THE ATLAS OF ITALIAN DIGITAL CULTURAL HERITAGE

Guidelines and best practices for creating FAIR research products in the Digital Humanities

Finanziato dall'Unione europea NextGenerationEU

Ministero dell'Università e della Ricerca

Italiadomani
PIANO NAZIONALE DI RIPRESA E RESILIENZA

Authors

Alessia Bardi (alessia.bardi@isti.cnr.it)
Marina Buzzoni (mbuzzoni@unive.it)
Marilena Daquino (marilena.daquino2@unibo.it)
Riccardo Del Gratta (riccardo.delgratta@ilc.cnr.it)
Angelo Mario Del Grosso (angelo.delgrosso@ilc.cnr.it)
Roberto Rosselli Del Turco (roberto.rosselli@unito.it)
Franz Fischer (franz.fischer@unive.it)
Sebastiano Giacomini (sebastiano.giacomin2@unibo.it)
Chiara Martignano (chiara.martignano@unive.it)
Giorgia Rubin (giorgia.rubin@ilc.cnr.it)
Francesca Tomasi (francesca.tomasi@unibo.it)

Contributors

Laurent Fintoni

Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA

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 and corpora, 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 FAIRness 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.

Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA

Table of contents

Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA

Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA

Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA

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 selected DH research projects on Italian Heritage and their outcomes. The results of this analysis are thoroughly presented in the ATLAS whitebook,[2] a key project outcome. Other outputs include:

- the ATLAS ontology,[3] which maps schemas and ontologies from pilot projects to enhance data and tool interoperability;
- a knowledge graph[4] documenting DH projects and scholarly data on Italian Cultural Heritage, accessible through the ATLAS web application[5] 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 the **whitebook** and the following publications:

- 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.

[1] https://dh-atlas.github.io/.
[2] https://zenodo.org/doi/10.5281/zenodo.14925266.
[3] https://w3id.org/dh-atlas/.
[4] https://doi.org/10.5281/zenodo.14058143.
[5] https://projects.dharc.unibo.it/atlas/.

Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA

- 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.

Existing recommendations and references for implementing FAIR data

The concept of "FAIR data" and its principles were developed within FORCE11[6]—an international community of scholars, librarians, archivists, publishers, and research funders—and were first introduced in 2016 in the article "**The FAIR Guiding Principles for scientific data management and stewardship**."

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.

Wilkinson, Mark D., Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, et al. '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.

[6] https://force11.org/.

Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA

Nearly a decade later, the concept of FAIR data has spread widely and gained worldwide recognition in the scientific community. Data FAIRness has become a key requirement for publicly funded research projects. The **Horizon2020** Programme, for instance, has developed specific **guidelines for FAIR data management**.

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** 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**, 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** 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** (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 developed by the Research Data Alliance**,[7] which covers topics such as data citation, certification of repositories, metadata management, and interoperability. **OpenAIRE's[8] quick guide on making FAIR data**, 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.

[7] 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/.

[8] 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/.

Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA

PARTHENOS'[9] **Guidelines to FAIRify data management and make data reusable** 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" 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 "FAIR data" 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, and OPERAS. Below, we present each of them with brief references to their key outputs related to FAIR principles implementation.

ARIADNE Research Infrastructure (ARIADNE RI)
is an international non-profit organisation founded in 2022, continuing the work of ARIADNEplus and ARIADNE. With over 25 founding members, ARIADNE RI promotes digital techniques in archaeology and heritage, develops tools, creates and integrates research archives, maintains infrastructure, and supports communication. As the largest archaeological network globally, it connects repositories, heritage

- **Catalogue**, main point of access for searching and browsing archeological datasets.
- **Reference model**, built on CIDOC-CRM and composed of multiple formal ontologies, such as CRMarcheo.
- **Training hub**, in particular the course "Applying open/the Fair Principles to archaeology".
- **Complete list of services**, including
  - Data Management Plan Tool

[9] 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/.

Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA

bodies, museums, and research institutions while fostering collaboration between archaeologists and information scientists.	and Templates ○ ARIADNEplus Lab, which also offers a Linked Open Data (GraphDB) catalogue ○ Visual Media Service for easy publication and presentation on the web of complex visual media assets.
CLARIN (Common Language Resources and Technology Infrastructure) is a digital infrastructure that provides language data and tools for humanities and social sciences research, including text, audio, and video resources. Established in 2012, CLARIN is European Research Infrastructure Consortium and a distributed digital infrastructure that connects European research centers and enables resource sharing across locations.	• **Centres providing a certified repository** for data deposit including support for persistent identifiers. • **Virtual Language Observatory (VLO)**, a user-friendly interface to search language resources across domains. • **Language Resource Switchboard**, online tools for finding language processing tools for your data. • **Recommendations on licences** • **Recommendations on formats and standards**
DARIAH (Digital Research Infrastructure for the Arts and Humanities) is a European network supporting digital arts and humanities research by connecting people, expertise, and technologies across member countries. Established in 2014 as a European Research Infrastructure Consortium, it now has 22 Members and several Cooperating Partners. DARIAH strengthens digital humanities research by providing tools, training, and infrastructure while preserving traditional humanities approaches in the digital age.	• **Catalogue of tools and services**. • **DARIAH-Campus**, a platform to access DARIAH and DARIAH-affiliated offerings in training and education. • **DARIAH Teach**: community-driven and multilingual learning and teaching materials.
E-RIHS (European Research Infrastructure for Heritage Science) Initiated in 1999 and established as a European Research Infrastructure Consortium (ERIC) in 2025, E-RIHS supports heritage science research in Europe. It operates on two levels, with a central hub and the national nodes, providing services and training.	• **Catalogue of services**. • **Training camps.**
OPERAS (Open Scholarly Communication in the European Research Area for Social Sciences and Humanities) is a Research Infrastructure that supports open scholarly communication in social sciences and humanities (SSH) across Europe. It coordinates resources to address scholarly communication needs of European SSH researchers, aiming to make Open Science	• **GoTriple**, a multilingual platform for social sciences and humanities that centralises access to publications, data, projects, and researcher profiles from diverse sources, making them discoverable and reusable. • **TRIPLE Ontologies**.

Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA

accessible to all without barriers.	• **TRIPLE Open Science Training Series**, including "TRIPLE Training on FAIR Data in SSH". • **OPERAS Common Standards White Paper, June 2021**.

Open science and the Social Sciences and Humanities Open Cloud (SSHOC)

Several FAIR principles—particularly accessibility and reuse—align with open science principles. While FAIR guidelines advocate making data "as open as possible and as closed as necessary," open science emphasizes collaborative work and comprehensive sharing of knowledge, tools, and results throughout the research process. The goal is to publish scientific outputs—from methodologies to final results—in ways that enable both access and free reuse by others. We provide references below for those wishing to explore open science further.

- UNESCO Recommendation on Open Science
- Leonelli, Sabina. *Philosophy of Open Science*. 1st edn. Cambridge University Press, 2023. https://doi.org/10.1017/9781009416368.

Finally, we briefly present the **Social Sciences & Humanities Open Cloud (SSHOC)**, an EU Horizon 2020 project uniting 47 partners to develop the SSH area of the **European Open Science Cloud (EOSC)**. The project includes research infrastructures—like CLARIN and DARIAH—and libraries with expertise spanning the full data lifecycle. From 2019-2022, 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**, an inventory of training materials relevant for the Social Sciences and Humanities;
- **the SSH Open Marketplace**, 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.

Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA

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**.

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 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.

