana
    ↪  del Medioevo / Archive of the Italian Latinity of the
    ↪  Middle Ages) ..."@en ;
    schema:foundingDate "1996-01-01"^^xsd:date ;
    schema:funder <https://w3id.org/dh-atlas/1729001319-1277974>
    ↪  ;
    schema:location <https://sws.geonames.org/2523920>,
        ... ;
    schema:member <https://w3id.org/dh-atlas/1729012784-159052>,
        ... ;
    schema:name "Archivio della Latinità Italiana del Medioevo
    ↪  (ALIM) Project"@en ;
    schema:url <https://alim.unisi.it/> .
```

Has Part

Value type: **URI**	*Domain:* **Dataset**
Cardinality: **MANY**	*Range:* **schema:Dataset**
RDF property: **schema:hasPart**	*Is inverse of:* **schema:isPartOf**

A research product that forms part of the catalogued one. This could be, for instance, a sub-collection of texts in a specific language within a larger text collection. An example is the 'Hellenica' corpus within the Musisque Deoque collection.

We recommend filling out this field, to give value to the single components of the research product.

```
<https://w3id.org/dh-atlas/1729160920-6521974> a schema:Dataset ,
        <http://purl.org/spar/fabio/Anthology> ,
        atlas:TextCollection ;
    rdfs:label "Musisque Deoque (MQDQ)"@en ;
```

```
    schema:hasPart <https://w3id.org/dh-atlas/1729012784-159053>;
... > .


<https://w3id.org/dh-atlas/1729012784-159053> a schema:Dataset ,
        <http://purl.org/spar/fabio/Anthology> ,
        atlas:TextCollection ;
    rdfs:label "Poeti d'Italia in lingua latina"@en ;
... > .
```

Is Part Of

Value type: **URI**	*Domain:* **Dataset**
Cardinality: **ONE**	*Range:* **schema:Dataset**
RDF property: **schema:isPartOf**	*Is inverse of:* **schema:hasPart**

Inverse of Has Part this field indicates the research product of which the catalogued one is a component.

```
<https://w3id.org/dh-atlas/1729160920-6521974> a schema:Dataset ,
        <http://purl.org/spar/fabio/Anthology> ,
        atlas:TextCollection ;
    rdfs:label "Musisque Deoque (MQDQ)"@en ;
... > .


<https://w3id.org/dh-atlas/1729012784-159053> a schema:Dataset ,
        <http://purl.org/spar/fabio/Anthology> ,
        atlas:TextCollection ;
    rdfs:label "Poeti d'Italia in lingua latina"@en ;
        schema:isPartOf
        ↪  <https://w3id.org/dh-atlas/1729160920-6521974> ;
... > .
```

Documentation

Value type: **URL**	*Domain:* **Dataset, Web Site, Computer Program**
Cardinality: **MANY**	
RDF property: **schema:usageInfo**	

The URL of the research product's documentation.

```
<https://w3id.org/dh-atlas/1728948052-580608> a schema:Dataset,
        <http://purl.org/spar/fabio/OntologyDocument> ,
```

```
        atlas:Ontology ;
    rdfs:label "Semantic Publishing and Referencing Ontologies
    ↪    (SPAR) "@en ;
    ...
    schema:usageInfo <http://www.sparontologies.net/examples> .
```

Research Activities   *Mandatory*

Value type: **URI**	*Domain:* **Dataset**,**Web Site**,
	Computer Program
Cardinality: **MANY**	
RDF property:	
schema:educationalUse	

Research activities enabled or supported by the research product.

Expected values are URIs to terms from the Taxonomy of Digital Research Activities in the Humanities (TaDiRAH).[187] Multiple activities can be specified, ranging from broad to specific levels of granularity.

```
<https://w3id.org/dh-atlas/1729187122-1272922> a schema:Dataset ,
        <http://purl.org/spar/fabio/ComputerProgram> ,
        atlas:Software ;
    rdfs:label "Voyant tools"@en ;
    schema:educationalUse
    ↪    <https://vocabs.dariah.eu/tadirah/contentAnalysis>,
        <https://vocabs.dariah.eu/tadirah/discovering>,
        <https://vocabs.dariah.eu/tadirah/structuralAnalysis>,
        <https://vocabs.dariah.eu/tadirah/visualAnalysis> ;
... > .
```

Academic Field

Value type: **URI**	*Domain:* **Dataset**
Cardinality: **MANY**	
RDF property: **schema:about**	

The academic field(s) to which the research product pertains.

Expected values are the IDs and/or names from an official classification system. We recommend using the classification established by CUN (National University

---

[187]https://vocabs.dariah.eu/tadirah/en/.

Council of the Italian Ministry of University and Research), which is also
available as a controlled vocabulary.[188]

```
<https://w3id.org/dh-atlas/1728995301-3097954> a schema:Dataset ,
        <http://purl.org/spar/fabio/DataFile> ,
        atlas:LinkedOpenData ;
    rdfs:label "DanteSources RDF Dataset"@en ;
    schema:about
    ↪  <https://w3id.org/italia/controlled-vocabulary/classifi
    cations-for-universities/academic-disciplines/
    SSD-L-FIL-LET-08>,
    ↪  <https://w3id.org/italia/controlled-vocabulary/classifi
    cations-for-universities/academic-disciplines/
    SSD-L-FIL-LET-10>,
    ↪  <https://w3id.org/italia/controlled-vocabulary/classifi
    cations-for-universities/academic-disciplines/
    SSD-M-STO-08> ;
... > .
```

Methodology

Value type: **URI**   *Domain:* **Dataset**
Cardinality: **MANY**
RDF property: **atlas:methodology**

Research activities conducted to create or develop the research product.

We recommend filling out this field to enhance transparency in the research
product's creation process and to share effective methodologies and workflows
with fellow scholars.

Expected values are URIs to terms from the Taxonomy of Digital Research
Activities in the Humanities (TaDiRAH). Multiple activities can be specified,
ranging from broad to specific levels of granularity.

```
<https://w3id.org/dh-atlas/1729019160-8370216> a schema:Dataset ,
        <http://purl.org/spar/fabio/MetadataDocument> ,
        atlas:DigitalScholarlyEdition ;
    rdfs:label "Codice Pelavicino Digital Edition"@en ;
    atlas:methodology
    ↪  <https://vocabs.dariah.eu/tadirah/annotating>,
        <https://vocabs.dariah.eu/tadirah/capturing>,
        <https://vocabs.dariah.eu/tadirah/transcribing> ;
... > .
```

---

[188]https://schema.gov.it/lodview/controlled-vocabulary/classifications-for-universities/aca
demic-disciplines.

Research Product Reuse

Value type: **URI**	*Domain:* **Dataset**
Cardinality: **MANY**	*Range:* **Dataset, Computer Program**
RDF property: **atlas:used**	

Software tools or other research products used to create or develop the research product.

We recommend filling out this field to enhance transparency in the research product's creation process and to share effective methodologies and workflows with fellow scholars.

For language models, use this field to describe any related datasets that were used for training or testing purposes.

Expected values are URIs of: software tools, software libraries, computer programs, digital editions, text collections, ontologies, and linked open datasets.

```
<https://w3id.org/dh-atlas/1729156722-0141878> a schema:Dataset ,
        <http://purl.org/spar/fabio/Anthology> ,
        atlas:TextCollection ;
    rdfs:label "Biblioteca Italiana"@en ;
    ...
    atlas:used <http://www.muruca.org/> .
<http://www.muruca.org/> rdfs:label "MURUCA"^^xsd:string .
```

```
<https://w3id.org/dh-atlas/1729019160-8370216> a schema:Dataset ,
        <http://purl.org/spar/fabio/MetadataDocument> ,
        atlas:DigitalScholarlyEdition ;
    rdfs:label "Codice Pelavicino Digital Edition"@en ;
...
atlas:used <https://w3id.org/dh-atlas/1729010538-925614> .
<https://w3id.org/dh-atlas/1729010538-925614> rdfs:label "Edition
↪  Visualization Technology (EVT)"^^xsd:string .
```

Bibliographic References

Value type: **URL** *Domain:* **Dataset**
Cardinality: **MANY** *Range:* **xsd:anyUri**
RDF property: **schema:citation**

URLs of publications that describe or review the research product.

We recommend linking open access publications when available. Ideally, use the publication's persistent identifier, such as DOI or Handle.

```
<https://w3id.org/dh-atlas/1729185052-855823> a schema:Dataset ,
        <http://purl.org/spar/fabio/MetadataDocument> ,
        atlas:DigitalScholarlyEdition ;
    rdfs:label "VaSto - Varchi, Storia fiorentina Digital
    ↪   Edition"@en ;
    schema:citation
    ↪   <http://doi.org/10.30687/mag/2724-3923/2021/03/006> ;
... > .
```

Note on Research Product

Value type: **String** *Domain:* **Dataset**
Cardinality: **ONE**
RDF property:
atlas:noteOnResearchProduct

Additional information about the research product.

We recommend using this field to clarify the product's methodology and provide guidance on its use and execution, particularly for **software tools** and **language models**. For **text collections** and **digital scholarly editions**, this field can include scholarly notes on topics such as peculiarities of the textual transmission of the edited texts, the editorial criteria applied, the work performed, and any caveats.

```
<https://w3id.org/dh-atlas/1729190822-8668764> a schema:Dataset ,
        <http://purl.org/spar/fabio/MetadataDocument> ,
        atlas:DigitalScholarlyEdition ;
    rdfs:label "National Edition of Aldo Moro's Works"@en ;
    atlas:noteOnResearchProduct "The current edition includes
    ↪   1,398 works, many of which are transcriptions of
    ↪   speeches, making it difficult to provide adequate
    ↪   references to the documents used and source works."@en ;
... > .
```

Mentions

Value type: **URI**   *Domain:* **_Dataset_**
Cardinality: **MANY**
RDF property: **_schema:mentions_**

A set of entities to which the research product contains a reference. For example, people, places, institutions, etc.

Recommended values are links to authority records in VIAF, Wikidata and other similar resources. These entities can be retrieved automatically from the research product thanks to the knowledge extraction feature (described in Chapter 3).

```
<https://w3id.org/dh-atlas/1729019160-8370216/> rdfs:label
↪  "Codice Pelavicino Digital Edition"@en ;
   schema:mentions
   ↪  <https://w3id.org/dh-atlas/1729019160-8370216/extraction
   -1755592940896-1/> ;
... > .
<http://www.wikidata.org/entity/Q18489802> a schema:Place ;
   rdfs:label "Pulica"^^xsd:string .
<http://www.wikidata.org/entity/Q270220> a schema:Place ;
   rdfs:label "Luni"^^xsd:string .
```

Text Collection

In the ATLAS ontology, a text collection is a subclass of `schema:Dataset` and `fabio:Anthology`. While the latter is defined as "a collection of selected literary or scholastic works, for example, poems, short stories, plays or research papers", in our model, Text Collection comprises a broader selection of research products, namely: digital libraries, such as ALIM and Biblioteca Italiana; linguistic corpora, e.g., ParlaMint;[189] collections of texts used for vocabularies and thesauri, e.g., Tesoro della Lingua Italiana delle Origini (TLIO).[190]

The line between text collections and digital scholarly editions often blurs. On one hand, texts published within a collection are often edited and curated, applying philological methods. On the other hand, there are many digital scholarly editions of multiple texts or text collections (e.g., Corpus Rhythmorum Musicum).[191] For this reason, data curators can choose to represent a research product either as a Text Collection or as a Digital Scholarly Edition, depending on their specific needs. The selected class determines which properties are available. The Digital Scholarly Edition class enables detailed documentation of sources, documents and witnesses, while the Text Collection class is designed for

---

[189] https://www.clarin.eu/parlamint.
[190] https://tlio.ovi.cnr.it/TLIO/.
[191] https://www.corimu.unisi.it/public/frontend/index.

broader information. Both classes share several core properties for describing works, authors, edition types, and genres of the texts.

Edition type *Mandatory*

Value type: **URI**	*Domain:* **Text Collection** (**Digital Scholarly Edition**)
Cardinality: **MANY**	*Super property:* **dcterms:type**
RDF property: **atlas:editionType**	

The editing methods applied to the texts in the collection.

Available values are:

- best-manuscript edition;

- critical edition;

- diplomatic edition;

- documentary edition;

- eclectic edition;

- facsimile edition;

- genetic edition;

- interpretative edition;

- multi-witnessed edition;

- single-witnessed edition;

- synoptic edition.

It is possible to select multiple values.

To further describe how the digital edition was implemented from a technical and practical point of view (e.g., semantic edition, LOD edition, crowdsourced edition) use the Research Product's field Methodology (e.g., https://vocabs.dariah.eu/tadirah/semantification,%5D(https://vocabs.dariah.eu/tadirah/semantification), https://vocabs.dariah.eu/tadirah/linkedOpenData, https://vocabs.dariah.eu/tadirah/crowdsourcing).

Note   The available values include types applicable to both digital and printed editions, since scholarly goals are typically consistent across formats. These types were carefully selected to represent the main categories of digital editions, based on the 'Parvum Lexicon Stemmatologicum' [Roelli and Macé 2015] and the 'Lexicon of Scholarly Editing' [Dillen 2020].

For complete definitions and examples, please consult the ATLAS Edition Type Vocabulary (ETV).[192]

```
<https://w3id.org/dh-atlas/1729156722-0141878> a schema:Dataset ,
        <http://purl.org/spar/fabio/Anthology> ,
        atlas:TextCollection ;
    rdfs:label "Biblioteca Italiana"@en ;
    atlas:editionType etv:FacsimileEdition etv:DiplomaticEdition;
... > .
```

Collected works

Value type: **URI**	*Domain:* **Text Collection** (**Digital Scholarly Edition**)
Cardinality: **MANY**	
RDF property: **bibframe:expressionOf**	

URI of the cataloguing records of the works included in the text collection, if applicable.

We recommend using records from VIAF. Another viable option is the Open Library.[193]

If no record exists, enter the title(s).

We recommend inserting the titles of the most relevant works included in the text collection.

```
<https://w3id.org/dh-atlas/1729156722-0141878> a schema:Dataset ,
        <http://purl.org/spar/fabio/Anthology> ,
        atlas:TextCollection ;
    rdfs:label "Musisque Deoque (MQDQ)"@en ;
    bibframe:expressionOf "http://viaf.org/viaf/294478900"
... > .
<http://viaf.org/viaf/294478900> rdfs:label
↪  "Georgiche"^^xsd:string .
```

---

[192]https://w3id.org/dh-atlas-vocabularies/etv/.
[193]https://openlibrary.org/.

Author of the texts

Value type: **URI**	*Domain:* **Text Collection** (***Digital Scholarly Edition***)
Cardinality: **MANY**	*Range:* **schema:Person**, **schema:Organization**
RDF property: **atlas:referencedAuthor**	*Super property:* **schema:contributor**

URI of the cataloguing record(s) of the author(s) of the texts included in the text collection. We recommend using VIAF.

```
<https://w3id.org/dh-atlas/1729156722-0141878> a schema:Dataset ,
        <http://purl.org/spar/fabio/Anthology> ,
        atlas:TextCollection ;
    rdfs:label "Biblioteca Italiana"@en ;
    atlas:referencedAuthor <https://viaf.org/viaf/39389587/> ;
... > .
<https://viaf.org/viaf/39389587/> rdfs:label "Alfieri,Vittorio"
↪  ^^xsd:string .
```

Genre of the collected texts

Value type: **URI**	*Domain:* **Text Collection** (***Digital Scholarly Edition***)
Cardinality: **MANY**	
RDF property: **schema:genre**	

The genre(s) of the collected texts.

Expected values are terms of the CWRC Genre Ontology.[194]

If specific genres such as madrigal are not available, we recommend inserting the generic categories in which they fit best, e.g., poetic genre (https://sparql.cwrc.ca/ontologies/genre.html#PoeticGenre).

```
<https://w3id.org/dh-atlas/1729160920-6521974> a schema:Dataset ,
        <http://purl.org/spar/fabio/Anthology> ,
        atlas:TextCollection ;
    rdfs:label "Musisque Deoque (MQDQ)"@en ;
    schema:genre
    ↪  <http://sparql.cwrc.ca/ontologies/genre#PoeticGenre> ;
... > .
```

---

[194]https://sparql.cwrc.ca/ontologies/genre.html.

Geographical Coverage

Value type: **Place**   *Domain:* **Text Collection**
Cardinality: **MANY**
RDF property:
schema:spatialCoverage

The place(s) which are the focus of the text collection's content. We recommend using GeoNames.[195]

```
<https://w3id.org/dh-atlas/1757506267-1739075 a schema:Dataset,
        atlas:TextCollection ;
    rdfs:label "Codice Diplomatico della Lombardia medievale
    ↪  (CDLM)"@en ;
schema:spatialCoverage <https://sws.geonames.org/3171364>,
        <https://sws.geonames.org/3173434>,
        <https://sws.geonames.org/3177837>,
... > .
<https://sws.geonames.org/3171364> rdfs:label "Province of
↪  Pavia"^^xsd:string .
<https://sws.geonames.org/3173434> rdfs:label "Province of
↪  Milan"^^xsd:string .
<https://sws.geonames.org/3177837> rdfs:label "Province of
↪  Cremona"^^xsd:string .
```

Temporal Coverage

Value type: **String**   *Domain:* **Text Collection**
Cardinality: **MANY**
RDF property:
schema:temporalCoverage

The period that the text collection's content applies to.

It is possible to specify values in the ISO standard date format YYYY-MM-DD,[196] by indicating the centuries in roman numbers (e.g., "XIX") or by referring to historical eras and periods (e.g., "Late Middle Ages").

```
<https://w3id.org/dh-atlas/1758628052-4316523 a schema:Dataset,
        atlas:TextCollection ;
```

---

[195]https://www.geonames.org/.
[196]https://www.iso.org/iso-8601-date-and-time-format.html.

```
    rdfs:label "Digital library of Late-Antique Latin Texts
    ↪  (digilibLT)"@en ;
schema:temporalCoverage <http://www.wikidata.org/entity/Q217050>,
        <http://www.wikidata.org/entity/Q8089>,
        <http://www.wikidata.org/entity/Q8090>,
... > .
<http://www.wikidata.org/entity/Q217050> rdfs:label "late
↪  antiquity"^^xsd:string .
<http://www.wikidata.org/entity/Q8089> rdfs:label "7th
↪  century"^^xsd:string .
<http://www.wikidata.org/entity/Q8090> rdfs:label "6th
↪  century"^^xsd:string .
```

Digital Scholarly Edition

The class `atlas:DigitalScholarlyEdition` is a subclass of `schema:Dataset` and of `fabio:MetadataDocument`. We chose this alignment for two main reasons.

First, while the FaBiO ontology provides a class for critical editions (`fabio:CriticalEdition`), it lacks a specific class for digital scholarly editions. Since digital scholarly editions are primarily TEI-encoded documents where critical annotations serve as metadata, we use `fabio:MetadataDocument`. Second, the ATLAS ontology models research products as types of expressions. Since `fabio:CriticalEdition` is a subclass of `frbr:Work` but `fabio:MetadataDocument` is a subclass of `frbr:Expression`, aligning `atlas:DigitalScholarlyEdition` with the latter maintains our desired hierarchy.

The class `atlas:DigitalScholarlyEdition` has been defined as follows: "An information resource which offers a critical representation of documents or texts of literary or historical value following a shared methodology determined by a digital paradigm."

Edition type   *Mandatory*

Value type: **URI**	*Domain:* **Digital Scholarly Edition**
Cardinality: **MANY**	*Super property:* **dcterms:type**
RDF property: **atlas:editionType**	

The type(s) of edition. This indicates the editorial approach taken by the team and the main objectives of the edition.

Available values are:

- best-manuscript edition;

- critical edition;

- diplomatic edition;

- documentary edition;

- eclectic edition;

- facsimile edition;

- genetic edition;

- interpretative edition;

- multi-witnessed edition;

- single-witnessed edition;

- synoptic edition.

It is possible to select multiple values.

To further describe how the digital edition was implemented from a technical and practical point of view (e.g., semantic edition, LOD edition, crowdsourced edition) use the Research Product's field Methodology (e.g., https://vocabs.dariah.eu /tadirah/semantification, https://vocabs.dariah.eu/tadirah/linkedOpenData, https://vocabs.dariah.eu/tadirah/crowdsourcing).

Note The available values include types applicable to both digital and printed editions, since scholarly goals are typically consistent across formats. These types were carefully selected to represent the main categories of digital editions, based on the 'Parvum Lexicon Stemmatologicum' [Roelli and Macé 2015] and the 'Lexicon of Scholarly Editing' [Dillen 2020].

For complete definitions and examples, please consult the ATLAS Edition Type Vocabulary (ETV).

```
<https://w3id.org/dh-atlas/1729019160-8370216> a schema:Dataset ,
        <http://purl.org/spar/fabio/MetadataDocument> ,
        atlas:DigitalScholarlyEdition ;
    rdfs:label "Codice Pelavicino Digital Edition"@en ;
    atlas:editionType etv:DocumentaryEdition ;
... > .
```

Edited work   *Mandatory*

Value type: **URI**   *Domain:* **Digital Scholarly Edition
                                        (Text Collection)**

Cardinality: **MANY**
RDF property:
bibframe:expressionOf

URI of the cataloguing record of the edited work.

We recommend using records from VIAF. Another viable option is the Open Library.

If no record exists, enter the title.

```
<https://w3id.org/dh-atlas/1729185052-855823> a schema:Dataset ,
        <http://purl.org/spar/fabio/MetadataDocument> ,
        atlas:DigitalScholarlyEdition ;
    rdfs:label "VaSto - Varchi, Storia fiorentina Digital
    ↪  Edition"@en ;
    bibframe:expressionOf
    ↪  "http://viaf.org/viaf/893145857890623020112" ;
... > .
```

Original document or witness available at   *Mandatory*

Value type: **URL**   *Domain:* **Digital Scholarly Edition**
Cardinality: **MANY**
RDF property: **bibframe:
associatedResource**

The URL of a web resource presenting as digital facsimile the document(s) or witness(es) on which the edition is based.

If an edition is based on more than one document/witness, we recommend inserting all documents/witnesses that are available online.

If the digital facsimile of the document/witness is not available, we recommend inserting the link to a cataloguing record in the Open Library or other resources.

We also recommend adding the bibliographic or archival reference of the document or witness as a label to the URL. For codices and manuscripts, insert the shelf mark.

```
<https://w3id.org/dh-atlas/1729185052-855823> a schema:Dataset ,
        <http://purl.org/spar/fabio/MetadataDocument> ,
        atlas:DigitalScholarlyEdition ;
    rdfs:label "VaSto - Varchi, Storia fiorentina Digital
    ↪  Edition"@en ;
    bibframe:associatedResource
    ↪  "https://www.mirabileweb.it/manuscript/roma-accademia-
    nazionale-dei-lincei-biblioteca-(bi-manuscript/19567"
    ↪  rdfs:label "Roma, Accademia Nazionale dei Lincei,
    ↪  Biblioteca (Biblioteca Corsiniana) 39.D.5 (Rossi 395;
    ↪  Cors. 1532)" ;
... > .
```

Author of the edited text

Value type: **URI**	*Domain:* **Digital Scholarly Edition (Text Collection)**
Cardinality: **MANY**	*Range:* **schema:Person**, **schema:Organization**
RDF property: **atlas:referencedAuthor**	*Super property:* **atlas:contributor**

URI of the cataloguing record of the edited text's author. We recommend using VIAF.

```
<https://w3id.org/dh-atlas/1729190822-8668764> a schema:Dataset ,
        <http://purl.org/spar/fabio/MetadataDocument> ,
        atlas:DigitalScholarlyEdition ;
    rdfs:label "National Edition of Aldo Moro's Works"@en ;
    atlas:referencedAuthor <http://www.viaf.org/viaf/73876435> ;
... > .
<http://www.viaf.org/viaf/73876435> rdfs:label "Aldo
↪  Moro"^^xsd:string .
```

Genre

Value type: **URI**	*Domain:* **Digital Scholarly Edition (Text Collection)**
Cardinality: **MANY**	
RDF property: **schema:genre**	

The genre(s) of the edited text.

Expected values are terms of the CWRC Genre Ontology.

If specific genres such as madrigal are not available, we recommend inserting the generic categories in which they fit best, e.g., poetic genre (https://sparql.cwrc.ca/ontologies/genre.html#PoeticGenre).

```
<https://w3id.org/dh-atlas/1729190822-8668764> a schema:Dataset ,
        <http://purl.org/spar/fabio/MetadataDocument> ,
        atlas:DigitalScholarlyEdition ;
    rdfs:label "National Edition of Aldo Moro's Works"@en ;
    schema:genre <http://sparql.cwrc.ca/ontologies/genre#speech>,
    ↪   <http://sparql.cwrc.ca/ontologies/genre#lecture> ;
... >
```

Facsimile available at

Value type: **URL**	*Domain:* ***Digital Scholarly Edition***
Cardinality: **MANY**	
RDF property:	
bibframe:accompaniedBy	

The URL to the documents' or witnesses' facsimile(s) provided within the catalogued edition, if applicable.

