Umanistica Digitale – ISSN: 2532-8816 – n. 22, 2026: Special Issue
S. Giacomini, C. Martignano, G. Rubin, A. Bardi, M. Buzzoni, M. Daquino, R. Del Gratta, A.M. Del Grosso, F. Fischer, R. Rosselli Del Turco, F. Tomasi – *ATLAS: A Knowledge Graph to Enhance the Findability of International Scholarly Research on the Italian Digital Cultural Heritage*
DOI: https://doi.org/10.60923/issn.2532-8816/21765

ATLAS: A Knowledge Graph to Enhance the Findability of International Scholarly Research on the Italian Digital Cultural Heritage

Sebastiano Giacomini

University of Bologna
sebastiano.giacomin2@unibo.it

Chiara Martignano

Ca' Foscari University of Venice
chiara.martignano@unive.it

Giorgia Rubin

Institute for Computational Linguistics "A. Zampolli" - National Research Council
giorgia.rubin@ilc.cnr.it

Alessia Bardi

Institute of Information Science and Technologies "A. Faedo" - National Research Council
alessia.bardi@isti.cnr.it

Marina Buzzoni

Ca' Foscari University of Venice
mbuzzoni@unive.it

Marilena Daquino

University of Bologna
marilena.daquino2@unibo.it

Riccardo Del Gratta

Institute for Computational Linguistics "A. Zampolli" - National Research Council
riccardo.delgratta@ilc.cnr.it

Angelo Mario Del Grosso

Institute for Computational Linguistics "A. Zampolli" - National Research Council
angelo.delgrosso@ilc.cnr.it

Franz Fischer

Ca' Foscari University of Venice
franz.fischer@unive.it

Roberto Rosselli Del Turco

University of Turin
roberto.rossellidelturco@unito.it

Francesca Tomasi

University of Bologna
francesca.tomasi@unibo.it

Abstract

In recent years, the abundance of scholarly information has requested constant development and revision of standardized models and shared guidelines. Based on these frameworks, the Digital Humanities (DH) landscape is represented in a variety of aggregators expected to enhance research data findability while promoting use and reuse. However, current semantic models fail to capture the specificity of DH research products, hindering data discovery and hampering the valorisation of Cultural Heritage. The ATLAS project addresses these key challenges by developing a unified framework for describing and aggregating scholarly outputs, particularly in the Italian Digital Cultural Heritage domain. This paper presents the initial versions of the ATLAS Ontology and Knowledge Graph, designed to model DH outcomes such as Digital Scholarly Editions, text collections, Linked Open Data, ontologies, and software. In so doing, ATLAS aims to enhance resource findability and reuse, paving the way for improved interoperability and future advancements in the field.

Keywords: Knowledge Graph, Knowledge Extraction, Semantic Web, Research Infrastructures, Italian Cultural Heritage, Web Application

Negli ultimi anni, l'abbondanza di dati scientifici ha reso necessario un continuo sviluppo e aggiornamento di modelli standardizzati e linee guida condivise. Sulla base di questi riferimenti, il panorama delle Digital Humanities (DH) viene rappresentato da una varietà di aggregatori, il cui compito consiste nel migliorare la reperibilità dei dati della ricerca, promuovendone al contempo l'uso e il riuso. Tuttavia, i modelli semantici attualmente disponibili non riescono a cogliere a pieno le specificità dei prodotti della ricerca nelle DH, ostacolando la scoperta dei dati e la valorizzazione del Patrimonio Culturale connesso. Il Progetto ATLAS affronta queste sfide cruciali attraverso lo sviluppo di un sistema unificato per la descrizione e aggregazione dei prodotti di ricerca, con particolare attenzione al Patrimonio Culturale Digitale Italiano. Il presente contributo introduce le versioni iniziali dell'Ontologia e del Knowledge Graph ATLAS, progettati per modellare prodotti delle DH come edizioni digitali, raccolte di testi, collezioni di Linked Open Data, ontologie e software. Così facendo, ATLAS intende migliorare la reperibilità e il riutilizzo delle risorse, aprendo la strada a un'accresciuta interoperabilità e a futuri sviluppi nel settore.

Parole chiave: Grafo della conoscenza, Estrazione della conoscenza, Web Semantico, Infrastrutture di Ricerca, Patrimonio Culturale Italiano, Applicazione Web

1. Introduction[1]

In recent years, Semantic Web technologies have significantly changed how scholarly activities in the Digital Humanities (DH) domain are carried out, offering unprecedented opportunities for preserving, sharing, and reusing research outputs and publications [41, 8]. The abundance of available scholarly information has required the iterative development and revision of standardised models and shared guidelines to ensure common rules for dissemination and long-term preservation across communities. Such frameworks have become the foundation for data aggregators and exploratory environments, such as Europeana[2] and OpenAIRE,[3] which are designed to collect documents and data from various research settings, including those entirely or partially focused on DH. Such initiatives have embraced Semantic Web technologies, particularly Linked Open Data, to unravel the complex relations between scholarly endeavours and the Cultural Heritage.

In particular, the DH landscape is scattered across a variety of aggregators, each focusing on different aspects of research activities. In so doing, they are expected to enhance data and metadata findability while promoting use and reuse [20]. By aggregating resources, such systems

[1] Authors' responsibilities: Sebastiano Giacomini is responsible for Section 4; Chiara Martignano is responsible for Sections 3 and 6; Giorgia Rubin is responsible for Sections 2. and 5; all authors contributed to Sections 1 and 7.

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

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

attempt to offer additional research value that conventional forms of retrieval and browsing cannot achieve [16]. However, to the best of our knowledge, despite recent efforts made by cultural institutions, the analysis of the Italian context reveals the lack of a unified research framework for Cultural Heritage and research data discoverability, as well as the lack of a comprehensive catalogue of DH scholarly data [21, 35, 26, 10], and domain-dependent best practices to foster data findability and reusability, ultimately hindering resource discovery. In other terms, (1) representative services for collecting, aggregating, and disseminating DH research products are missing, and (2) domain-specific ontologies and vocabularies are not easily adaptable to describe the heterogeneous nature of Digital Cultural Heritage outputs (e.g., digital editions, text collections).

In this article, we present the initial results of the ATLAS project, which aim at coping with the aforementioned problems. Results include the ATLAS ontology, the ATLAS knowledge graph, and the ATLAS platform for data entry and dissemination. The ATLAS ontology has been developed to meet the main challenges posed by the description of DH research activities and products. These include Digital Scholarly Editions, text collections, Linked Open Data datasets, RDF vocabularies, and software. To populate the ontology and test the proposed model, an initial knowledge graph has been created by extracting, structuring, and enriching high-quality data from potentially unstructured or semi-structured digital sources. To achieve this goal, the ATLAS project has extended the functionalities of CLEF[4] (Crowdsourcing Linked Entities via Web Form) to develop a collaborative web platform for data entry that facilitates curation and dissemination of LOD collections. Notably, the latest version of CLEF allows data curators to semi-automatically extract knowledge from various sources, including APIs, SPARQL endpoints, and static files (.csv, .json, and .xml formats) and populate the descriptive record of a research product. To support the ontology design and the collection of technical requirements for the ATLAS platform, we analysed a set of pilot projects on the Italian Digital Cultural Heritage, as well as ontological models for describing scholarly data, that have been reviewed and mapped to highlight classes and properties currently lacking.

The article extends our previous work [21] by addressing the following aspects: (1) it describes the recently published ATLAS platform for crowdsourcing scholarly data according to the ATLAS ontology, and provides extensive examples illustrating the usage of both the ATLAS Ontology and web application; (2) it presents the results of a user study conducted to validate both the semantic model via the evaluation of the platform; and (3) it discusses current challenges, limitations and future steps for scaling proposed solutions beyond the Italian landscape.

Accordingly, the paper is structured as follows. Section 2 reviews scholarly aggregators of DH research activities and outputs, with specific considerations on Italian Digital Cultural Heritage, as well as existing semantic models and their main properties, so as to highlight the motivation for our work. Section 3 outlines the methodology and approach used to develop the ontology and populate it through a knowledge graph. Section 4 presents the initial versions of both the ATLAS Ontology and the related knowledge graph, including an illustrative example from the described pilot resources. Section 5 examines the findings of an application-based evaluation of the semantic model and its implementation in the web platform. Finally, Section 6 discusses the outcomes and limitations, while Section 7 outlines future directions of the ATLAS project.

[4] https://polifonia-project.github.io/clef/.

2. State of the Art

Over the last few years, GLAM institutions (Galleries, Libraries, Archives, and Museums) have increasingly promoted initiatives aimed at sharing their holdings across the web. While these efforts have significantly broadened access to invaluable Cultural Heritage resources, they have also resulted in the proliferation of new models, schemas, and vocabularies, leading to uncontrolled growth of metadata standards across the Web [29]. Amidst this complex and fragmented landscape, a number of aggregators have recently emerged, highlighting the fundamental role of such services in providing homogeneous access to heterogeneous metadata collections [7].

Within the Italian scenario, institutions have invested in digitising and aggregating cultural holdings, making them available as Linked Open Data collections. Projects like dati.culturaitalia,[5] the Linked Open Data platform by the Italian Ministry of Culture, exemplify the recent commitment to making Italian Cultural Heritage data interoperable with some prominent digitisation efforts within the European landscape, including ARIADNE[6] and Europeana [15]. Similarly, the ArCO[7] project has developed a Knowledge Graph from the General Catalog of Italian Cultural Heritage, offering reusable Linked Open Data collections based on the official institutional database of Italian Cultural Heritage [9]. Despite these efforts and other notable initiatives for collecting DH research data,[8] there are either no representative, comprehensive catalogues tailored to DH projects, or they do not allow the retrieval of research products relevant to the Italian Cultural Heritage. Additionally, no structured collections on DH projects and artefacts leveraging Semantic Web technologies are available [11]. The broader scholarly landscape presents several platforms that play a crucial role in providing persistent identification, long-term preservation, and enhanced findability of research data [38]. Prominent services include Zenodo[9] and OpenAIRE [30]. The OpenAIRE network integrates several services, including community web portals like the Digital Humanities and Cultural Heritage gateway,[10] which facilitate the discovery and sharing of research outcomes and Open Science practices.

However, despite targeted attempts to highlight DH research activities, aggregators like Zenodo and OpenAIRE serve as broad data collectors on various disciplines, often lacking references to the Cultural Heritage sources that drove the creation of DH scholarly data. In addition, the absence of domain-specific vocabularies hampers the identification of resources produced by DH practices, e.g., digital editions.