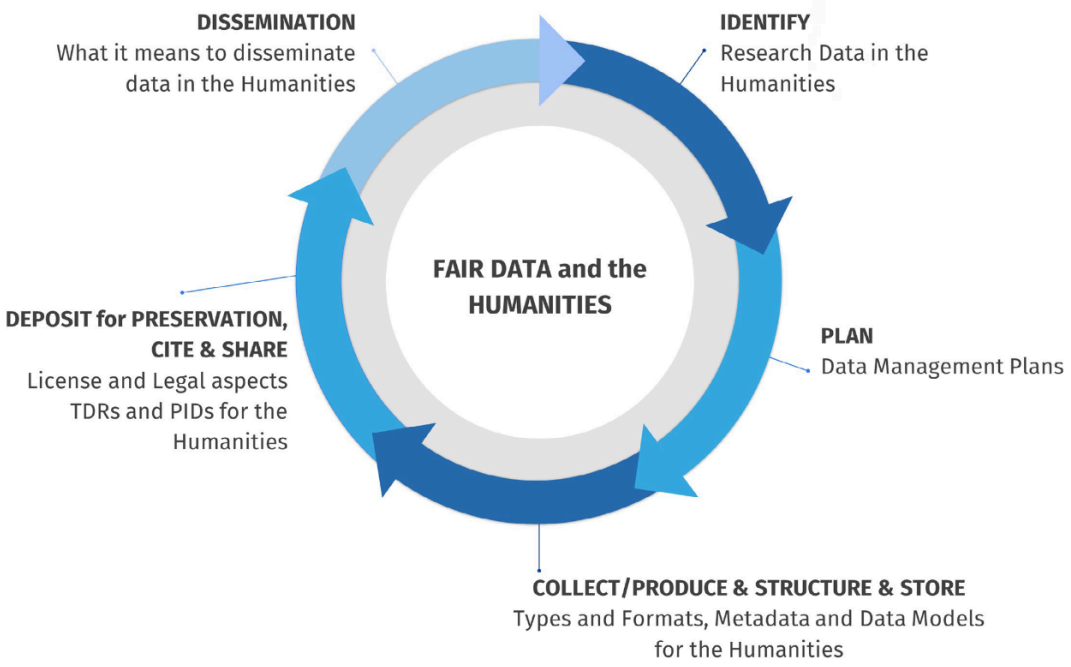

Figure 1 - Life cycle of FAIR research data in the humanities. Credits: ALLEA E-Humanities Working Group.

Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA

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, best practices, and existing resources such as tools and datasets that may be useful;
- **PLAN** - recommendations about planning your research product lifecycle 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 data, including considerations for licences and legal aspects;
- **DISSEMINATE** - recommendations on how to effectively reach your audience and make your data (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 ALLEA 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.

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 library/repository staff and consider RDM costs early on.

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 (Persistent Identifiers), and make metadata rich and machine-readable.

Data Models

Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA

- 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.

Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA

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**) and based on the XML markup language. The TEI guidelines 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 and the TEIGarage tool, for automatic conversion between numerous formats and TEI.
- Roma tool, for creating custom encoding schemas
- The TEI Archiving, Publishing, and Access Service (TAPAS), 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), for musical documents;
- Epidoc (Epigraphic Documents in TEI XML), for epigraphs and inscriptions;
- CEI (Charters Encoding Initiative), for medieval diplomatic documents (charters).

Since many digital scholarly editions need to provide readers with digital images of the documents or manuscripts on which they are based, the most effective way is through the International Image Interoperability Framework (**IIIF**).
Introductory guide to understanding how IIIF works.

Many libraries and cultural institutions (for example, the Vatican Library and the French National Library) have adopted this set of standards and APIs, allowing scholars to access and reuse photographic reproductions of cultural objects from their collections.
List of all collections available through IIIF.

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). Alternatively, other standard metadata schemas such as

Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA

Dublin Core and the FRBR/LRM model can be used and also included in TEI documents using the xenoData element.

The RIDE journal's criteria for reviewing and evaluating digital critical editions highlight which aspects need to be addressed or given particular attention to create reliable and high-quality editions. RIDE Criteria for Reviewing Scholarly Digital Editions, version 1.1

Recommendations

1) 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.
- ★ To discover tools, you can use curated lists and dedicated catalogues such as:
  - TaPoR;
  - DARIAH's catalogue of tools and services;
  - SSH Open Marketplace.

2) 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, PRISMS, FairCopy) and digital publishing (e.g., EVT) are compatible with IIIF, allowing images to be retrieved directly from their repositories. Alternatively, IIIF-compatible viewers, such as OpenSeadragon, Universal Viewer, and Mirador, 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"

Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA

3) 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 CWRC-Writer, an online XML and RDF editor that automatically detects named entities and links them to records in Linked Open Data authorities.

Example: the digital semantic edition of Paolo Bufalini's notebook.

4) Use XML/TEI or another appropriate standard schema to 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., TEI Garage).
  - Consider using annotation tools based on Domain-Specific Languages (Zenzaro, Boschetti and Del Grosso, 2025).

5) 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) 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.

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

Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA

achieve full FAIR compliance, an edition's encoding schema should 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 TEI Roma tool.
  - Follow the TEI customisation guidelines.
  - Use TEI doumentation elements to thoroughly describe new elements.
- ★ Publish encoding schemas in a trustworthy repository, alongside the edition's files.