```
<https://w3id.org/dh-atlas/:1729019160-8370216 a schema:Dataset,
        atlas:DigitalScholarlyEdition ;
    rdfs:label "Codice Pelavicino Digital Edition"@en ;
bibframe:accompaniedBy <https://pelavicino.labcd.unipi.it/evt> ;
... >
```
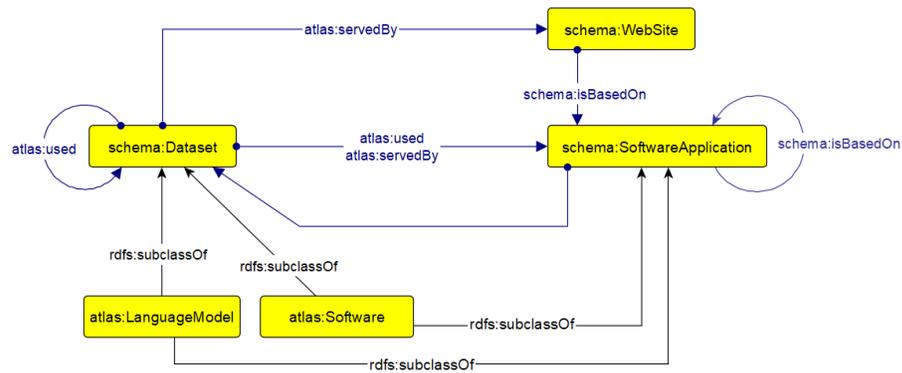
Software

In the ATLAS ontology, a piece of software is a subclass of schema:Dataset[197] and `schema:SoftwareApplication`. This class encompasses various kinds of computer programmes and software artefacts both in source or compiled form.

---

[197]The rationale behind the decision to use `schema:Dataset` to represent all types of search products in our model, including software, is presented in this chapter at the beginning of the description section.

Figure 13: overview of the classes used within the ATLAS ontology to represent software tools as research products.

Programming Language   *Mandatory*

Value type: **URI**	*Domain:* **Computer Program**
Cardinality: **MANY**	
RDF property:	
osw:hasImplementationLanguage	

The programming language(s) employed in the software's development.

Recommended values are links to authority records in Wikidata and other similar resources.

```
<https://w3id.org/dh-atlas/1729187122-1272922> a schema:Dataset ,
        <http://purl.org/spar/fabio/ComputerProgram> ,
        atlas:Software ;
    rdfs:label "Voyant tools"@en ;
    osw:hasImplementationLanguage
    ↪   <http://www.wikidata.org/entity/Q2005>,
        <http://www.wikidata.org/entity/Q251>,
        <http://www.wikidata.org/entity/Q32110>,
        <http://www.wikidata.org/entity/Q46441> ;
... > .
<http://www.wikidata.org/entity/Q2005> rdfs:label
↪   "JavaScript"^^xsd:string .
<http://www.wikidata.org/entity/Q251> rdfs:label
↪   "Java"^^xsd:string .
<http://www.wikidata.org/entity/Q32110> rdfs:label
↪   "XSLT"^^xsd:string .
```

```
<http://www.wikidata.org/entity/Q46441> rdfs:label "Cascading
↪    Style Sheets"^^xsd:string .
```

Code Repository URL   *Mandatory*

Value type: **URL**	*Domain:* **Software**
Cardinality: **ONE**	*Range:* **xsd:anyUri**
RDF property: **osw:hasCodeLocation**	

The URL of the source code's repository (e.g., GitHub, GitLab, etc.).

```
<https://w3id.org/dh-atlas/1729010538-925614> a schema:Dataset ,
        <http://purl.org/spar/fabio/ComputerProgram> ,
        atlas:Software ;
    rdfs:label "Edition Visualization Technology (EVT)"@en ;
    osw:hasCodeLocation
    ↪    <https://github.com/evt-project/evt-viewer/> ;
... > .
```

Input Format

Value type: **URI**	*Domain:* **Software**
Cardinality: **MANY**	
RDF property: **osw:hasInput**	

The file format(s) of data the software can process as input, if applicable.

Expected values are URIs to the terms of the EU File Type Vocabulary[198] or of the IANA Media Types list.[199]

```
<https://w3id.org/dh-atlas/1729187122-1272922> a schema:Dataset ,
        <http://purl.org/spar/fabio/ComputerProgram> ,
        atlas:Software ;
    rdfs:label "Voyant tools"@en ;
    osw:hasInput
    ↪    <https://publications.europa.eu/resource/authority/
    file-type/HTML>,
<https://publications.europa.eu/resource/authority/file-type/TXT>,
<https://publications.europa.eu/resource/authority/file-type/XML>
↪    ;
... > .
```

---

[198]http://publications.europa.eu/resource/dataset/file-type.

[199]https://www.iana.org/assignments/media-types/media-types.xhtml

Output Format

Value type: **URI**   *Domain:* **Software**
Cardinality: **MANY**
RDF property: **osw:hasOutput**

The file format(s) of data the software can generate as output, if applicable.

Expected values are URIs to the terms of the EU File Type Vocabulary or of the IANA Media Types list.

```
<https://w3id.org/dh-atlas/1729187122-1272922> a schema:Dataset ,
        <http://purl.org/spar/fabio/ComputerProgram> ,
        atlas:Software ;
    rdfs:label "Voyant tools"@en ;
    osw:hasOutput
    ↪  <http://publications.europa.eu/resource/authority/
    file-type/HTML>,
<http://publications.europa.eu/resource/authority/file-type/JSON>,
<http://publications.europa.eu/resource/authority/file-type/RTF>,
<https://publications.europa.eu/resource/authority/file-type/PNG>,
<https://publications.europa.eu/resource/authority/file-type/TSV>,
<https://publications.europa.eu/resource/authority/file-type/TXT>
↪  ;
... > .
```

Based on

Value type: **URI**   *Domain:* **Computer Program**, **Web Site**, **Dataset**
Cardinality: **MANY**
RDF property: **schema:isBasedOn**

Software component(s) used or extended in the catalogued software.

While the software reuse property indicates a more general relationship of reuse, this property may be used from a more technical perspective to indicate a derivative relationship.

Expected values include catalogued research products or URLs of the software's landing pages.

```
<https://w3id.org/dh-atlas/1729003408-5163774> a
↪  <http://purl.org/spar/fabio/ComputerProgram> ;
    rdfs:label "DanteSources RDF Dataset - SPARQL Endpoint"@en ;
```

```
    schema:isBasedOn <https://virtuoso.openlinksw.com/> ;
... > .
```

Linked Open Data

`atlas:LinkedOpenData` is a subclass of `schema:Dataset` and of `void:Dataset`,[200] which is defined as follows:

> "A dataset is a set of RDF triples that are published, maintained or aggregated by a single provider. [... T]he term dataset has a social dimension: we think of a dataset as a meaningful collection of triples, that deal with a certain topic, originate from a certain source or process, are hosted on a certain server, or are aggregated by a certain custodian. Also, typically a dataset is accessible on the Web, for example through resolvable HTTP URIs or through a schema, and it contains sufficiently many triples that there is benefit in providing a concise summary. Since most datasets describe a well-defined set of entities, datasets can also be seen as a set of descriptions of certain entities, which often share a common URI prefix".

Vocabularies and Ontologies

Value type: **URI**　　　　　　　　　*Domain:* **Linked Open Data**
Cardinality: **MANY**
RDF property: **atlas:vocabulary**

URIs of the ontologies and vocabularies employed in data modelling.

Expected values can be URIs of catalogued research products or URIs drawn from LOV (Linked Open Vocabularies) service.

```
<https://w3id.org/dh-atlas/1728943937-3007112> a schema:Dataset ,
        <http://purl.org/spar/fabio/DataFile> ,
        atlas:LinkedOpenData ;
    rdfs:label "Zeri Photo Archive RDF Dataset"@en ;
        <https://www.w3.org/RDF/> ;
    atlas:vocabulary
    ↪  <https://w3id.org/dh-atlas/1728948052-580608>,
... > .
<https://w3id.org/dh-atlas/1728948052-580608> rdfs:label
↪  "Semantic Publishing and Referencing Ontologies
↪  (SPAR)"^^xsd:string .
```

---

[200]Vocabulary of Interlinked Datasets (VoID): https://www.w3.org/TR/void/.

Ontology

In the ATLAS model an ontology is a subclass of `schema:Dataset` and `fabio:OntologyDocument`, which represents documents containing ontologies, such as OWL files.

Namespace URI   *Mandatory*

Value type: **URI**   *Domain:* **Ontology**
Cardinality: **ONE**
RDF property:
vann:preferredNamespacePrefix

The namespace used by the terms of the ontology.

```
<https://w3id.org/dh-atlas/1728948736-9982028> a schema:Dataset,
        <http://purl.org/spar/fabio/OntologyDocument> ,
        atlas:Ontology ;
    rdfs:label "CIDOC Conceptual Reference Model (CRM)"@en ;
    vann:preferredNamespaceUri
    ↪   <http://www.cidoc-crm.org/cidoc-crm/> ;
... > .
```

Imported Models

Value type: **URI**   *Domain:* **Ontology**
Cardinality: **MANY**
RDF property: **dcterms:references**

Ontologies or vocabularies imported or partially reused by the catalogued ontology.

Expected values can be URIs of catalogued research products or URIs drawn from LOV (Linked Open Vocabularies) service.

```
<https://w3id.org/dh-atlas/1728948776-872414> a schema:Dataset,
        <http://purl.org/spar/fabio/OntologyDocument> ,
        atlas:Ontology ;
    rdfs:label "Historical Context Ontology (HiCO)"@en ;
    dcterms:references <http://purl.org/spar/cito>
... > .
```

Language Model

In the ATLAS ontology a language model is a subclass of `schema:Dataset` and `schema:SoftwareApplication`, and is defined as "a machine learning model designed to predict and generate language".

Task   *Mandatory*

Value type: **URI**   *Domain:* **Language Model**
Cardinality: **MANY**
RDF property: **schema:**
applicationCategory

URI(s) of the task(s) that can be carried out by employing the language model.

Expected values are URIs of the tasks[201] listed in the Hugging Face platform, such as https://huggingface.co/tasks/token-classification and https://huggingface.co/tasks/feature-extraction.

```
<https://w3id.org/dh-atlas/1757973188-9464774 a schema:Dataset,
        :LanguageModel ;
    rdfs:label "Minerva 7B"@en ;
    schema:applicationCategory
    ↪  <https://huggingface.co/models?pipeline_tag=text-
    generation> ;
... > .
```

Used Libraries

Value type: **URI**   *Domain:* **Language Model,**
                                                        Software
Cardinality: **MANY**
RDF property: **schema:isBasedOn**

URIs of the machine learning libraries utilised for the development, training, evaluation, or deployment of the model.

Expected values include catalogued research products or URLs of the libraries' landing pages. We recommend using stable and permanent identifiers if possible.

```
<https://w3id.org/dh-atlas/1757973188-9464774 a schema:Dataset,
        :LanguageModel ;
    rdfs:label "Minerva 7B"@en ;
```

---

[201]https://huggingface.co/tasks.

```
    schema:isBasedOn
    ↪   <https://huggingface.co/datasets/Babelscape/ALERT>,
... > .
```

3D Digital Twin

In the ATLAS model a 3D Digital Twin is a research product, i.e., a subclass of
`schema:Dataset`, and is defined as a 3D digital replica of a Cultural Heritage
object (e.g., an artwork, an archaeological site, or a museum environment)
allowing cultural operators to explore, preserve, and share Cultural Heritage in
innovative ways [Niccolucci and Felicetti 2024].

Has 3D Model *Mandatory*

Value type: **URI**	*Domain:* **3D Digital Twin**
Cardinality: **MANY**	
RDF property:	
schema:associatedMedia	

The URL to one or multiple files providing a 3D representation of the Cultural
Heritage Object.

```
<https://w3id.org/dh-atlas/1758060615-906372 a schema:Dataset,
        :3DDigitalTwin ;
    rdfs:label "DiSpLay - Digital Spolia Layering"@en ;
    schema:associatedMedia <https://sketchfab.com/spolia_unive> ;
... > .
```

Reproduces creative work *Mandatory*

Value type: **URI**	*Domain:* **3D Digital Twin**
Cardinality: **MANY**	
RDF property:	
schema:exampleOfWork	

The Cultural Heritage Object represented by the 3D Digital Twin.