At the core of the information retrieval problem outlined above, we find the lack of a comprehensive data model that allows one to describe the peculiarities of the DH research products in the first place. While several data models exist and are shared in the broader scholarly community, they describe research outputs in general terms, without considering the diversity

[5] https://dati.culturaitalia.it.

[6] https://ariadne-infrastructure.eu.

[7] https://w3id.org/arco/.

[8] These include catalogues of Digital Scholarly Editions [32, 18], heterogeneous projects gathered by national associations (AIUCD), research centres (/DH.arc, VeDPH, DH@FBK), international associations (EADH), disciplinary surveys [24, 23].

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

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

and specificity of DH outputs. Notable examples include the OpenAIRE Graph,[11] which provides a Scholarly Knowledge Graph [1] collecting metadata on the following core entities: Research products, Data sources, Organisations, Projects, and Communities. Research products include "Publication", "Data", "Software", and "Other research product". RO-Crate (Research Object Crate)[12] offers another approach for packaging research data along with its metadata and associated component files [39]. RO-Crates are based on the concept of Research Object (RO), defined as a semantically rich aggregation of resources [6], and serve data according to Schema.org[13] in JSON-LD format. The current data model (v1.1.3) distinguishes between Data entities (e.g., directories, files) and Contextual entities (person, organisations, equipment) [37]. Within this framework, an RO-crate resource is treated as a root data entity with type `schema:Dataset`. The SKG-IF (Scientific Knowledge Graph Interoperability Framework)[14] Working Group has recently developed a metadata model targeting interoperability among Scientific Knowledge Graphs and their usability [2]. The model (v1.1) is structured around six core entities: Research product, Agent, Grant, Venue, Topic, and Data source. Research products are described via the FaBiO Ontology (FRBR-aligned Bibliographic Ontology) [28]; namely, `fabio:Dataset` (research data), `fabio:ScholarlyWork` (literature), and `fabio:Software` (software). Lastly, the KNOT[15] project aims to showcase the Digital Cultural Heritage of Italian universities [17]. The ontology[16] (v1.2) leverages entities from DCAT, PROV-O, and CIDOC-CRM, and the KNOT knowledge graph mainly focuses on Research Projects, Digital Objects (e.g., Datasets, Knowledge Graphs, Ontologies), and Web Services (e.g., Digital Editions, Digital Libraries, Endpoints). However, the model does not focus on identifying Cultural Heritage artefacts, using the generic `dcterms:subject` property to broadly indicate related disciplines and Wikidata keywords. In addition, no information is retrieved directly from available sources (e.g., datasets, TEI encodings).

In conclusion, despite such remarkable achievements, the models fall short of addressing all complexities set by the current DH landscape. Even advanced schemas, such as OpenAIRE Graph and SKG-IF, which introduce higher levels of granularity, fail to capture the heterogeneity of research outputs in the Digital Cultural Heritage domain. In fact, diverse projects can result in a variety of outcome types—such as text collections, Digital Scholarly Editions, Linked Open Data datasets, RDF vocabularies, and software—, each of which deserves to be described accordingly. Firstly, specialised terminologies are needed to identify the different products, particularly those peculiar to Digital Cultural Heritage, such as digital textual archives and Digital Scholarly Editions. Secondly, the existing models lack semantic attributes and controlled resources designed to adequately describe the methodological aspects of DH research. Crucial issues, such as textual typologies and edition criteria, which are critical for a comprehensive representation of peculiar outcomes and research practices, remain insufficiently addressed. Lastly, existing models do not provide adequate solutions for linking research activities to their corresponding Cultural Heritage objects, despite the potential offered by Linked Open Data. This results in two main consequences, namely: (1) it limits users and researchers in discovering

[11] https://graph.openaire.eu/.

[12] https://researchobject.org/ro-crate/.

[13] https://schema.org/.

[14] https://skg-if.github.io/.

[15] https://projects.dharc.unibo.it/knot/records.

[16] http://purl.org/knot/ontology.

products and perspectives on Digital Cultural Heritage resources, and (2) hinders Cultural Heritage resources retrieval and valorisation.

Further limitations derive from services and websites that do not include such information when providing access to research products metadata. These shortcomings affect both the data collection processes, due to the lack of suitable tools for extracting meaningful entities from available resources, and the dissemination stage, where the absence of dedicated systems for data visualisation hampers discovery. To address the challenges, the CLEF application is actively working on developing novel solutions, including data entry and exploration services such as Intermediate Templates, Advanced Knowledge Extraction, and Data Visualisation tools.

While hindering findability, current limitations prevent serendipitous discoveries and limit the effective reuse of research outputs in Humanities research. Bridging this gap requires the development of a semantic model that accommodates the diversity of DH outputs while facilitating the integration of Cultural Heritage metadata into services. To this extent, existing software solutions for cataloguing scholarly data lack the means to (1) leverage complex data models, and (2) automatically extract information from data sources (e.g., extracting the Cultural Heritage resources mentioned in a research product). Moreover, (3) they lack web-based solutions for performing data analysis without requiring users' advanced technical skills [14, 12].

3. Methodology and Approach

In the ATLAS project we investigated a number of pilots representative of DH projects and resources relevant to the Italian Cultural Heritage [11] and we classified them into five main groups, namely:

- Text collections: ALIM (Archive of the Italian Latinity of the Middle Ages); Biblioteca Italiana; BUP - Digital Humanities; Musisque Deoque
- Digital Scholarly Editions: VaSto (VArchi STOria fiorentina); Codice Pelavicino Digitale; Leges Langobardorum; Digital Edition of Aldo Moro's works
- Software: EVT (Edition Visualisation Technology); Voyant Tools
- Linked Open Data: Zeri & LODE; DanteSources; LiLa - Linking Latin; Biflow - Toscana Bilingue Catalogue
- Ontologies: CIDOC-CRM; SPAR; HiCO

These projects were selected as representative case studies because they are widely recognised as reference models in the Italian Digital Humanities domain. They exemplify best practices in adopting community standards (e.g., TEI/XML, Linked Open Data, CIDOC-CRM, SPAR) and serve as authoritative benchmarks for producing new research outputs and for the creation of a descriptive model. In particular, pilots served two main purposes, namely (1) identifying essential metadata for building the ATLAS catalogue and its semantic model, and (2) validating and populating the ontology with scholarly data resulting in a knowledge graph. Additionally, this analysis also aimed to produce a set of guidelines to help improve data management practices in the Digital Humanities projects.

The results of the pilot analysis offered an initial base for evaluating existing standards for the description of research products. Metadata from pilot projects were systematically collected, assigning a label and corresponding values to each piece of information. Labels provided a starting point for a preliminary mapping of existing data models and frameworks, enabling a semantic alignment and arrangement of identified metadata. Detailed mapping tables are provided in the supplementary materials of the ATLAS Ontology and include the following vocabularies and frameworks: RO-Crate,[17] KNOT,[18] OpenAIRE Graph,[19] OpenAIRE Application Profile,[20] SKG-IF,[21] IRIS.[22]

The preliminary analysis revealed the need for a novel data model capable of addressing the current issues highlighted in the state of the art, ensuring a nuanced representation of research outputs, enhancing metadata completeness, and improving accessibility. The resulting ATLAS Ontology[23] imports several models. The backbone is based on classes and properties from Schema.org (v28.0),[24] a vocabulary that has already proved to be suitable for describing and aggregating Cultural Heritage objects metadata [19]. However, the complexity of the Digital Cultural Heritage research domain required integrating other models, particularly those offering granularity concerning the DH domain. Among these, particular attention was paid to FaBiO, the FRBR-aligned Bibliographic Ontology [28], and DC Terms,[25] both suggesting the importance of working on multiple levels of cultural objects [40].

To first test and validate the newly created model, metadata collected from the preliminary analysis of pilot resources were reused to develop a first Knowledge Graph populating the novel ontology. Subsequently, a further application-based evaluation phase was conducted through a hands-on-session, held during the ATLAS Workshop.[26] This second validation stage sought to engage scholars, researchers, and students in the cataloguing process, allowing them to contribute through feedback and insights. Participants were assigned the task of describing a Research Product of the international scholarly research on Italian Digital Cultural Heritage. A dedicated survey aimed to assess the usability of the ATLAS platform.

To populate the preliminary Knowledge Graph and further expand it during the hands-on-session, the ATLAS platform was developed on top of the CLEF web application [22], providing users with a system to verify the adequacy of the semantic schema and to streamline data entry activities. CLEF supports the collaborative creation of Linked Open Data collections through customisable "Templates" corresponding to ontological classes and rendered as user-friendly Web Forms. The platform's key features, including automatic Entity Reconciliation and Knowledge Extraction features, enable the development of a Knowledge Graph of interlinked

---

17 https://w3id.org/ro/crate/1.1.

18 See footnote 15.

19 https://graph.openaire.eu/docs/.

20 https://openaire-guidelines-for-literature-repository-managers.readthedocs.io/en/v4.0.0/.

21 https://w3id.org/skg-if/context/docs/skg-if.json.

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

23 https://w3id.org/dh-atlas/.

24 https://github.com/schemaorg/schemaorg/tree/main/data/releases/28.0/.

25 http://purl.org/dc/terms/.

26 https://dh-atlas.github.io/workshop.html.

records, managed by the Blazegraph[27] triplestore and simultaneously serialised in Turtle format for milestones data publication and versioning purposes.

To meet the granularity requirements of the ATLAS Ontology and make proper use of the content in available resources (e.g., datasets and TEI documents), ATLAS worked on extending CLEF functionalities. This effort focused on three key areas, namely: innovative solutions for representing complex data models in data entry, streamlining data entry processes, and providing data processing tools to enhance user experience and catalogue exploration and visualisation.

4. Results

4.1 ATLAS Ontology

The ATLAS Ontology is an OWL 2 DL ontology [5] designed to represent scholarly research projects on the Italian Cultural Heritage and their outcomes. Its primary goal is to describe features of DH research products, highlighting their unique attributes to the broader landscape of scholarly artefacts. As aforementioned, the ATLAS Ontology leverages terms from different existing models to facilitate the alignment between the ATLAS catalogue and existing data sources. Schema.org (prefix schema, https://schema.org) serves as the backbone of the vocabulary, and it is enriched with terms from DCTerms, and FaBiO (prefix fabio, http://purl.org/spar/fabio/) [28]. To enhance granularity and be representative of the terminology used by practitioners in the DH, ATLAS has also introduced new Classes and Properties (prefix atlas, https://w3id.org/dh-atlas/), aligned to existing models. In Figure 1, we show an overview of classes and properties.

[27] https://blazegraph.com/.

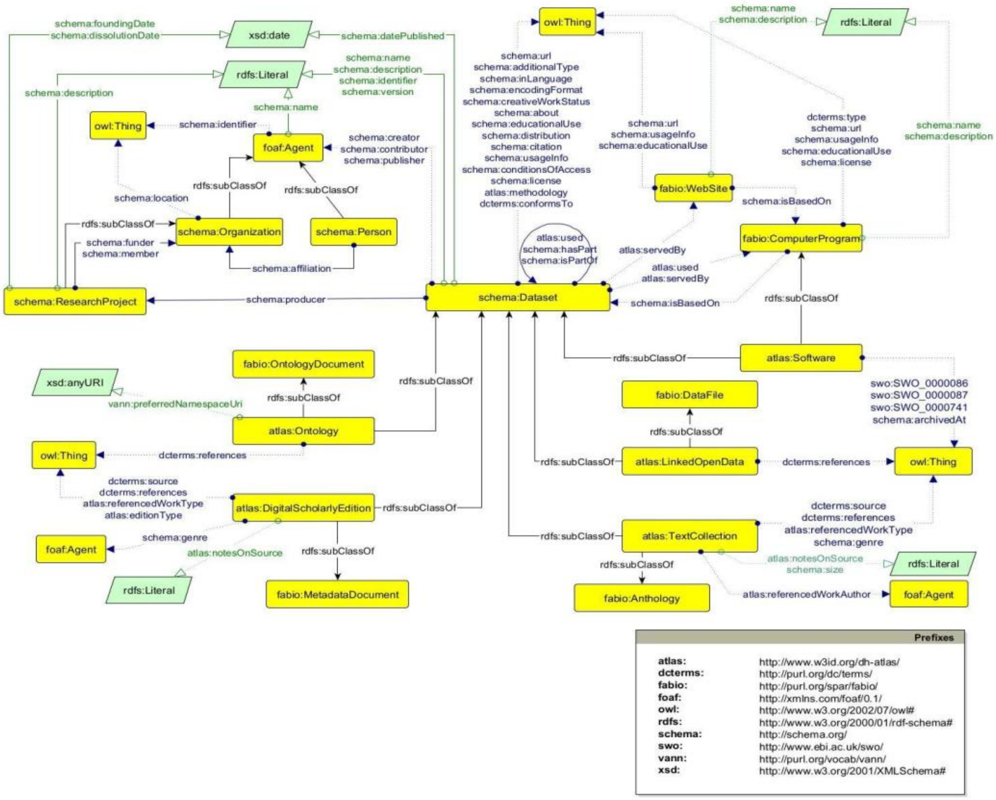


Figure 1: A visual diagram of the ATLAS Ontology: classes and properties.

4.1.1 Research Product

The results of research activities are first-class citizens in many reviewed models. The ATLAS Ontology follows this approach and makes research products the core of the vocabulary, represented by the class `schema:Dataset`, a subclass of `schema:CreativeWork`. While Schema.org broadly defines `schema:Dataset` as any "body of structured information describing some topic(s) of interest", additional specifications clarify its intended applications [36]. Usage examples include collections of packaged data, such as those "published in scientific, scholarly or governmental open data repositories", as well as "data that is stored in collections of spreadsheet files, or as digital images, or in dedicated scientific, geospatial and engineering file formats".

This view is aligned with the DCAT vocabulary[28], where `dcat:Dataset` is defined as "a collection of data, published or curated by a single agent, and available for access or download in one or more representations." The notion of dataset in DCAT is intentionally broad and inclusive, accommodating diverse resource types arising from different communities.

[28] https://www.w3.org/TR/2024/REC-vocab-dcat-3-20240822/.

To better frame the nature of scholarly outcomes in the DH, the property `schema:additionalType` allows us to associate instances of `schema:Dataset` with subclasses of the class `frbr:Expression`, namely: `atlas:TextCollection`, `atlas:DigitalScholarlyEdition`, `atlas:LinkedOpenData`, `atlas:Ontology`, and `atlas:Software`. Following the broad interpretations of both Schema.org and DCAT, software is modelled as a subclass of `schema:Dataset` because it constitutes a scholarly output that produces, organises, or enables access to structured information. This approach allows software to be treated consistently with other research outputs in the knowledge graph, while the `atlas:Software` subclass clearly distinguishes it from other products types.

Depending on the associated class, additional properties can be used to describe scholarly products. In ATLAS we distinguish artefact-dependent properties from general properties. General properties include information such as the title (`schema:name`), a description (`schema:description`), the release date (`schema:datePublished`), the current version (`schema:version`), the current work status (`schema:creativeWorkStatus`), external identifiers (`schema:identifier`), the resource link (`schema:url`), and links to distributions (`schema:distribution`).

Further details focus on the technical content of the resource, such as the subject matter (`schema:about`), used languages (`schema:inLanguage`), the encoding format (`schema:encodingFormat`), bibliographic references (`schema:citation`), adopted standards (`dcterms:conformsTo`), and documentation web pages (`schema:usageInfo`).

To refine the description of DH artefacts and allow a more practical use of ATLAS cataloguing data, two properties describe the research activities afforded by the research product (`schema:educationalUse`) and those performed during the production of the outcome at hand (`atlas:methodology`): in both cases, values are expected to be taken from the TaDiRAH taxonomy.[29] The properties `schema:license` and `schema:conditionsOfAccess` are expected to provide information on the license and access rights respectively.

Relations between artefacts and people/organisations, i.e., instances of the class `foaf:Agent`, include authors (`schema:creator`), contributors (`schema:contributor`), publishers (`schema:publisher`), and the Research Project the object is a result of (`schema:producer`). Relations between Research Products can be expressed through `schema:hasPart`, `schema:isPartOf`, and `atlas:used`, the latter specifying external resources reused to generate the product although not being part of it. At the same time, the `atlas:isServedBy` property introduces those services and tools that make available the content of the Research Products (e.g., Visualisation Software, SPARQL endpoints). In Table 1, we summarise properties associated with the five classes defined in the ATLAS Ontology.

[29] Taxonomy of Digital Research Activities in the Humanities, https://vocabs.dariah.eu/tadirah/en/.

Table 1: Classes and properties for describing Research Products in ATLAS.

ATLAS Type	RDF Property	Property Description
atlas:TextCollection, atlas:DigitalScholarlyEdition	dcterms:source	The cataloguing record of the main edited work(s)
	dcterms:references	The URL of a web resource that presents the main edited source(s)
	atlas:notesOnSource	Additional information on the edited text(s)
	atlas:referencedAuthor	The main author(s) of the edited text(s)
	atlas:referencedWorkType	The type of the edited text(s)
	schema:genre	The genre of the edited text(s)
atlas:DigitalScholarlyEdition	atlas:editionType	The type of edition
atlas:TextCollection	schema:size	The number of collected items
atlas:LinkedOpenData, atlas:Ontology	dcterms:references	Imported ontologies or vocabularies
atlas:Ontology	vann:preferredNamespaceUri	The preferred namespace URI to use terms from this vocabulary
atlas:Software	schema:archivedAt	The URL of the software's repository
	swo:0000086	The format of input data
	swo:0000087	The format of output data
	swo:0000741	Used programming language(s)
	schema:isBasedOn	Reused or extended software component(s)

4.1.2 People & Organisations

Identifying communities and scholars involved in scholarly outcomes represents one of the desiderata of the ATLAS Ontology. ATLAS distinguishes between `schema:Person` and `schema:Organisation`, allows users to record their current or most recent affiliation (`schema:affiliation`) and differentiates contribution roles to research outputs (see Research Product above). Common attributes of agents include their name (`schema:name`), external identifiers (`schema:identifier`), such as ORCID,[30] and links to authority records (`schema:sameAs`), e.g., Wikidata entities. For Organisations, additional details, such as their landing page (`schema:url`) and location (`schema:location`), are also captured.

4.1.3 Research Project

All reviewed models provide information on research activities supporting the production of an outcome. However, the focus is usually set on specific aspects, such as funding agencies, grants, and open-access mandates. ATLAS attempts to combine all such aspects and identify the main actors. To represent Research Projects, the class `schema:ResearchProject` is used. Following the hierarchical arrangement by Schema.org, this is a subtype of `schema:Organisation`, thus it inherits all its properties. In ATLAS we are interested in the following attributes: description (`schema:description`), start date (`schema:foundingDate`), end date (`schema:dissolutionDate`), organisations part of the project (`schema:member`), and funding entities (`schema:funder`).

4.1.4 Website & Computer program

Websites and tools that expose access points to research data play a pivotal role in enhancing the findability and reusability of scholarly outcomes. To provide an effective representation of these services, the ATLAS Ontology introduces two types: Websites (`fabio:WebSite`) and Computer Programs (`fabio:ComputerProgram`).

Computer Programs were previously mentioned in the context of Research Product subtypes. Specifically, `fabio:ComputerProgram` is one of the two parent types for `atlas:Software`. The description of a Computer Program includes the type of provided service (`dcterms:type`), the title (`schema:name`), a description (`schema:description`), the access URL (`schema:url`), a URL for a documentation page (`schema:usageInfo`), afforded research activities (`schema:educationalUse`), the license (`schema:license`), and links to other software components that the described program extends or reuses (`schema:isBasedOn`). A similar set of attributes is also available for Websites, except for `dcterms:type` and `schema:license`. In this context, the `schema:isBasedOn` expresses connections to domain-relevant tools (i.e., Computer Programs), such as deployed Visualisation software to present Digital Scholarly Editions.

The review of the current landscape of controlled vocabularies for scholarly data highlighted the lack of taxonomies to describe a few aspects relevant to DH resources. The ATLAS Ontology introduces several terms (named individuals) to address such an issue. For instance, we collected a preliminary list of different types of Digital Scholarly Editions (e.g., `atlas:BestManuscriptEdition`, `atlas:DiplomaticEdition`, `atlas:DocumentaryEdition`), created from the *Parvum Lexicon Stemmatologicum* [31], and categories of textual resources (e.g., `atlas:CollectedWorks`, `atlas:Paper`,

[30] https://orcid.org/.

`atlas:SingleManuscript`, etc.) from Patrick Sahle's Catalog of Digital Scholarly Editions [32].

4.2 ATLAS Knowledge Graph

The ATLAS Ontology has been populated with a preliminary Knowledge Graph (ATLAS-KG) [4] describing selected pilot projects and resources. The ATLAS-KG also served as a testing ground for validating the semantic model outlined in the previous paragraphs and testing the functionalities of the ATLAS platform. ATLAS-KG leverages SKOS Thesauri and Authority Records used in the DH community, such as the above-mentioned TaDiRAH and EU Vocabularies,[31] but also national controlled vocabularies (Schema.gov),[32] COAR,[33] Linked Open Vocabularies (LOV),[34] Wikidata,[35] VIAF,[36] Geonames,[37] ORCID, ROR.[38] The Knowledge Graph is organised in a number of Named Graphs, each corresponding to the content of a record in the ATLAS platform, filled in using a template, which in turn corresponds to a class/concept described above, namely: Research Product, Research Project, Person, Organisation, Computer program, and Website. Created data are available in their Turtle serialisations and accessible through the ATLAS Platform. To date, the graph accounts for 236 records, including 37 Research Products, 14 Research Projects, 96 instances of Person, 69 Organisations, and 20 Websites and Computer Programs.

Figure 2 provides a graphical example of the description of a Research Product, i.e., the Zeri Photo Archive RDF Dataset [13], the primary research outcome of the Zeri & LODE project. For the sake of brevity, only a few core statements are presented here, while a complete serialisation is available in the graph repository. Pink circles represent instances of ATLAS classes, with their types represented in yellow boxes.

[31] https://op.europa.eu/en/web/eu-vocabularies/controlled-vocabularies.

[32] https://schema.gov.it/.

[33] https://vocabularies.coar-repositories.org/.

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

[35] https://wikidata.org/.

[36] https://viaf.org/.

[37] https://geonames.org/.

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

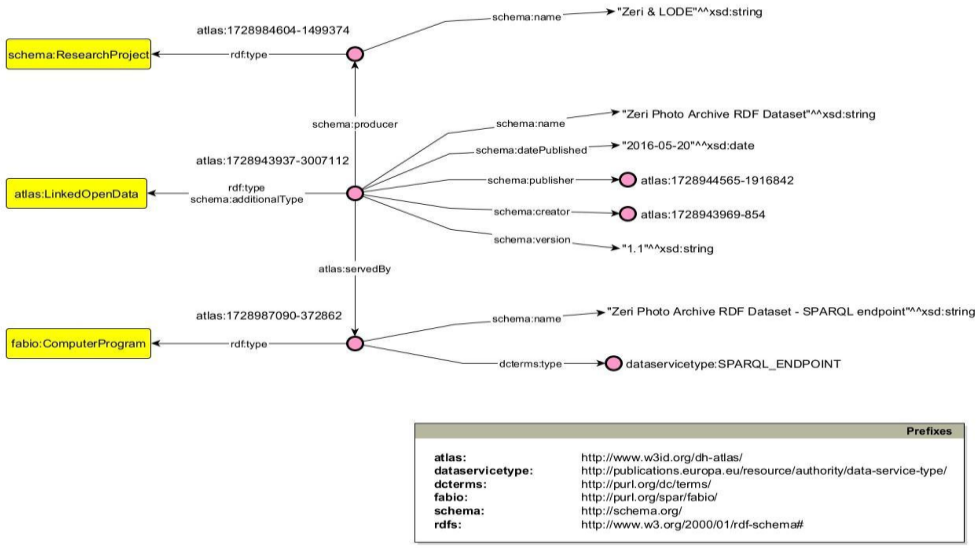


Figure 2: A visual diagram exemplifying the description of a Research Product and related entities.

Black arrows indicate predicates connecting entities to either entities or literal values. In the example, the Research Product named "Zeri Photo Archive RDF Dataset" (`atlas:1728943937-3007112`) is an instance of `atlas:LinkedOpenData` through the `rdf:type` and `schema:additionalType` properties. The relation with the Research Project responsible for its creation (`atlas:1728984604-1499374`), named "Zeri & LODE", is represented using the `schema:producer` property. Two object properties link the Research Product (`schema:publisher` and `schema:creator`) to the Agents (Person and Organisation) who contributed to its realisation. Lastly, `atlas:servedBy` connects the artefact to one of its access points, that is, an instance of `fabio:ComputerProgram` (`atlas:1728987090-372862`), labelled "Zeri Photo Archive RDF Dataset - SPARQL endpoint".

4.3 CLEF v3.0

The first version of the Knowledge Graph was created by leveraging the new functionalities provided by the latest release of the software CLEF. This tool is designed to facilitate the collaborative creation of LOD collections, thus providing a solid foundation for the development of the ATLAS web platform. Although the contribution here presented does not aim to address all potential technical requirements underlying a catalogue of scholarly data, it provides a number of features that current solutions have so far overlooked [14, 22], namely: (1) the usage of intermediate templates to prevent users from delving into the complexities of an ontology while entering data, (2) the possibility to fill in the record by semi-automatically extracting data from online data sources, and (3) provide customisable data visualisations based on the data created.

4.3.1 Intermediate Templates

CLEF supports Linked Open Data crowdsourcing by streamlining data entry processes. Users can create LOD by filling a user-friendly web form, wherein fields correspond to RDF properties and the record is an entity of a class. Each record complies with a template, i.e., a set of mandatory and optional fields/properties to be filled with appropriate values.

However, implementing complex data models could result in intricate templates and describing a single resource often requires creating and linking several records. For instance, in ATLAS, when creating the record of a Research Product, users must also define (1) Organisation and Person instances for related creators, contributors, and publishers, (2) the corresponding Research Project, and (3) available Computer Programs and Websites serving as access points. While in existing systems this would require users to create preliminary records for such secondary entities, and only then recall these entities in the main record, CLEF allows users to create multiple records at the same time using a mechanism of subtemplates, which graphically include fields for describing the secondary, ancillary entity along with the main one. Notably, the mechanism underlying this functionality is ontology-independent, and can be reused in any new template.

While this solution facilitates the implementation of complex data models on a practical level, other updates have focused on knowledge engineering improvements. These include allowing the association of multiple OWL classes with the same Template as well as the integration of Subclasses.

4.3.2 Enhanced Knowledge Extraction

The 2.0 version of CLEF introduced a working area for Knowledge Extraction, allowing users to retrieve named entities or Linked Open Data from various types of sources, including SPARQL endpoints, API services, and Static Files (.csv and .json formats) [12]. To query Static Files, CLEF 2.0 relies on SPARQL Anything,[39] a reengineering tool that facilitates SPARQL interrogations on diverse data formats and returns RDF data regardless of the input format.

ATLAS seeks to (gradually) make Knowledge Extraction accessible to users with more or less technical background, therefore overcoming the barrier posed by query languages. To achieve this goal, a Manual Extraction option has been introduced. This feature enables contributors to provide the URL of a document (i.e., a .json, .csv, or .xml file), which is automatically parsed to identify JSON keys, CSV columns, or XML tags. Users can then select desired elements through a suggestion dropdown to extract corresponding values. Additionally, filtering options can be specified, such as a minimum number of occurrences and regular expressions. In the end, provided parameters are automatically converted into a SPARQL Anything query.

To complete the Extraction process and return LOD, template creators can now configure fields by associating them with an automatic Entity Reconciliation system. So doing, extracted terms are matched to the most relevant URI in selected sources like Wikidata and VIAF.

4.3.3 Data visualisation

CLEF integrates new explorative tools for improving user interaction with cataloguing data. Specifically, the updated platform introduces a new Charts Template section, designed to support the editorial board in creating customised data visualisation interfaces. This feature allows one to combine and arrange several presentations, enriched with textual description. For

[39] https://sparql-anything.cc/.

greater customisation, contributors can use HTML tags and attributes can be used to modify captions, ensuring design flexibility. Available visualisations rely on SPARQL queries to extract data from the catalogue and showcase it by leveraging the amCharts js library.[40] Key options include a) Counters, displaying some key metrics as standalone numerical values associated with customisable labels (Figure 3), b) Charts, visualising trends and data distributions through a variety of chart types, including bar graphs, pie charts, and doughnut charts (Figure 4), c) Maps, providing geographic representations of data by plotting resource distribution on interactive maps (Figure 5).

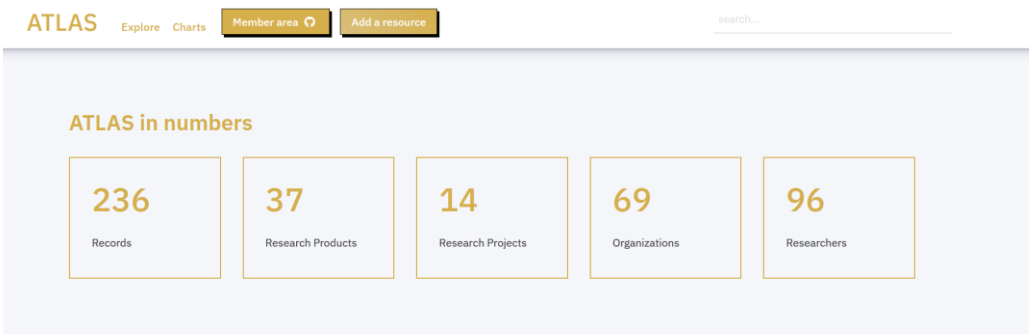


Figure 3: An example of counters displaying statistical data from the ATLAS Knowledge Graph.

Figure 4: A bar chart illustrating statistics extracted from the ATLAS Knowledge Graph.

[40] https://amcharts.com/.

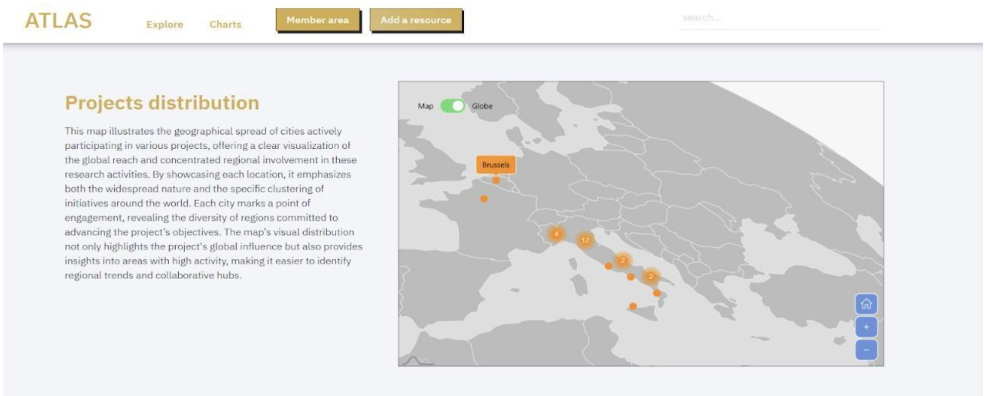


Figure 5: An example of an interactive map based on the ATLAS Knowledge Graph.

4.4 The ATLAS platform

The ATLAS platform is built to fulfill two fundamental aims: (1) increasing the findability of described resources and (2) streamlining user participation in the growth of the ATLAS Knowledge Graph through user-friendly Web Forms. In this Section, we illustrate these goals in practice by analysing a case study from the aforementioned pilot resources: the *Codice Pelavicino Digital Edition*.[41] This is a TEI-XML encoded digital scholarly edition of a 13th-century manuscript preserved in Sarzana (Italy) produced by a research project launched in 2014 by prof. Enrica Salvatori. The edition is published as a website via the EVT visualisation tool and it presents both transcriptions and images of the codex [33]. The digital edition, along with all its related entities—the research project, the software, the website, as well as the organisations and individuals involved—are documented in the ATLAS catalogue.

This example highlights the descriptive and exploratory solutions implemented by the ATLAS catalogue, along with their contribution to improving data usability and accessibility.

4.4.1 Expanding the ATLAS knowledge graph

As outlined in Section 4.1, the ATLAS ontology is designed with a major emphasis on Research Products resulting from international scholarly research on Italian Digital Cultural Heritage. While standard Content Management Systems (CMS) like Omeka S[42] support the individual definition of semantic entities, Research Products require linking them to multiple contextual entities (e.g., people and organisations participating in the creation of the product) that may or may not already exist at the time a user starts the description. This creates the need for a more agile solution that allows users to define new entities and interlink various resources within a single record. By leveraging the new Intermediate Templates feature developed in CLEF 3.0, the ATLAS platform aims to meet this requirement and make research outcomes the first-class citizen of the data-entry tasks, while preventing cumbersome operations such as opening several tabs and creating multiple records simultaneously.

[41] https://pelavicino.labcd.unipi.it/.

[42] https://omeka.org/s/.

Users can add a new resource by navigating to the dedicated section of the platform. A dropdown, as shown in Figure 6, allows them to select the Research Product template from the available ones. Once the template is selected, the corresponding Web Form is displayed, with each input field matching one of the RDF properties introduced in Section 4.1.

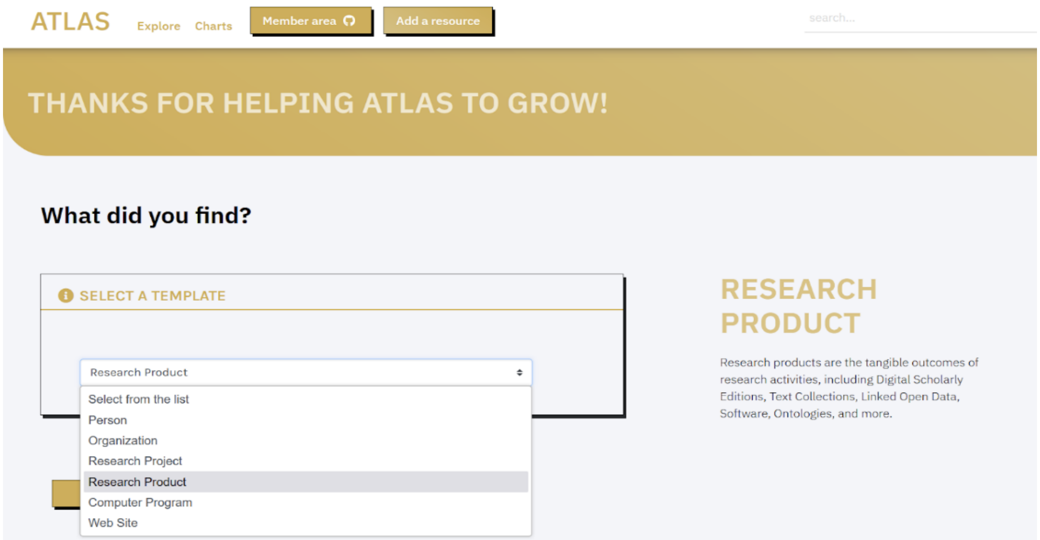


Figure 6: Example of selecting a Template to define a new ATLAS resource.

Figure 7 presents the interface designed for describing the Research Product at hand, featuring some input fields that are shared across the types of research products. To ensure accessibility for all levels of users, no URIs or technical details are shown–only labels and brief descriptions are provided to guide contributors in completing the form. Moreover, the data entry process is supported by a set of automatic suggestion systems, specifically devised to facilitate data reuse and entity reconciliation from both the ATLAS catalogue and external Linked Open Data resources (e.g., Wikidata, VIAF, and SKOS Thesauri) while preventing duplication of already described resources.

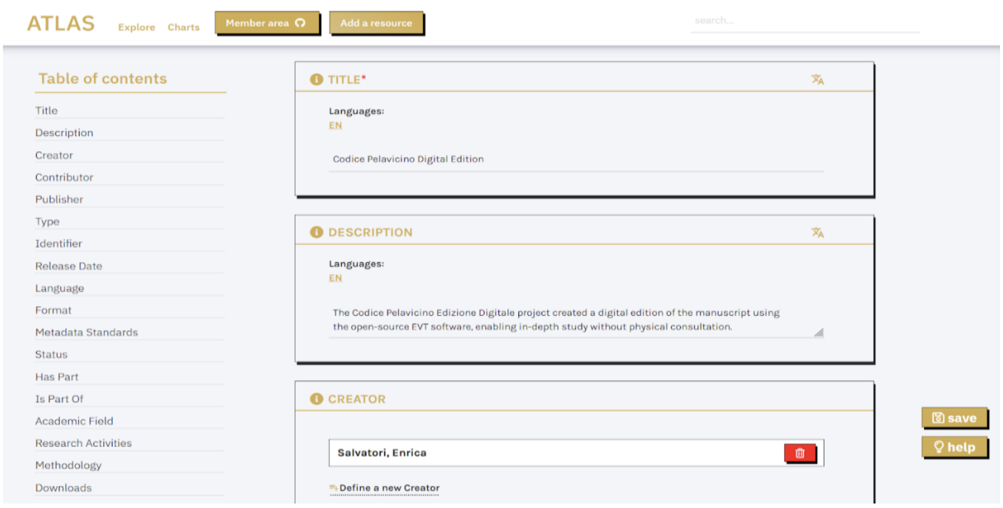


Figure 7: The interface of a Web Form designed for describing a Research Product.

For instance, the "Creator" input field included in the template of the Research Product (Figure 8) corresponds to the `schema:creator` property and exemplifies the use of Intermediate Templates. It is intended to be populated with values representing either Organization or Person entities that may already exist in the catalogue or that have to be created *ex novo*. Specifically, the button "Define a new Creator" enables users to dynamically import required subforms for the description of contextual entities. In this case, the button imports in the current record the template for describing a Person.

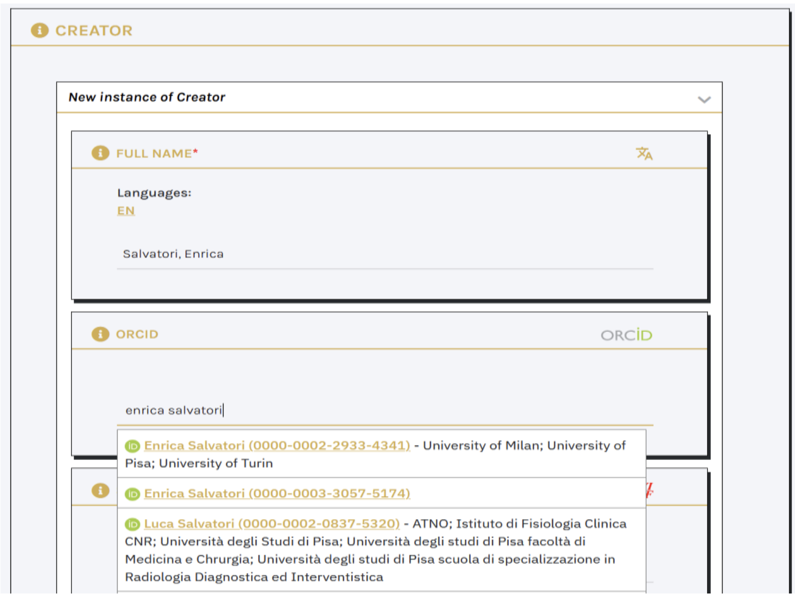


Figure 8: Example of subrecord creation through Intermediate Templates.

After filling in fields shared across any type of research product, an input field is dedicated to the classification of the described Research Product (`schema:additionalType`). Available values, as shown in Figure 9, include the current five subclasses of `schema:Dataset`, as defined by the ATLAS Ontology. Upon selection, the system dynamically displays a new set of additional input fields to provide detailed information specific to the selected type of outcome.

Figure 9: Example of selecting the subclass of a Research Product.

For instance, in the case of the *Codice Pelavicino*, which is a Digital Scholarly Edition, users can describe the particular type of edition and enter a reference to the edited text. Additional elements also include a Knowledge Extraction field, designed to retrieve meaningful entities from various static files (e.g., .csv and .json), including XML/TEI-encoded texts from digital editions [34]. Figure 10 illustrates the parameters for performing this extraction. Users must provide the URL of the XML file containing the resource. Based on this, the system performs an initial parsing of the file to identify all XML tags in the text. These tags are then automatically suggested, and users can select them to specify which data to extract. Optional filters can be applied, such as regular expressions or a minimum number of occurrences, to refine the extraction.

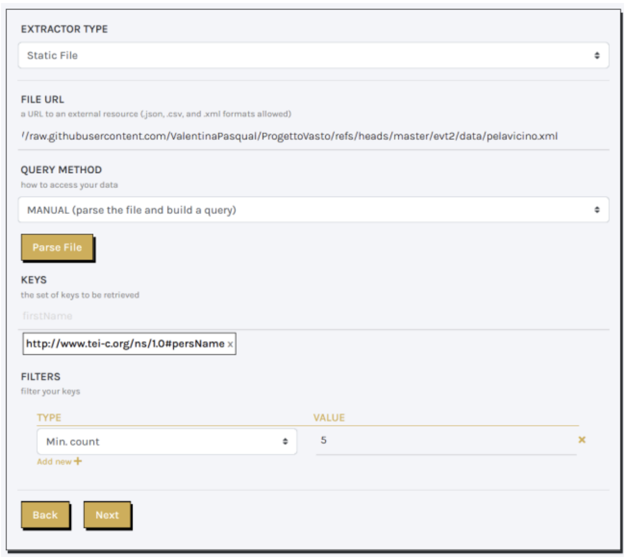
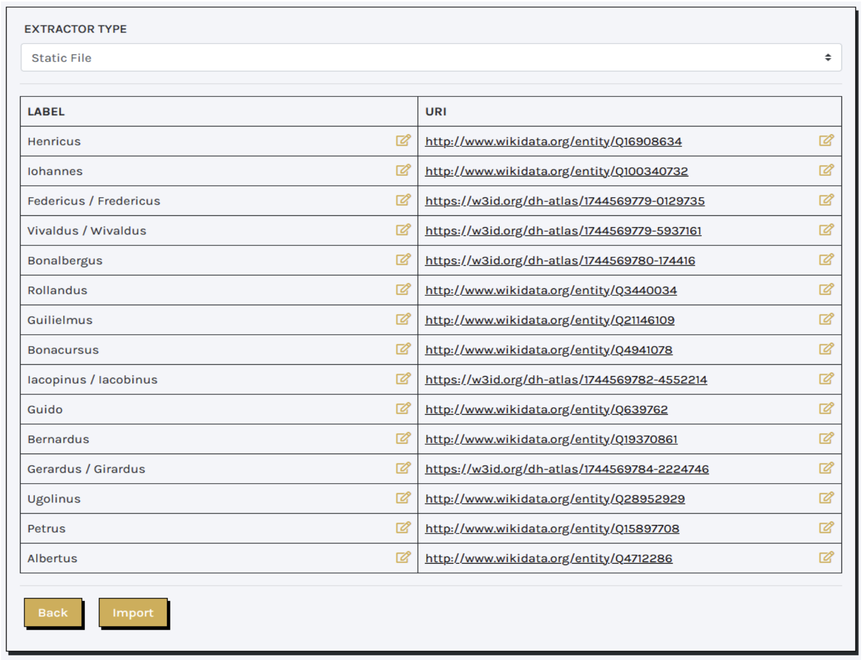


Figure 10: Example of extracting Knowledge from an XML/TEI document.

Extracted strings are automatically aligned with best matches from Wikidata or other reconciliation services, assigning the URI of a controlled entity to each entry. As illustrated in Figure 11, users can review and manually adjust these values before completing the extraction by importing the results. Through this instrument, the content of large documents can be easily captured without extensive curatorial effort.



LABEL	URI
Henricus	http://www.wikidata.org/entity/Q16908634
Iohannes	http://www.wikidata.org/entity/Q100340732
Federicus / Fredericus	https://w3id.org/dh-atlas/1744569779-0129735
Vivaldus / Wivaldus	https://w3id.org/dh-atlas/1744569779-5937161
Bonalbergus	https://w3id.org/dh-atlas/1744569780-174416
Rollandus	http://www.wikidata.org/entity/Q3440034
Guilielmus	http://www.wikidata.org/entity/Q21146109
Bonacursus	http://www.wikidata.org/entity/Q4941078
Iacopinus / Iacobinus	https://w3id.org/dh-atlas/1744569782-4552214
Guido	http://www.wikidata.org/entity/Q639762
Bernardus	http://www.wikidata.org/entity/Q19370861
Gerardus / Girardus	https://w3id.org/dh-atlas/1744569784-2224746
Ugolinus	http://www.wikidata.org/entity/Q28952929
Petrus	http://www.wikidata.org/entity/Q15897708
Albertus	http://www.wikidata.org/entity/Q4712286

Figure 11: Knowledge Extraction results on an XML/TEI document.

4.4.2 Exploring the ATLAS knowledge graph

When all information has been entered into the web form and saved, a new Named Graph is generated and uploaded to the triplestore for each newly described entity. All new records must be reviewed by an ATLAS project member before being published and made accessible on the catalogue's "Explore" page. This section of the platform includes all reviewed records described in the Knowledge Graph, grouped by ontological class and accessible through filters and facets.

Figure 12 shows three types of filters. First, Research Products are filtered based on the subclass of `schema:Dataset` associated with the `schema:additionalType` property. This facet prunes results of the following filters, which show products sorted alphabetically (Title), or based on specific values (e.g., Creator). For example, users can access the record of the *Codice Pelavicino Digital Edition*, a Digital Scholarly Edition, by selecting its title's initial letter or its creator's name, "Salvatori, Enrica".

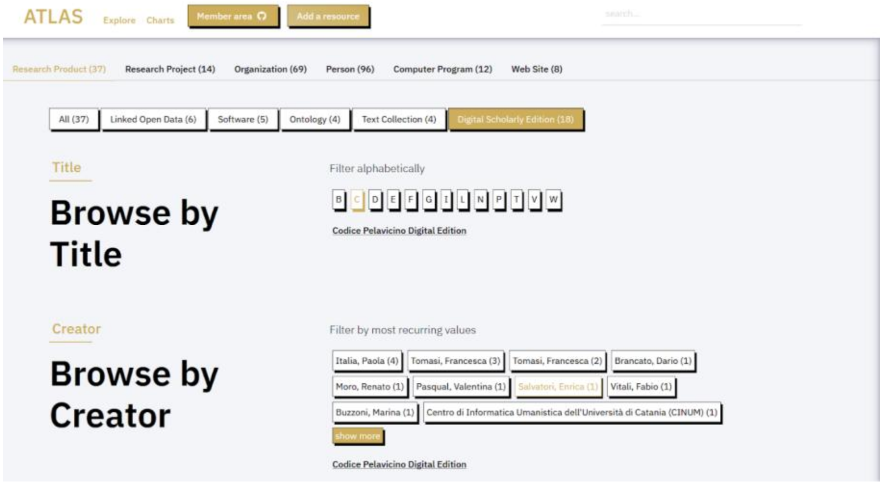


Figure 12: View of the ATLAS Explore page.

The record for the *Codice Pelavicino Digital Edition*, shown in Figure 13, includes texts and linked entities, the latter characterised by a number of actions. Properties like Creator and Contributor address entities that are either imported from external resources (e.g., Wikidata) or created by ATLAS users. For these values, expandable boxes allow users to view detailed information. Alternatively, by clicking the external link icon next to the entity's label, users are directed to an intermediate page. As shown in Figure 14, this interface offers an overview of the selected term, including a link to the entity's dedicated record as well as a list of all records in which the selected entity appears.

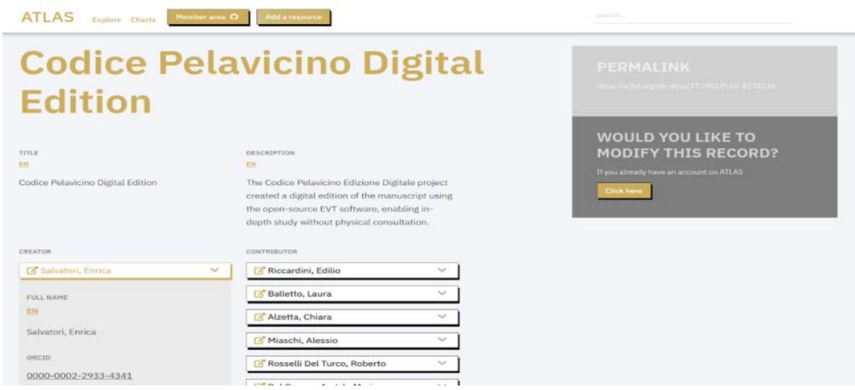


Figure 13: View of a record on the ATLAS platform.

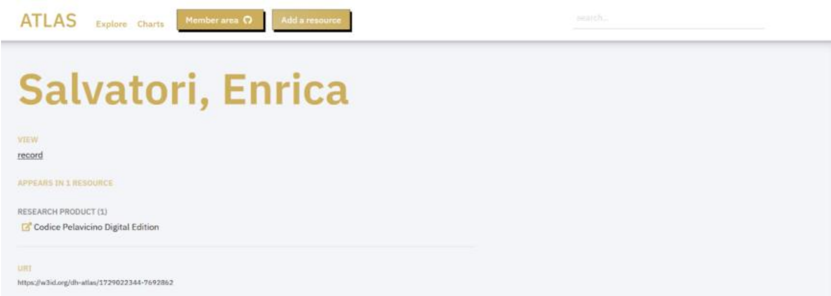


Figure 14: View of a term-page on the ATLAS platform.

5. User Evaluation

To evaluate the validity of the first version of the ATLAS ontology, we adopted an application-based approach, where a group of users completed a task, i.e., the creation of a record for a Research Product, and provided feedback on the practical use of the ATLAS platform. Answers were collected via web form, including both open-ended and closed answers. Thanks to this approach, the ontology effectiveness is evaluated through its integration and functionality within the application. The understandability, and usefulness of the ontology emerge directly from the practical user experience provided by the platform, and the feedback provided on the latter by users can be inherited by the ontology itself.

The evaluation form [3] was designed to collect structured feedback, enabling both quantitative and qualitative analyses across three fundamental dimensions of usability defined by ISO [25]:

- Efficacy: the application's ability to fulfill the objectives for which it was designed;
- Efficiency: the speed and ease with which useful results are obtained;
- Satisfaction: users' perceived satisfaction while interacting with the system.

The questionnaire was conducted with 18 users—including professors (2), researchers (5), as well as master's (1) and PhD students (10)—during a hands-on session as part of the ATLAS Workshop.

With regard to efficacy, users answered two questions, both of which are summarised in Figure 15. The first was a closed-ended question aimed at directly measuring whether they achieved the intended goal. The second asked participants to rate, on a 5-point Likert Scale[43], how easily they could gather information from the initial resource.

For the first question, results show that 72.2% of users were able to achieve their task, while 22.2% faced some challenges due to missing information about the described resource or application bugs during form completion. One participant (5.6%) failed to complete the task because of insufficient information.

[43] A Likert Scale is a measurement scale used to assess the intensity of agreement or disagreement of a respondent with a statement, typically on a 5-point scale where 1 indicates "strongly disagree", 5 "strongly agree" and 3 represents a neutral position.

Regarding the second question, 55.6% of participants (10) gave a neutral rating (3 on the 5-point Likert scale) for the manual retrieval of information needed to create a record in ATLAS. Difficulties were mainly caused by the unavailability of some required information in the original resource documentation, the complex or poorly organized structure of the available materials, and the challenges in accurately identifying the publication and release dates of the resources. At the same time, 27.8% of participants found it easy, and 11.1% found it very easy, to collect information from the original data sources.

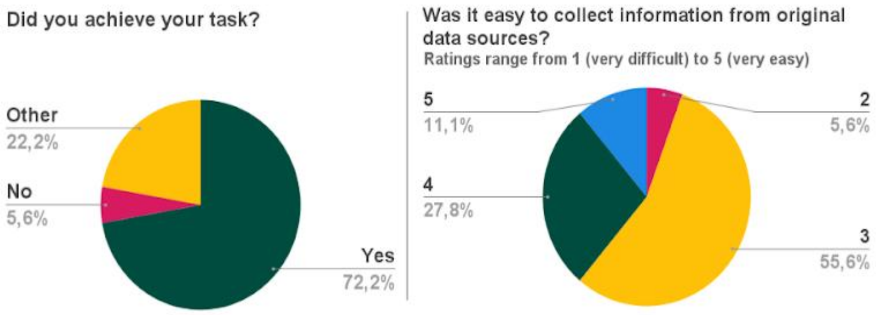


Figure 15: Efficacy evaluation. The charts measure the ease perceived by ATLAS users in achieving the objectives and retrieving the necessary information from the initial resources.

Efficiency was assessed through two questions aimed at measuring the clarity and usability of the data entry interface. Users were asked to indicate how understandable they found the interface and how easy it was to use (Figure 16). Around 61% deem the interface understandable, while the remaining 38.9% do not provide a strong opinion. No negative impressions are recorded. Likewise, 66.7% of users believe the interface is easy to use, while 5.6% report issues.

Figure 16: Efficiency evaluation. The charts measure the clarity and usability of the ATLAS data entry interface.

Finally, the satisfaction dimension explored the overall level of user satisfaction and the perceived usefulness of the platform. To this end, two questions were asked concerning the overall user experience and the perceived value of a tool like ATLAS (Figure 17). 83.3% of participants record a high level of satisfaction, and the totality of participants believe the platform is useful.

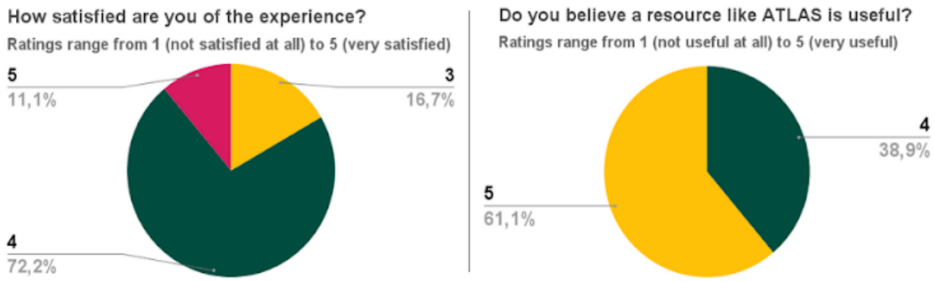


Figure 17: Satisfaction evaluation. The charts measure the level of satisfaction expressed by ATLAS users and the perceived usefulness of the platform.

Qualitative feedback highlighted several strengths. Users particularly appreciated the clarity of the interface, its exploratory features, the data entry system, the open-source and collaborative catalog. Other comments mentioned the importance of the White Book [27]. However, some concerns on the behaviour of the interface emerged, e.g., occasional slow response when changing language, overlay elements bothering in smaller screens or the lack of confirmation after saving the record, as well as the design limitation that prevents users from reviewing or modifying the entered data after submission. A few comments can be directly related to the use of the ontology. While we have already mentioned a general appreciation and perception of its usefulness, some difficulties arose in understanding certain specific input requirements, e.g., distinguishing between research activities and methodologies, and separating research projects from research products.

6. Discussion

This feedback provides valuable insights for further improving both the ATLAS platform and its underlying ontological model, while overall confirming the effectiveness and usefulness of both the platform and the ontology, although for a selected set of scholarly outcomes.

A key strength of the ATLAS ontology is that each type of research product is formalized as a dedicated class. This modelling approach allows research products to be represented at very fine levels of granularity while making the model flexible and easily expandable. To represent other aspects of these products, we will need to add new properties within the already defined framework. Another strength of the ontology is its novel approach to describe digital scholarly editions and text collections, treating them primarily as datasets. This emphasizes features and methodologies specific to the digital paradigm that distinguish these research products from their print-oriented versions. In contrast, traditional catalogues typically present textual archives and (digital) scholarly editions through content-focused editorial information—which remains essential and is included in our model.

The ATLAS ontology needs further refinements to effectively represent the Italian landscape of Digital Cultural Heritage and its research products. To enhance and refine the ontology, we will extend our current set of research projects and their outcomes, then incorporate their definitions into the knowledge graph. While we have so far focused on scholarly projects with a strong focus on the Italian Cultural Heritage, we will now select pilot projects from a broader range of DH fields, using national and international standards and best practices as guidelines.

In particular, the soon-to-be-released version of the ontology will include a wider selection of scholarly output types. As described in section 3, the ontology was initially designed using a selected pool of pilot projects as reference, which led to the creation of five subclasses describing specific types of research products: Digital Scholarly Edition, Ontology, Software, Text Collection, and Linked Open Data. While these categories cover a significant portion of scholarly outcomes in the DH landscape, additional types are needed for a complete catalogue. For example, the ATLAS ontology should include models used in computational linguistics and natural language processing, 3D models created in the archaeological field, images and digital art objects, audio resources and different kinds of structured datasets. The inclusion of new research product categories is therefore necessary to effectively scale the ATLAS ontology and knowledge graph.

Additional types of research products can be easily included in the model by formalizing them as new subclasses of Research Product (`schema:Dataset`). However, we are reassessing the definition of the Linked Open Data class to find a more flexible and scalable way to model all kinds of datasets. In particular, we are considering creating a broadly applicable class to describe any type of structured data collection and specifying the type of data it represents with a dedicated property. The types[44] used within the ARIADNE infrastructure can guide the selection. Further refinements are needed to improve the completeness of the ontology in representing the types of scholarly outcomes already included in the model. For instance, when describing text collections and digital scholarly editions, the `schema:genre` property should be bound to a controlled vocabulary of terms from the CWRC genre ontology.[45] Though not comprehensive, it offers a practical framework for describing main textual genres—from journalism to poetry, from drama to advertisement. Moreover, as mentioned above, the DH field still lacks a complete formal definition of edition types. Consequently, we need to investigate how to appropriately identify the value set for the `atlas:editionType` property. Lastly, while some text collection properties currently capture detailed information about individual text sources (`dcterms:references` and `atlas:notesOnSource`), this granular approach has proven impractical and redundant for manual data entry. The next version of the ontology will remove these properties, and include two broader ones instead to describe collections' geographical areas and temporal coverage.

Additionally, the ontology will be extended to include relations between data sources and entities extracted from data sources themselves, such as places, people, and organisations. We will assign

[44] https://rdf-vocabulary.ddialliance.org/ddi-cv/GeneralDataFormat/2.0.3/GeneralDataFormat.html.

[45] The CWRC Genre Ontology is used by the Canadian Writing Research Collaboratory to assign genres to different types of cultural objects (https://sparql.cwrc.ca/ontologies/genre.html).

a class to each type of extracted entity and link records to entities using the `schema:mentions` property.

7. Conclusions

The ATLAS Ontology seeks to enhance the description of Digital Cultural Heritage projects and their related outcomes by leveraging the potential of Linked Open Data. To this end, it integrates properties and entities from some of the most relevant semantic models within the DH domain and Schema.org, and provides terms to address the description of peculiarities relevant to scholars in the Humanities.

To evaluate the model, we extended the functionalities of CLEF and developed the ATLAS platform, through which we created the ATLAS Knowledge Graph, including metadata of selected pilot projects. The newly implemented features, including intermediate templates, advanced knowledge extraction, and data visualisation tools, provided us with the instruments for populating and validating the ontology through the creation of a Knowledge Graph.

The level of granularity introduced by the ATLAS Ontology shows great potential for performing detailed data analyses on the Italian Cultural Heritage and its relation with Digital Humanities outcomes. In particular, its terminology has proven to effectively capture and describe different types of Research Products among selected resources, covering peculiar aspects such as DH methodologies. However, while the ontology provides a solid base for addressing a shared terminology, we will perform further user tests to prove the goodness of our solutions and improve the terminology with user-contributed terms, so as to allow diversity and richness in the way scholars describe their results. Future developments will indeed expand ATLAS vocabularies, enabling better handling of this crucial gap and increasing the coverage of underrepresented concepts.

The extension of CLEF functionalities with scalable methods for Knowledge Extraction effectively simplifies this descriptive process by leveraging the Linked Open Data potential. Nonetheless, the road to facilitate LOD generation via user-friendly interfaces still poses a number of challenges, due to the variety of technical skills of scholars who would provide descriptions of their data. For this reason, the next stages of the ATLAS project will focus on extending the current Knowledge Graph through the analysis of new research initiatives. The insights and issues emerging from this process will inform the efforts to consolidate the developed model, while further usability tests will contribute to delivering a refined crowdsourcing platform. In so doing, ATLAS aims to offer an increasingly comprehensive tool, capable of advancing research in the DH domain and fostering the full valorisation of Italian Cultural Heritage.

8. Acknowledgments

Funded by the European Union- Next Generation EU, Mission 4 Component 1 CUP J53D23013000006.

9. References

[1] Amodeo, Stefania, Alane Brunschweiger, Franziska Krauss, Giulia Malaguarnera, and OpenAIRE Non-Profit Civil Partnership. 'The OpenAIRE Graph: Empowering Research through Open Science', 3 October 2024. https://doi.org/10.5281/ZENODO.13885430.

[2] Amodeo, Stefania, Thanasis Vergoulis, Elli Papadopoulou, Matthew Buys, Andrea Mannocci, and Giulia Malaguarnera. 'EOSC Collaborative Frontiers to Achieve Interoperability and Enhance Scholarly Data'. 22 October 2024. https://doi.org/10.5281/zenodo.14055894.

[3] Bardi, Alessia, Marina Buzzoni, Marilena Daquino, Riccardo Del Gratta, Angelo Mario del Grosso, Franz Fischer, Sebastiano Giacomini, et al. 'DH ATLAS: Datathon Survey Responses v1.0'. Zenodo, 8 April 2025. https://doi.org/10.5281/ZENODO.15173474.

[4] Bardi, Alessia, Marina Buzzoni, Marilena Daquino, Riccardo Del Gratta, Angelo Mario del Grosso, Franz Fischer, Roberto Rosselli Del Turco, et al. 'DH ATLAS: Knowledge Graph v1.0'. Zenodo, 8 November 2024. https://doi.org/10.5281/zenodo.14058144.

[5] Bardi, Alessia, Marina Buzzoni, Marilena Daquino, Riccardo Del Gratta, Angelo Mario del Grosso, Franz Fischer, Roberto Rosselli Del Turco, et al. 'DH ATLAS: Ontology v1.0', 8 November 2024. https://zenodo.org/records/14058232.

[6] Bechhofer, Sean, Iain Buchan, David De Roure, Paolo Missier, John Ainsworth, Jiten Bhagat, Philip Couch, et al. 'Why Linked Data Is Not Enough for Scientists'. *Future Generation Computer Systems* 29, no. 2 (2013): 599–611. https://doi.org/10.1016/j.future.2011.08.004.

[7] Brogan, Martha L. 'Survey of Digital Library Aggregation Services', Digital Library Federation. 2003. http://old.diglib.org/pubs/dlf101/dlf101.htm.

[8] Brügger, Niels, and Niels Ole Finnemann. 'The Web and Digital Humanities: Theoretical and Methodological Concerns'. *Journal of Broadcasting & Electronic Media* 57, no. 1 (2013): 66–80. https://doi.org/10.1080/08838151.2012.761699.

[9] Carriero, Valentina Anita, Aldo Gangemi, Maria Letizia Mancinelli, Ludovica Marinucci, Andrea Giovanni Nuzzolese, Valentina Presutti, and Chiara Veninata. 'ArCo: The Italian Cultural Heritage Knowledge Graph'. In *The Semantic Web – ISWC 2019*, edited by Chiara Ghidini, Olaf Hartig, Maria Maleshkova, Vojtěch Svátek, Isabel Cruz, Aidan Hogan, Jie Song, Maxime Lefrançois, and Fabien Gandon, 11779:36–52. Cham: Springer International Publishing, 2019. https://doi.org/10.1007/978-3-030-30796-7_3.

[10] Ciotti, Fabio. 'A landscape of Digital Humanities in Italy: backgrounds, institutions and infrastructures.' Testo e senso 16 (2015): 195-213.

[11] 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 Di Silvestro, Antonio and Daria Spampinato (edited by), *Me.Te. Digitali. Mediterraneo in rete tra testi e contesti, Proceedings del XIII Convegno Annuale AIUCD2024*, pp. 588–592, 2024. https://doi.org/10.6092/unibo/amsacta/7927.

[12] Daquino, Marilena, Laurent A. Fintoni, Sebastiano Giacomini, and Francesca Tomasi. 'CLEF 2.0. Soluzioni per la catalogazione nativa Linked Data del patrimonio digitale culturale italiano'. In Di Silvestro, Antonio and Daria Spampinato (edited by), *Me.Te. Digitali. Mediterraneo in rete tra testi e contesti, Proceedings del XIII Convegno Annuale AIUCD2024*, pp. 588–592, 2024. https://doi.org/10.6092/unibo/amsacta/7927.

[13] Daquino, Marilena, Francesca Mambelli, Silvio Peroni, Francesca Tomasi, and Fabio Vitali. 'Zeri Photo Archive RDF Dataset'. Centro Risorse per la Ricerca (CRR-MM), Università di Bologna, 18 May 2016. https://doi.org/10.6092/unibo/amsacta/5497.

[14] Daquino, Marilena, Mari Wigham, Enrico Daga, Lucia Giagnolini, and Francesca Tomasi. 'CLEF. A Linked Open Data Native System for Crowdsourcing'. *Journal on Computing and Cultural Heritage* 16, no. 3 (30 September 2023): 1–17. https://doi.org/10.1145/3594721.

[15] Di Giorgio, Sara. 'Gli archivi del MiBACT L'integrazione dei dati archeologici digitali'. In *Proceedings del Workshop L'integrazione dei dati archeologici digitali - Esperienze e prospettive in Italia* (Lecce, Italia, 1-2 ottobre, 2015), pp. 47–55. https://ceur-ws.org/Vol-1634/paper6.pdf.

[16] Fenlon, Katrina, Jacob Jett, and Carole L. Palmer. 'Digital Collections and Aggregations', in *Digital Humanities Data Curation*. 2017. https://guide.dhcuration.org/contents/digital-collections-and-aggregations/.

[17] Fintoni, Laurent A. 'Rethinking Scholarly Digital Objects as Cultural Heritage: The KNOT Project'. In Di Silvestro, Antonio and Daria Spampinato (edited by), *Me.Te. Digitali. Mediterraneo in rete tra testi e contesti, Proceedings del XIII Convegno Annuale AIUCD2024*, pp. 588–592, 2024. https://doi.org/10.6092/unibo/amsacta/7927.

[18] Franzini, Greta, Melissa Terras, and Simon Mahony. 'A Catalogue of Digital Editions'. In Driscoll, Matthew James, and Elena Pierazzo, eds. *Digital Scholarly Editing: Theories and Practices*. 1st ed. Vol. 4 (pp. 161–182). Digital Humanities Series. Cambridge, UK: Open Book Publishers, 2016. https://doi.org/10.11647/OBP.0095.

[19] Freire, Nuno, Valentine Charles, and Antoine Isaac. 'Evaluation of Schema.Org for Aggregation of Cultural Heritage Metadata'. In *The Semantic Web*, edited by Aldo Gangemi, Roberto Navigli, Maria-Esther Vidal, Pascal Hitzler, Raphaël Troncy, Laura Hollink, Anna Tordai, and Mehwish Alam, 225–39. Cham:

Springer International Publishing, 2018. https://doi.org/10.1007/978-3-319-93417-4_15.

[20] Frosini, Luca, Alessia Bardi, Paolo Manghi, and Pasquale Pagano. 'An Aggregation Framework for Digital Humanities Infrastructures: The PARTHENOS Experience'. *SCIRES-IT - SCIentific RESearch and Information Technology* 8, no. 1 (11 July 2018). https://doi.org/10.2423/i22394303v8n1p33.

[21] Giacomini, Sebastiano, Alessia Bardi, Marina Buzzoni, Marilena Daquino, Riccardo Del Gratta, Angelo Mario Del Grosso et al. 'ATLAS: Towards a Knowledge Graph of International Scholarly Research on the Italian Digital Cultural Heritage'. 2025. https://ceur-ws.org/Vol-3937/paper4.pdf.

[22] Giacomini, Sebastiano, Marilena Daquino, Francesca Tomasi, and Laurent Antoine Fintoni. 'CLEF 2.0. Solutions for Native Linked Data Cataloguing of Italian Digital Cultural Heritage'. *JLIS.It* 16, no. 1 (14 January 2025): 108–26. https://doi.org/10.36253/jlis.it-611.

[23] Griseldaonline. 'Gli strumenti dell'Italianistica digitale', 2020. https://site.unibo.it/griseldaonline/it/strumenti/strumenti-italianistica-digitale.

[24] Hall, Crystal. 'Digital Humanities and Italian Studies: Intersections and Oppositions'. *Italian Culture* 37, no. 2 (3 July 2019): 97–115. https://doi.org/10.1080/01614622.2019.1717754.

[25] International Organization for Standardization. *ISO 9241-11:2018 - Ergonomics of human-system interaction — Part 11: Usability: Definitions and concepts*. ISO, 2018. https://www.iso.org/standard/63500.html.

[26] Lorenzini, Matteo, Marco Rospocher, and Sara Tonelli. 'On assessing metadata completeness in digital cultural heritage repositories.' *Digital Scholarship in the Humanities* 36, no. Supplement_2 (2021): ii182-ii188.

[27] Martignano, Chiara, Alessia Bardi, Marina Buzzoni, Marilena Daquino, RICCARDO DEL GRATTA, ANGELO MARIO DEL GROSSO, Franz Fischer, et al. 'DH ATLAS: White Book v1.2'. Zenodo, 25 February 2025. https://doi.org/10.5281/ZENODO.14925266.

[28] Peroni, Silvio, and David Shotton. 'FaBiO and CiTO: Ontologies for Describing Bibliographic Resources and Citations'. *Journal of Web Semantics* 17 (2012): 33–43. https://doi.org/10.1016/j.websem.2012.08.001.

[29] Peroni, Silvio, Francesca Tomasi, and Fabio Vitali. 'The Aggregation of Heterogeneous Metadata in Web-Based Cultural Heritage Collections: A Case Study'. *International Journal of Web Engineering and Technology* 8, no. 4 (2013): 412. https://doi.org/10.1504/IJWET.2013.059107.

[30] Rettberg, Najla, and Birgit Schmidt. 'OpenAIRE - Building a Collaborative Open Access Infrastructure for European Researchers'. *LIBER Quarterly: The Journal of the Association of European Research Libraries* 22, no. 3 (29 November 2012): 160–75. https://doi.org/10.18352/lq.8110.

[31] Roelli, Philipp, and Caroline Macé. 'Parvum Lexicon Stemmatologicum. A Brief Lexicon of Stemmatology'. Helsinki: Helsinki University Homepage, 1 November 2015. https://doi.org/10.5167/uzh-121539.

[32] Sahle, Patrick et al. 'A catalog of Digital Scholarly Editions'. v.4.112 2020ff. 2020. https://www.digitale-edition.de.

[33] Salvatori, Enrica, Roberto Rosselli Del Turco, Chiara Alzetta, Chiara Di Pietro, Chiara Mannari, and Alessio Miaschi. 'Il Codice Pelavicino tra edizione digitale e Public History'. *Umanistica Digitale*, no. 1 (1 October 2017). https://doi.org/10.6092/issn.2532-8816/7232.

[34] Salvatori, Enrica, and Vittore Casarosa Vittore. 'Codice Pelavicino', ILC-CNR for CLARIN-IT repository hosted at Institute for Computational Linguistics "A. Zampolli", National Research Council, in Pisa, 2024, http://hdl.handle.net/20.500.11752/OPEN-1012.

[35] Scarpa, Erica, and Riccardo Valente. "A RESOURCE HUB FOR INTEROPERABILITY AND DATA INTEGRATION IN HERITAGE RESEARCH: THE H-SETIS DATABASE." Archeologia e Calcolatori 35, no. 1 (2024).

[36] Schema.org. 'Data and Datasets overview'. 2024. https://schema.org/docs/data-and-datasets. html.

[37] Sefton, Peter, Eoghan Ó Carragáin, Stian Soiland-Reyes, Oscar Corcho, Daniel Garijo, Raul Palma, Frederik Coppens, et al. 'RO-Crate Metadata Specification 1.1.3', 26 April 2023. https://doi.org/10.5281/zenodo.7867028.

[38] Sicilia, Miguel-Angel, Elena García-Barriocanal, and Salvador Sánchez-Alonso. 'Community Curation in Open Dataset Repositories: Insights from Zenodo'. *Procedia Computer Science* 106 (2017): 54–60. https://doi.org/10.1016/j.procs.2017.03.009.

[39] Soiland-Reyes, Stian, Peter Sefton, Mercè Crosas, Leyla Jael Castro, Frederik Coppens, José M. Fernández, Daniel Garijo, et al. 'Packaging Research Artefacts with RO-Crate'. Edited by Silvio Peroni. *Data Science* 5, no. 2 (20 July 2022): 97–138. https://doi.org/10.3233/DS-210053.

[40] Tomasi, Francesca. 'Digital humanities e organizzazione della conoscenza: una pratica di insegnamento nel LODLAM'. *AIB Studi* 60, no. 2 (28 May 2020). https://doi.org/10.2426/aibstudi-12068.

[41] Tomasi, Francesca. *Organizzare la conoscenza: Digital Humanities e Web semantico*. IT: Editrice Bibliografica, 2022. https://doi.org/10.53134/9788893573573