6) 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, Lexicon of Scholarly Editing).

If you develop solutions to common challenges while working on your edition—whether it's a custom encoding schema or new software tools—document and share them in detail to help other scholars facing similar issues.

- Include references to other research products, such as software tools, datasets and previous DSEs, used to prepare the edition.

7) 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 provides guidance for describing each contributor's specific role in the edition.

Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA

8) 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, 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).

9) 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.

- 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: digital edition of De rebus Siculis Carmen edited by Fulvio delle Donne available in XML/TEI and in PDF.

Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA

10) Store the edition in a trustworthy 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 (CoreTrustSeal, Nestor, etc.) ensures its long-term accessibility, even if the visualisation system becomes obsolete.

- ★ To maximise discoverability, index the edition in searchable resources like OpenAIRE (which happens automatically when using repositories like Zenodo) and in relevant field-specific catalogues and lists of digital critical editions (see point 2).

11) For visualising the edition, use non-proprietary tools or adopt existing publishing solutions.

IDENTIFY **DISSEMINATE**

Various tools for visualizing DSEs—including TEIPublisher, TEIBoilerplate, 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.

Ready-to-use publishing solutions allow editors to focus solely on preparing the edition while delegating publication tasks. Examples are the Micro-Editions of the Scholarly Editing journal and the editions in the Digital Humanities series by BUP.

12) 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.

Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA

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. 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).

For collection metadata, several standards and formats are available alongside TEI:

- CIDOC CRM, 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, a well-established standard for exchanging bibliographic records, developed and maintained by the library community;
- Metadata Encoding and Transmission Standard (METS), an XML schema for encoding structural metadata about complex digital objects;
- MODS (Metadata Object Description Schema), an XML schema for 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.

For presenting texts and documents as digital facsimiles, IIIF is recommended, particularly for institutions digitizing their textual heritage.
Guide for IIIF implementers.

For developing a complete IT environment to create and manage digital collections, the Reference model for an Open Archival Information System (OAIS) serves as the standard model. However, this technical and complex work typically extends beyond individual scholars' scope.
OAIS Introductory Guide (2nd Edition)

The **National Information Standards Organization (NISO) Framework of Guidance for Building Good Digital Collections (3rd edn, 2007)** provides a comprehensive set of principles that in 2007 anticipated the FAIR principles. The NISO principles are organised

Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA

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.

Collections	Objects
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 the collection's authenticity, integrity, and interpretation. 3. A good collection is curated, which is to say, its resources are actively managed during their entire lifecycle. 4. A good collection is broadly available and avoids unnecessary impediments to use. Collections should be accessible to persons with disabilities, and usable effectively in conjunction with adaptive technologies. 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 to be recorded. 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. 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 that can be resolved to the current address of the object. 5. A good object can be authenticated. 6. A good object has associated metadata.
Metadata	**Initiatives**
1. Good metadata conforms to community standards in a way that is appropriate to the materials in the collection, users of the collection, and current and potential future uses of the collection. 2. Good metadata supports interoperability. 3. Good metadata uses authority control and content standards to describe objects and collocate related objects. 4. Good metadata includes a clear statement of the conditions and terms of use for the digital object. 5. Good metadata supports the long-term curation and preservation of objects in	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 to achieve its objectives. 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 the initiative's process and outcomes. 6. A good digital initiative considers the entire lifecycle of the digital collection and

Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA

collections. 6. Good metadata records are objects themselves and therefore should have the qualities of good objects, including authority, authenticity, archivability, persistence, and unique identification.	associated services.

Finally, the RIDE journal offers quality evaluation criteria for digital text collections, providing guidance for collection preparation and management. RIDE Criteria for Reviewing Digital Text Collections, version 1.0.

Recommendations

1) 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) 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>` element in the header).
- ★ Prefer links to digital libraries, online catalogues and other similar resources that provide persistent identifiers for their objects.

3) Link authors' and works' records to corresponding authority records if available, e.g., VIAF, Wikidata.

PLAN **PRODUCE**

Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA

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.

4) 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) Assign each text a persistent identifier.

DEPOSIT

This enables users to easily cite and reuse individual texts from the collection.