```
<https://w3id.org/dh-atlas/1758060615-906372 a schema:Dataset,
        :3DDigitalTwin ;
    rdfs:label "BrancacciPOV"@en ;
    schema:exampleOfWork
    ↪   <http://www.wikidata.org/entity/Q1996075> ;
... > .
```

Chapter 3
ATLAS Catalogue

Abstract

The third and final chapter of the whitebook presents the ultimate outcome: the ATLAS knowledge graph and its associated services. This chapter serves as a guide for potential users and curators of the catalogue. It begins with a quick overview of the ATLAS web application, followed by a set of guidelines for data entry, including instructions on creating new records and using the support systems. The chapter concludes with detailed information about the data access services—namely, the GUI, schema, API, and data dump.

The ATLAS Web Application

For the implementation of the ATLAS knowledge graph, we used CLEF (Crowdsourcing Linked Entities via web Form) v. 3.0,[202] a LOD-native crowdsourcing platform for collaborative data collection, peer review, and publication. Developed by an international team of researchers within the ERC-funded project Polifonia,[203] CLEF offers a highly configurable, web-ready solution for producing linked open data through a user interface.

At CLEF's core is the templating system. When setting up the application, administrators define templates for describing their resources. Each field in the template maps to an ontology predicate, ensuring consistent data entry and validation. These templates guide the peer-review process and enable data exploration through actionable filters. While administrators can specify custom ontology terms, CLEF encourages the reuse of established vocabularies by harmonising terms with labels from LOV Linked Open Vocabularies. Users then contribute data by completing the resulting web forms, aided by autocomplete suggestions from Wikidata, Geonames, and the existing catalogue.

CLEF supports both anonymous and authenticated contributions to the data catalogue. The editorial workflow consists of three steps: record creation, peer review, and publication. Each record is represented as an RDF named graph, with editing activities (such as dates and agents involved) automatically documented through RDF statements.

Data entered in CLEF becomes immediately accessible through the automatically generated schema and 'Explore' page, which offers filtered views of the collected data. The platform integrates with GitHub for user authentication, version control, and data backup. CLEF can run either locally or on a server, and its Python-based source code is available on both GitHub and Zenodo.

---

[202]https://zenodo.org/badge/latestdoi/479251315. CLEF's GitHub repository: https://github.com/polifonia-project/clef.
[203]https://polifonia-project.eu/.

The ATLAS web application[204] comprises four main pages:

1. 'Explore', which enables users to browse the catalogue and access the knowledge graph data via an intuitiveGUI;

2. The 'Charts' page provides users with a numerical overview of the catalogued data;

3. 'Member Area', where authenticated project members can edit and review entered data;

4. 'Add a resource', where all authenticated users can enter new data into the catalogue.

Figure 14: the home page of the ATLAS web application.

In the footer, users can find links to the schema and important documentation, such as the data model (the ATLAS ontology) and editorial guidelines.

Figure 15: the footer of the ATLAS web application, containing the links to access the data model and the schema.

---

[204]https://projects.dharc.unibo.it/atlas/.

Data entry

Editorial workflow

The ATLAS web application allows both project members (PM) and other users to enter new data in the catalogue.

Although all users can contribute to the ATLAS catalogue, the application can only keep track of records created by authenticated users thanks to synchronisation between the ATLAS platform and its corresponding GitHub repository. To log in to the application, users must click on the 'Member area' button in the navigation bar and enter their credentials on the GitHub login page.

When first accessing the member area, users are prompted to a list of all records. Each record in the ATLAS web application corresponds to an RDF named graph. The list can be sorted according to the record status: edited, reviewed and published.

Figure 16: the list of the ATLAS catalogue's records in the CLEF application's member area.
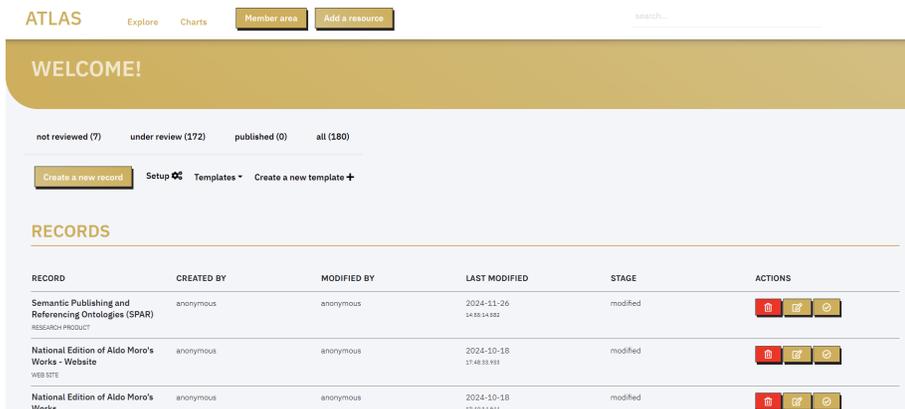

All users can create new records to describe research products, by filling out a user-friendly web form, wherein fields correspond to RDF properties and the record is an entity of a class.

To add a new record, users can click on the 'Add a resource' button in the navigation bar and are prompted to a web page, where they can select the type of record they want to create. Available options are: Research Product, Research Project, Person, Organization, Computer Program, and Website.

After selecting the type of record, users are prompted to the corresponding template and set of input fields. For more information, see the templates section.

After editing, users can save the record, which is then automatically ready for revision. While only project members (PMs) can review and publish all

Figure 17: the template selector.
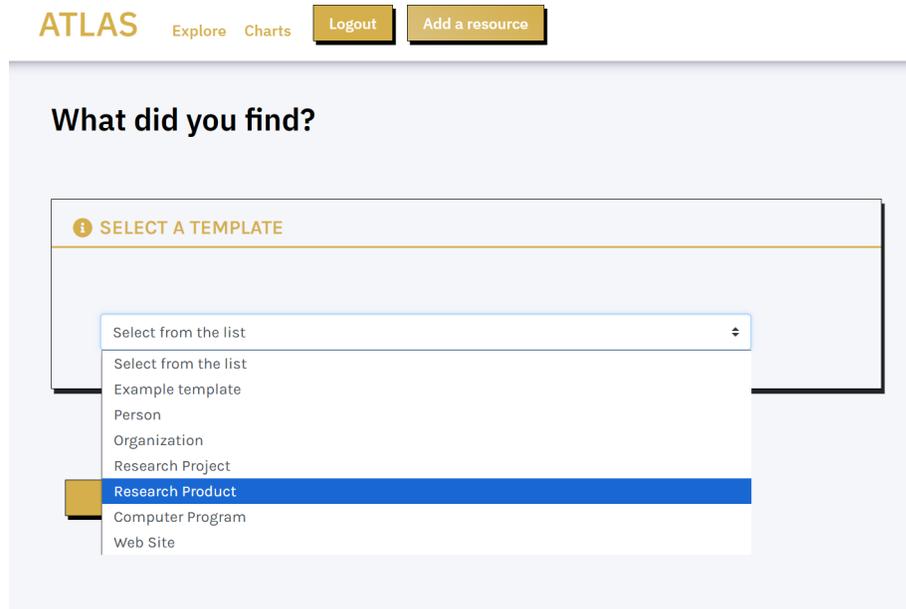
records, other authenticated users can review and edit their own before a PM revises them. After the revision records can be published. It is possible to browse a published record from the 'Explore' page, search it via text search, and retrieve it as Linked Open Data from the schema via the REST API at https://projects.dharc.unibo.it/atlas/sparql. It is not possible to unpublish a published record. This restriction prevents inconsistencies between records relying on other records' identifiers.

Templates

Each record in CLEF complies with a template, i.e., a set of mandatory and optional fields/properties to be filled with appropriate values. ATLAS's main template in CLEF describes research products, while the other templates are used to represent creators, contributors, publishers, research projects, and access points of research products. Each template is associated with a class of the

Figure 18: an example of how a record is displayed in the ATLAS web application.

ATLAS ontology, namely: Person,[205] Organization,[206] Research Project,[207] Website[208] and Computer Program.[209]

CLEF simplifies data entry by allowing users to create multiple entities of the ancillary classes alongside the main research product entity. Through a system of subtemplates (or intermediate templates), the fields for each class appear within the Research Product template. In this way, users can easily create new records for describing creators, access points, research projects, etc. of a research product. This section presents each template and its fields.

Research Product   The template first displays fields that are common across all research product types. When a specific type is selected in the "Type" field—Digital Scholarly Edition, Text Collection, Software, Linked Open Data, Ontology, Language Model, or 3D Digital Twin—additional fields specific to that type appear. The type "Other" is not associated with further specific fields, but allows users to categorise the research product by selecting a custom type, as shown in Chapter 2.

The research product's web form comprises a set of fields, each corresponding to one of the properties of the Research Product class, as described in the data

---

[205]https://dh-atlas.github.io/deliverables/ontology/2.0/index-en#https://schema.org/Person.
[206]https://dh-atlas.github.io/deliverables/ontology/2.0/index-en#https://schema.org/Organization.
[207]https://dh-atlas.github.io/deliverables/ontology/2.0/index-en#https://schema.org/ResearchProject.
[208]https://dh-atlas.github.io/deliverables/ontology/2.0/index-en#https://schema.org/WebSite.
[209]This template is called Computer Program, because it was originally linked to the fabio:ComputerProgram class. We later switched to Schema.org's Software Application class (https://dh-atlas.github.io/deliverables/ontology/2.0/index-en#https://schema.org/SoftwareApplication) to improve interoperability. However, due to time constraints, we did not update the name of the corresponding template within the ATLAS application.

model (Chapter 2). The interactive list on the left of the form helps users navigate between the properties.

Figure 19: the template for the research products.

Person and Organization   To represent individual scholars or organisations such as universities, research centres and cultural institutions involved in a research product as creators, collaborators or publishers, the 'Person' and the 'Organization templates are used, respectively.

For each person, data curators must specify the full name and may optionally indicate the person's affiliation, ORCID identifier, and link to an authority record (e.g., Wikidata, VIAF) that unambiguously identifies the person.

The 'Organization' template requires curators to specify their name and location, optionally including the organisation's website URL, persistent identifier, and a link to an authority record (e.g., Wikidata, VIAF).

Research Project   Most research products are developed within research projects. To link a product to other project outcomes and better understand its context and purpose, users can describe the associated project using the 'Research Project' template.

Data curators must provide the research project's full title (including any abbreviations or aliases), a brief description, the project's country location, funding agency, and website link. They may also include additional details such as the project's start and end dates, member names, identifiers (e.g., Grant number), and a link to an authority record (e.g., Wikidata, VIAF) that uniquely identifies the research project.

Figure 20: selection of the Person or Organization intermediate template within the Creator field.
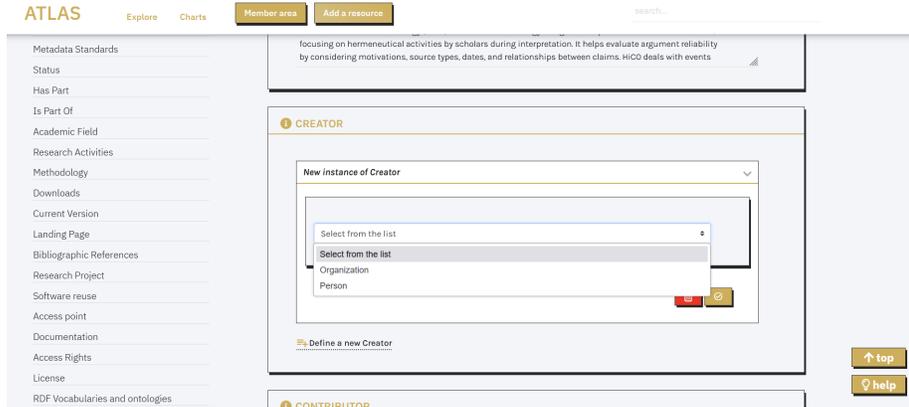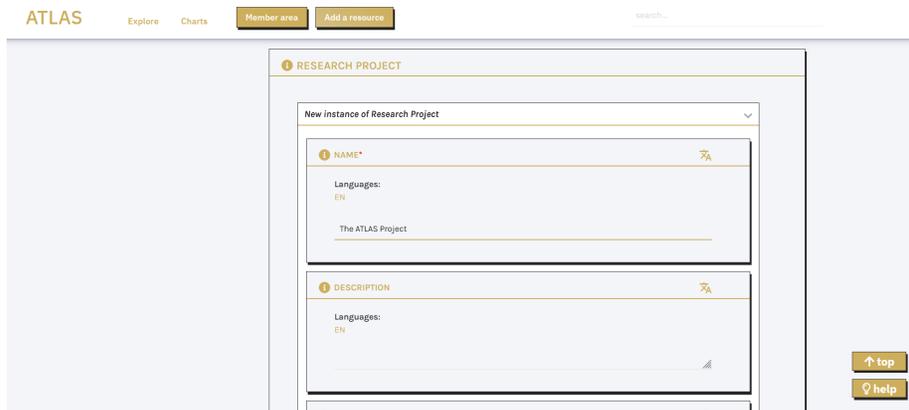
Figure 21: the template for research projects.

Computer Program and Website   The intermediate templates 'Computer Program' and 'Website' are used to describe software tools, data services, and websites that provide access to research products. Both templates require curators to specify the title and URL of the website or the access URL for the research product. In addition to these mandatory fields, curators may include:

- a brief description of the software tool's or website's functionality;

- URL to the research product's documentation;

- research activities enabled by the service or software in relation to the dataset;

- the licence terms;

- software components that are reused or extended in the current tool or website, which can include research products or URLs to official software pages.

For computer programs, curators can also classify the type according to the EU Data service type Vocabulary,[210] which includes: API, Download service, Human interaction service, and schema.

Figure 22: selection of the Web Site or Computer Program intermediate template within the Access point field.


---

[210]http://publications.europa.eu/resource/authority/data-service-type.
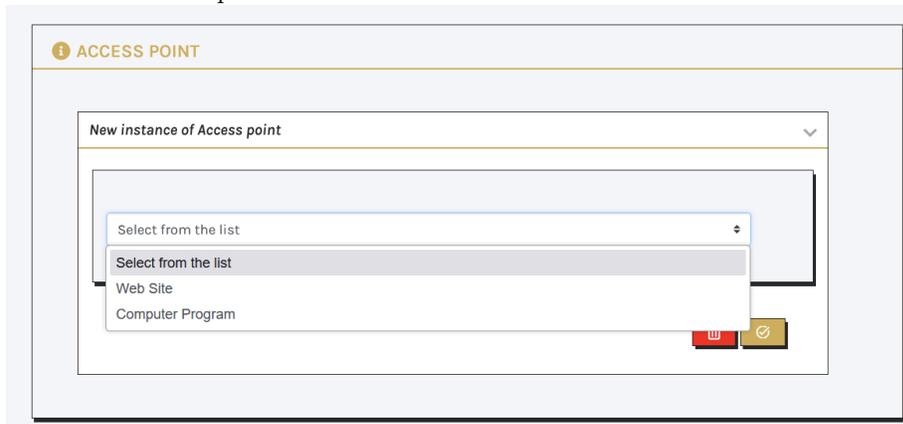
Features

While editing a research product record, contributors are supported in tasks relevant to the data reusability, namely: data reconciliation, duplicate avoidance, keyword extraction, and knowledge extraction.

Data reconciliation When field values refer to real-world entities (e.g., Collected works, Geographical Coverage, Genre) or concepts that may appear in multiple records (e.g., Creator), the system provides autocomplete suggestions by querying both external sources (in real-time) and the ATLAS SPARQL endpoint. These suggestions appear as lists of terms, with each term showing a label, a short description (to help distinguish between similar items), and a link to the online record (such as a Wikidata page or an existing project record). If no matches are found, collaborators can create new entities that are added to the knowledge base and will appear in future suggestion lists.

Figure 23: an example of the autocomplete feature associated with the Same as field in the Person template.
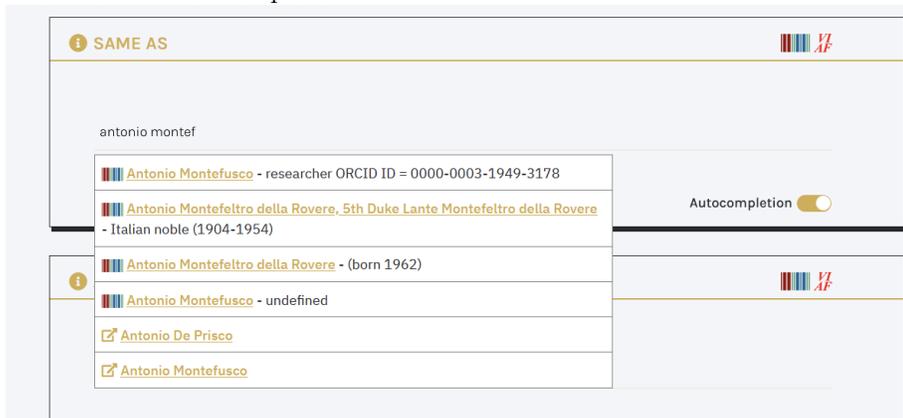
The system retrieves suggestions primarily from Wikidata and VIAF for personal names, Geonames for locations, and specialised SKOS vocabularies like TaDiRAH. Users have the option to disable autocomplete functionality for any field where it is available.

Duplicate avoidance   When creating a new record, the application alerts users of potential duplicates already existing in the catalogue in order to prevent involuntary inconsistencies. Contributors may accept or ignore the recommendation. A possible scenario for this feature is when different contributors describe the same creator of multiple research products.

Keyword extraction   Long textarea input fields (e.g. Description or Note on Research Product) are equipped with a keyword extraction function that uses SpaCy[211] APIs to identify entity names (e.g., people, places or subjects) in the text. The extracted entities are reconciled with the corresponding entities in the Wikidata database and the keywords are bound to the relevant QIDs. The extracted entities and keywords are then presented to users for approval or rejection. Approved terms are included in the data as machine-readable keywords associated with the record.

Figure 24: an example of keywords extracted from a text input.

Knowledge extraction   For all research products, it is possible to use the knowledge extraction function provided in the Mentions field, which retrieves named entities and Linked Open Data from schemas, APIs, static files (in CSV, JSON, and XML format), and web pages. It uses SPARQL Anything[212] to query these static files and convert them to RDF format.

To make this feature more user-friendly, users can rely on manual extraction, simply providing a document URL, and the system automatically identifies its structure. Through a dropdown menu, users can select elements to extract and set basic filters like minimum occurrences. The system then handles the technical SPARQL query conversion. The extraction process includes automatic entity reconciliation: extracted entities are reconciled to URIs in sources like Wikidata and VIAF and are shown to users to approve/discard. Approved terms are included in the data as machine-readable keywords associated with the subject entity.

---

[211]https://spacy.io/api.
[212]https://sparql-anything.cc/.

Figure 25: the GUI for the knowledge extraction feature.

Data access

The ATLAS knowledge graph is accessible through a user-friendly "Explore" interface, a schema, and an API. The catalogue's data can also be visualised in the "Charts" page. Last but not least, the knowledge graph is also available as a data dump to support data reuse and integration.

'Explore' page

The 'Explore' page[213] is accessible from the web application's homepage through the corresponding navigation bar button.

Users can filter entities in the knowledge graph by type—including research product, person, organisation, research project, computer program, and website—by selecting the appropriate tab at the top of the 'Explore' page. Research products can be further refined by sub-classes.

Once a type of entity is selected, the 'Explore' page displays an alphabetical index and relevant filters. Research products, for instance, can be filtered by creator, type, language, and encoding format.

Figure 26: the 'Explore' page.
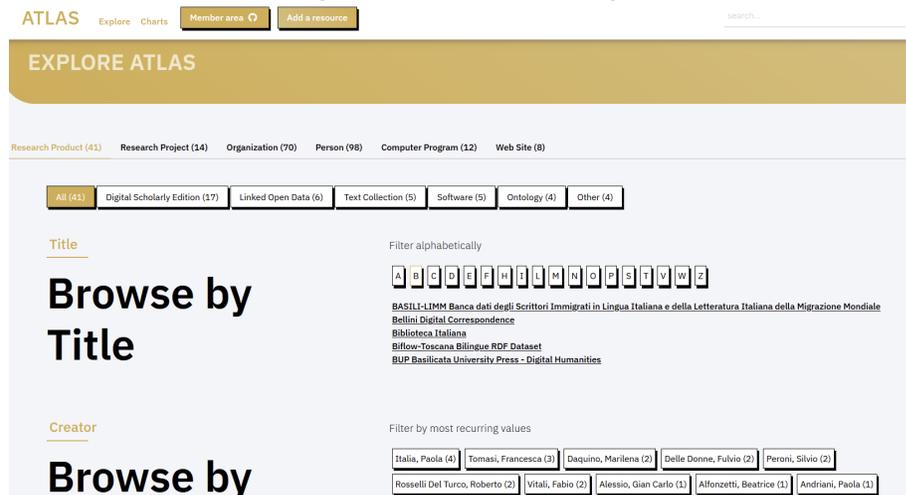
The navigation bar on the right includes a search field where users can query the catalogue's entities by name or title. As users type their search terms, the field displays a list of relevant entity suggestions.

---

[213]https://projects.dharc.unibo.it/atlas/records.

Figure 27: the search field located in the navigation bar.

'Charts' page

In the 'Charts' page,[214] we provide an overview of the catalogue's size—including the number of entities, research products, and research projects available— through a set of counters. We also present maps showing the geographical distribution of scholars and organisations involved in creating research products, as well as the distribution of research projects. An interactive timeline shows the historical periods to which the catalogued text collections pertain.

The data displayed through maps, counters, timelines, and interactive charts are retrieved via the schema. Each graphic component includes a brief comment explaining the displayed data.

Figure 28: a timeline showing the Research Products related to the selected historical period.

---

[214]https://projects.dharc.unibo.it/atlas/charts-visualization.

Figure 29: a map showing the locations of the organizations recorded within the ATLAS catalogue.

Figure 30: a chart showing the percentage of different research product types included in the ATLAS catalogue.

SPARQL endpoint

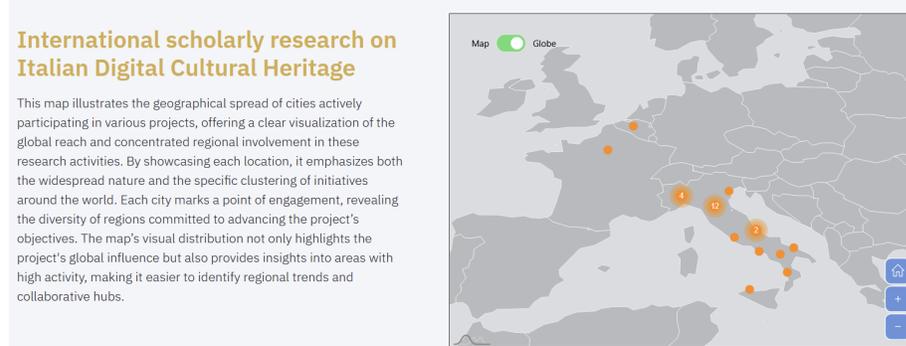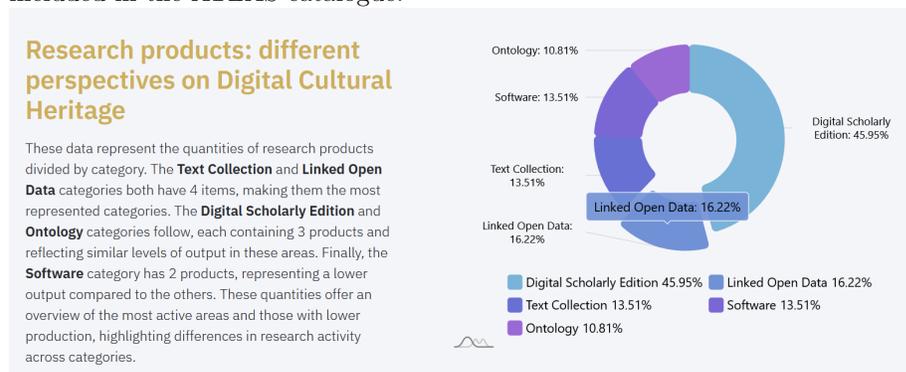ATLAS's schema is accessible from the web application homepage through the corresponding button in the footer, as shown above (see section Editorial workflow).

The schema relies on Blazegraph triplestore.[215] The application provides users with a graphic interface, based on Yasgui,[216] for the editing of SPARQL queries and the visualisation of the search results.

Query examples are listed in the table below. For detailed indications on how to use ATLAS's SPARQL editor, please refer to Yasgui's documentation.[217]

Retrieve all research products in Latin with status 'complete'.

```
PREFIX schema: <https://schema.org/>
SELECT DISTINCT ?research_product WHERE {
  ?research_product schema:inLanguage
  ↪  <http://publications.europa.eu/resource/authority/language/
  LAT>;
    schema:creativeWorkStatus
    ↪  <http://publications.europa.eu/resource/authority/dataset
    -status/COMPLETED> .
}
```

Retrieve all types of digital scholarly editions.

```
PREFIX schema: <https://schema.org/>
PREFIX atlas: <https://w3id.org/dh-atlas/>
SELECT DISTINCT ?research_product ?edition_type WHERE {
  ?research_product schema:additionalType
  ↪  atlas:DigitalScholarlyEdition;
    atlas:editionType ?edition_type .
}
```

Retrieve all collections of poetic texts.

```
PREFIX schema: <https://schema.org/>
PREFIX atlas: <https://w3id.org/dh-atlas/>
SELECT DISTINCT ?research_product WHERE {
  ?research_product schema:additionalType atlas:TextCollection;
    schema:genre
    ↪  <https://sparql.cwrc.ca/ontologies/genre.html#PoeticGenre>
    ↪  .
}
```

Retrieve all software tools that provide XML/TEI outputs.

---

[215]https://blazegraph.com/.
[216]https://yasgui.triply.cc/.
[217]https://github.com/TriplyDB/Yasgui.

```
PREFIX schema: <https://schema.org/>
PREFIX atlas: <https://w3id.org/dh-atlas/>
PREFIX osw: <http://ontosoft.org/software#>
SELECT DISTINCT ?research_product WHERE {
  ?research_product schema:additionalType atlas:Software;
    osw:hasOutput
    ↪  <https://www.iana.org/assignments/media-types/application
    /tei+xml> .
}
```

Retrieve all research products that mention Dante Alighieri.

```
PREFIX schema: <https://schema.org/>
PREFIX atlas: <https://w3id.org/dh-atlas/>
PREFIX osw: <http://ontosoft.org/software#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
SELECT DISTINCT ?research_product WHERE {
  ?research_product a schema:Dataset ;
    schema:mentions ?extraction_graph .
  GRAPH ?extraction_graph {
      <http://www.wikidata.org/entity/Q1067> rdfs:label "Dante
      ↪  Alighieri"
  }
}
```

API

The ATLAS project team has developed RESTful APIs to access the knowledge graph, ensuring machine actionability and encouraging data reuse. These APIs were created using RAMOSE (Restful API Manager Over SPARQL Endpoints),[218] a Python tool that enables the development and publication of documented REST APIs over SPARQL endpoints [Daquino et al. 2022]. Thanks to the APIs, external platforms and services can programmatically access the ATLAS knowledge graph, enabling automated querying and retrieval of structured data in JSON format. These APIs serve as a key interoperability layer, allowing third-party applications to access and reuse the graph's contents in real time. Here follows a full list of available API operations.

`/records` retrieves descriptive metadata for each dataset included in the DH ATLAS knowledge graph.

```
GET https://atlas-prin2022-api.cloud.d4science.org/api/v1/records
```

`/classes/\{pid\}` retrieves the RDF classes to which a given resource belongs within the DH ATLAS knowledge graph. The input parameter pid must be a valid URI identifying a resource in the graph.

---

[218]https://github.com/opencitations/ramose.

Parameter: entity URI

```
GET https://atlas-prin2022-api.cloud.d4science.org/api/v1/
classes/<entityURI>
```

`/labels/\{pid\}` retrieves the label (i.e., the title or name) of a resource in the DH ATLAS knowledge graph. The input parameter pid must be a valid URI identifying a resource in the graph.

Parameter: entity URI

```
GET https://atlas-prin2022-api.cloud.d4science.org/api/v1/labels/
<entityURI>
```

`/stats/entities` retrieves the total number of distinct entities recorded in the DH ATLAS knowledge graph.

```
GET https://atlas-prin2022-api.cloud.d4science.org/api/v1/stats/
entities
```

`/stats/projects` retrieves the total number of research projects recorded in the DH ATLAS knowledge graph.

```
GET https://atlas-prin2022-api.cloud.d4science.org/api/v1/stats/
projects
```

`/stats/products` retrieves the total number of research products recorded in the DH ATLAS knowledge graph.

```
GET https://atlas-prin2022-api.cloud.d4science.org/api/v1/stats/
products
```

`/stats/organizations` retrieves the total number of organisations recorded in the DH ATLAS knowledge graph.

```
GET https://atlas-prin2022-api.cloud.d4science.org/api/v1/stats/
organizations
```

`/stats/persons` retrieves the total number of researchers recorded in the DH ATLAS knowledge graph.

```
GET https://atlas-prin2022-api.cloud.d4science.org/api/v1/stats/
persons
```

`/stats/organizations/by-location` retrieves the number of organizations recorded in the DH ATLAS knowledge graph, grouped by their associated geographical location.

```
GET https://atlas-prin2022-pi.cloud.d4science.org/api/v1/stats/
organizations/by-location
```

`/stats/products/by-additional-type` retrieves the number of research products, grouped by their respective class ordered in descending order by count.

```
GET https://atlas-prin2022-api.cloud.d4science.org/api/v1/stats/
products/by-additional-type
```

`/stats/products/by-organization` retrieves the number of research products recorded in the DH ATLAS knowledge graph, grouped by the organization responsible for their production, ordered in descending order by count.

```
GET https://atlas-prin2022-api.cloud.d4science.org/api/v1/stats/
products/by-organization
```

OpenAIRE Italian DH-CH Gateway

The ATLAS project was conceived from the outset to be interoperable with major European infrastructures for Open Science, particularly OpenAIRE, in order to ensure maximum visibility, reusability, and integration of digital humanities research outputs related to Italian cultural heritage.

To enable interoperability between the ATLAS Scholarly Knowledge Graph and the OpenAIRE knowledge graph, the project adopted the OAI-PMH (Open Archives Initiative Protocol for Metadata Harvesting) protocol, version 2.0.[219] This protocol allows for the automated harvesting of metadata from the ATLAS graph, enabling OpenAIRE to integrate its contents into its own semantic graph.

A particularly relevant aspect of this interoperability concerns the representation of *product type*, which reflects the different modeling assumptions underlying the ATLAS Ontology and the OpenAIRE data model. In ATLAS, research outputs related to italian digital humanities are modeled as datasets and described through a rich, type-specific ontology that captures fine-grained distinctions among scholarly resources (such as digital scholarly editions, ontologies, text collections, linked open data, software, 3D digital twins, and language models). This approach enables a high level of semantic granularity, supporting detailed documentation of cultural heritage-specific features that are often absent from generic research metadata schemas. By contrast, OpenAIRE adopts a more abstract and transversal notion of research product, based on DataCite[220] and Dublin Core and governed by Coalition of Open Access Repositories (COAR) controlled vocabularies,[221] prioritizing interoperability, aggregation, and cross-domain reuse over domain-specific expressiveness. To enable exposure of ATLAS resources through the OpenAIRE Graph via OAI-PMH, a selective alignment process was required: fine-grained, type-dependent properties defined in ATLAS were simplified, while core, transversal metadata elements were retained and structural properties –such as creator subproperties, affiliations, and identifier specifications– were expanded to comply with OpenAIRE's mandatory requirements defined in the Application Profile. In this process, ATLAS product types were mapped to corresponding COAR resource types, ensuring semantic consis-

---

[219]https://www.openarchives.org/OAI/openarchivesprotocol.html.

[220]https://schema.datacite.org/.

[221]https://vocabularies.coar-repositories.org/.

tency while preserving the ability of OpenAIRE to distinguish between different classes of research products:

- `atlas:DigitalScholarlyEdition` → `coar:encodedData`;

- `atlas:Ontology` → `coar:KnowledgeOrganizationSystem`;

- `atlas:TextCollection` → `coar:collection`;

- `atlas:LinkedOpenData` → `coar:dataset`;

- `atlas:Software` → `coar:software`;

- `atlas:3DDigitalTwin` → `coar:dataset`;

- `atlas:LanguageModel` → `coar:software`.

This interoperability between the two models is implemented through a search portal built on the OpenAIRE CONNECT[222] Gateway, a customised access point to the OpenAIRE[223] infrastructure designed to aggregate, visualise, and interlink research outputs from a specific scientific community. The result is the creation of an Italian subcommunity within the thematic DH-CH (Digital Humanities and Cultural Heritage) Gateway, which aggregates resources relevant to international DH and CH research focused on Italian cultural heritage.

This collaborative virtual environment promotes semantic interoperability by using the OpenAIRE graph to connect publications, datasets, software, projects, and authors via common metadata standards and persistent identifiers (e.g., DOI, ORCID, ROR). It supports European Open Access policy[224] compliance, helping meet funding requirements like Horizon Europe through metadata monitoring and validation. Research outputs gain visibility through OpenAIRE indexing, increasing discoverability and impact. The gateway also provides reporting tools, dashboards, and metrics for analyzing dissemination and usage of scientific results.

Within the Italian DH-CH Gateway,[225] users can explore the resources connected to the ATLAS graph. The portal allows users to filter resources according to various criteria: Access type (Open Access, embargoed, closed, restricted); Product type (Publications, Research Data, Software, Other research products); Year of publication; Field of Science; Funder; Sustainable Development Goals (SDG); Country; Language; Source; Research community.

Once a resource is selected, users can view its bibliometric information, a summary of its contents, and analytical metrics calculated by OpenAIRE based on its semantic graph. These metrics offer a dynamic and multidimensional view

---

[222]https://connect.openaire.eu/.

[223]https://www.openaire.eu/

[224]https://www.openaire.eu/ec-policies-and-mandates.

[225]https://dh-ch.openaire.eu/search/advanced/research-outcomes?f0=conceptname&fv0= Italian%20DH-CH.

Figure 31: interface of the OpenAIRE Italian Digital Humanities and Cultural Heritage Gateway.
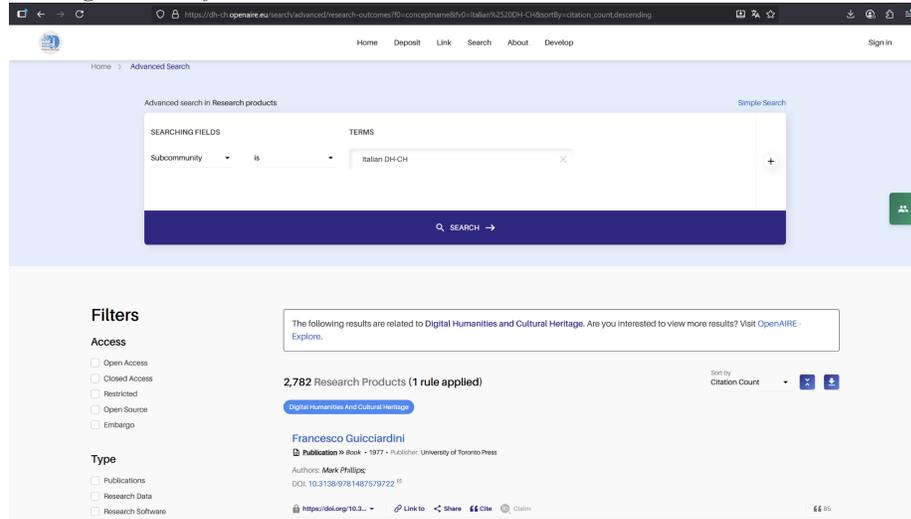
of a research output's impact within the scientific community, supporting both qualitative and quantitative evaluations.

Data dump

The latest version of the DH ATLAS knowledge graph is available as a set of Turtle files deposited in Zenodo (https://doi.org/10.5281/zenodo.14058143) and in the ILC4CLARIN repository (http://hdl.handle.net/20.500.11752/ILC-1040). It contains metadata from the pilot projects discussed in Chapter 1 and their related entities. The data structure adheres to the ATLAS ontology model (Chapter 2). The package also includes demo files that showcase the structure of the main entities within the model.

To access previous versions of the knowledge graph and for a full list of all ATLAS project outputs, please refer to ATLAS Outputs.

Figure 32: focus on the information about a resource contained inside the OpenAIRE Italian Digital Humanities and Cultural Heritage Gateway.
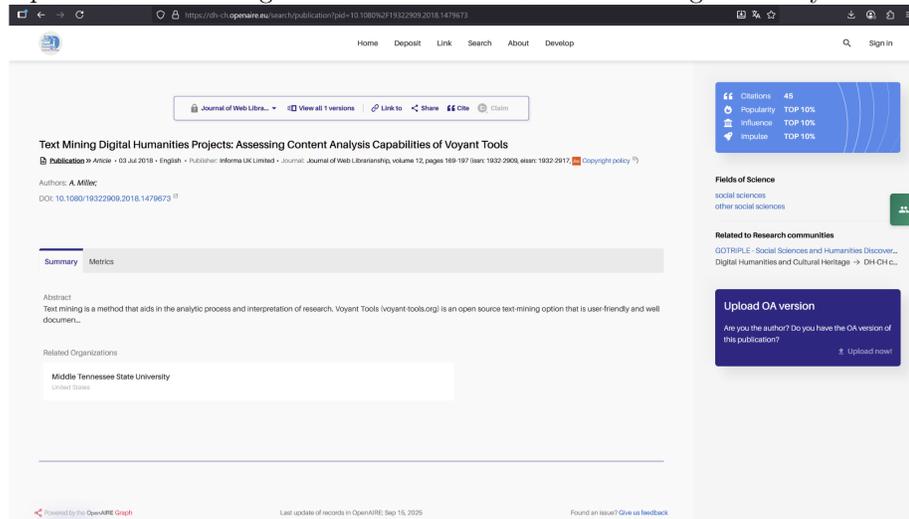
Conclusions

The ATLAS project addresses a critical gap in the Digital Humanities landscape by creating a comprehensive knowledge graph of digital research products related to Italian Cultural Heritage. Through careful analysis of selected pilot projects spanning text collections, digital scholarly editions, software tools, linked open data, and ontologies, the project has established a framework for cataloguing, preserving, and enhancing the discoverability of valuable scholarly outputs.

The development of the ATLAS ontology and knowledge graph represents a significant advancement in the application of Semantic Web technologies to the humanities domain. By mapping and interlinking diverse digital artifacts, the project facilitates greater interoperability between previously isolated resources and enables researchers to discover connections that might otherwise remain hidden.

Key Achievements of the ATLAS project are:

- the evaluation of multiple pilot projects according to existing standards and the FAIR principles, identifying metadata for knowledge representation of different types of digital research outputs;

- the development of the ATLAS ontology, mapping existing schemas and models to enhance metadata interoperability across the Digital Humanities domain;

- the creation of the ATLAS knowledge graph, which documents DH projects and scholarly data related to Italian Cultural Heritage, accessible through the ATLAS web application and preserved in certified repositories (Zenodo and ILC4CLARIN);

- the publication of a set of guidelines which provide detailed recommendations for producing FAIR research products in the Digital Humanities, based on practical analysis rather than purely theoretical approaches.

The ATLAS project demonstrates that enhancing the FAIRness of Digital Humanities research products requires both technical infrastructure, community-driven standards, and collaborative approaches.

While the ATLAS project has made significant strides in cataloguing and enhancing Digital Humanities research outputs, several challenges remain. The diversity of digital artifacts in the humanities makes comprehensive standardisation difficult. Additionally, ensuring long-term sustainability of the knowledge graph requires ongoing community engagement and institutional support.

The ATLAS project represents an important step toward ensuring that the valuable scholarly outputs produced by Digital Humanities projects focused on Italian Cultural Heritage remain discoverable, accessible, and reusable for future generations of researchers. By combining rigorous analysis of existing practices with Semantic Web technologies, we have created a framework that not only preserves these resources but enhances their value through meaningful interconnections.

As the digital landscape continues to evolve, the principles and approaches established through the ATLAS project will hopefully serve as a foundation for ongoing efforts to document, preserve, and enhance the rich landscape of digital scholarship on Italian Cultural Heritage.

Part 2
Guidelines

Abstract

These guidelines provide practical recommendations for researchers in Digital Humanities (DH) seeking to develop or improve the FAIRness (Findability, Accessibility, Interoperability, Reusability) of their research products. Building upon the comprehensive ALLEA E-Humanities Working Group recommendations, this document offers specific guidance for five common DH output types: digital scholarly editions, text collections, software tools, ontologies, and linked open datasets. Each section focuses on an output type and begins with a survey of existing standards and guidelines, followed by recommendations organised around the fundamental phases of digital object creation: identifying and planning, modelling and formatting, publishing and depositing.

Rather than replacing existing disciplinary standards, these guidelines are designed to complement them by focusing specifically on FAIR principles. The document includes additional recommendations for enhancing research product quality beyond basic FAIR compliance, making outputs easier to cite, use, and evaluate. By implementing these guidelines, scholars can create research products that are not only technically compliant with FAIR principles but also genuinely valuable to the broader research community, ensuring their work remains accessible, usable, and relevant in the evolving digital landscape of humanities research.

Introduction

The ATLAS project

These guidelines were developed within the PRIN 2022 ATLAS project,[1] a joint effort of the University of Bologna, the Ca' Foscari University of Venice, and the CNR Unit of Pisa (comprising the Institute of Information Science and Technologies "Alessandro Faedo"—ISTI—and the Institute for Computational Linguistics "A. Zampolli"—ILC). ATLAS has two main objectives:

1) creating a catalogue of Digital Humanities Research on Italian Cultural Heritage, using semantic web technologies to establish a framework that captures diverse DH research outputs, and

2) developing clear and practical guidelines for creating high-quality scholarly data.

While best practices and guidelines exist alongside FAIR principles, these resources require expansion to cover the full range of DH resources and need systematic organisation. These guidelines are based on the analysis of existing standards, guidelines, and best practices, as well as on the examination of se-

---

[1] https://dh-atlas.github.io/.

lected DH research projects on Italian Heritage and their outcomes. The results of this analysis are thoroughly presented in Part 1. Other outputs include:

- the ATLAS ontology (Part 1, Chapter 2), which maps schemas and ontologies from pilot projects to enhance data and tool interoperability;

- a knowledge graph (Part 1, Chapter 3), documenting DH projects and scholarly data on Italian Cultural Heritage, accessible through the ATLAS web application[2] and stored in a trusted repository;

- a search portal built on the OpenAIRE CONNECT Gateway that focuses on scholarly literature and data related to the pilots and beyond.

For a full presentation of the ATLAS project, its outputs, and methodologies, refer to Part 1 and the publications of the ATLAS team.

Existing recommendations and references for implementing FAIR data

The concept of FAIR data and its principles were developed within FORCE11[3]—an international community of scholars, librarians, archivists, publishers, and research funders—and were first introduced in Wilkinson et al. 2016 **'The FAIR Guiding Principles for Scientific Data Management and Stewardship'.** *Scientific Data* 3, no. 1 (15 March 2016): 160018. https://doi.org/10.1038/sdata.2016.18.2016.

The key points of the article are:
- **Purpose of FAIR Principles.** The FAIR principles aim to improve the infrastructure supporting the reuse of scholarly data. They provide guidelines to ensure that data and associated metadata are well-managed and can be easily found, accessed, integrated, and reused by humans and machines.
- **Applicability Beyond Data.** While initially focused on data, the principles are also applicable to algorithms, tools, and workflows, recognizing that all digital research outputs should adhere to these standards to facilitate transparency and reproducibility.
- **Stakeholder Benefits.** Implementing FAIR principles benefits various stakeholders, including researchers, data publishers, software developers, funding agencies, and data scientists, by promoting efficient data sharing and reuse.
- **Emphasis on Machine-Actionability.** A significant aspect of the FAIR principles is the emphasis on machine-actionability, ensuring that computational systems can automatically find and use the data, which is crucial in the era of big data and complex analyses.

Nearly a decade later, the concept of FAIR data has spread widely and gained worldwide recognition in the scientific community. Data FAIRness has become

---

[2]https://projects.dharc.unibo.it/atlas/.
[3]https://force11.org/.

a key requirement for publicly funded research projects. The **Horizon2020** Programme, for instance, has developed specific **guidelines for FAIR data management**.[4]

Despite the success of the FAIR principles, creating findable, accessible, interoperable and reusable digital research products can be challenging. To address this, many institutions have developed resources and materials to guide both individual scholars and organisations in creating FAIR data and implementing FAIR data services. Here are some notable examples:

- **How to FAIR**[5] is ideal for individual scholars who want to learn FAIR principles from scratch. This user-friendly portal offers a 60-minute video course on **Research Data Management and FAIR**,[6] practical guides with examples and videos—covering topics like file formats, metadata, persistent identifiers, and data licences—and a quiz to test your knowledge of FAIR principles.

- **GO-FAIR**[7] provides a quick guide on implementing FAIRness with a focus on machine-actionability. The guide is organised by operative points and includes practical examples.

- **FAIRsFAIR**[8] (Fostering FAIR Data Practices in Europe) is a comprehensive portal for universities and research organisations, offering resources to help data stewards implement FAIR data repositories and support educators in training personnel and students on FAIR principles.

Many recommendations and guides on implementing data FAIRness are aimed at scholars across different fields and disciplines, including the humanities. An example is the set of **recommendations**[9] **developed by the Research Data Alliance**,[10] which covers topics such as data citation, certification of repositories, metadata management, and interoperability. **OpenAIRE's**[11] **quick guide on**

---

[4]H2020 Programme. Guidelines on FAIR Data Management in Horizon 2020. V. 3.0. 26 July 2016. https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/o a_pilot/h2020-hi-oa-data-mgt_en.pdf.

[5]Deutz et al. 2020. How to FAIR: a Danish website to guide researchers on making research data more FAIR. https://doi.org/10.5281/zenodo.3712065. https://howtofair.dk/.

[6]https://howtofair.dk/what-is-fair/#elearning-course-research-data-management-.

[7]https://www.go-fair.org/fair-principles/.

[8]https://www.fairsfair.eu/.

[9]https://www.rd-alliance.org/recommendations-and-outputs/.

[10]The Research Data Alliance is a community-driven initiative launched in 2013 with a mission to enable researchers to openly share and reuse data across technologies, disciplines, and countries. The RDA builds social and technical infrastructure to reduce data sharing barriers through Working Groups, Interest Groups and Communities of Practice that bring together scientists, researchers and technical experts. https://www.rd-alliance.org/.

[11]OpenAIRE AMKE is a non-profit organisation promoting open scholarship and improving research data management globally. It operates a European e-infrastructure with public services for Open Science adoption, supported by National Open Access Desks across Europe. OpenAIRE serves researchers, policymakers, organisations, universities, libraries, and citizen scientists as a key implementer of the European Open Science Cloud (EOSC). https://www. openaire.eu/.

making FAIR data,[12] included in their Research Data Management guides, is another example.

Some recommendations are specifically developed for particular research areas, adapting FAIR principles to the unique requirements of specific domains.

PARTHENOS'[13] Guidelines to FAIRify data management and make data reusable[14] serve data producers, archivists, and users in humanities and social sciences who want to make research data as reusable as possible. The PARTHENOS guidelines are written in clear, accessible language for audiences with varying technical expertise. Each guideline includes specific recommendations for both individual scholars and research institutions and infrastructures.

The Recommendations of the ALLEA Working Group for E-Humanities "Sustainable and FAIR Data Sharing in the Humanities"[15] were published in 2020, providing the most thorough guide for implementing FAIR principles in the humanities. These recommendations walk users through each phase of a digital object's lifecycle—from planning through dissemination and preservation. Each recommendation set begins with a detailed introduction explaining the rationale and benefits, followed by a curated list of web resources and references for deeper learning. **We highly recommend these guidelines**, as they combine clear language with practical implementation tools while offering **a comprehensive roadmap for creating truly FAIR data**.

Research Infrastructures

Research Infrastructures (RIs) play a key role in both disseminating and implementing FAIR principles. Whether established before or after the concept of data FAIRness spread across scholarly communities, these infrastructures are guided by the same goals of enhancing the production and sharing of high-quality research data. As a result, RIs have developed various resources—from recommendations and training materials to services and tools—that help both individual scholars and institutions create and use FAIR data.

In Europe, the main Research Infrastructures working in Humanities, Cultural Heritage, and Social Sciences are ARIADNE, CLARIN, DARIAH, E-RIHS,

---

[12] https://www.openaire.eu/how-to-make-your-data-fair.

[13] PARTHENOS (Pooling Activities, Resources and Tools for Heritage E-research Networking, Optimization and Synergies) is a collaborative project that strengthens research connections in Linguistics, Humanities, Cultural Heritage, History, and Archaeology across European Research Infrastructures. Through partnerships with DARIAH and CLARIN, it creates unified standards and services for data management while promoting cross-disciplinary integration. The project, coordinated by PIN (Italy), involves 16 partners. https://www.parthenos-project.eu/.

[14] PARTHENOS, Hollander, H., Morselli, F., Uiterwaal, F., Admiraal, F., Trippel, T., & Di Giorgio, S. (2019). PARTHENOS Guidelines to FAIRify data management and make data reusable. Zenodo. https://doi.org/10.5281/zenodo.3368858.

[15] Harrower, Natalie, Maciej Maryl, Timea Biro, Beat Immenhauser, and ALLEA Working Group E-Humanities. 'Sustainable and FAIR Data Sharing in the Humanities: Recommendations of the ALLEA Working Group E-Humanities'. Digital Repository of Ireland, 2020. https://doi.org/10.7486/DRI.TQ582C863.

OPERAS, and CESSDA. Below, we present each of them with brief references to their key outputs related to FAIR principles implementation.

Infrastructure & Description	Key Services & Resources
ARIADNE Research Infrastructure (https://ariadne-infrastructure.eu/) is an international non-profit organisation founded in 2022. It connects repositories, heritage bodies, and research institutions, fostering collaboration between archaeologists and information scientists.	• **Catalogue**: Main point of access for searching datasets. • **Reference model**: Built on CIDOC CRM. • **Training hub**: Courses on FAIR principles in archaeology. • **Services**: DMP Tool, ARIADNEplus Lab, Visual Media Service.
CLARIN (https://www.clarin.eu/) Provides language data and tools for SSH research. Established in 2012, it enables resource sharing across European research centers.	• **Certified repositories** for data deposit. • **Virtual Language Observatory (VLO)**. • **Language Resource Switchboard**. • Recommendations on **licences, formats, and standards**.
DARIAH (https://www.dariah.eu/) A European network supporting digital arts and humanities by connecting people, expertise, and technologies.	• **Catalogue of tools and services**. • **DARIAH-Campus**: Training and education platform. • **DARIAH Teach**: Multilingual learning materials.
E-RIHS (https://www.e-rihs.eu/) Supports heritage science research in Europe via a central hub and national nodes.	• **Catalogue of services**. • **Training camps**.
OPERAS (https://operas-eu.org/) Coordinates resources for open scholarly communication in SSH, aiming to make Open Science accessible to all.	• **GoTriple**: Multilingual discovery platform. • **TRIPLE Ontologies**. • **Open Science Training Series**. • **Common Standards White Paper**.
CESSDA (https://www.cessda.eu/) Promotes social science research results and interoperability among scholars and institutions.	• **Data catalogue**: >40,000 collections. • **ELSST**: Multilingual social science thesaurus. • **Vocabulary service** and **Resource Directory**. • **Euro Question Bank** and training resources.

Open science and the Social Sciences and Humanities Open Cloud (SSHOC)

Several FAIR principles—particularly accessibility and reuse—align with open science principles...

- UNESCO Recommendation on Open Science. 2021.
  https://doi.org/10.54677/MNMH8546.
- Leonelli, Sabina. *Philosophy of Open Science.* Cambridge University Press, 2023.
  https://doi.org/10.1017/9781009416368.

Finally, we briefly present the **Social Sciences & Humanities Open Cloud (SSHOC)**,[16] an EU Horizon 2020 project uniting 47 partners to develop the SSH area of the **European Open Science Cloud (EOSC)**.[17] The project includes research infrastructures, such as CLARIN, CESSDA and DARIAH, as well as universities and computing centres from across Europe. From 2019, SSHOC transformed siloed data facilities into an integrated cloud network. The infrastructure provides tools and training for researchers to access, process, analyse, enrich and compare data across repositories. SSHOC aligns with EOSC requirements to ensure service sustainability.

Two key tools provided by the SSHOC infrastructure are:

- **the SSH Training Discovery Toolkit**,[18] an inventory of training materials relevant for the Social Sciences and Humanities;

- **the SSH Open Marketplace**,[19] a discovery portal including tools, services, training materials, datasets, publications and workflows. It showcases solutions for the SSH research data life cycle and serves as an aggregator of curated resources and an entry point to EOSC.

About these guidelines

As mentioned above, the Digital Humanities community already has comprehensive guidelines for implementing FAIR principles: **Sustainable and FAIR Data Sharing in the Humanities: Recommendations of the ALLEA Working Group E-Humanities**.[20]

While the ALLEA guidelines are broad in scope, we created **a complementary tool** that offers **specific, practical advice** for applying FAIR principles to common DH research outputs: digital scholarly editions, text collections, software tools, linked open data, and ontologies. Through our research in the

---

[16]https://www.sshopencloud.eu/.
[17]https://open-science-cloud.ec.europa.eu/.
[18]https://sshopencloud.eu/training.
[19]https://marketplace.sshopencloud.eu/.
[20]https://doi.org/10.7486/DRI.tq582c863.

ATLAS project, we identified best practices, reference standards, and tools for implementing FAIR principles in the design, development, and maintenance of these outputs.

These guidelines are primarily aimed at Digital Humanities scholars seeking to learn about or deepen their understanding of FAIR principles, focusing on how to put them into practice when creating their research outputs. Additionally, the guidelines are also suitable for scholars who are (relatively) new to DH and wish to explore this field's distinctive research products, along with their respective best practices and reference standards.

The guidelines are divided into sections by output type. Each section begins with an overview of the reference standards, followed by recommendations that cover the phases of a digital object's life cycle—aligning with the framework proposed in the ALLEA recommendations, as illustrated below.

Figure 33: life cycle of FAIR research data in the humanities. Credits: ALLEA E-Humanities Working Group.
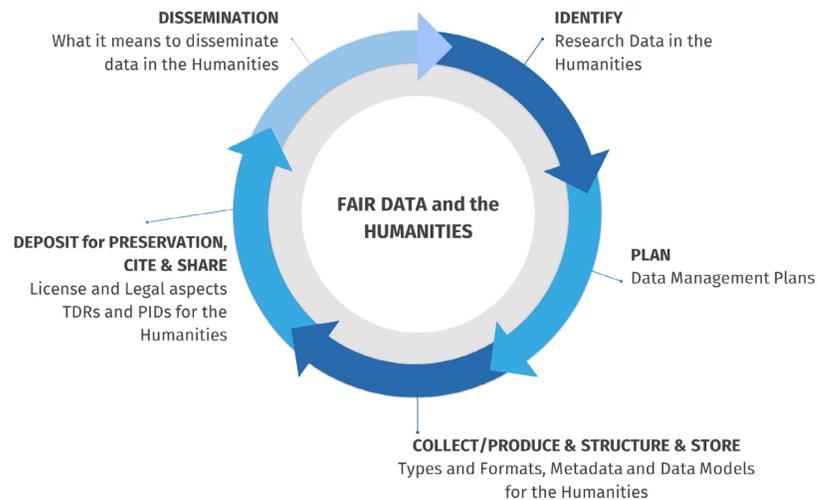
To help you identify which ALLEA phases each recommendation corresponds to, you will find the following tags next to each recommendation:

- **IDENTIFY** - recommendations focused on identifying your own research data, community standards and best practices, as well as existing resources, such as tools and datasets, that could be useful for creating the research product;

- **PLAN** - recommendations about planning your research product life-cycle and what to consider from creation through publication and maintenance;

- **PRODUCE** - recommendations for creating and formatting your research product, with particular attention to formats, metadata, and data models;

- **DEPOSIT** - recommendations for properly depositing and preserving your research product, including considerations for licences and legal aspects;

- **DISSEMINATE** - recommendations on how to effectively reach your audience and make your research product (re-)usable.

We conclude with additional recommendations **to enhance the quality** of research outputs beyond FAIR principles, making them easier to **cite, use, and assess**. These recommendations address common shortcomings we discovered while reviewing research outputs during our project.

Before exploring our guidelines, **we strongly recommend reading the AL-LEA recommendations first**, especially if you are new to FAIR principles. To help you get started, we provide a concise summary of the ALLEA recommendations below, giving you an overview of the key concepts and helping you identify which areas require deeper study.

FAIR DATA WORKFLOW

IDENTIFY
- Consider all your research assets as potentially reusable data.
- Learn the FAIR Data Principles.
- Document your research digitally from the start.
- Use recognised tools and browse existing humanities datasets.
- Aim for data to be as open as possible, as closed as necessary.

PLAN
- Create a Data Management Plan (DMP) before collecting data.
- Use funder templates or tools like DMPOnline.
- Plan for metadata documentation using standard schemas and controlled vocabularies.
- Keep DMPs updated as research progresses.
- Involve data support staff and consider RDM costs early on.

FAIR DATA WORKFLOW (Continued)

COLLECT, PRODUCE, STRUCTURE, STORE

Data Types & Formats
- Choose community-accepted formats and those preferred by preservation repositories.
- Check what other researchers use for similar data.

Metadata
- Follow metadata standards and ensure consistency.
- Use controlled vocabularies, PIDs, and make metadata rich and machine-readable.

Data Models
- Apply FAIR principles to data modeling.
- Use open, human- and machine-readable standards (e.g., XML, RDF).
- Normalise data and use identifiers like DOI, VIAF.
- Align data models with the DMP and document thoroughly.

Legal Aspects
- Address legal issues early, including consent, copyright, and anonymisation.
- Get legal support from your institution or library.

Licences
- Identify data ownership before licensing.
- Prefer open licences (e.g., CC BY, CC0), avoid overly restrictive ones (e.g., NC, ND).
- Use licence selector tools and make licences machine-readable.

Trusted Repositories & PIDs
- Use certified repositories (e.g., CoreTrustSeal).
- Repositories should assign PIDs and allow rich metadata.
- Link publications and datasets using PIDs.

DISSEMINATE
- Use networks, portals, and researcher profiles (e.g., ORCID).
- Share data and supporting materials online.
- Consider data papers to increase visibility and reuse.
- Engage broader audiences using non-traditional formats (infographics, apps, exhibitions).
- Promote open data for education and outreach (e.g., Hackathons).
- Use trusted repositories for self-archiving.

LEGACY DATA
- Curate data to prevent it becoming legacy at risk.
- Address licensing for older data.
- Advocate for funding for digitisation and infrastructure.
- Make legacy data open and FAIR whenever possible.

Digital scholarly editions

Standards and guidelines

The standard document format for text encoding is the one developed and promoted for more than 30 years by the Text Encoding Initiative (**TEI**)[21] and based on the XML markup language. The TEI guidelines[22] are organised in modules and allow to represent various text types and literary genres, from manuscripts to dictionaries, from performance texts to speech transcriptions. The current version of the guidelines is P5, which is freely available online in HTML and PDF formats, and is also available in different languages.

The TEI consortium offers a series of tools to facilitate text encoding:

- Default encoding schemas, including TEI All, TEI Lite.

- XSLT stylesheets[23] and the TEIGarage[24] tool, for automatic conversion between numerous formats and TEI.

- Roma[25] tool, for creating custom encoding schemas

- The TEI Archiving, Publishing, and Access Service (TAPAS),[26] which allows users to deposit and publish their TEI-encoded texts in a single online environment.

Given the growing adoption of TEI as a standard for creating digital scholarly editions, many editing tools are compatible with TEI or allow data export in TEI format.

Over the years, building on TEI guidelines and the experience of its promoting consortium, new encoding schemas and standards have been developed to represent specific types of texts and documents, namely:

- MEI (Music Encoding Initiative),[27] for musical documents;

- Epidoc (Epigraphic Documents in TEI XML),[28] for epigraphs and inscriptions;

- CEI (Charters Encoding Initiative),[29] for medieval diplomatic documents (charters).

Since many digital scholarly editions (DSEs) need to provide readers with digital images of the documents or manuscripts on which they are based, the most

---

[21] https://tei-c.org/.
[22] https://www.tei-c.org/release/doc/tei-p5-doc/en/html/index.html.
[23] https://github.com/TEIC/Stylesheets.
[24] https://teigarage.tei-c.org/.
[25] https://roma.tei-c.org/.
[26] https://tapasproject.org/.
[27] https://music-encoding.org/.
[28] https://epidoc.stoa.org/.
[29] https://www.cei.lmu.de/.

effective way is through the International Image Interoperability Framework (**IIIF**).[30] The IIIF website offers an introductory guide to help you understand how the IIIF works.[31]

Many libraries and cultural institutions (for example, the Vatican Library and the French National Library)[32] have adopted this set of standards and APIs, allowing scholars to access and reuse photographic reproductions of cultural objects from their collections.

For metadata, the most convenient solution is to implement them within the same document, always using the TEI standard, as it allows for describing both the digital edition itself and how it was prepared, as well as the documents/witnesses on which it is based (TEI Header Module).[33] Alternatively, other standard metadata schemas such as Dublin Core[34] and the FRBR[35]/LRM[36] model can be used and also included in TEI documents using the xenoData[37] element.

The RIDE[38] journal's criteria for reviewing and evaluating digital critical editions[39] highlight which aspects need to be addressed or given particular attention to create reliable and high-quality editions.

Recommendations

1) Planning the editorial workflow

Plan an editorial workflow keeping the FAIR principles in mind and taking advantage of existing tools, to optimise editing and data management.

IDENTIFY   **PLAN**

Numerous tools are available to assist philologists throughout the editing process, from transcribing primary sources to lemmatisation, making edition creation more efficient and cost-effective. With the goal of producing a FAIR-compliant digital edition in mind, selecting appropriate tools and establishing a workflow becomes more straightforward, for example by favoring tools that support standard formats.

- Review existing software tools that help editors make digital scholarly editions.

---

[30]https://iiif.io/.

[31]https://iiif.io/get-started/how-iiif-works/.

[32]The IIIF website provides a list of all collections available through IIIF: https://iiif.io/guides/finding_resources/.

[33]https://www.tei-c.org/release/doc/tei-p5-doc/en/html/HD.html.

[34]https://www.dublincore.org/.

[35]https://repository.ifla.org/handle/20.500.14598/830,

[36]https://repository.ifla.org/handle/20.500.14598/40.2.

[37]https://tei-c.org/release/doc/tei-p5-doc/en/html/ref-xenoData.html.

[38]https://ride.i-d-e.de/.

[39]https://www.i-d-e.de/publikationen/weitereschriften/criteria-version-1-1/.

- To discover tools, you can use curated lists and dedicated catalogues such as:

  - TaPoR,[40] a gateway to tools used for scholarly research and activities;

  - DARIAH's catalogue of tools and services;[41]

  - SSH Open Marketplace.[42]

2) Including photographic reproductions of the documents

Include photographic reproductions of the documents/witnesses when available, preferably through IIIF.

PLAN   **PRODUCE**

Manuscript and document images can be used both for the preparation and the publication of the edition. The inclusion of photographic reproductions within the editions allows readers to verify the editor's readings first-hand.

- Many tools for both digital scholarly editing (e.g., eScriptorium,[43] PRISMS,[44] FairCopy)[45] and digital publishing (e.g., EVT)[46] are compatible with IIIF, allowing images to be retrieved directly from their repositories. Alternatively, IIIF-compatible viewers, such as OpenSeadragon,[47] Universal Viewer,[48] and Mirador,[49] may be integrated into the edition's website.

- If it is not possible to include the documents'/witnesses' digital facsimiles in the DSE, we recommend providing links to the facsimiles or the cataloguing records offered online.

Guide: "How to use IIIF resources and image viewers".[50]

---

[40]https://tapor.ca/.
[41]https://www.dariah.eu/tools-services/tools-and-services/.
[42]https://marketplace.sshopencloud.eu/.
[43]https://escriptorium.rich.ru.nl/.
[44]https://www.prisms.digital/.
[45]https://faircopyeditor.com/.
[46]https://evt.labcd.unipi.it/.
[47]https://openseadragon.github.io/.
[48]https://universalviewer.io/.
[49]https://projectmirador.org/.
[50]https://iiif.io/guides/using_iiif_resources/.

3) Highlighting named entities with semantic web technologies

For texts rich in references to people, places, and named entities, consider creating a semantic edition to highlight these aspects.

PLAN   **PRODUCE**

semantic editions enable automated text analysis, including the identification of recurring themes and character relationships. These editions can be integrated with other databases and research tools, enabling broader historical, cultural, and linguistic analysis of the text.

- To easily create semantic editions, you can use tools like LEAF-Writer,[51] an online XML and RDF editor that automatically detects named entities and links them to records in a Linked Open Data authority.

Example: digital semantic edition of Paolo Bufalini's notebook.[52]

4) Adopting standard formats

Use XML/TEI or another appropriate standard schemato ensure interoperability.

PRODUCE

- To facilitate encoding, consider these strategies:
  - Employ an editor that exports data as XML/TEI or other appropriate formats.
  - Convert from other data formats to desired format using conversion tools (e.g., TEIGarage).
  - Consider using annotation tools based on Domain-Specific Languages [Zenzaro et al. 2025].

5) Creating and sharing encoding schemas

If necessary, create and publish custom encoding schemas, starting from existing ones.

PRODUCE   **DEPOSIT**

Existing digital scholarly editions may have developed custom encoding schemas that can be directly reused or used as a reference for modelling your domain.

- To identify existing domain-relevant editions, it is possible to use community-driven lists (e.g., AIUCD's projects list,[53] list of tools, portals

---

[51]https://leaf-writer.leaf-vre.org/.

[52]https://projects.dharc.unibo.it/bufalini-notebook/.

[53]https://www.aiucd.it/progetti/.

and editions on Griseldaonline)[54] and two well-known catalogues of digital scholarly editions:

- Franzini, G. (2012-) Catalogue of Digital Editions, https://doi.org/10.5281/zenodo.1161425.

- Sahle, Patrick et al., a catalog of Digital Scholarly Editions, v.4.112 2020ff, last change 2024-06-06, https://www.digitale-edition.de/exist/apps/editions-browser/index.html.

Digital editions often describe their underlying data models in written documentation. While these descriptions are very useful for users, they are not machine-readable. To achieve full FAIR compliance, an edition's encoding schemashould be formalised and published with a retrievable URL.

- When using XML/TEI, you can apply the following strategies to easily create custom schemas:

  - Use the TEIRoma tool.

  - Follow the TEI customisation guidelines.[55]

  - Use TEI documentation elements[56] to thoroughly describe new elements.

- Publish encoding schemas in a certified repository, alongside the edition's files.

6) Documenting textual tradition and the editorial process

Provide a detailed description of the textual tradition, the editorial process and applied methodologies, accompanied by links to relevant web resources.

PRODUCE

This will allow readers to clearly understand the edition's objectives and how it was prepared, and to verify the editor's work.

- Specify the provided edition type(s), such as diplomatic, interpretative, critical, etc. To clarify the meaning of these labels, provide the link to their definition within a shared lexicon (e.g., Parvum Lexicon Stemmatologicum,[57] Lexicon of Scholarly Editing).[58]

If you develop solutions to common challenges while working on your edition—whether it's a custom encoding schemaor new software tools—document and share them in detail to help other scholars facing similar issues.

---

[54]https://site.unibo.it/griseldaonline/it/strumenti/strumenti-italianistica-digitale.

[55]https://www.tei-c.org/release/doc/tei-p5-doc/it/html/USE.html#MD.

[56]https://www.tei-c.org/release/doc/tei-p5-doc/it/html/TD.html.

[57]https://wiki.helsinki.fi/xwiki/bin/view/stemmatology/Parvum%20lexicon%20stemmatologicum/.

[58]https://lexiconse.uantwerpen.be/.

- Include references to other research products, such as software tools, datasets and previous DSEs, used to prepare the edition.

7) Giving credit to all editors and contributors

Give credit to all editors and contributors, including those responsible for markup and technical implementation.

PRODUCE

Markup is a scholarly activity that significantly impacts the expressiveness and quality of a digital edition. Choosing elements and attributes to best represent textual phenomena and creating new custom elements are complex tasks that deserve recognition. Similarly, the technical implementation of a digital scholarly edition is crucial—developers' work is essential for producing usable and accessible DSEs.

- The Contributor Role Taxonomy[59] provides guidance for describing each contributor's specific role in the edition.

8) Making the editorial process transparent

Make your editorial process transparent by publishing different versions of the edition and its by-products.

PRODUCE   **DEPOSIT**

Creating a digital edition is a long, iterative process where editors continually refine the text and its accompanying materials (critical apparatuses, introduction, etc.). When users consult an online edition, they often cannot tell whether they are viewing the final version or a work in progress that may soon change. Publishing without clear version status makes digital editions appear unstable and unreliable. Therefore, we recommend publishing intermediate versions before the final edition, clearly documenting changes and work status. This approach allows intermediate versions to be cited and reused while making the edition more trustworthy. During the early stages, editors can also publish other standard components of a critical edition, such as the bibliography, list of witnesses/documents, and their transcriptions and/or translations.

- Within trusted repositories like Zenodo,[60] editors can upload new versions of the same file. The application automatically generates a PID for each version, maintaining links between them.

- The TEI guidelines allow for detailed description of the various revisions made to a document (TEI's Revision Description).[61]

---

[59]https://credit.niso.org/.

[60]https://zenodo.org/.

[61]https://www.tei-c.org/release/doc/tei-p5-doc/en/html/HD.html#HD6.

9) Including a downloadable version

Enable users to download the edition, including a print-ready PDF version when appropriate, to facilitate reuse.

PRODUCE   **DISSEMINATE**

The practice of reuse in philology is far more common than many realise. Editions frequently build upon previous editions, serve as foundations for linguistic studies, or provide source material for dictionaries. In the digital age, FAIR principles have expanded the possibilities for reuse - editions can now serve as training data for machine learning models (such as automatic transcription and linguistic analysis tools) and enable sophisticated intertextual analysis.

Examples of reuse scenarios: Spadini and Palenzuela, 'Re-Using Data from Editions', 2025, https://doi.org/10.62637/sup.GHST9020.8.

- Where there are no copyright restrictions, make the edition freely downloadable in a standard format such as XML/TEI.

- Consider other formats as well, including TXT, ODT and JSON. The more formats available, the easier the edition is to reuse.

- The PDF format can be more accessible for some users.

Example: 'De rebus Siculis Carmen'[62] edited by Fulvio delle Donne available in XML/TEI and in PDF.

10) Depositing and indexing the edition

Store the edition in a certified repository and ensure it is indexed in relevant field-specific catalogues to guarantee long-term preservation and enhance its discoverability.

DEPOSIT

Depositing the edition in a certified repository ensures its long-term accessibility, even if the visualisation system becomes obsolete. Well-known certifications are CoreTrustSeal[63] and Nestor.[64]

- To maximise discoverability, index the edition in searchable resources like OpenAIRE Explore[65] (which happens automatically when using repositories like Zenodo) and in relevant field-specific catalogues and lists of digital critical editions (see point 2).

---

[62]https://bup-unibas.atcult.it/omp/index.php/bup/catalog/book/978-88-31309-02-8.

[63]https://www.coretrustseal.org/.

[64]https://www.langzeitarchivierung.de/Webs/nestor/EN/Zertifizierung/zertifizierung_node.html.

[65]https://explore.openaire.eu/.

- Check if any research infrastructures, associations or other institutions in your field compile lists or portals to which you can submit your digital edition.

11) Adopting non-proprietary and existing publishing solutions

For visualising the edition, use non-proprietary tools or adopt existing publishing solutions.

IDENTIFY **DISSEMINATE**

Various tools for visualizing DSEs—including TEIPublisher,[66] TEIBoilerplate,[67] and EVT—allow you to customise text formatting. When you need specialised visualisation features, instead of building from scratch, use existing open-source tools. You can often work directly with tool development teams to tailor these solutions to your needs.

Example: the collaboration between the digital edition of the Codice Pelavicino and the visualisation software EVT, https://pelavicino.labcd.unipi.it/il-progetto/il-software-evt/.

Ready-to-use publishing solutions allow editors to focus solely on preparing the edition while delegating publication tasks. Examples are the Micro-Editions[68] of the Scholarly Editing journal and the editions in the Digital Humanities series by Basilicata University Press (BUP).[69]

12) Providing indexes and a search function

For extensive and rich texts, provide indexes and a search function to improve discoverability.

DISSEMINATE

Example: analytical indexes of the National Edition of Aldo Moro's works, https://doi.org/10.6092/unibo/aldomoro.

---

[66] https://teipublisher.com/exist/apps/tei-publisher-home/index.html.
[67] https://dcl.luddy.indiana.edu/teibp/index.html.
[68] https://scholarlyediting.org/contributing/.
[69] https://bup-unibas.atcult.it/omp/index.php/bup/catalog/series/DH.

Text collections

Standards and guidelines

Based on a collection's scientific objectives and technical requirements, texts can be prepared in different formats. A common approach, especially for Natural Language Processing (NLP), is **plain text** with accompanying metadata and annotations in structured formats like CSV and JSON. In these cases, **Unicode** character encoding is essential to ensure full accessibility, reusability, and interoperability.

XML/TEI is another widely used format, as it provides a comprehensive framework for encoding and describing textual data. Its guidelines include a module dedicated to linguistic corpora.[70] For more information about TEI, its tools, and other specialised encoding schemas based on or inspired by TEI, please refer to the Digital Scholarly Editions section. Another standard is CES (Corpus Encoding Standard).[71]

For collection metadata, several standards and formats are available alongside TEI:

- CIDOC-CRM,[72] an ISO standard and reference ontology that provides definitions and formal structure for describing cultural heritage documentation concepts and relationships;

- Dublin Core, a simple, generic metadata element set for various digital object types. This standard is widely adopted world-wide;

- MARC 21,[73] a well-established standard for exchanging bibliographic records, developed and maintained by the library community;

- Metadata Encoding and Transmission Standard (METS),[74] an XML schemafor encoding structural metadata about complex digital objects;

- MODS (Metadata Object Description Schema),[75] an XML schemafor descriptive metadata compatible with MARC 21 bibliographic format.

For more information about metadata standards, refer to the arts and humanities standards listed in the Metadata Standards Catalog.[76]

For presenting texts and documents as digital facsimiles, IIIF is recommended, particularly for institutions digitizing their textual heritage. The IIIF website offers a guide for IIIF implementers.[77]

---

[70]https://www.tei-c.org/release/doc/tei-p5-doc/en/html/CC.html.
[71]https://tei-c.org/activities/projects/corpus-encoding-standard-ces/.
[72]https://cidoc-crm.org/.
[73]https://www.loc.gov/marc/bibliographic/.
[74]https://www.loc.gov/standards/mets/.
[75]https://www.loc.gov/standards/mods/.
[76]https://rdamsc.bath.ac.uk/.
[77]https://iiif.io/guides/guide_for_implementers/.

For developing a complete IT environment to create and manage digital collections, the Reference model for an Open Archival Information System (OAIS)[78] serves as the standard model. This technical and complex work typically extends beyond individual scholars' scope. However, you can access the OAIS introductory guide via the link in the footnote.[79]

The **National Information Standards Organization (NISO) Framework of Guidance for Building Good Digital Collections (3rd edn, 2007)**[80] provides a comprehensive set of principles that in 2007 anticipated the FAIR principles. The NISO principles are organised around four core entities: collections, collected objects, metadata, and "initiatives" (programmes or projects for creating and managing collections). We offer the complete list of principles below. Please refer to the document linked above, for practical guidance on how to achieve these quality requirements.

---

[78]https://ccsds.org/wp-content/uploads/gravity_forms/5-448e85c647331d9cbaf66c096458bdd5/2025/01//650x0m3.pdf.

[79]http://www.dpconline.org/component/docman/doc_download/1359-dpctw14-02.

[80]https://www.niso.org/sites/default/files/2017-08/framework3.pdf.

Collections	Metadata
1. A good digital collection is created according to an explicit collection development policy. 2. Collections should be described so that a user can discover characteristics of the collection, including scope, format, restrictions on access, ownership, and any information significant for determining authenticity and interpretation. 3. A good collection is curated: its resources are actively managed during their entire lifecycle. 4. A good collection is broadly available and avoids unnecessary impediments to use. 5. A good collection respects intellectual property rights. 6. A good collection has mechanisms to supply usage data and other data that allows standardised measures of usefulness. 7. A good collection is interoperable. 8. A good collection integrates into the users own workflow. 9. A good collection is sustainable over time.	1. Good metadata conforms to community standards in a way that is appropriate to the materials, users, and future uses. 2. Good metadata supports interoperability. 3. Good metadata uses authority control and content standards to describe and collocate objects. 4. Good metadata includes a clear statement of the conditions and terms of use. 5. Good metadata supports the long-term curation and preservation of objects. 6. Good metadata records are objects themselves and should have qualities like authority, persistence, and unique identification.

Objects	Initiatives
1. A good object exists in a format that supports its intended current and future use. 2. A good object is preservable. 3. A good object is meaningful and useful outside of its local context. 4. A good object will be named with a persistent, globally unique identifier. 5. A good object can be authenticated. 6. A good object has associated metadata.	1. A good digital initiative has a substantial design and planning component. 2. A good digital initiative has an appropriate level of staffing with necessary expertise. 3. A good digital initiative follows best practices for project management. 4. A good digital initiative has an evaluation component. 5. A good digital initiative markets itself and broadly disseminates information about outcomes. 6. A good digital initiative considers the entire lifecycle of the digital collection and services.

Finally, the RIDE journal offers quality evaluation criteria for digital text collections,[81] providing guidance for collection preparation and management.

Recommendations

1) Adopting standard formats

Encode texts in standard formats, preferring XML/TEI or other appropriate schemas for literary texts, to ensure interoperability and facilitate re-use.

PRODUCE

- When detailed encoding is impractical due to time or resource constraints, consider implementing a light encoding first (e.g., ALIM, Biblioteca Italiana) and different levels of encoding after, progressively covering more aspects and phenomena.

- If custom encoding schemas were used, publish and describe them in the documentation.

2) Citing the sources

Always cite the sources used for text preparation by providing complete bibliographic references and links to descriptive web resources, if available.

PRODUCE

A comprehensive set of metadata describing the sources used for text preparation enables users to verify both the editorial work and text quality.

- Use a metadata standard such as Dublin Core or XML/TEI (<sourceDesc>[82] element in the header).

- Prefer links to digital libraries, online catalogues and other similar resources that provide persistent identifiers for their objects.

---

[81] https://www.niso.org/sites/default/files/2017-08/framework3.pdf.
[82] https://www.tei-c.org/release/doc/tei-p5-doc/en/html/HD.html#HD3.

3) Linking to authority records

Link authors' and works' records to corresponding authority records if available, e.g., VIAF, Wikidata.

PLAN **PRODUCE**

By linking the collected works and their authors to authority records, researchers can search across multiple datasets using standardised identifiers, particularly when the collection offers API access.

Examples in Spadini, Elena, and José Luis Losada Palenzuela. "Re-Using Data from Editions." *Digital Editing and Publishing in the Twenty-First Century*, edited by James O'Sullivan et al., 1st ed., Scottish Universities Press, 2025, https://doi.org/10.62637/sup.GHST9020.8.

Example: The Perseus Catalog, https://catalog.perseus.org/.

4) Documenting collection criteria and the editorial process

In the documentation, specify collection criteria and editorial criteria, stating the philological methodologies applied and the edition types.

PRODUCE

This information helps users understand the relationship between the collected text and its sources, as well as the editorial preparation process. For example, it clarifies whether the text is a transcription of an audio file or performance, a document processed through OCR software, or a critical edition.

- To specify the edition type, you can reference established definitions from scholarly literature or online resources like the Parvum Lexicon Stemmatologicum and the Lexicon of Scholarly Editing.

5) Assigning persistent identifiers

Assign a persistent identifier to each text in the collection to make it easier to access and reuse, and to increase its citability.

DEPOSIT

- When texts are published as individual units in repositories like ILC4CLARIN[83] and Zenodo, PIDs are automatically assigned.

---

[83]https://dspace-clarin-it.ilc.cnr.it/.

6) Depositing the text collection

Store the text collection in a certified repository and upload the latest version of the collection regularly.

DEPOSIT

As mentioned in the previous section, depositing the text collection in a certified repository ensures its long-term accessibility, even if the visualisation system becomes obsolete.

- While the collection is still being developed, you can publish and deposit the intermediate outcomes progressively.

- Once the collection is complete, we recommend depositing the latest version after any revisions or relevant changes.

Following these recommendations will make your editorial process more transparent, giving users access to different versions of the collection and preliminary research results.

7) Documenting changes and final goals

Document changes and current status thoroughly, indicating the number of available texts, the collection's completeness relative to its scientific objectives and, if applicable, a roadmap about the evolution of the text collection.

DEPOSIT   **DISSEMINATE**

Specifying the number of texts in a collection enables users to assess both the corpus's completeness for their research goals and its overall representativeness.

For collections containing large volumes of texts, editorial work typically relies on time-limited funding, which leads to periodic additions to and revisions of the collection. Such cases require detailed documentation of the work's status, including clear descriptions of previous work and future plans. This documentation gives users a clear understanding of the stability and reliability of the collection's texts.

- You can format this part of the documentation as a changelog, following the guiding principles of the "keep a changelog" project,[84] in particular marking changes to the texts in the collection as added, removed, changed, or fixed.

---

[84]https://keepachangelog.com/en/1.1.0/.

8) Facilitating text exploration

Facilitate text exploration through search functionalities, indexes and sub-collections.

DISSEMINATE

A well-designed search functionality can help users easily find meaningful content in the text collection, while indexes provide quick access to individual texts or sub-collections.

Sub-collections can showcase the information contained in the collection, by organising texts by theme, topic, author, genre, etc. To help users engage with the collection, sub-collections should model how users can approach the search functionality with a question or theme to produce meaningful results.

Examples in Chapman, Alison, et al. "Browse, Search and Serendipity: Building Approachable Digital Editions." *Digital Editing and Publishing in the Twenty-First Century*, edited by James O'Sullivan et al., 1st ed., Scottish Universities Press, 2025, https://doi.org/10.62637/sup.GHST9020.6.

Software tools

Standards and guidelines

The FAIR Principles for Research Software (FAIR4RS Principles),[85] developed by the FAIR for Research Software Working Group within the Research Data Alliance framework, were designed to improve the sharing and reuse of research software. They expand the FAIR principles, in order to address specific characteristics of software — such as its executability, composite nature, and continuous evolution and versioning — and namely are:

> "F: Software, and its associated metadata, is easy for both humans and machines to find.
>
> A: Software, and its metadata, is retrievable via standardised protocols.
>
> I: Software interoperates with other software by exchanging data and/or metadata, and/or through interaction via application programming interfaces (APIs), described through standards.
>
> R: Software is both usable (can be executed) and reusable (can be understood, modified, built upon, or incorporated into other software)."

The Research Software MetaData Guidelines (RSMD)[86] is a comprehensive set of guidelines aimed at enhancing the findability, accessibility, interoperability, and reusability of research software artifacts, developed within the FAIR-IMPACT

---

[85]https://doi.org/10.15497/RDA00068.

[86]https://fair-impact.github.io/RSMD-guidelines/.

project[87] for the European Open Science Cloud (EOSC). The RSMD guidelines[88] are organised into seven distinct aspects — accessibility and preservation, reference and identification, description and classification, credit and attribution, reuse and legal, re-execute, and general remarks — each with a clear objective and a set of actionable recommendations. Recommendations are categorised into three priority levels: essential, important, and useful. This prioritisation helps emphasise the critical recommendations and ensure the guidelines address key areas effectively.

To assess the FAIRness level of a software artifact, you can use the metrics[89] outlined within the FAIR-IMPACT project. These metrics emphasise key aspects such as the importance of clearly describing the software's purpose, defining its development status, and enabling the identification and reuse of individual software components. You may find the complete list below.

---

[87]https://fair-impact.eu/.

[88]RSMD checklist, a quick overview of the RSMD guidelines: https://fair-impact.github.io/RSMD-guidelines/8.rsmd_checklist/.

[89]https://fair-impact.eu/metrics-software.

1. Does the software have a globally unique and persistent identifier?

2. Do the different components of the software have their own identifiers?

3. Does each version of the software have a unique identifier?

4. Does the software include descriptive metadata which helps define its purpose?

5. Does the software include development metadata which helps define its status?

6. Does the software include metadata about the contributors and their roles?

7. Does the software metadata include the identifier for the software?

8. Does the software have a publicly available, openly accessible and persistent metadata record?

9. Is the software developed in a code repository / forge that uses standard communications protocols?

10. Are the formats used by the data consumed or produced by the software open and a reference provided to the format?

11. Does the software use open APIs that support machine-readable interface definition?

12. Does the software provide references to other objects that support its use?

13. Does the software describe what is required to use it?

14. Does the software come with test cases to demonstrate it is working?

15. Does the software source code include licensing information for the software and any bundled external software?

16. Does the software metadata record include licensing information?

17. Does the software include provenance information that describe the development of the software?

Chue Hong, Neil, et al. *D5.2 - Metrics for Automated FAIR Software Assessment in a Disciplinary Context.* Oct. 2023, https://doi.org/10.5281/ZENODO.10047401.

In order to be FAIR, software artefacts must be published under clear standard licences. We recommend using licenses approved by the OSI (Open Source Initiative).[90] A complete list of licences used for software artefacts is available on the SPDX specification website.[91]

Software Heritage[92] is an international organisation that collects and preserves software in source code form, recognizing software as a valuable part of our cultural heritage that embodies technical and scientific knowledge. You can use Software Heritage services to store and preserve your own software or to search and access archived software.[93]

---

[90]https://opensource.org/licenses.

[91]https://spdx.org/licenses/.

[92]https://www.softwareheritage.org/.

[93]Software Heritage documentation: https://docs.softwareheritage.org/.

Last but not least, developers working in the DH field can follow and join the activities of DHTech,[94] an ADHO special interest group aimed at supporting the development and reuse of software in the Digital Humanities.

Recommendations

Disclaimer

The field of software engineering offers numerous guidelines and standards for developing robust and usable software. Your choice of architectures, technologies, and development tools should align with your software's intended purpose and development context. While our recommendations are general and primarily focus on ensuring software FAIRness—particularly regarding metadata and reusability—we have included references to widely accepted best practices and standards to guide scholars with limited software development experience who wish to enter this field.

1) Reusing and enhancing existing software and other resources.

Before creating new software from scratch, investigate existing similar solutions and explore opportunities to further develop or adapt them, promoting the reuse and enhancement of existing resources.

IDENTIFY

Example: forks on GitHub of the EVT software, https://github.com/evt-project/evt-viewer.

2) Involving domain experts and applying software engineering best practices

Involve domain experts in software design and apply software engineering methodologies and best practices, in order to create robust software that is easy to use, maintain and further develop.

PLAN   **PRODUCE**

- Utilise documented and shared design patterns [Gamma et al. 1994].

- When applying the object-oriented programming paradigm, follow the SOLID principles [Silén 2024].

- For complex software, implement the "domain-driven design" approach [Evans 2004].

- Organise code into modules to facilitate the reuse of individual components.

---

[94]https://dh-tech.github.io/.

- Adopt DevOps practices to streamline development and deployment processes (plan, code, build and test, release, deploy, operate, and monitor) [Silén 2024].

- Ensure that all software dependencies, whether libraries, frameworks, or operating system components, are clearly documented and managed. This also includes defining the operational requirements, such as minimum and optimal hardware resources (e.g., CPU, RAM, disk space) needed to ensure that the software works properly.

- Integrate a structured testing phase as part of the software development process, establishing clear metrics and goals to determine testing success.

3) Defining software integration strategies

Define and implement software integration strategies with the goal of achieving a cohesive, scalable and maintainable software ecosystem, minimising the risks of incompatibility and the efforts required for adaptation.

PLAN   **PRODUCE**

This also allows easy handling of format migration and can be achieved with the following steps:

- Define integration approaches: whether these will be API-based or exchange files, for example, and prepare standard protocols to facilitate communication.

- Ensure interoperability and compatibility between different systems by considering standard data formats and structured schemas.

- Plan strategies for handling errors and malfunctions.

- Ensure scalability and the ability to handle increased load without compromising overall performance.

4) Employing standard and non-proprietary programming languages

Employ standard and non-proprietary programming languages and technologies to develop tools, ensuring greater longevity and easier maintainability.

PRODUCE

- W3C standards for web development.[95]

- Community Development of Java Technology Specifications.[96]

---

[95]https://www.w3.org/TR/?status%5B0%5D=standard.

[96]https://www.jcp.org/en/home/index.

Choose a programming language with mature libraries that can ease the development and maintenance of your software. For example, use Python for NLP software development to easily integrate available tools.

5) Open source and collaborative development

When possible, develop in open source and foster collaborative development.

PRODUCE   **DEPOSIT**

- Write clear, comprehensive code comments.

- Provide guidelines for contributing to software development.

- Utilise platforms such as GitHub that foster collaboration among developers.

- Follow shared methodologies and strategies for versioning (e.g., Semantic Versioning)[97] and branching (e.g., GitFlow workflow).[98]

6) Releasing software officially

Release software officially through freely accessible channels (e.g.,GitHub), providing detailed and user-friendly documentation.

PRODUCE   **DEPOSIT**

With each released version, always attach a changelog document that provides a clear and organised chronology of updates, improvements, bug fixes, and other changes.

7) Publishing your released software in trusted repositories

Publish your released research software in a trusted scholarly repository (e.g., Zenodo, HAL) with rich metadata to ensure citability and credit to the development team.

DEPOSIT

- To prepare metadata, follow the Research Software MetaData Guidelines.

- Consider depositing your software source code in the Software Heritage Archive.

---

[97]https://semver.org/.

[98]https://nvie.com/posts/a-successful-git-branching-model/.

Ontologies

Standards and guidelines

As the primary developer of Semantic Web technologies, W3C has created several standard formats for expressing ontologies:

- **RDF XML** (Resource Description Framework XML Syntax)[99] serves as the foundation for many ontologies in the Semantic Web. It uses XML syntax to express relationships through triples (subject, predicate, object).

- **OWL** (Web Ontology Language)[100] represents complex knowledge about things and their relationships. Based on computational logic, it enables programs to verify consistency and reveal implicit knowledge.

- **Turtle**[101] provides a simplified, human-readable syntax for RDF that is more concise than RDF-XML.

- **JSON-LD** (A JSON-based Serialisation for Linked Data)[102] integrates Linked Data into web environments through JSON compatibility.

Other important W3C Semantic Web Standards[103] include:

- **SKOS** (Simple Knowledge Organization System)[104] for creating vocabularies and taxonomies;

- **SPARQL** (SPARQL Query Language for RDF)[105] for querying diverse RDF data sources and retrieving both result sets and RDF graphs.

The scientific literature offers several publications with clear, precise guidelines for producing FAIR ontologies. These cover essential topics like prefix and namespace conventions, as well as documentation publishing methods, including:

- Garijo, Daniel, and María Poveda-Villalón. *Best Practices for Implementing FAIR Vocabularies and Ontologies on the Web.* arXiv, 2020, https://doi.org/10.48550/ARXIV.2003.13084.

- Hugo, Wim, et al. *D2.5 FAIR Semantics Recommendations Second Iteration.* Dec. 2020, https://doi.org/10.5281/ZENODO.4314320.

- Poveda-Villalón, María, et al. "Coming to Terms with FAIR Ontologies." *Knowledge Engineering and Knowledge Management*, edited by C. Maria Keet and Michel Dumontier, vol. 12387, Springer International Publishing, 2020, pp. 255–70, https://doi.org/10.1007/978-3-030-61244-3_18.

---

[99] https://www.w3.org/TR/rdf-syntax-grammar/.
[100] https://www.w3.org/2001/sw/wiki/OWL.
[101] https://www.w3.org/TR/turtle/.
[102] https://www.w3.org/2001/sw/wiki/JSON-LD.
[103] https://www.w3.org/2001/sw/wiki/Main_Page.
[104] https://www.w3.org/2001/sw/wiki/SKOS.
[105] https://www.w3.org/2001/sw/wiki/SPARQL.

To evaluate whether your ontology adheres to the good practices outlined in these publications, you can use tools such as O'FAIRe (Ontology FAIRness Evaluator)[106] and FOOPS! (Ontology Pitfall Scanner for FAIR).[107]

Another effective tool designed specifically for SKOS vocabularies is SKOS Play!,[108] which enables users to validate and convert SKOS files to PDF and HTML formats, and also generates SKOS files from Excel spreadsheets.

Recommendations

1) Reusing existing ontologies

Consider reusing existing ontologies to build upon established knowledge structures.

IDENTIFY   **PLAN**

Numerous ontologies have been developed for cultural heritage and humanities disciplines in recent years. Before creating your own ontology from scratch, examine existing ones to evaluate whether and how they could serve your research needs. This approach not only speeds up development but also results in a more robust final product.

- Domain-relevant models, ontologies and vocabularies, used on a national and international level: CIDOC-CRM, FOAF,[109] Schema.org,[110] Dublin Core,[111] LRM$_{OO}$,[112] DataCite,[113] DCAT,[114] SPAR Ontologies,[115] and ArCO.[116]

- To explore existing models, use the LOV (Linked Open Vocabularies)[117] database, BARTOC (Basic Registry of Thesauri, Ontologies and Classifications)[118] or the H-SeTIS (Heritage – Semantic Tools and Interoperability Survey)[119] database specific for the heritage domain.

---

[106]https://github.com/agroportal/fairness.
[107]https://w3id.org/foops.
[108]https://skos-play.sparna.fr/play/.
[109]http://xmlns.com/foaf/spec/.
[110]http://schema.org/.
[111]https://www.dublincore.org/specifications/dublin-core/dcmi-terms/.
[112]https://cidoc-crm.org/lrmoo/home-0.
[113]https://schema.datacite.org/.
[114]https://www.w3.org/TR/vocab-dcat-2/#Property:resource_type.
[115]http://www.sparontologies.net/.
[116]https://dati.beniculturali.it/arco/index.php.
[117]https://lov.linkeddata.es/dataset/lov.
[118]https://bartoc.org/.
[119]https://h-setis.cnr.it/ontologies/.

2) Engaging with the user community and domain experts.

Engage the user community and domain experts throughout the design phase to create more robust and relevant ontologies.

PLAN **PRODUCE**

The first step in creating an ontology is defining its domain—the scope of reality it must represent—and identifying competency questions. These are user-oriented questions that help scope the ontology by determining what users want to learn when exploring and querying the ontology and its knowledge base.

Creating an effective ontology requires deep domain knowledge. Involving domain experts and community members early in the process helps identify precise terminology and appropriate properties to describe domain entities, their relationships, and competency questions. This collaborative approach ensures the ontology will be both useful and reusable.

Surveys and focus groups serve as effective tools for gathering domain expert feedback.

3) Adhering to shared methodologies

Adhere to shared methodologies for ontology design and implementation to ensure consistency and best practices.

PLAN **PRODUCE**

The literature offers several established methodologies and best practices for creating ontologies:

- Ontology Development 101: A guide to creating your first ontology,[120] a comprehensive guide for beginners learning to model domains through ontologies;

- Simplified Agile Methodology for Ontology Development (SAMOD),[121] an advanced approach developed by the SPAR Ontologies team;

- LOT (Linked Open Terms) Methodology[122] an industrial method for developing ontologies and vocabularies, that encompasses four main phases: requirements specification, implementation, publication and maintenance;

- The NeOn Methodology for Ontology Engineering[123] proposes nine scenarios for building ontologies and networks, focusing on resource reuse, reengineering, and collaborative development.

---

[120]https://protege.stanford.edu/publications/ontology_development/ontology101.pdf.
[121]https://essepuntato.it/papers/samod-owled2016.html.
[122]https://lot.linkeddata.es/.
[123]https://lot.linkeddata.es/.

When facing common challenges during domain modeling, you can apply Ontology Design Patterns[124] as proven solutions.

For the practical development of ontologies, several graphical editors are available, including Protégé,[125] Fluent Editor,[126] and OWLGrEd.[127] Another popular development platform for managing OWL ontologies is VocBench.[128] For more options, see BARTOC's comprehensive list of editors and development environments.[129]

4) Facilitating integration through conceptual alignment

Foster interoperability and facilitate integration by producing a comprehensive mapping that aligns your ontology with existing models and ontologies.

PRODUCE

Conceptual alignment involves associating each concept expressed in your ontology — whether as a class or a property — with an equivalent concept in other models. This process enables the production of a comprehensive mapping, enabling the easy integration of data modelled according to your ontology into other datasets. Such a mapping is also useful for facilitating data maintenance and handling legacy data.

- Many W3C models, such as DCAT, include mapping recommendations.

- Standards:

  - RDF Mapping Language (RML)[130]

  - Simple Standard for Sharing Ontological Mappings (SSSOM)[131]

  - InteroperAble Descriptions of Observable Property Terminology WG (RDA I-ADOPT WG)[132]

- Article: "Moving towards FAIR mappings and crosswalks".[133]

---

[124]http://ontologydesignpatterns.org/index.php/Ontology_Design_Patterns_._org_(ODP).

[125]https://protege.stanford.edu/.

[126]https://www.cognitum.eu/semantics/FluentEditor/.

[127]https://owlgred.lumii.lv/.

[128]https://vocbench.uniroma2.it/.

[129]https://gbv.github.io/bartoc-vocabulary-software/.

[130]https://rml.io/specs/rml/.

[131]https://mapping-commons.github.io/sssom/.

[132]https://i-adopt.github.io/.

[133]https://research.manchester.ac.uk/en/publications/moving-towards-fair-mappings-and-crosswalks.

5) Providing user documentation

Provide users with detailed documentation, complete with practical usage examples and intuitive graphical representations of the ontology.

PRODUCE **DISSEMINATE**

Documentation should be complete and accurate, beginning with clear definitions of all ontology elements. These definitions can incorporate references to external sources and specific examples to clarify concepts and properties.

To clearly demonstrate the ontology's utility, include practical examples showing how it represents real-world objects familiar to your community.

- Web environment for visualising and customising ontology documentation: LODE,[134] WebVOWL,[135] Widoco.[136] A comprehensive list of visualisation tools is available in the semantic web standards wiki.[137]

Graphical representations effectively showcase your ontology's classes and their relationships. You can produce simple UML class diagrams or use specific tools for ontologies such as Graffoo[138] and Fluent Editor.[139]

6) Ensuring the permanence of URIs

Ensure the permanence of URIs, considering the use of services for long-term stability. Conduct regular maintenance checks to ensure the continued functionality of links to your ontology.

DEPOSIT

Ensuring your ontology's long-term accessibility and reusability depends critically on functioning URIs. Regular maintenance checks are essential to keep the ontology ready to use.

W3id,[140] developed by the W3C Permanent Identifier Community Group, provides secure, permanent URL redirection services for web applications. As a result, the ontology maintains its stability and reliability, making it more appealing for reuse.

---

[134] https://essepuntato.it/lode/.
[135] https://github.com/VisualDataWeb/WebVOWL.
[136] https://dgarijo.github.io/Widoco/.
[137] https://www.w3.org/2001/sw/wiki/Category:Visualizer.
[138] https://essepuntato.it/graffoo/.
[139] https://www.cognitum.eu/semantics/FluentEditor/.
[140] https://w3id.org/.

7) Depositing the ontology

Store the ontology in multiple formats within certified repositories to guarantee easier access and long-term preservation.

DEPOSIT

By depositing your ontology in a repository such as Zenodo, it is automatically indexed in searchable resources. This allows a wider audience to find your ontology and possibly reuse it for their own modelling purposes.

Linked open data

Standards and guidelines

Linked data are another key building block of the semantic web and are empowered by the same technologies outlined in the Ontologies section, in particular RDF, SPARQL, JSON-LD, OWL, and SKOS. Another common specification is RDF-a (RDF in Attributes),[141] which allows the use of attributes within HTML5, XHTML and XML documents to express structured data.

The 5-star guideline[142] is the main guideline for linked open data publication, comprising five steps:

- *"make your stuff available on the Web (whatever format) under an **open licence***

- *make it available as **structured data** (e.g., Excel instead of image scan of a table)*

- *make it available in a **non-proprietary open format** (e.g., CSV instead of Excel)*

- ***use URIs** to denote things, so that people can point at your stuff*

- ***link your data to other data** to provide context"*

---

[141] https://www.w3.org/2001/sw/wiki/RDFa.
[142] https://5stardata.info/en/.

Recommendations

1) Reusing existing resources

Reuse existing resources, such as ontologies, taxonomies, thesauri, tools etc., clearly specifying them in the documentation.

IDENTIFY   **PRODUCE**

- Explore existing linked open datasets using the Linked Open Data Cloud.[143]

- Domain-relevant vocabularies: the Getty Vocabularies,[144] CLARIN Concept Registry,[145] DARIAH Vocabs services.[146]

- Domain-relevant models and ontologies, used on a national and international level: CIDOC-CRM, FOAF, Schema.org, Dublin Core, LRM_{OO}, DataCite, DCAT, SPAR Ontologies, and ArCO.

- To explore existing resources, use the LOV (Linked Open Vocabularies) database, BARTOC (Basic Registry of Thesauri, Ontologies and Classifications), EU vocabularies[147] or the H-SeTIS (Heritage – Semantic Tools and Interoperability Survey) database specific for the heritage domain.

2) Formalising and releasing the data model

Formalise and release the data model as an ontology, in case it comprises new classes and properties that may be useful for other scholars.

PRODUCE

In this case, follow the recommendations of the Ontologies section.

3) Linking data to external resources

Link your data extensively to external resources such as authority records and vocabularies to reconcile entities and harness the semantic web's potential.

PRODUCE

- Check records in existing wide-spread resources such as: Wikidata,[148] VIAF,[149] Open Library,[150] and WorldCat.[151]

---

[143]https://lod-cloud.net/.

[144]https://www.getty.edu/research/tools/vocabularies/index.html.

[145]https://www.clarin.eu/content/clarin-concept-registry.

[146]https://vocabs.dariah.eu/en/.

[147]https://op.europa.eu/en/web/eu-vocabularies/controlled-vocabularies.

[148]https://www.wikidata.org/wiki/Wikidata:Main_Page.

[149]https://viaf.org/.

[150]https://openlibrary.org/.

[151]https://search.worldcat.org/it.

4) Ensuring data provenance

Ensure data provenance, storing provenance information along with content data in order to prevent inconsistencies when integrating sources, emphasise content responsibility, and foster data credibility.

PRODUCE

- Provenance data can be modeled according to the PROV ontology.[152]

5) Providing the data model and the URI pattern

Provide the data model and URI pattern to facilitate data reuse.

DEPOSIT   **DISSEMINATE**

- Create persistent URIs, using services such as W3ID.

6) Depositing in certified and relevant repositories

Publish your data in certified repositories in different standard formats and register your linked open dataset in LOD-specific resources like the Linked Open Data Cloud or other domain-relevant repositories.

DEPOSIT

- Druid[153] repository for Linked Data.

7) Offering a schema and a user-friendly GUI

Offer a schema with accompanying search examples. Consider implementing a user-friendly GUI for data visualisation and navigation, catering to users unfamiliar with SPARQL.

DISSEMINATE

While a schema is the preferred tool for exploring a linked open dataset, a SPARQL graphical user interface (GUI) or a simpleGUI with browsing and searching functionalities can help a wider audience access and use your data.

- To easily create a graphical user interface for your schema, you can use tools such as Sparnatural,[154] that adapt to your knowledge graph ontology.

- For publishing your ontology, you can use visualisation tools such as Widoco, LodView,[155] and LODE. A comprehensive list of visualisation tools is available on the Bartoc website.[156]

---

[152]https://www.w3.org/TR/prov-o/.
[153]https://druid.datalegend.net/.
[154]https://sparnatural.eu/.
[155]https://lodview.it/.
[156]https://gbv.github.io/bartoc-vocabulary-software/.

From FAIR to top-notch: tips to further enhance research products' quality

While analysing the pilot research products of the ATLAS project, we identified common shortcomings in how the products were published and released to the audience, making it difficult for users to find key information such as the licence, and to understand how the product could be cited and used. These final recommendations are designed to help researchers overcome these problems and are applicable to all research product types, offering further tips that are not in the ALLEA guidelines.

Easy to find

- Choose a clear, descriptive title for your research product, including relevant keywords. If using an acronym, provide the full name for clarity.

- Create a landing page for the research product, where alongside the licence and other key metadata the following information is prominently displayed:

    - access points to the data;

    - version number;

    - status (e.g., completed, under development).

While all FAIR guidelines, including those of ALLEA, emphasise the importance of accurately licensing data and making it available to users, we believe that the information listed above is also essential to ensure easier access and the full findability of data. Furthermore, knowing the status of the research product (whether it is still under development, has been completed, or is no longer maintained) allows users to assess its usability.

You don't need to create a website if you can use existing platforms and solutions, such as the description associated with each record in Zenodo or a README file for GitHub repositories. Choose a platform that allows you to edit this information easily, so you can keep your landing page updated at all times.
If you wish to create a website, practical and low-cost solutions are GitHub Pages,[157] Google Sites,[158] and Wordpress.[159] Other tools are listed in TAPoR.

---

[157]https://pages.github.com/.
[158]https://www.newgooglesites.com/home.
[159]https://wordpress.org/.

Easy to cite

- Provide complete credits, including names, affiliations and persistent identifiers—when available—of creators, collaborators, and institutional partners. Include contact information for users to report issues or suggest collaborations.

- Enhance citability by providing citation guidelines.

- For products developed within a research project, describe the project or provide a link to its landing page, explaining how the product contributes to the project's goals. For time-limited projects, specify the end date and outline future plans for preservation and maintenance.

- Clearly state the licences of your research product's website/landing page, and all its by-products.

Easy to use

- In the documentation describe the research product's design and development process, including used tools and applied methodologies. If you have published this information as articles or development reports, provide links to these publications.

- Provide a user guide and examples demonstrating how to utilise the research product.

- To ensure efficient documentation management, maintain a single source document in a repository and generate different format versions from it as needed.
  Common tools for building documentation for a research project are Git-Book[160] and WikiMedia.[161]

Offer metadata and documentation in English (in addition to other languages) to reach a broader audience.

Quick Guide

Digital scholarly editions

1. Plan an editorial workflow keeping the FAIR principles in mind and taking advantage of existing tools.

2. Include photographic reproductions of the documents/witnesses, preferably through IIIF.

3. Consider creating a semantic edition to highlight people, places, and named entities.

---

[160]https://www.gitbook.com/.
[161]https://www.mediawiki.org/wiki/MediaWiki.

4. Use XML/TEI or another appropriate standard schema(e.g., Epidoc, MEI, CEI).

5. Create and publish custom encoding schemas, starting from existing ones.

6. Provide a detailed description of the textual tradition and the editorial process.

7. Give credit to all editors and contributors, including those responsible for markup and technical implementation.

8. Make your editorial process transparent by publishing different versions of the edition and its by-products.

9. Enable users to download the edition, including a downloadable version.

10. Store the edition in a certified repository and ensure it is indexed in relevant field-specific catalogues.

11. Use non-proprietary visualisation tools or adopt existing publishing solutions.

12. Provide indexes and a search function.

Text collections

1. Choose standard formats, such as XML/TEI or other appropriate schemas.

2. Always cite the sources used for text preparation by providing complete bibliographic references and links to descriptive web resources.

3. Link authors' and works' records to corresponding authority records.

4. In the documentation, specify collection criteria and editorial criteria.

5. Assign each text a persistent identifier.

6. Store the text collection in a certified repository and upload the latest version of the collection regularly.

7. Document changes and current status thoroughly, with a roadmap about the evolution of the text collection.

8. Facilitate text exploration through search functionalities, indexes and sub-collections.

Software tools

1. Review existing software tools and resources and explore opportunities to further develop or adapt them.

2. Involve domain experts in software design and apply software engineering methodologies and best practices.

3. Define and implement software integration strategies.

4. Employ standard and non-proprietary programming languages and technologies.

5. When possible, develop in open source and foster collaborative development.

6. Release software officially through freely accessible channels, providing detailed and user-friendly documentation.

7. Publish your released research software in a trusted scholarly repository with rich metadata and credit to the development team.

Ontologies

1. Reuse existing ontologies.

2. Engage the user community and domain experts throughout the design phase.

3. Adhere to shared methodologies for ontology design and implementation.

4. Foster interoperability with existing ontologies and facilitate integration through comprehensive mapping.

5. Provide users with detailed documentation, complete with practical usage examples and intuitive graphical representations of the ontology.

6. For URIs, use services for long-term stability and conduct regular maintenance checks.

7. Store the ontology in multiple formats within certified repositories.

Linked open data

1. Reuse existing resources, such as ontologies and taxonomies, clearly specifying them in the documentation.

2. Formalise and release the data model as an ontology, in case it comprises new classes and properties.

3. Link your data extensively to external resources such as authority records and vocabularies to reconcile entities.

4. Ensure data provenance, storing provenance information along with content data.

5. Provide the data model and URI pattern to facilitate data reuse.

6. Publish your data in certified repositories in different standard formats and register your linked open dataset in LOD-specific resources and domain-relevant repositories.

7. Offer a schema with accompanying search examples. Consider implementing a user-friendlyGUI for data visualisation and navigation.

From FAIR to top-notch

- Choose a clear, descriptive title for your research product.

- Create a landing page for the research product, where the access points to the data, the licence, version number, and the status are prominently displayed.

- Provide complete credits, including names, affiliations and persistent identifiers. Include contact information.

- Provide citation guidelines.

- For products developed within a research project, describe the project or provide a link to its landing page.

- Clearly state the licences of your research product's website/landing page, and all its by-products.

- Document the research product's design and development process, including used tools and applied methodologies.

- Provide a user guide and examples on how to utilise the research product.

- For efficient documentation management, maintain a single source document in a repository and generate different format versions from it as needed.

- Offer metadata and documentation in English.

Bibliography

[Alhudithi 2021] Alhudithi, Ella. 2021. 'Review of Voyant Tools: See through Your Text'. *Language Learning & Technology* 25 (3): 43–50.

[ALLEA 2023] ALLEA, All European Academies. 2023. *Recognising Digital Scholarly Outputs in the Humanities.* ALLEA. https://doi.org/10.26356/O UTPUTS-DH.

[Amdouni et al. 2022] Amdouni, Emna, Syphax Bouazzouni, and Clement Jonquet. 2022. 'O'FAIRe Makes You an Offer: Metadata-Based Automatic FAIRness Assessment for Ontologies and Semantic Resources'. *International Journal of Metadata, Semantics and Ontologies* 16 (1): 16–46. https://doi.org/10.1504/IJMSO.2022.131133.

[Amendola 2021] Amendola, Cristiano. 2021. 'Editoria Universitaria, Open Access e Nuove Frontiere Del Lavoro Umanistico: La Basilicata University Press e La Collana "Digital Humanities"'. *DigItalia* 16 (2): 135–42. https://doi.org/10.36181/digitalia-00042.

[Amico and Felicetti 2021] Amico, Nicola, and Achille Felicetti. 2021. 'Ontological Entities for Planning and Describing Cultural Heritage 3D Models Creation'. arXiv, June. https://doi.org/10.48550/arXiv.2106.07277.

[Ayris 2017] Ayris, P. 2017. 'Delivering the European Open Science Cloud (EOSC): Principle and Practice in Delivering Open Science'. In *In: LEARN Toolkit of Best Practice for Research Data Management. (Pp. 87-89). Leaders Activating Research Networks (LEARN) (2017).* Leaders Activating Research Networks (LEARN). https://doi.org/10.14324/000.learn.17.

[Barabucci et al. 2017] Barabucci, Gioele, Elena Spadini, and Magdalena Turska. 2017. *Data vs. Presentation. What Is the Core of a Scholarly Digital Edition?* Sidestone Press. https://serval.unil.ch/notice/serval:BIB_09C6C598108A.

[Barker et al. 2022] Barker, Michelle, Neil P. Chue Hong, Daniel S. Katz, et al. 2022. 'Introducing the FAIR Principles for Research Software'. *Scientific Data* 9 (1): 622. https://doi.org/10.1038/s41597-022-01710-x.

[Bartalesi et al. 2016] Bartalesi, Valentina, Paola Andriani, Daniele Metilli, Carlo Meghini, and Mirko Tavoni. 2016. 'DanteSources : Una Biblioteca

Digitale Delle Fonti Dantesche'. *Le Forme e La Storia : Rivista Di Filologia Moderna : IX, 1, 2016*, no. IX. https://doi.org/10.1400/256610.

[Bartalesi et al. 2018] Bartalesi, Valentina, Carlo Meghini, Daniele Metilli, Mirko Tavoni, and Paola Andriani. 2018. 'A Web Application for Exploring Primary Sources: The DanteSources Case Study'. *Digital Scholarship in the Humanities* 33 (4): 705–23. https://doi.org/10.1093/llc/fqy002.

[Barzaghi 2021] Barzaghi, Sebastian. 2021. 'National Edition of Aldo Moro's Works (RDF Dataset)'. Zenodo, October 25. https://doi.org/10.5281/ZENODO.5592157.

[Barzaghi et al. 2024] Barzaghi, Sebastian, Francesco Paolucci, Francesca Tomasi, and Fabio Vitali. 2024. 'KwicKwocKwac, a Tool for Rapidly Generating Concordances and Marking up a Literary Text'. Preprint, arXiv. https://doi.org/10.48550/ARXIV.2410.06043.

[Bekiari et al. 2024] Bekiari, Chryssoula, George Bruseker, Erin Canning, et al. 2024. 'Definition of the CIDOC Conceptual Reference Model (Version 7.1.3)'. February. https://cidoc-crm.org/Version/version-7.1.3.

[Berners Lee 2009] Berners Lee, Tim. 2009. 'Linked Data - Design Issues'. Accessed 11 January 2021. https://www.w3.org/DesignIssues/LinkedData.html.

[Bizer et al. 2009] Bizer, Christian, Tom Heath, and Tim Berners-Lee. 2009. 'Linked Data - The Story So Far'. *International Journal on Semantic Web and Information Systems* 5 (3): 1–22. https://doi.org/10.4018/jswis.2009081901.

[Bollini et al. 2016] Bollini, Andrea, Michele Mennielli, Susanna Mornati, and David T Palmer. 2016. 'IRIS: Supporting & Managing the Research Life-Cycle'. *Universal Journal of Educational Research* 4 (4): 738–43. https://doi.org/10.13189/ujer.2016.040410.

[Borgman 1999] Borgman, Christine L. 1999. 'What Are Digital Libraries? Competing Visions'. *Information Processing & Management* 35 (3): 227–43. https://doi.org/10.1016/S0306-4573(98)00059-4.

[Boschetti et al. 2021] Boschetti, Federico, Riccardo Del Gratta, Monica Monachini, Marina Buzzoni, Paolo Monella, and Roberto Rosselli Del Turco. 2021. '"Tea for Two": The Archive of the Italian Latinity of the Middle Ages Meets the CLARIN Infrastructure'. June 22, 37–46. https://doi.org/10.3384/ecp1805.

[Brancato et al. 2021] Brancato, Dario, Milena Corbellini, Paola Italia, Valentina Pasqual, and Roberta Priore. 2021. '*VaSto* : Un'edizione Digitale Interdisciplinare'. *Magazén*, no. 1 (June): JournalArticle_5057. https://doi.org/10.30687/mag/2724-3923/2021/03/006.

[Brogan 2003] Brogan, Martha L. 2003. *Survey of Digital Library Aggregation Services*. Digital Library Federation. http://old.diglib.org/pubs/dlf101/dlf 101.htm.

[Brügger and Finnemann 2013] Brügger, Niels, and Niels Ole Finnemann. 2013. 'The Web and Digital Humanities: Theoretical and Methodological Concerns'. *Journal of Broadcasting & Electronic Media* 57 (1): 66–80. https://doi.org/ 10.1080/08838151.2012.761699.

[Buzzoni and Rosselli Del Turco 2015] Buzzoni, Marina, and Roberto Rosselli Del Turco. 2015. *Verso un'edizione digitale dell'Editto di Rotari*. ITA. https: //iris.unito.it/handle/2318/1575103.

[Buzzoni and Rosselli Del Turco 2024a] Buzzoni, Marina, and Roberto Rosselli Del Turco. 2024a. *Edizione Diplomatico-Interpretativa Con Facsimile Digitale Dell'* Edictum *Rothari , Ivrea, Biblioteca Capitolare, XXXIV (5)*. Fondazione Università Ca' Foscari. https://doi.org/10.30687/978-88-6969- 889-7.

[Buzzoni and Rosselli Del Turco 2024b] Buzzoni, Marina, and Roberto Rosselli Del Turco. 2024b. *Edizione Diplomatico-Interpretativa Con Facsimile Digitale Dell'* Edictum Rothari *, Vercelli, Biblioteca e Archivio Capitolare, CLXXXVIII*. Fondazione Università Ca' Foscari. https://doi.org/10.30687 /978-88-6969-888-0.

[Carriero et al. 2019a] Carriero, Valentina Anita, Aldo Gangemi, Maria Letizia Mancinelli, et al. 2019. 'ArCo: The Italian Cultural Heritage Knowledge Graph'. In *The Semantic Web – ISWC 2019*, edited by Chiara Ghidini, Olaf Hartig, Maria Maleshkova, et al., vol. 11779. Springer International Publishing. https://doi.org/10.1007/978-3-030-30796-7_3.

[Castro et al. 2024] Castro, Leyla-Jael, Daniel Garijo, Dietrich Rebholz-Schuhmann, Dhwani Solanki, Research Data Alliance FAIR4ML Task Force et al. 2024. 'FAIR4ML Metadata Schema'. https://rda-fair4ml.github.io/ FAIR4ML-schema/release/0.1.0/index.html.

[Chapman et al. 2025] Chapman, Alison, Martin Holmes, Kaitlyn Fralick, Kailey Fukushima, Narges Montakhabi Bakhtvar, and Sonja Pinto. 2025. 'Browse, Search and Serendipity: Building Approachable Digital Editions'. In *Digital Editing and Publishing in the Twenty-First Century*, 1st ed., edited by James O'Sullivan, Michael Pidd, Sophie Whittle, Bridgette Wessels, Michael Kurzmeier, and Órla Murphy. Scottish Universities Press. https://doi.org/ 10.62637/sup.GHST9020.6.

[Chue Hong et al. 2021] Chue Hong, Neil, et al. 2021. *FAIR Principles for Research Software (FAIR4RS Principles)*. https://doi.org/10.15497/RDA0006 8.

[Chue Hong et al. 2023] Chue Hong, Neil, Elena Breitmoser, Mario Antonioletti, et al. 2023. *D5.2 - Metrics for Automated FAIR Software Assessment in a*

Disciplinary Context. October 27. https://doi.org/10.5281/ZENODO.10047 401.

[Ciula et al. 2023] Ciula, Arianna, Øyvind Eide, Cristina Marras, and Patrick Sahle. 2023. *Modelling Between Digital and Humanities: Thinking in Practice*. 1st ed. Open Book Publishers. https://doi.org/10.11647/OBP.0369.

[Cohen et al. 2021] Cohen, Jeremy, Daniel S. Katz, Michelle Barker, Neil Chue Hong, Robert Haines, and Caroline Jay. 2021. 'The Four Pillars of Research Software Engineering'. *IEEE Software* 38 (1): 97–105. https://doi.org/10.1 109/MS.2020.2973362.

[Courbebaisse et al. 2023] Courbebaisse, Guy, Bernd Flemisch, Kay Graf, Uwe Konrad, Jason Maassen, and Raphael Ritz. 2023. *Research Software Lifecycle*. September 7. https://doi.org/10.5281/ZENODO.8324828.

[Daquino 2021] Daquino, Marilena. 2021. 'Linked Open Data Native Cataloguing and Archival Description'. *JLIS*, no. 3. https://doi.org/10.4403/jlis.it-12703.

[Daquino et al. 2022] Daquino, Marilena, Ivan Heibi, Silvio Peroni, and David Shotton. 2022. 'Creating RESTful APIs over SPARQL Endpoints Using RAMOSE'. *Semantic Web* 13 (2): 195–213. https://doi.org/10.3233/SW-210439.

[Daquino et al. 2017] Daquino, Marilena, Francesca Mambelli, Silvio Peroni, Francesca Tomasi, and Fabio Vitali. 2017. 'Enhancing Semantic Expressivity in the Cultural Heritage Domain: Exposing the Zeri Photo Archive as Linked Open Data'. *Journal on Computing and Cultural Heritage* 10 (4): 1–21. https://doi.org/10.1145/3051487.

[Daquino and Tomasi 2015] Daquino, Marilena, and Francesca Tomasi. 2015. 'Historical Context Ontology (HiCO): A Conceptual Model for Describing Context Information of Cultural Heritage Objects'. In *Metadata and Semantics Research*, edited by Emmanouel Garoufallou, Richard J. Hartley, and Panorea Gaitanou, vol. 544. Springer International Publishing. https://doi.org/10.1007/978-3-319-24129-6_37.

[Daquino et al. 2023a] Daquino, Marilena, Mari Wigham, Enrico Daga, Lucia Giagnolini, and Francesca Tomasi. 2023. 'CLEF. A Linked Open Data Native System for Crowdsourcing'. *Journal on Computing and Cultural Heritage* 16 (3): 1–17. https://doi.org/10.1145/3594721.

[DDI Training Group 2023] DDI Training Group. 2023. 'Achieving FAIR with DDI'. November 1. https://doi.org/10.5281/zenodo.10064667.

[Del Grosso et al. 2016] Del Grosso, Angelo Mario, Davide Albanesi, Emiliano Giovannetti, and Simone Marchi. 2016. 'Defining the Core Entities of an Environment for Textual Processing in Literary Computing'. *Digital Humanities 2016: Conference Abstracts* (Kraków), 771–75. https://dh2016 .adho.org/abstracts/425.

[Del Grosso et al. 2017] Del Grosso, Angelo Mario, Emiliano Giovannetti, and Simone Marchi. 2017. 'The Importance of Being... Object-Oriented: Old Means for New Perspectives in Digital Textual Scholarship'. In *Advances in Digital Scholarly Editing*. Sidestone Press.

[Dillen 2020] Dillen, Wout. 2020. 'Lexicon of Scholarly Editing'. http://uahost.u antwerpen.be/lse/.

[Directorate-General for Research and Innovation 2018] Directorate-General for Research and Innovation (European Commission). 2018. *Prompting an EOSC in Practice: Final Report and Recommendations of the Commission 2nd High Level Expert Group on the European Open Science Cloud (EOSC), 2018*. Publications Office of the European Union. https://data.europa.eu/doi/10.2777/112658.

[Doerr 2005] Doerr, Martin. 2005. *The CIDOC CRM, an Ontological Approach to Schema Heterogeneity*. 5 pages, 175090 bytes. https://doi.org/10.4230/DAGSEMPROC.04391.22.

[Driscoll and Pierazzo 2016] Driscoll, Matthew James, and Elena Pierazzo, eds. 2016. *Digital Scholarly Editing: Theories and Practices*. 1st ed. Vol. 4. Digital Humanities Series. Open Book Publishers. https://doi.org/10.11647/OBP.0095.

[Eide 2014] Eide, Øyvind. 2014. 'Ontologies, Data Modeling, and TEI'. *Journal of the Text Encoding Initiative*, no. Issue 8 (December). https://doi.org/10.4000/jtei.1191.

[Euzenar and Shvaiko 2007] Euzenar, Jérôme, and Shvaiko. 2007. *Ontology Matching*. Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-49612-0.

[Evans 2004] Evans, Eric. 2004. *Domain-Driven Design: Tackling Complexity in the Heart of Software*. Addison-Wesley.

[Fenlon et al. 2017] Fenlon, Katrina, Jacob Jett, and Carole L. Palmer. 2017. 'Digital Collections and Aggregations'. *Digital Humanities Data Curation*. https://guide.dhcuration.org/contents/digital-collections-and-aggregations/.

[Ferrarini 2017] Ferrarini, Edoardo. 2017. 'ALIM ieri e oggi'. *Umanistica Digitale*, October 1, No 1 (2017). https://doi.org/10.6092/ISSN.2532-8816/7193.

[Filiposka et al. 2025] Filiposka, Sonja, Dominique Green, Anastas Mishev, et al. 2025. *FAIR-by-Design Methodology for Learning Materials*. January 21. https://doi.org/10.5281/ZENODO.14711524.

[Franzini et al. 2016] Franzini, Greta, Simon Mahony, and Melissa Terras. 2016. 'A Catalogue of Digital Edition'. In *Digital Scholarly Editing: Theories and Practices*, vol. 4. Digital Humanities Series. Open Book Publishers. https://doi.org/10.11647/OBP.0095.

[Freire et al. 2018] Freire, Nuno, Valentine Charles, and Antoine Isaac. 2018. 'Evaluation of Schema.Org for Aggregation of Cultural Heritage Metadata'. In *The Semantic Web*, edited by Aldo Gangemi, Roberto Navigli, Maria-Esther Vidal, et al., vol. 10843. Springer International Publishing. https://doi.org/10.1007/978-3-319-93417-4_15.

[Gamma et al. 1994] Gamma, Erich, Richard Helm, Ralph Johnson, and John Vlissides. 1994. *Design Patterns: Elements of Reusable Object-Oriented Software*. Pearson Education.

[García et al. 2017] García, Jordi Gràcia, Ilan Kernerman, and Julia Bosque Gil. 2017. 'Toward Linked Data-Native Dictionaries'. *Electronic Lexicography in the 21st Century: Proceedings of eLex 2017 Conference*, 550–59.

[Garijo and Poveda-Villalón 2020] Garijo, Daniel, and María Poveda-Villalón. 2020. 'Best Practices for Implementing FAIR Vocabularies and Ontologies on the Web'. Preprint, arXiv. https://doi.org/10.48550/ARXIV.2003.13084.

[Giacomini et al. 2025] Giacomini, Sebastiano, Marilena Daquino, Francesca Tomasi, and Laurent Antoine Fintoni. 2025. 'CLEF 2.0. Solutions for Native Linked Data Cataloguing of Italian Digital Cultural Heritage'. *JLIS.It* 16 (1): 108–26. https://doi.org/10.36253/jlis.it-611.

[Gonano et al. 2014] Gonano, Ciro Mattia, Francesca Tomasi, Francesca Mambelli, Fabio Vitali, and Silvio Peroni. 2014. 'Zeri e LODE. Extracting the Zeri Photo Archive to Linked Open Data: Formalizing the Conceptual Model'. *IEEE/ACM Joint Conference on Digital Libraries* (London, United Kingdom), September, 289–98. https://doi.org/10.1109/JCDL.2014.6970182.

[IFLA 2020] IFLA Functional Requirements for Bibliographic Records (FRBR) Review Group, Pat Riva, Patrick Le Boeuf, and Maja Žumer. 2020. *IFLA Library Reference Model: Un modello concettuale per le informazioni bibliografiche*. November. https://repository.ifla.org/handle/20.500.14598/44.

[Gruenpeter et al. 2020] Gruenpeter, Morane, Roberto Di Cosmo, Hylke Koers, et al. 2020. *M2.15 Assessment Report on 'FAIRness of Software'*. October 16. https://zenodo.org/records/4095092.

[Grünpeter et al. 2024] Grünpeter, Morane, Sabrina Granger, Alain Monteil, et al. 2024. *D4.4 - Guidelines for Recommended Metadata Standard for Research Software within EOSC*. March 6. https://doi.org/10.5281/ZENODO.10786147.

[Guarino et al. 2009] Guarino, Nicola, Daniel Oberle, and Steffen Staab. 2009. 'What Is an Ontology?' In *Handbook on Ontologies*, edited by Steffen Staab and Rudi Studer. Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-92673-3_0.

[Harrower et al. 2020] Harrower, Natalie, Maciej Maryl, Timea Biro, Beat Immenhauser and ALLEA Working Group E-Humanities. 2020. 'Sustainable

and FAIR Data Sharing in the Humanities: Recommendations of the AL-LEA Working Group E-Humanities'. Preprint, Digital Repository of Ireland. https://doi.org/10.7486/DRI.TQ582C863.

[Horik 2019] Horik, René Van. 2019. *The Research Data Alliance and the Humanities.* Zenodo. https://doi.org/10.5281/ZENODO.3355144.

[Hugo et al. 2020] Hugo, Wim, Yann Le Franc, Gerard Coen, Jessica Parland-von Essen, and Luiz Bonino. 2020. *D2.5 FAIR Semantics Recommendations Second Iteration.* December 21. https://doi.org/10.5281/ZENODO.4314320.

[ISO/IEC 2015] *ISO/IEC TR 12182:2015 - Systems and Software Engineering — Framework for Categorization of IT Systems and Software, and Guide for Applying It.* 2015. Version 2nd edition. https://www.iso.org/standard/636 11.html.

[Johnston and Robinson 2002] Johnston, Pete, and Bridget Robinson. 2002. *Collections and Collection Description.* Collection Description Focus Briefing Paper, 1. UKOLN. https://resources.ials.sas.ac.uk/calim/collections-and-collection-description.

[Kurzmeier et al. 2024] Kurzmeier, Michael, James O'Sullivan, Mike Pidd, Orla Murphy, and Bridgette Wessels. 2024. 'Visualising the Catalogues of Digital Editions'. *The Journal of Electronic Publishing* 26 (1). https://doi.org/10.3 998/jep.3569.

[LEARN 2017] LEARN. 2017. *LEARN Toolkit of Best Practice for Research Data Management.* https://doi.org/10.14324/000.learn.00.

[Leonelli 2023] Leonelli, Sabina. 2023. *Philosophy of Open Science.* 1st ed. Cambridge University Press. https://doi.org/10.1017/9781009416368.

[Mambrini and Passarotti 2023] Mambrini, Francesco, and Marco Carlo Passarotti. 2023. 'The LiLa Lemma Bank: A Knowledge Base of Latin Canonical Forms'. *Journal of Open Humanities Data* 9 (November): 28. https://doi.org/10.5334/johd.145.

[Mancinelli and Montefusco 2020] Mancinelli, Tiziana, and Antonio Montefusco. 2020. 'The BIFLOW-Toscana Bilingue Catalogue: A Digital Representation of the Socio-Cultural History of Translation in the Tuscan Middle Ages (1260–1430)'. *Textual Cultures* 13 (2). https://doi.org/10.14434/textual.v13 i2.31597.

[Mancinelli and Pierazzo 2020] Mancinelli, Tiziana, and Elena Pierazzo. 2020. *Che cos'è un'edizione scientifica digitale.* Carocci.

[Manghi et al. 2019] Manghi, Paolo, Alessia Bardi, Claudio Atzori, et al. 2019. *The OpenAIRE Research Graph Data Model.* Zenodo. https://doi.org/10.5 281/ZENODO.2643199.

[Mannocci and Baglioni 2023] Mannocci, Andrea, and Miriam Baglioni. 2023. "SKG-IF Documentation v. 1.0". https://skg-if.readthedocs.io/en/v1.0/index.html.

[Marden et al. 2013] Marden, Julia, Carolyn Li-Madeo, Noreen Whysel, and Jeffrey Edelstein. 2013. 'Linked Open Data for Cultural Heritage: Evolution of an Information Technology'. *Proceedings of the 31st ACM International Conference on Design of Communication*, September 30, 107–12. https://doi.org/10.1145/2507065.2507103.

[Martignano 2023] Martignano, Chiara. 2023. 'Un modello concettuale per le edizioni critiche digitali'. University of Siena. https://dx.doi.org/10.25434/chiara-martignano_phd2023.

[Martínková et al. 2024] Martínková, Jana, Nick Juty, Alejandra Gonzalez-Beltran, Carole Goble, and Yann Le Franc. 2024. 'Moving towards FAIR Mappings and Crosswalks.' Paper presented at FAIR principles for Ontologies and Metadata in Knowledge Management. July 8. https://research.manchester.ac.uk/en/publications/moving-towards-fair-mappings-and-crosswalks.

[Mastandrea 2007] Mastandrea, Paolo. 2007. *Musisque Deoque. Un archivio digitale di poesia latina, dalle origini al Rinascimento italiano.* https://iris.unive.it/handle/10278/30263.

[Matentzoglu et al. 2013] Matentzoglu, Nicolas, Samantha Bail, and Bijan Parsia. 2013. 'A Snapshot of the OWL Web'. In *Advanced Information Systems Engineering*, edited by David Hutchison, Takeo Kanade, Josef Kittler, et al., vol. 7908, edited by Camille Salinesi, Moira C. Norrie, and Óscar Pastor. Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-41335-3_21.

[Münster 2022] Münster, Sander. 2022. 'Digital 3D Technologies for Humanities Research and Education: An Overview'. *Applied Sciences* 12 (5): 2426. https://doi.org/10.3390/app12052426.

[Murano et al. 2023] Murano, Francesca, Valeria Quochi, Angelo Mario Del Grosso, Luca Rigobianco, and Mariarosaria Zinzi. 2023. 'Describing Inscriptions of Ancient Italy. The ItAnt Project and Its Information Encoding Process'. *Journal on Computing and Cultural Heritage* 16 (3): 1–14. https://doi.org/10.1145/3606703.

[Murano et al. 2024] Murano, Francesca, Luca Rigobianco, Valeria Quochi, and Angelo Mario Del Grosso. 2024. *ItAnt EpiDoc Schema.* September 30. https://zenodo.org/records/13862559.

[NISO 2013] National Information Standards Organization. 2013. *ANSI/NISO Z39.7-2013, Information Services and Use: Metrics and Statistics for Libraries and Information Providers Data Dictionary.* NISO. https://doi.org/10.3789/ansi.niso.z39.7-2013.

[Naveen et al. 2024] Naveen, Palanichamy, Rajagopal Maheswar, and Pavel Trojovský. 'GeoNLU: Bridging the Gap between Natural Language and Spatial

Data Infrastructures'. *Alexandria Engineering Journal* 87 (January 2024): 126–47. https://doi.org/10.1016/j.aej.2023.12.027.

[Niccolucci and Felicetti 2024] Niccolucci, Franco, and Achille Felicetti. 2024. 'Digital Twin Sensors in Cultural Heritage Applications'. *Sensors* 24 (12): 3978. https://doi.org/10.3390/s24123978.

[Noy n.d.] Noy, Natalya F. n.d. 'Ontology Development 101: A Guide to Creating Your First Ontology'. *Linked Data for Professional Education.* Accessed 15 April 2025. https://ld4pe.dublincore.org/learning_resource/ontology-development-101-a-guide-to-creating-your-first-ontology/.

[OCLC and Lavoie 2014] OCLC, and Brian Lavoie. 2014. *The Open Archival Information System (OAIS) Reference Model: Introductory Guide (2nd Edition).* Digital Preservation Coalition. https://doi.org/10.7207/twr14-02.

[Orlando et al. 2024] Orlando, Riccardo, Luca Moroni, Pere-Lluís Huguet Cabot, et al. 2024. 'Minerva LLMs: The First Family of Large Language Models Trained from Scratch on Italian Data'. In *Proceedings of the 10th Italian Conference on Computational Linguistics (CLiC-It 2024)*, edited by Felice Dell'Orletta, Alessandro Lenci, Simonetta Montemagni, and Rachele Sprugnoli. CEUR Workshop Proceedings. https://aclanthology.org/2024.clicit-1.77/.

[Pareti 2016] Pareti, Paolo. 'Human-Machine Collaboration over Linked Data.' Paper presented at the 15th International Semantic Web Conference (ISWC), 17-21 October 2016, Kobe. 2016. https://ceur-ws.org/Vol-1690/paper28.pdf.

[Passarotti 2023] Passarotti, Marco. 2023. *La Knowledge Base LiLa. Interoperabilità tra risorse testuali e lessicali per il latino.* June 24. https://doi.org/10.5281/ZENODO.8076839.

[Peroni and Shotton 2018] Peroni, Silvio, and David Shotton. 2018. 'The SPAR Ontologies'. In *The Semantic Web – ISWC 2018*, edited by Denny Vrandečić, Kalina Bontcheva, Mari Carmen Suárez-Figueroa, et al., vol. 11137. Springer International Publishing. https://doi.org/10.1007/978-3-030-00668-6_8.

[Peroni et al. 2013] Peroni, Silvio, Francesca Tomasi, and Fabio Vitali. 2013. 'The Aggregation of Heterogeneous Metadata in Web-Based Cultural Heritage Collections: A Case Study'. *International Journal of Web Engineering and Technology* 8 (4): 412. https://doi.org/10.1504/IJWET.2013.059107.

[Pierazzo 2014] Pierazzo, Elena. 2014. 'Digital Scholarly Editing: Theories, Models and Methods'. September 28. https://hal.univ-grenoble-alpes.fr/hal-01182162.

[Pierazzo 2019] Pierazzo, Elena. 2019. 'What Future for Digital Scholarly Editions? From Haute Couture to Prêt-à-Porter'. *International Journal of Digital Humanities* 1 (2): 209–20. https://doi.org/10.1007/s42803-019-00019-3.

[Poveda-Villalón et al. 2020] Poveda-Villalón, María, Paola Espinoza-Arias, Daniel Garijo, and Oscar Corcho. 2020. 'Coming to Terms with FAIR Ontologies'. In *Knowledge Engineering and Knowledge Management*, edited by C. Maria Keet and Michel Dumontier, vol. 12387. Springer International Publishing. https://doi.org/10.1007/978-3-030-61244-3_18.

[Quondam 2021] Quondam, Amedeo. 2021. 'Memorie per una storia dell'italianistica digitale: «Biblioteca italiana»'. *Griseldaonline*, December 15, 137-147 Paginazione. https://doi.org/10.6092/ISSN.1721-4777/12360.

[Ramsay and Rockwell 2012] Ramsay, Stephen, and Geoffrey Rockwell. 2012. 'Developing Things: Notes toward an Epistemology of Building in the Digital Humanities'. In *Debates in the Digital Humanities*, edited by Matthew K. Gold. University of Minnesota Press. https://doi.org/10.5749/minnesota/9780816677948.003.0010.

[RDA/FORCE11 2020] Research Data Alliance/FORCE11 Software Source Code Identification WG, Alice Allen, Anita Bandrowski, et al. 2020. *Use Cases and Identifier Schemes for Persistent Software Source Code Identification (V1.0)*. https://doi.org/10.15497/RDA00053.

[Rettberg and Schmidt 2012] Rettberg, Najla, and Birgit Schmidt. 2012. 'OpenAIRE - Building a Collaborative Open Access Infrastructure for European Researchers'. *LIBER Quarterly: The Journal of the Association of European Research Libraries* 22 (3): 160–75. https://doi.org/10.18352/lq.8110.

[Roelli and Macé 2015] Roelli, Philipp, and Caroline Macé. 2015. "Parvum Lexicon Stemmatologicum. A Brief Lexicon of Stemmatology". https://doi.org/10.5167/UZH-121539.

[Rosselli Del Turco 2019] Rosselli Del Turco, Roberto. 2019. 'Designing an Advanced Software Tool for Digital Scholarly Editions: The Inception and Development of EVT (Edition Visualization Technology)'. *Textual Cultures: Texts, Contexts, Interpretation* 12 (2): 91–111.

[Rosselli Del Turco et al. 2019] Rosselli Del Turco, Roberto, Chiara Di Pietro, and Chiara Martignano. 2019. 'Progettazione e implementazione di nuove funzionalità per EVT 2: lo stato attuale dello sviluppo'. *Umanistica Digitale*, December, No 7 (2019). https://doi.org/10.6092/ISSN.2532-8816/9322.

[Sahle 2016] Sahle, Patrick. 2016. 'What Is a Scholarly Digital Edition?' In *Digital Scholarly Editing*, 1st ed., edited by Matthew James Driscoll and Elena Pierazzo, vol. 4. Theories and Practices. Open Book Publishers. https://www.jstor.org/stable/j.ctt1fzhh6v.6.

[Salvatori et al. 2017] Salvatori, Enrica, Roberto Rosselli Del Turco, Chiara Alzetta, Chiara Di Pietro, Chiara Mannari, and Alessio Miaschi. 2017. 'Il Codice Pelavicino tra edizione digitale e Public History'. *Umanistica Digitale*, October 1, No 1 (2017). https://doi.org/10.6092/ISSN.2532-8816/7232.

[Scarpa and Valente 2024] Scarpa, Erica, and Riccardo Valente. 2024. 'Heritage – Semantic Tools and Interoperability Survey'. Preprint, Heritage – Semantic Tools and Interoperability Survey. https://doi.org/10.71795/8T2Z-HH65.

[Schiltz et al. 2025] Schiltz, Arthur, Emidio Stani, and Ine Weyts. 2025. 'MLDCAT-AP (Machine Learning DCAT Application Profile)'. Application Profile. Version 2.0.0. February 18. https://semiceu.github.io/MLDCAT-AP/releases/2.1.0/.

[Schreibman and Hanlon 2010] Schreibman, Susan, and Ann M. Hanlon. 2010. 'Determining Value for Digital Humanities Tools: Report on a Survey of Tool Developers'. *Digital Humanities Quarterly* 004 (2).

[Silén 2024] Silén, Petri. 2024. *Clean Code Principles and Patterns, 2nd Edition.* Leanpub. http://leanpub.next/cleancodeprinciplesandpatterns2ndedition.

[Soiland-Reyes et al. 2022] Soiland-Reyes, Stian, Peter Sefton, Mercè Crosas, et al. 2022. 'Packaging Research Artefacts with RO-Crate'. *Data Science* 5 (2): 97–138. https://doi.org/10.3233/DS-210053.

[Spadini and Losada Palenzuela 2025] Spadini, Elena, and José Luis Losada Palenzuela. 2025. 'Re-Using Data from Editions'. In *Digital Editing and Publishing in the Twenty-First Century*, 1st ed., edited by James O'Sullivan, Michael Pidd, Sophie Whittle, Bridgette Wessels, Michael Kurzmeier, and Órla Murphy. Scottish Universities Press. https://doi.org/10.62637/sup.GHST9020.8.

[SSHOC 2022] SSHOC. 2022. *SSHOCingly Good and Sustainable Tools.* Zenodo. https://doi.org/10.5281/ZENODO.6394622.

[Suárez-Figueroa et al. 2012] Suárez-Figueroa, Mari Carmen, Asunción Gómez-Pérez, and Mariano Fernández-López. 2012. 'The NeOn Methodology for Ontology Engineering'. In *Ontology Engineering in a Networked World*, edited by Mari Carmen Suárez-Figueroa, Asunción Gómez-Pérez, Enrico Motta, and Aldo Gangemi. Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-24794-1_2.

[Szleszyński et al. 2025] Szleszyński, Bartłomiej, Agnieszka Szulińska, and Marta Błaszczyńska. 2025. 'Predicting the Future of Digital Scholarly Editions in the Context of FAIR Data Principles'. In *Digital Editing and Publishing in the Twenty-First Century*, 1st ed., edited by James O'Sullivan, Michael Pidd, Sophie Whittle, Bridgette Wessels, Michael Kurzmeier, and Órla Murphy. Scottish Universities Press. https://doi.org/10.62637/sup.GHST9020.7.

[CodeMeta] 'The CodeMeta Project'. n.d. Accessed 20 June 2025. https://codemeta.github.io/.

[Tomasi 2022] Tomasi, Francesca. 2022. *Organizzare la conoscenza: Digital Humanities e Web semantico. Un percorso tra archivi, biblioteche e musei.* https://cris.unibo.it/handle/11585/848605.

[Tomasi et al. 2013] Tomasi, Francesca, Fabio Ciotti, Maurizio Lana, Fabio Vitali, Silvio Peroni, and Diego Magro. 2013. *Dialogue and Linking between TEI and Other Semantic Models*. ITA. https://cris.unibo.it/handle/11585/185113.

[UNESCO 2021] UNESCO. 2021. *UNESCO Recommendation on Open Science*. UNESCO. https://doi.org/10.54677/MNMH8546.

[Venuti et al. 2023] Venuti, Martina, Angelo Mario Del Grosso, Federico Boschetti, et al. 2023. 'La 'Galassia MQDQ:' Un Concetto Di Filologia Tradizionale, Digitale, Sostenibile'. *Magazén*, no. 1 (August): JournalArticle_12673. https://doi.org/10.30687/mag/2724-3923/2023/07/003.

[Vogt et al. 2024] Vogt, Lars, Philip Strömert, Nicolas Matentzoglu, et al. 2024. *FAIR 2.0: Extending the FAIR Guiding Principles to Address Semantic Interoperability*. https://doi.org/10.48550/ARXIV.2405.03345.

[Welsh 2014] Welsh, Megan. 2014. 'Review of Voyant Tools'. *Collaborative Librarianship* 6 (2). https://digitalcommons.du.edu/collaborativelibrarianship/vol6/iss2/8.

[Wilkinson et al. 2016] Wilkinson, Mark D., Michel Dumontier, IJsbrand Jan Aalbersberg, et al. 2016. 'The FAIR Guiding Principles for Scientific Data Management and Stewardship'. *Scientific Data* 3 (1): 160018. https://doi.org/10.1038/sdata.2016.18.

[Zenzaro 2024] Zenzaro, Simone. 2024. 'Models for Digital Humanities Tools: Coping with Technological Changes and Obsolescence'. *International Journal of Information Science and Technology* 8 (2): 1–10. https://doi.org/10.57675/IMIST.PRSM/ijist-v8i2.283.

[Zenzaro et al. 2025] Zenzaro, Simone, Federico Boschetti, and Angelo Mario Del Grosso. 2025. 'Making Digital Scholarly Editions Based on Domain Specific Languages'. In *Digital Editing and Publishing in the Twenty-First Century*, 1st ed., edited by James O'Sullivan, Michael Pidd, Sophie Whittle, Bridgette Wessels, Michael Kurzmeier, and Órla Murphy. Scottish Universities Press. https://doi.org/10.62637/sup.GHST9020.9.

[Almas and Beaulieu 2013] Almas, Bridget, and Marie-Claire Beaulieu. 'Developing a New Integrated Editing Platform for Source Documents in Classics'. *Literary and Linguistic Computing* 28, no. 4 (2013): 493–503. https://doi.org/10.1093/llc/fqt046.

[Shillingsburg 2015] Shillingsburg, Peter. 2014. 'Development Principles for Virtual Archives and Editions'. In *The Journal of the European Society for Textual Scholarship* . Brill. https://doi.org/10.1163/9789401212113_002.

[Robinson and Bordalejo 2016] Robinson, Peter, and Barbara Bordalejo. 2016. 'Textual Communities.' In *Digital Humanities 2016: Conference Abstracts*, 876-877. Kraków: Jagiellonian University and Pedagogical University. https://dh2016.adho.org/abstracts/384.

[Ciotti 2014] Ciotti, Fabio. 2014. 'Digital Literary and Cultural Studies: State of the Art and Perspectives.' Between 4. 8. https://art.torvergata.it/handle /2108/123046.

[Schmidt2014] Schmidt, Desmond. 2014. 'Towards an Interoperable Digital Scholarly Edition.' *Journal of the Text Encoding Initiative*, no. Issue 7 (November). https://doi.org/10.4000/jtei.979.

[Ackerman and Gonzalez 2011] Ackerman, Lee, and Celso Gonzalez. 2011. *Patterns-Based Engineering: Successfully Delivering Solutions Via Patterns.* Addison-Wesley.

Glossary

ADHO Alliance of Digital Humanities Organizations. 164

API (Application Programming Interface), it is a set of rules and protocols that lets one program request services or data from another via a standard protocol, usually HTTP. 12, 16, 18, 22, 26, 27, 29, 30, 53, 72, 78, 86, 115, 118, 122, 124, 126, 130, 148, 159, 163

authority refers to a system that provides authority control—a way to ensure consistency and reliability in naming and identifying concepts, entities, or resources across datasets. 150

authority record a structured data entry that uniquely identifies and describes an entity—such as a person, organization, place, or concept—using standardised formats and persistent identifiers. Its purpose is to ensure consistency, disambiguation, and interoperability across datasets. 39, 41, 44, 62, 98, 120, 159, 173, 177

branching in version control systems, it is a technique that allows developers to diverge from the main codebase to work on changes independently. 166

changelog a file or document that records all the notable changes made to a product over time. 16, 27, 160, 166

CSV (Comma-Separated Values), open and standard text format used to store tabular data, like spreadsheets. 53, 57, 124, 155, 172

design pattern in software engineering, it is a reusable solution to a common problem that arises during software design. 164

DevOps combination of Development and Operations, it is a cultural and technical movement in software engineering that aims to bridge the gap between software developers and IT operations teams. 165

DOI (Digital Object Identifier), a unique, persistent code for digital scholarly content, such as articles, reports and datasets. It acts as a permanent ID,

195

ensuring that the content can always be found and cited, even if its web address (URL) changes.. 30, 42, 43, 46, 68, 83, 97, 133, 146

domain expert a person with deep knowledge and experience in the specific field or subject area that a software is being built for. 48, 65, 164, 169, 178

domain-driven design a software development approach that focuses on deeply understanding the real-world area the software is meant to serve and modeling it directly in the code. 164

Domain-Specific Language (DSL) in software engineering, it is a programming or scripting language designed to solve problems in a specific domain—a narrow area of interest—rather than being general-purpose like Python or Java. 150

Dublin Core a widely used metadata standard for describing various types of digital and physical resources. Dublin Core is also an ISO-standard. 38, 46, 47, 72, 74, 155, 158, 173

entity in the Semantic Web, an entity is a uniquely identifiable thing that is described using structured data so that machines can understand and reason about it. 54, 117, 119, 124, 126, 131

FAIR principles guidelines for making data findable (F), accessible (A), interoperable (I) and reusable (R). 3, 14, 59, 64, 138

FRBR (Functional Requirements for Bibliographic Records), a conceptual model developed by the International Federation of Library Associations and Institutions (IFLA) to help organize and retrieve bibliographic information more effectively. The latest version of this model was renamed Library Reference Model (LRM). 78

GitHub a web-based platform that helps developers collaborate on code, track changes, and manage software projects using Git. 18, 38, 46, 50–52, 59, 67, 109, 115, 117, 166

GUI (Graphical User Interface), a visual way for users to interact with software using elements like windows, buttons, and icons, rather than typing commands in a text-based interface. 116, 174, 179

HAL (Hyper Articles en Ligne), an open-access repository based in France used by researchers to share publications. 166

IIIF (International Image Interoperability Framework), a set of open standards that enables interoperable access to high-resolution images and other media across institutions, platforms, and tools. 24, 26, 33, 38, 40, 41, 148, 149, 155, 176

indexing refers to the process of creating data structures that improve the speed and efficiency of data retrieval. 133

IRI (Internationalized Resource Identifier), a standardized way to identify resources on the internet, similar to a URI (Uniform Resource Identifier), but with support for a wider range of characters. 64, 68, 69, 73

JSON (JavaScript Object Notation), a text-based format for storing and sharing data, especially used on the web. 47, 124, 130, 153, 155, 167

mapping in the Semantic Web, it refers to the process of connecting or aligning different data models, vocabularies, or ontologies so that they can work together seamlessly. 12, 56, 63, 66, 71, 73, 135, 170, 178

markup in computer technology, it refers to a system of symbols or tags inserted into text to define its structure, presentation, or meaning. 19, 44, 63, 147, 152, 177

NLP (Natural Language Processing), a field of artificial intelligence (AI) that focuses on enabling computers to understand, interpret, generate, and respond to human language—whether spoken or written. 59, 155, 166

OAI-PMH (Open Archives Initiative Protocol for Metadata Harvesting), a low-barrier protocol that enables interoperability between repositories by allowing service providers to harvest structured metadata from data providers using six standardized HTTP-based operations. 132

object-oriented programming a programming paradigm that organizes software design around objects—self-contained units that combine data and behavior. 48, 164

OCR (Optical Character Recognition), a technology that enables computers to read and convert text from images, scanned documents, or handwritten notes into machine-readable data. 159

ODT (Open Document Text), an open and standard file format for word processing documents. 153

ORCID (Open Researcher and Contributor ID), is a unique identifier that individuals can use when engaging in research and scholarly activities. Provided by the eponymous organisation, these identifiers enable transparent and trustworthy connections between researchers, their contributions and their affiliations. 46, 74, 81, 82, 120, 133, 146

PID (persistent identifier), a long-lasting reference to a digital object or resource. It ensures that even if the location changes, the identifier remains stable. Examples are: DOI (Digital Object Identifier), Handle, ARK. 146, 152, 159

Python a high-level, general-purpose programming language widely used in web development, data science, AI, automation, and digital humanities. 115, 130, 166

README file a text file included in software projects, to explain what the project does, how to install and use it, and who maintains it. It is often written in Markdown and displayed automatically on platforms like GitHub. 50, 175

RIDE a review journal for Scholarly Digital Editions and other digital resources. 158

schema is a file that defines rules for structuring and validating data, especially in XML. It ensures that data conforms to expected formats. 21, 23–27, 33, 38, 39, 41, 46, 47, 56, 58, 63, 111, 115, 116, 118, 122, 124, 126, 127, 129, 150, 151, 155, 174, 177, 179

semantic edition a digital scholarly edition enriched with semantic annotations. It uses technologies like ontologies, and Linked Data to make texts machine-readable and interoperable. Usually, it also enables advanced search, visualization, and integration with other datasets. 150, 176

SKOS (Simple Knowledge Organization System), a W3C standard used to represent semi-structured knowledge, such as thesauri, taxonomies, and classification schemes, in a machine-readable format. 123, 168, 172

SPARQL endpoint a web-accessible interface that allows users to query RDF data using the SPARQL language. 56, 58, 61, 62, 78, 86

taxonomy a structured classification system that organizes concepts into hierarchical relationships, typically using broader and narrower terms. It helps machines and humans understand how ideas relate to one another. 76

TEI (Text Encoding Initiative), a consortium and a set of guidelines for representing texts in digital form, especially for scholarly research in the humanities. 18, 21, 25, 26, 33–35, 38–41, 44, 46, 50, 147, 148, 151, 152

TXT refers to a plain text file and is one of the simplest and most universal file formats used in computing. 20, 153

UML (Unified Modeling Language), a world-wide used standard for representing information systems. UML provides a set of diagram types, such as class diagrams, that model different aspects of a system. 171

Unicode a universal character encoding standard that assigns a unique number to every character across languages and scripts. 155

URI (Uniform Resource Identifier), is a string of characters used to uniquely identify a resource on the internet or within a system. 53, 54, 56, 58–61, 63, 66, 69, 86, 88, 89, 94–96, 109–113, 124, 130, 131, 171, 172, 174, 178, 179

versioning the practice of tracking and managing changes to digital content or code over time. 18, 161, 166

VIAF (Virtual International Authority File), a global authority file that aggregates name records from national libraries. 56, 63, 98, 100, 101, 105, 106, 120, 123, 124, 146, 159, 173

viewer type of software application conceived to visualise and interact with images, 3D models or other types of digital objects. 33, 149

vocabulary in the semantic web, it refers to a controlled set of terms used to describe and classify data consistently. 43, 54, 63, 64, 72, 75, 85–90, 95

W3C (World Wide Web Consortium), the main international standards organization for the web. 53, 56–58, 60, 63, 69, 72, 74, 165, 167, 170, 171

W3ID a service that provides permanent, stable URLs for web applications, especially those using Linked Data and semantic web technologies. 83, 174

Wikidata a collaborative knowledge base that stores structured data for Wikimedia projects and beyond. 43, 56, 98, 115, 120, 123, 124, 159, 173

XML (Extensible Markup Language), a markup language used to structure and encode data in a hierarchical format. 18, 20, 33, 35, 51, 66, 124, 146, 147, 150, 167, 172

XSLT (Extensible Stylesheet Language Transformations), a language for transforming XML documents into other formats such as HTML and RDF. 49, 52

Zenodo an open-access repository for research outputs, including datasets, publications, and software. 26, 43, 44, 59, 60, 73, 75, 115, 134, 136, 141, 152, 153, 159, 166, 172, 175

ATLAS Outputs

Digital Resources

Resource	Identifier / URL
Knowledge Graph • Download all versions	http://hdl.handle.net/20.500.11752/ILC-1040 https://doi.org/10.5281/zenodo.14058143
Ontology • Download v. 2.0 • Download v. 1.0	http://www.w3id.org/dh-atlas/ https://doi.org/10.5281/zenodo.16780282 https://doi.org/10.5281/zenodo.14058231
Edition Type Vocabulary • Download	https://lab.dharc.unibo.it/skosmos/etv/en/ https://doi.org/10.5281/zenodo.16883994
Mapping	https://doi.org/10.5281/zenodo.16780282
Web Application • Source code	https://projects.dharc.unibo.it/atlas/ https://doi.org/10.5281/zenodo.15482297
SPARQL endpoint	https://projects.dharc.unibo.it/atlas/sparql
API • Source code	https://atlas-prin2022-api.cloud.d4science.org/api/v1 https://doi.org/10.5281/zenodo.17361169
Whitebook previous versions	https://doi.org/10.5281/zenodo.14169356
Guidelines • Download and website source code	https://dh-atlas.github.io/guidelines/ https://doi.org/10.5281/zenodo.17167252

Resource	Identifier / URL
ATLAS VRE	https: //services.d4science.org/group/atlas/home
OpenAIRE Gateway	https://dh-ch.openaire.eu/
OAI-PMH Source code	https://doi.org/10.5281/zenodo.17360920
Project Website • Source code	https://dh-atlas.github.io/ https://doi.org/10.5281/zenodo.13993056
ATLAS Workshop Survey	https://doi.org/10.5281/zenodo.15173473
ATLAS Final Workshop Survey	https://doi.org/10.5281/zenodo.17332837
ATLAS Seminar Survey	https://doi.org/10.5281/zenodo.17332765

Publications

Bardi, Alessia, Marina Buzzoni, Marilena Daquino, Riccardo Del Gratta, Angelo Mario Del Grosso, Roberto Rosselli Del Turco, Franz Fischer, Sebastiano Giacomini and Francesca Tomasi. 'The ATLAS of Italian Digital Humanities: a knowledge graph of digital scholarly research on Italian Cultural Heritage', in *Workflows: Digital Methods for Reproducible Research Practices in the Arts and Humanities*, Lisbon, Portugal. DARIAH, 2024. https: //doi.org/10.5281/zenodo.11569280.

Daquino, Marilena, Alessia Bardi, Marina Buzzoni, Riccardo Del Gratta, Angelo Mario Del Grosso, Franz Fischer, Francesca Tomasi, and Roberto Rosselli Del Turco. 'The ATLAS: A Knowledge Graph of Digital Scholarly Research on Italian Cultural Heritage', in *Me.Te. Digitali. Mediterraneo in rete tra testi e contesti*. Catania: AIUCD, 2024. https://doi.org/10.6092/unibo/amsacta/7927.

Giacomini, Sebastiano, Alessia Bardi, Marina Buzzoni, Marilena Daquino, Riccardo Del Gratta, Angelo Mario Del Grosso, Franz Fischer, et al. 'ATLAS: Towards a Knowledge Graph of International Scholarly Research on the Italian Digital Cultural Heritage'. In CEUR Workshop Proc., Vol. 3937. CEUR-WS, 2025. https://ceur-ws.org/Vol-3937/paper4.pdf.

Martignano, Chiara, Giorgia Rubin, Sebastiano Giacomini, Alessia Bardi, Marina Buzzoni, Marilena Daquino, Riccardo Del Gratta, et al. 'ATLAS: A Data Model for Describing FAIR Digital Humanities Research Outcomes'. In *Diversità, Equità e Inclusione: Sfide e Opportunità per l'Informatica Umanistica Nell'Era Dell'Intelligenza Artificiale*. Verona: AIUCD, 2025. https://doi.org/https: //doi.org/10.6092/unibo/amsacta/8380.