- ★ When texts are published as individual units in repositories like ILC4CLARIN and Zenodo, PIDs are automatically assigned.

6) 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.

Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA

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", in particular marking changes to the texts in the collection as added, removed, changed, or fixed.

7) 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.

Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA

Software tools

Standards and guidelines

The FAIR Principles for Research Software (FAIR4RS Principles), 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) is a comprehensive set of guidelines aimed at enhancing the findability, accessibility, interoperability, and reusability of research software artifacts, developed within the FAIR-IMPACT project for the European Open Science Cloud (EOSC). The RSMD guidelines 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.
RSMD checklist: quick overview of the RSMD guidelines.

To assess the FAIRness level of a software artifact, you can use the metrics 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.

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?

Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA

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.
OSI (Open Source Initiative) Approved licences.
SPDX licence List: a complete list of licences used for software artefacts.

Software Heritage 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.
Software Heritage Documentation.

Last but not least, developers working in the DH field can follow and join the activities of DHTech, 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.

Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA

1) 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.

2) 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.
- 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) 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.

Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA

★ Ensure scalability and the ability to handle increased load without compromising overall performance.

4) Employ standard and non-proprietary programming languages and technologies to develop tools, ensuring greater longevity and easier maintainability.

PRODUCE

★ W3C standards for web development.
★ Community Development of Java Technology Specifications.

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) 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 repositories such as GitHub that foster collaboration among developers.
★ Follow shared methodologies and strategies for versioning (e.g., Semantic Versioning) and branching (e.g., GitFlow workflow).

6) 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) 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.

Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA

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) 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) represents complex knowledge about things and their relationships. Based on computational logic, it enables programs to verify consistency and reveal implicit knowledge.
- **Turtle** provides a simplified, human-readable syntax for RDF that is more concise than RDF-XML.
- **JSON-LD** (A JSON-based Serialisation for Linked Data) integrates Linked Data into web environments through JSON compatibility.

Other important W3C Semantic Web Standards include:

- **SKOS** (Simple Knowledge Organization System) for creating vocabularies and taxonomies;
- **SPARQL** (SPARQL Query Language for RDF) 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.

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) and FOOPS! (Ontology Pitfall Scanner for FAIR).

Another effective tool designed specifically for SKOS vocabularies is SKOS Play!, which enables users to validate and convert SKOS files to PDF and HTML formats, and also generates SKOS files from Excel spreadsheets.

Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA

Recommendations

1) 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, schema.org, Dublin Core, LRMOO, DataCite, DCAT, SPAR Ontologies, and ArCO.
- ★ To explore existing models, use the LOV (Linked Open Vocabularies) database, BARTOC (Basic Registry of Thesauri, Ontologies and Classifications) or the H-SeTIS (Heritage – Semantic Tools and Interoperability Survey) database specific for the heritage domain.

2) 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) 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:

Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA

- ★ Ontology Development 101: A guide to creating your first ontology, a comprehensive guide for beginners learning to model domains through ontologies;
- ★ Simplified Agile Methodology for Ontology Development (SAMOD), an advanced approach developed by the SPAR Ontologies team;
- ★ LOT (Linked Open Terms) Methodology 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 proposes nine scenarios for building ontologies and networks, focusing on resource reuse, reengineering, and collaborative development.

When facing common challenges during domain modeling, you can apply Ontology Design Patterns as proven solutions.

For the practical development of ontologies, several graphical editors are available, including Protégé, Fluent Editor, and OWLGrEd. Antoher popular development platform for managing OWL ontologies is VocBench**.** For more options, see this comprehensive list of editors and development environments.

4) Foster interoperability with existing ontologies and facilitate integration through comprehensive mapping.

PRODUCE

By associating each element of your ontology to the elements of other existing models, you enable easy integration of data modeled according to your ontology in other datasets. Comprehensive mapping is also useful to facilitate data maintenance and the handling of legacy data.

- ★ Many W3C models, such as DCAT, include mapping recommendations.
- ★ Standards:
  - RDF Mapping Language (RML)
  - Simple Standard for Sharing Ontological Mappings (SSSOM)
  - InteroperAble Descriptions of Observable Property Terminology WG (RDA I-ADOPT WG)
- ★ Article: "Moving towards FAIR mappings and crosswalks".

5) Provide users with detailed documentation, complete with practical usage examples and intuitive graphical representations of the ontology.

PRODUCE DISSEMINATE

Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA

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, WebVOWL, Widoco. List of visualisation tools.

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 and Fluent Editor.

6) 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, 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.

7) Utilise external services for ontology publication to guarantee long-term accessibility and preservation.

DEPOSIT

To publish your ontology, you can leverage existing online tools. This approach allows researchers to concentrate on development while simplifying long-term maintenance.

- ★ Workbench for storing and searching: graphdb, Virtuoso. List of Triple Stores.

Additionally, we recommend storing the ontology in multiple formats within trustworthy repositories.

Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA

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), which allows to use attributes within HTML5, XHTML and XML documents to express structured data.

List of tools that can be used to create a Triple Store, i.e., an RDF-based database.

The 5-star guideline is the main guideline for linked open data publication, comprising five steps:

1. "make your stuff available on the Web (whatever format) under an **open licence**
2. make it available as **structured data** (e.g., Excel instead of image scan of a table)
3. make it available in a **non-proprietary open format** (e.g., CSV instead of Excel)
4. **use URIs** to denote things, so that people can point at your stuff
5. **link your data to other data** to provide context"

Recommendations

1) 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.
- ★ Domain-relevant vocabularies: the Getty Vocabularies, CLARIN Concept Registry, DARIAH Vocabs services.
- ★ Domain-relevant models and ontologies, used on a national and international level: CIDOC-CRM, FOAF, schema.org, Dublin Core, LRMOO, 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 or the H-SeTIS (Heritage – Semantic Tools and Interoperability Survey) database specific for the heritage domain.

2) 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

Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA

In this case, follow the recommendations of the Ontologies section.

3) 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, VIAF, Open Library, and WorldCat.

4) 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.

5) Provide the data model and URI pattern to facilitate data reuse.

DEPOSIT **DISSEMINATE**

★ Create persistent URIs, using services such as W3ID.

6) Publish your data in trustworthy 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

7) Offer a SPARQL endpoint with accompanying search examples. Consider implementing a user-friendly GUI for data visualisation and navigation, catering to users unfamiliar with SPARQL.

DISSEMINATE

While a SPARQL endpoint is the preferred tool for exploring a linked open dataset, a graphical user interface with browsing and searching functionalities can help a wider audience access and use your data.

★ For publishing your ontology, you can use visualisation tools such as Widoco and LODE. List of available tools.

Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA

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 the following information is prominently displayed:
    - access points to the data;
    - licence;
    - version number;
    - status (e.g., completed, under development).

  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, Google Sites, and Wordpress. Explore other tools in TAPoR.

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

Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA

- 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 GitBook and WikiMedia.
- Offer metadata and documentation in English (in addition to other languages) to reach a broader audience.

Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA

References

ALLEA | All European Academies. *Recognising Digital Scholarly Outputs in the Humanities*. ALLEA, 2023, https://doi.org/10.26356/OUTPUTS-DH.

Amdouni, Emna, et al. "O'FAIRe Makes You an Offer: Metadata-Based Automatic FAIRness Assessment for Ontologies and Semantic Resources." *International Journal of Metadata, Semantics and Ontologies*, vol. 16, no. 1, 2022, pp. 16–46, https://doi.org/10.1504/IJMSO.2022.131133.

Ayris, P. "Delivering the European Open Science Cloud (EOSC): Principle and Practice in Delivering Open Science." *In: LEARN Toolkit of Best Practice for Research Data Management. (Pp. 87-89). Leaders Activating Research Networks (LEARN) (2017)*, Leaders Activating Research Networks (LEARN), 2017, pp. 87–89, https://doi.org/10.14324/000.learn.17.

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.

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.

Courbebaisse, Guy, et al. *Research Software Lifecycle*. Sept. 2023, https://doi.org/10.5281/ZENODO.8324828.

Directorate-General for Research and Innovation (European Commission). *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, 2018, https://data.europa.eu/doi/10.2777/112658.

Evans, Eric. *Domain-Driven Design: Tackling Complexity in the Heart of Software*. Addison-Wesley, 2004.

FAIR4ML. https://rda-fair4ml.github.io/FAIR4ML-schema/release/0.1.0/index.html. Accessed 20 June 2025.

Filiposka, Sonja, et al. *FAIR-by-Design Methodology for Learning Materials*. Jan. 2025, https://doi.org/10.5281/ZENODO.14711524.

Gamma, Erich, et al. *Design Patterns: Elements of Reusable Object-Oriented Software*. Pearson Education, 1994.

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.

Finanziato dall'Unione europea NextGenerationEU

Ministero dell'Università e della Ricerca

Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA

Group, DDI Training. *Achieving FAIR with DDI*. https://doi.org/10.5281/zenodo.10064667.

Gruenpeter, Morane, Sabrina Granger, et al. *D4.4 - Guidelines for Recommended Metadata Standard for Research Software within EOSC*. Mar. 2024, https://doi.org/10.5281/ZENODO.10786147.

Gruenpeter, Morane, Roberto Di Cosmo, et al. *M2.15 Assessment Report on "FAIRness of Software."* Oct. 2020, https://zenodo.org/records/4095092.

Harrower, Natalie, et al. *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.

Horik, René Van. *The Research Data Alliance and the Humanities*. Zenodo, 30 July 2019, https://doi.org/10.5281/ZENODO.3355144.

Hugo, Wim, et al. *D2.5 FAIR Semantics Recommendations Second Iteration*. Dec. 2020, https://doi.org/10.5281/ZENODO.4314320.

LEARN Toolkit of Best Practice for Research Data Management. LEARN, 3 Apr. 2017, https://doi.org/10.14324/000.learn.00.

Martínková, Jana, Nick Juty, Alejandra Gonzalez-Beltran, Carole Goble, and Yann Le Franc. 'Moving towards FAIR Mappings and Crosswalks.', 2024. https://research.manchester.ac.uk/en/publications/moving-towards-fair-mappings-and-crosswalks.

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.

Research Data Alliance/FORCE11 Software Source Code Identification WG, et al. *Use Cases and Identifier Schemes for Persistent Software Source Code Identification (V1.0)*. 2020, https://doi.org/10.15497/RDA00053.

Scarpa, Erica, and Riccardo Valente. *Heritage – Semantic Tools and Interoperability Survey*. Heritage – Semantic Tools and Interoperability Survey, 2024, https://doi.org/10.71795/8T2Z-HH65.

Schiltz, Arthur, et al. *MLDCAT-AP (Machine Learning DCAT Application Profile)*. 2.0.0, 18 Feb. 2025, https://semiceu.github.io/MLDCAT-AP/releases/2.1.0/.

Silén, Petri. *Clean Code Principles and Patterns, 2nd Edition*. Leanpub, 2024, http://leanpub.next/cleancodeprinciplesandpatterns2ndedition.

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.

SSHOC. *SSHOCingly Good and Sustainable Tools*. Zenodo, 29 Mar. 2022, https://doi.org/10.5281/ZENODO.6394622.

Suárez-Figueroa, Mari Carmen, et al. "The NeOn Methodology for Ontology Engineering." *Ontology Engineering in a Networked World*, edited by Mari Carmen Suárez-Figueroa et al., Springer Berlin Heidelberg, 2012, pp. 9–34, https://doi.org/10.1007/978-3-642-24794-1_2.

The CodeMeta Project. https://codemeta.github.io/. Accessed 20 June 2025.

UNESCO. *UNESCO Recommendation on Open Science*. UNESCO, 2021, https://doi.org/10.54677/MNMH8546.

Vogt, Lars, et al. *FAIR 2.0: Extending the FAIR Guiding Principles to Address Semantic Interoperability*. 2024, https://doi.org/10.48550/ARXIV.2405.03345.

Wilkinson, Mark D., et al. "The FAIR Guiding Principles for Scientific Data Management and Stewardship." *Scientific Data*, vol. 3, no. 1, Mar. 2016, p. 160018, https://doi.org/10.1038/sdata.2016.18.

Zenzaro, Simone, Federico Boschetti, and Angelo Mario Del Grosso. 'Making Digital Scholarly Editions Based on Domain Specific Languages'. In *Digital Editing and Publishing in the Twenty-First Century*, edited by James O'Sullivan, Michael Pidd, Sophie Whittle, Bridgette Wessels, Michael Kurzmeier, and Órla Murphy, 1st edn. Scottish Universities Press, 2025. https://doi.org/10.62637/sup.GHST9020.9.

Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA