info@atrium-research.eu
atrium-research.eu

ATRIUM – Advancing Frontier Research in the Arts and Humanities

Work Package WP3

Facilitating Discoverability of and Access to Humanities Resources

Deliverable D3.1

Overview of models and formats in the field

Funding Instrument:	Horizon Europe
Call:	HORIZON-INFRA-2023-SERV-01
Call Topic:	Research infrastructure services to support health research, accelerate the green and digital transformation, and advance frontier knowledge (2023)
Project Start:	1 Jan 2024
Project Duration:	48 months
Beneficiary in Charge:	DARIAH
Document Identifier:	https://doi.org/10.5281/zenodo.14575783

Funded by the European Union

Funded by the European Union. Grant Agreement number 101132163. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Executive Agency. Neither the European Union nor the granting authority can be held responsible for them.

Deliverable Information

Action Number:	101132163
Action Acronym:	ATRIUM
Action title:	Advancing FronTier Research In the Arts and hUManities
Deliverable Number:	D3.1
Deliverable Full Title:	Overview of models and formats in the field
Document Identifier:	ATRIUM-D3.1-Overview-of-models-and-formats-in-the-field
Beneficiary in Charge:	ARIADNE RI
Report Version:	v1.0
Contractual Date:	31/12/2024
Report Submission Date:	31/12/2024
Dissemination Level:	PU
Nature:	Report
Lead Author(s):	Canan Arıkan-Caba, Massimiliano Carloni, Vera Maria Charvát, Matej Ďurčo, Achille Felicetti
Co-author(s):	Laure Barbot, Ceri Binding, Luca De Santis, Vicky Dritsou, Twan Goosen, Maxim Kupreyev, Michael Kurzmeier, Béatrice Markhoff, Maria Theodoridou
Keywords:	Data models, Metadata formats, Ontologies, Vocabularies, FAIR Principles, Interoperability, Arts and Humanities

Change Log

Date	Version	Author/Editor	Summary of Changes made
18/09/2024	v0.1	Achille Felicetti	Draft report template
06/11/2024	v0.2	Achille Felicetti	Draft template for resources description
30/11/2024	v0.3	Canan Arıkan-Caba, Laure Barbot, Ceri Binding, Maxim Kupreyev, Luca De Santis, Vicky Dritsou, Twan Goosen, Michael Kurzmeier, Béatrice Markhoff, Maria Theodoridou	Content provision and control from WP3 participants
13/12/2024	v0.4	Massimiliano Carloni, Vera Maria Charvát, Matej Ďurčo, Achille Felicetti	Document submission for internal review
19/12/2024	v0.5	Ulf Jakobsson, Valentina Vassallo, Megan Black	External Review and Proofreading
31/12/2024	v1.0	Massimiliano Carloni, Matej Ďurčo, Achille Felicetti	Approved Version & Final Edits

Table of Contents

List of Abbreviations

4CH	Competence Centre for the Conservation of Cultural Heritage
AAT	Getty Art and Architecture Thesaurus
ACDH-CH	Austrian Centre for Digital Humanities and Cultural Heritage
AO-Cat	ARIADNEplus Ontology Catalogue
ARIADNE	Advanced Research Infrastructure for Archaeological Dataset Networking in Europe
ATRIUM	Advancing fronTier Research In the Arts and hUManities
BT	Broader terms
CARARE	Connecting Archaeology and Architecture in Europeana
CCI	Catalogue collectif indexé de Frantiq
CCR	CLARIN Concept Registry
CESSDA	Consortium of European Social Science Data Archives
CH	Cultural Heritage
CIDOC CRM	CIDOC Conceptual Reference Model
CLARIN	Common Language Resources and Technology Infrastructure
CLS INFRA	Computational Literary Studies Infrastructure
CMDI	Component Metadata Infrastructure
COAR	Confederation of Open Access Repositories
CRIS	Current Research Information Systems
CRM	Conceptual Reference Model
CSV	Comma-Separated Values
DARIAH	Digital Research Infrastructure for the Arts and Humanities
DETECt	Detecting Transcultural Identity in European Popular Crime Narratives
DH	Digital Humanities
DTD	Document Type Definition
EDM	Europeana Data Model
ESE	Europeana Semantic Elements
E-RIHS	European Research Infrastructure for Heritage Science
FAIR	Findable, Accessible, Interoperable, Reusable
FISH	Forum on Information Standards in Heritage
HBIM	Heritage Building Information Models
IFLA LRM	IFLA Library Reference Model
JSON	JavaScript Object Notation
JSON-LD	JSON for Linked Data
LCSH	Library of Congress Subject Headings
LIDO	Lightweight Information Describing Objects
LOD	Linked Open Data

LRMoo	Library Reference Model Object-Oriented
NBVT	Narrative Building and Visualising Tool
NOnt	Narrative Ontology
NT	Narrower terms
OpenAIRE	Open Access Infrastructure for Research in Europe
OPERAS	Open Scholarly Communication in the European Research Area for Social Sciences and Humanities
PACTOLS	Peuples, Anthroponymes, Chronologie, Toponymes, Œuvres, Lieux, Sujets
PARTHENOS	Pooling Activities, Resources and Tools for Heritage E-research Networking, Optimization and Synergies
PeriodO	Period Ontology
PROV-O	Provenance Ontology
RDA	Research Data Alliance
RDBMS	Relational Database Management System
RDF	Resource Description Framework
RDFS	RDF Schema
RHDTO	Reactive Heritage Digital Twin Ontology
RO-Crate	Research Object Crate
RS	Research Spotlight
RT	Related terms
SKOS	Simple Knowledge Organization System
SO	Scholarly Ontology
SSH	Social Sciences and Humanities
SSHOC	Social Sciences and Humanities Open Cluster
SSHOMP	Social Sciences and Humanities Open Marketplace
STARC	Science and Technology in Archaeology and Culture
SWRL	Semantic Web Rule Language
TaDiRAH	Taxonomy of Digital Research Activities in the Humanities
TEI	Text Encoding Initiative
TGN	Getty Thesaurus of Geographic Names
Turtle	Terse RDF Triple Language
TSV	Tab-Separated Values
WP	Work Package
XML	eXtensible Markup Language
XSD	XML Schema Definition

Executive Summary

This deliverable D3.1 provides an overview of data models, ontologies, controlled vocabularies, and formats adopted in both humanities research and cultural heritage. Over 130 resources have been compiled/gathered from a number of sources, primarily from the insights of the four major research infrastructures involved in ATRIUM: ARIADNE, CLARIN, DARIAH, and OPERAS. These resources have been assessed for their relevance to ATRIUM, focusing on their maturity, community adoption and interoperability. Based on this assessment, 40 resources have been described in more detail in the document. The full list of gathered resources is available in Appendix A.

This overview should serve as a basis for improving the discoverability and (semantic) interoperability of cultural heritage resources and research objects.

1. Introduction

1.1 Purpose and Scope of the Document

The purpose of this document is to provide an extensive and in-depth overview of the data models, metadata schemas, ontologies, controlled vocabularies, and formats relevant to the Arts and Humanities research domains (referred to collectively as "resources" henceforth). This overview should serve as a basis to enhance the discoverability, accessibility, and semantic interoperability of cultural heritage as well as research data in the Arts and Humanities across heterogeneous institutional and disciplinary contexts. The document offers a brief description of selected resources, considered most relevant for communities targeted by the project, along with an examination of their practical applications and potential benefits within the ATRIUM framework.

Further exploration is provided into how these resources can facilitate semantic interoperability among different systems, ensuring that collected data can be easily shared, accessed, and understood across platforms, institutions, and disciplines.

The document is aligned with the broader objectives of the ATRIUM project, which include fostering more interconnected, integrated, and resilient research infrastructures in the Arts and Humanities. In particular, the document constitutes a valuable resource for researchers, practitioners, and stakeholders involved in research and cultural data management, offering insights into how these resources can be effectively leveraged.

This deliverable also aims to clarify the evolving standards landscape and its direct impact on cultural heritage documentation and preservation, by also drawing from related initiatives and resources, such as deliverables from 4CH and PARTHENOS, which offer insights into existing standards and best practices. The document aims to guide the adoption of common solutions in research data and cultural heritage management as a basis for the implementation of principles of findability, accessibility, interoperability, and reusability (FAIR). Given the dynamics of the ecosystem, however, this document only represents a momentary snapshot in the ongoing work of collaboratively building a shared knowledge base which will serve as topical/up-to-date reference also in the future.

1.2 Structure of the Document

This document is structured as follows: Section 2 outlines the methodological approach to resource collection. Section 3 provides a general overview on data formats. Section 4

presents relevant metadata schemas and ontologies. Section 5 provides details on vocabularies, thesauri, and gazetteers. Section 6 describes additional resources for retrieving further information. The deliverable concludes in Section 7. Appendix A contains the full list of collected resources, with basic information for each of them. On top of this, a List of Abbreviations is provided at the beginning of the document.

2. Methodological Considerations

This report outlines resources that can be used to manage and organise data and metadata in the Arts and Humanities. Below, we provide definitions for the different types of resources discussed throughout the document:

- **Data formats:** Resources for encoding and representing data, ensuring interoperability and compatibility between different applications and platforms. They define how data is structured, stored, and exchanged, ranging from simple formats like CSV to complex structures such as JSON-LD or RDF/XML. Semantic resources listed in the following sections of the document are typically expressed using one of the data formats described here.
- **Metadata schemas:** Resources for defining and organising metadata elements by specifying their attributes, relationships, and usage rules. They are essential to ensure consistency in describing and managing resources across different systems.
- **Ontologies:** Resources for formalising knowledge within a specific domain by defining concepts, entities, their attributes and their interrelations. They enable a shared understanding across systems, supporting semantic integration and interoperability.
- **Vocabularies:** Set of terms or concepts used within a specific domain, often accompanied by definitions or descriptions. They serve as basic resources for consistent communication and understanding in specific domains, without necessarily organising the terms into hierarchical or relational structures.
- **Thesauri:** Structured, controlled vocabularies where terms are not only defined but also interconnected through relationships such as synonyms, broader terms (BT), narrower terms (NT), and related terms (RT). Unlike a simple vocabulary, a thesaurus provides a semantic network that enables users to navigate concepts systematically and supports advanced information retrieval by highlighting connections between terms.
- **Gazetteers:** Resources that function as geographical dictionaries or directories, containing information about place names and their corresponding coordinates or attributes. They support geospatial data management and enable the integration of geographic information into digital systems.

To ensure a comprehensive collection of data formats, metadata schemas, ontologies and vocabularies, we consulted experts from the four key research infrastructures

involved in ATRIUM (ARIADNE, CLARIN, DARIAH, and OPERAS). These experts were asked to specify the resources used within their respective research areas, as well as any additional resources they were aware of that could be beneficial for the ATRIUM project or that they deemed relevant in their respective communities. We also conducted a comprehensive survey to identify general-purpose models, formats, and vocabularies that are widely recognised and commonly used across various domains, not limited to the Arts and Humanities.

Additionally, we reviewed the extensive documentation produced by the four ATRIUM infrastructures and other related projects, such as 4CH, PARTHENOS, and E-RIHS. This documentation included deliverables, technical reports, recommendations, best practices, and other relevant materials, providing valuable insights into the standards, vocabularies, formats and other resources adopted in the field. Furthermore, we conducted an extensive literature review and bibliometric analysis to get insights into the current state of the art in the field, and we explored various online resources, including metadata registries and inventories (such as the CLARIN Standards Information System and the catalogue of the RDA Metadata Working Group), to gather additional relevant information.

All the data collected through these methods were ingested into a shared database implemented using Baserow, an open-source, no-code database tool that facilitates collaborative collection and curation of structured/relational data offering a user-friendly spreadsheet-like user interface and allowing on-the-fly changes to the data model. The database serves as an internal tool for data curation.

This systematic approach has enabled us to create a robust and comprehensive overview of the data formats, metadata schemas, vocabularies, and ontologies used in the Arts and Humanities domain, encompassing roughly 130 resources.

In a subsequent step, we ranked the collected resources based on their "relevance" in the context of ATRIUM. Defining "relevance" is in itself a challenging task, as it is difficult to base it on clear-cut "objective" criteria.
Thus we adopted a less formal approach. We identified a number of criteria, not to be applied strictly to include or exclude resources, but rather as soft "aspects" to be considered when assessing the relevance of a resource. Multiple members of the team, representing different roles and all the involved research infrastructures, then voted on resources they considered relevant based on their professional experience. The resulting ranking (based on the number of votes) was used to select a subset of resources to be described in more detail in the document.

In the following, we will elaborate on the criteria applied, focusing on resources that are useful in meeting the diverse needs of the field, supporting both interoperability and the effective management of research data. Criteria are mainly concerned with the relevance and applicability of the resources to the specific contexts of Arts and Humanities. Given the extension of this domain, which encompasses areas as varied as art history, archaeology, architectural heritage, history, linguistics, or literary studies, one criterion was the versatility and extensibility of the models, allowing a wide range of resource types to be modelled and described. We considered resources according to their ability to accommodate the cataloguing and organisation of materials such as manuscripts, visual artworks, artefacts, and complex architectural heritage. This flexibility was a key factor in selecting models like CRMtex and CIDOC CRM, which have demonstrated their ability to describe and interrelate textual as well as museum collections. However, the focus remained on identifying models and vocabularies that fit particular resource types within specific contexts, rather than requiring a universal solution for all scenarios.

Equally important in the selection process was the level of adoption and recognition within the community. Resources widely embraced within Arts and Humanities institutions offer several advantages, such as their established credibility and broad support base. This widespread usage ensures that they benefit from community-driven development, ongoing updates, and a wealth of practical implementations, making them stable and reliable tools for practitioners. However, the emphasis was not solely on models with high levels of adoption; emerging or specialised models were also considered where they presented innovative approaches or unique benefits. In this sense, community adoption was viewed more as an indicator of potential utility and longevity rather than a strict requirement for inclusion.

Likewise, models that ensure consistency and support for interoperability, whether through adherence to established standards or their capacity for linked data, were considered especially useful and thus prioritised in our review. The capacity of each model to work in conjunction with other standards was especially regarded as an asset that could support the ongoing development of a more integrated and coherent data ecosystem. Models that align with well-known standards, such as those used in digital libraries or archival settings, are recognised for their potential to contribute to the broader goals of data interoperability. Consequently, their ability to facilitate interoperability and integration within the ATRIUM framework was taken into account, to ensure that the selected models would contribute meaningfully to the broader objectives of the project.

3. Overview of Data Formats

This section provides information on the principal data formats commonly used for encoding data in the Arts and Humanities, which are therefore relevant to the ATRIUM project. While there are of course many more data formats available (including image and audio formats), the focus here has been on those that are commonly used for encoding metadata and semantic resources.

The information below includes a brief description of each format, its typical applications, advantages, limitations, and support in various programming languages or tools. The section aims to assist in selecting appropriate data formats for research and data management within the ATRIUM project.

CSV

Extension: .csv
MIME type: text/csv

Description: Comma-Separated Values (CSV) is a text-based format for representing tabular data. Each line represents one row in the table, while specific cells in a row are delimited by a separator, in this case a comma.
CSV is typically used for exporting and importing data, especially in databases, spreadsheets, and data analysis tools. It is also one of the preferred formats for making tabular datasets available for public use, such as in large data repositories.
CSV has many advantages. It is lightweight, which makes it fast to process because the files are typically small. It is simple, human-readable, and easy to manipulate without specialised tools. It is largely cross-platform compatible.
However, CSV has some limitations when it comes to representing more complex data structures, such as hierarchical or nested structures. It also has limited character encoding support, which can cause problems when handling data with special characters. It does not allow the definition of schemas, which makes it difficult to validate the data.
CSV is supported in several programming languages, such as Python.
A data format that is similar to CSV but uses tabs as separators is Tab-Separated Values (TSV). TSV might be a better choice when dealing with data that contains many commas, since in CSV such commas require the use of escape characters or the inclusion of strings between quotes.

Support for schema language: No

Standard(s): RFC 4180 - "Common Format and MIME Type for Comma-Separated Values (CSV) Files"

JSON

Extension: .json
MIME type: application/json

Description: JavaScript Object Notation is a lightweight, text-based data interchange format. While initially developed in the context of the JavaScript programming language, it has found wider acceptance as a format for exchanging data over the web, especially when requesting and serving data from RESTful APIs.
It is primarily based on objects described by attribute-value pairs. Values can be associated with various data types, including strings, numbers, booleans, and arrays. The simple structure of JSON makes it both human-readable and easily parsable by machines, resulting in very good parsing performance.
JSON is supported by a wide range of programming languages, including Java, JavaScript, PHP, and Python. It is also supported in popular relational database systems, such as PostgreSQL, through the use of specialised modules. It can be used to represent Linked Open Data in its variant JSON-LD.
JSON is very flexible, allowing data to be described for which no schema exists (or for which the schema is being developed), but it also allows schemas to be expressed according to the JSON Schema conventions. A JSON schema is also expressed in JSON and allows for data validation through the use of specific tools and libraries.

Support for schema language: Yes – JSON Schema

Website: https://www.json.org

Standard(s): ECMA-404 (and others: STD 90, RFC 8259, ISO/IEC 21778:2017)

Documentation available at: https://www.json.org/json-en.html

Notation3

Extension: .n3
MIME type: text/n3;charset=utf-8

Description: Notation3, or N3, is a syntax and text-based file format to represent RDF triples. It was developed with human readability and easy parsability in mind. It is a superset of Turtle. It should not be confused with N-Triples, which is a subset of Turtle.
It allows for the use of prefixes for URIs, shortcut syntax, and blank nodes. It also enables the definition of properties and their relationships to classes (e.g., the definition of domain and range), thus including some of the features of languages used for defining schemas and ontologies (such as RDFS and OWL).
Notation3 has never found wide acceptance. In comparison, Turtle, which has a more limited set of features but is easier to use and read, is more widely used.

Website: www.w3.org/TeamSubmission/n3/

Standard(s): W3C Team Submission 28 March 2011

Documentation available at: http://www.w3.org/TeamSubmission/n3/

N-Triples

Extension: .nt
MIME type: application/n-triples

Description: N-Triples is a syntax and text-based file format to represent RDF triples. It is a subset of the Turtle language. Due to its lack of abbreviation mechanisms (which are instead present in Turtle), it is more difficult to use and read than Turtle when dealing with a large quantity of data. The N-Quads extension allows for the inclusion of a fourth contextual element in a triple.

Website: https://www.w3.org/TR/n-triples/

Standard(s): W3C Recommendation 25 February 2014

Documentation available at: https://www.w3.org/TR/n-triples/

Turtle

Extension: .ttl
MIME type: text/turtle

Description: Terse RDF Triple Language (Turtle) is a syntax and text-based file format to represent RDF triples. It was developed as a subset of Notation3 and has caught on as an alternative to the more complex RDF/XML for expressing and reading RDF data. A triple

can be expressed as a sequence of subject - predicate - object, separated by whitespaces and terminated by a period.
It allows for a more compact representation of RDF triples in comparison to Notation3. In particular, it has an abbreviation mechanism that allows for the omission of elements that are common to more than one triple (for example, subjects and predicates shared by several consecutive triples). The TriG format, which is an extension of Turtle, allows for the representation of named graphs.
Turtle is supported by several Semantic Web tools and frameworks, such as the triplestore Jena Fuseki and Python's RDFLib library.

Website: https://www.w3.org/TR/turtle/

Standard(s): W3C Recommendation 25 February 2014

Documentation available at: www.w3.org/TR/turtle/

XML

Extension: .xml
MIME type: application/xml, text/xml

Description: XML is a subset of SGML (Standard Generalised Markup Language), from which HTML is also derived. It provides a way of representing data and metadata through the use of tags surrounded by angle brackets (<...>) and organised in a hierarchical structure.
XML is not just a file format, but a general concept that brings together several related technologies. These include XPath (a language for referencing data elements within an XML document), XSLT (which allows documents to be transformed from one format to another), and XQuery (which allows data to be searched within XML documents). In particular, XQuery uses XPath expressions but integrates them into more complex queries similar to those of SQL.
Some of the strengths of XML are that it is cross-platform; it is both human- and machine-readable; and it separates content from presentation in an efficient way (allowing different visual representations of the same data). Also, and most importantly, schemas, according to the XML Schema Definition (XSD) or other languages, can be defined to express the rules that particular XML documents must follow. This allows for easy customisation to specific needs, clear communication of the requirements of a particular format/community, and easy validation using widely available tools. As a result, XML often acts as a serialisation standard for more specific formats and guidelines, such

as the Text Encoding Initiative (TEI), MARCXML (in library science), Scalable Vector Graphics (SVG), and RDF/XML (for representing RDF-based data).
While in recent years JSON has gained more traction as a data exchange format, XML is still widely used in the Humanities, especially for encoding and tagging texts (for example, according to the TEI guidelines). Some of the disadvantages of XML are that it is not designed for fast access (which makes large XML databases rather difficult to handle), and its rigid structure can complicate some modelling choices. For example, XML does not allow overlapping tags, and its hierarchical tree-based structure is not well suited to representing RDF.

Support for schema language: Yes. Several languages are available, with the main three as XML Schema Definition (XSD), Document Type Definition (DTD), and Regular Language for XML Next Generation (RELAX NG).

Website:

- (v1.0) http://www.w3.org/TR/xml/
- (v1.1) http://www.w3.org/TR/xml11

Standard(s):

- (v1.0) W3C Recommendation 26 November 2008
- (v1.1) W3C Recommendation 16 August 2006, edited in place 29 September 2006

Documentation available at:

- (v1.0) http://www.w3.org/TR/xml/
- (v1.1) http://www.w3.org/TR/xml11

4. Overview of Metadata Schemas and Ontologies

This section provides an overview of various metadata schemas, ontologies and semantic models relevant to the ATRIUM project. Each entry comes with specific fields to elucidate their nature and applicability, including the type of resource and a brief description, its domain of applicability, its maintainer(s) and major adopters, and — where available — some basic technical information (namespace; encoding, encompassing both the modelling language used and the specific formats in which a resource is serialised). Links to access each resource and its accompanying documentation are also provided. Additionally, a specific 'Relevance to ATRIUM' field explains how each resource aligns with and supports the objectives of the ATRIUM project, highlighting its significance within the project's framework.

4.1 Metadata Schemas

CARARE

Type: Metadata Schema
Domain: Archaeology, Architecture, Cultural Heritage

Description: The CARARE (Connecting Archaeology and Architecture in Europeana) metadata schema is designed to support the delivery of metadata relating to the archaeological and architectural heritage. It builds on existing standards and best practices from various countries, including the CIDOC Archaeological Sites Core Data Standard, the Core Data Index to Historic Buildings and Monuments of the Architectural Heritage, CIDOC CRM, MIDAS Heritage, POLIS DTD, and LIDO. The schema's strength lies in its ability to support a comprehensive range of descriptive information about monuments, buildings, landscape areas, and their representations.

Website:
https://www.carare.eu/en/services/carare-aggregation-services/carare-metadata-schema

Maintainer: CARARE Association

Major Adopters: Europeana, 3D-ICONS Project

Encoded in: XSD

Relevance to ATRIUM: The CARARE metadata schema offers a specialised framework tailored for the detailed documentation of archaeological and architectural heritage assets, including monuments, buildings, and landscape areas. This specificity aligns with ATRIUM's objective to enhance data services in the Arts and Humanities by providing structured and detailed metadata standards. By integrating the CARARE schema, ATRIUM can facilitate the aggregation and interoperability of diverse datasets, enabling researchers to access rich, standardised information pertinent to their studies. This integration supports the advancement of frontier research by ensuring that complex heritage data is accurately represented and easily accessible across various platforms.

Model available at:
https://www.carare.eu/en/services/carare-aggregation-services/carare-metadata-schema/carare-version-20/

Documentation available at:
https://www.carare.eu/en/services/carare-aggregation-services/carare-metadata-schema/

Component Metadata Infrastructure (CMDI)

Type: Metadata framework
Domain: Generic

Description: The Component Metadata Infrastructure (CMDI) offers a standard for metadata within CLARIN. CLARIN does not offer a fixed model for *metadata authoring* but rather for *metadata modelling* – in other words, it defines a meta-model. Together with a set of core services hosted by CLARIN, the meta-model forms the basis for the Component Metadata Infrastructure (CMDI), which can be thought of as a metadata framework rather than a single standard. At the time of writing, there are 278 "metadata profiles" which have been published in the CLARIN Component Registry. Each profile offers a distinct metadata blueprint, which can be instantiated as a CMD record that can be validated against the automatically generated, profile-specific metadata schema. A common *CLARIN Concept Registry* forms the basis for semantic interoperability across profiles. Conversions from various common metadata standards to CMD metadata have been implemented by CLARIN.
Several commonly used vocabularies within CLARIN, such as the ISO 639-3 language codes and ISO 3166-1 country codes have been implemented as CMD components for use

within CMDI.
A more complete description of CMDI can be found in Windhouwer & Goosen (2022).

Website: https://www.clarin.eu/cmdi

Maintainer: CLARIN ERIC

Major Adopters: Amongst other CLARIN centres, ARCHE (ACDH-CH, ÖAW), Max Planck Institute for Psycholinguistics, Royal Netherlands Academy of Arts and Sciences Humanities Cluster, Zentrum Sprache at the BBAW, Institute of Formal and Applied Linguistics (Charles University, Prague).
See https://centres.clarin.eu/oai_pmh for a list of metadata-providing CLARIN centres.

Encoded in: XSD

Official namespace: http://www.clarin.eu/cmd/1 (record envelope)

Relevance to ATRIUM: CMDI is the core technology that enables the Virtual Language Observatory, CLARIN's metadata catalogue. Existing models can be implemented in CMDI, which facilitates interoperability within and across research infrastructures.

Model available at:
- https://www.clarin.eu/content/cmdi-specification-version-12
- https://catalog.clarin.eu/ds/ComponentRegistry

Documentation available at: https://www.clarin.eu/cmdi

DCAT

Type: Metadata schema
Domain: Data catalogues and datasets

Description: DCAT is an RDF-based vocabulary to describe catalogues and the resources they contain (datasets, data services, etc.). It makes a distinction between the abstract form of a dataset and its different distributions, corresponding to the different serialistions of the data. By offering a common language for describing catalogues and resources, it lays an important foundation for enhancing interoperability across different data portals.

Website: https://www.w3.org/TR/vocab-dcat/

Maintainer: W3C

Major Adopters: Interoperable Europe, Europeana

Encoded in: RDF/XML, JSON-LD, Turtle

Official namespace: http://www.w3.org/ns/dcat#

Relevance to ATRIUM: Some of the possible use cases for DCAT fit perfectly with the objectives of ATRIUM, and in particular WP3 'Facilitating Discoverability of and Access to Humanities Resources' and WP6 'Service Interoperability and EOSC Integration'. DCAT can be used to compare datasets and services across repositories and enhance the findability of resources, while still retaining the possibility of describing format-specific features by integrating complementary vocabularies. DCAT was a major inspiration for AO-Cat, one of the ontologies used in ATRIUM.

Model available at:

- https://www.w3.org/ns/dcat.rdf (RDF/XML)
- https://www.w3.org/ns/dcat.jsonld (JSON-LD)
- https://www.w3.org/ns/dcat.ttl (Turtle)

Documentation available at: https://www.w3.org/TR/vocab-dcat/

LIDO

Type: Metadata schema
Domain: Museums, Cultural Heritage

Description: LIDO is a schema that enables the sharing of museum and cultural heritage object information. It is designed to support the aggregation of museum data for portals and applications. LIDO standardises the exchange of metadata about museum objects across different platforms by providing a comprehensive range of descriptive information for diverse object types, including art, architecture, cultural history, technology, and natural science, and is optimised for multilingual environments. Its structure is based on the CIDOC Conceptual Reference Model (CRM), ensuring semantic interoperability and alignment with international standards.

Website: https://cidoc.mini.icom.museum/working-groups/lido/

Maintainer: ICOM CIDOC

Major Adopters: Europeana, German Digital Library, Yale Center for British Art, and several museum institutions (using LIDO for sharing and aggregating museum data)

Encoded in: XSD

Official namespace: http://www.lido-schema.org

Relevance to ATRIUM: LIDO aims to enhance interoperability and composability of services within museums and research infrastructures, ensuring that data can be accurately exchanged and interpreted across diverse systems. LIDO's alignment with the CIDOC CRM and its capacity to represent complex event-centric metadata make it a valuable tool for ATRIUM's objectives in facilitating consistent metadata representation, improving data discoverability, and supporting complex workflows involving cultural heritage information.

Model available at: http://www.lido-schema.org/schema/latest/lido.xsd

Documentation available at: http://www.lido-schema.org/schema/latest/lido.html

OpenAIRE guidelines

Type: Metadata schema
Domain: Scholarly domain

Description: The OpenAIRE Metadata guidelines provide a framework for exchanging bibliographic information about research products of any type (publications, data, software, other). They cover various types of repositories, including literature, data archives, and Current Research Information Systems (CRIS), ensuring that metadata aligns with Open Science mandates and supports the FAIR principles. By implementing these guidelines, repositories enhance the discoverability and accessibility of research outputs, aiding compliance with European Commission Open Access requirements and promoting seamless data exchange across platforms.

Website: https://guidelines.openaire.eu

Maintainer: OpenAIRE AMKE

Major Adopters: OpenAIRE, LA Referencia (aggregator of South America)

Relevance to ATRIUM: The OpenAIRE guidelines may enable ATRIUM to standardise metadata exposure across its repositories, ensuring seamless integration with the broader OpenAIRE infrastructure. This alignment facilitates efficient metadata harvesting, improves the discoverability of research outputs, and supports compliance with Open Science mandates. By implementing these guidelines, ATRIUM can ensure that its data repositories adhere to the FAIR principles, promoting data sharing and reuse within the research community.

Documentation available at: https://guidelines.openaire.eu

Schema.org

Type: Metadata Schema
Domain: Structured data on the web

Description: Schema.org is a collaborative initiative launched in 2011 by major search engines, including Google, Bing, Yahoo!, and Yandex, to create and support a common set of schemas for structured data markup on web pages. These schemas enable webmasters to embed structured data into their web pages, enhancing the ability of search engines and other applications to understand and display content more effectively. The vocabulary encompasses a wide range of entities and relationships, facilitating improved search result presentation and richer user experience.

Website: https://schema.org/

Maintainer: Schema.org

Major Adopters: Google, Bing, Yahoo!, Yandex, and numerous websites implementing structured data

Encoded in: RDF/XML, JSON-LD, Turtle

Official namespace: https://schema.org/

Relevance to ATRIUM: Schema.org provides a standardised vocabulary for structured data markup, which can be instrumental in harmonising data descriptions across various platforms. By adopting Schema.org vocabularies, ATRIUM can facilitate consistent metadata representation, improving data discoverability and integration within the research ecosystem. This alignment supports ATRIUM's objectives of enabling complex

workflows and promoting seamless data exchange among participating research infrastructures.

Model available at:

- https://schema.org/docs/developers.html
- https://github.com/schemaorg/schemaorg

Documentation available at: https://schema.org/docs/documents.html

STARC Model

Type: Metadata Model
Domain: Archaeology, Cultural Heritage

Description: The STARC Metadata Model, developed by the Science and Technology in Archaeology and Culture Research Center (STARC) at The Cyprus Institute, is designed to describe in great detail and standardise the documentation and management of archaeological assets. This model facilitates the integration and interoperability of diverse datasets, enabling comprehensive analysis and preservation of archaeological and cultural heritage information. By employing this metadata framework, researchers can ensure consistent and accurate representation of cultural artefacts, enhancing data sharing and collaborative research efforts.

Website: http://public.cyi.ac.cy/starcRepo/explore/objects

Maintainer: Science and Technology in Archaeology and Culture Research Center (STARC), The Cyprus Institute

Major Adopters: The Cyprus Institute, various collaborative cultural heritage projects

Encoded in: XSD

Relevance to ATRIUM: The STARC Metadata Model provides a standardised framework for the complete documentation of archaeological assets (objects, ancient buildings and archaeological sites). By integrating this model, ATRIUM can harmonise data descriptions related to archaeological and cultural heritage objects, facilitating data discovery and aggregation. Moreover, the model's emphasis on standardisation aligns with ATRIUM's objective of improving data interoperability within the research ecosystem, enabling the integration of archaeological data into broader Arts and Humanities studies.

Model and documentation available at: http://public.cyi.ac.cy/starcRepo/

4.2 Ontologies and Semantic Models

AO-Cat

Type: Ontology
Domain: Archaeology

Description: The AO-Cat Ontology is a formal ontology of the resources managed by the ARIADNE Research Infrastructure, ARIADNE RI. AO-Cat was developed within the ARIADNEplus project. AO-Cat inherits the CIDOC CRM model and philosophy to provide the community with a powerful and efficient integration tool, capable of capturing the meaning of a diverse range of archaeological datasets offered by content providers and integrating their data in a unique and coherent way. The aim of AO-Cat is to present a core ontology to which data providers map their individual database schema, thereby ensuring interoperability between diverse resources. It is capable of being deployed at the "site" or "monument" level, but also effectively implemented at the "item" level, e.g. for individual artefacts or graves, providing a much more granular level of interoperability. ARIADNE resource types can be used in the portal to filter at different levels.

Website: https://github.com/ARIADNE-Infrastructure/AO-Cat-Ontology

Maintainer: ARIADNE RI

Major Adopters: ARIADNE RI, 4CH

Encoded in: RDFS

Official namespace: https://www.ariadne-infrastructure.eu/resource/ao/cat/

Relevance to ATRIUM: AO-Cat can provide a specialised framework for the integration and management of archaeological data within the ATRIUM project. Its interoperability with CIDOC CRM makes it a valuable resource for ensuring consistency and compatibility across different cultural heritage domains. AO-Cat will also be used to catalog the resources collected by ATRIUM, enhancing the project's ability to handle this data effectively. It is recommended to use AO-Cat in conjunction with CIDOC CRM to develop specific ontologies for ATRIUM, ensuring interoperability with other standards.

Model and documentation available at: https://zenodo.org/records/7818375

CIDOC CRM

Type: Ontology
Domain: Cultural Heritage, Museums, Historical Studies, Archaeology

Description: The CIDOC Conceptual Reference Model (CRM) is a conceptual model used in cultural heritage to enable information integration. It provides definitions and a structured framework to describe concepts, relationships, and data used in cultural heritage documentation. The CIDOC CRM, developed and maintained over a period of more than 20 years, was initially proposed by the CIDOC Documentation Standards Working Group and is currently managed by the CIDOC CRM SIG, both of which are working groups of the CIDOC Council. In December 2006, it received official recognition as an ISO standard, with its most recent version being reaffirmed in 2023 as ISO 21127:2023.

Website: https://www.cidoc-crm.org

Maintainer: CIDOC - ICOM International Committee for Documentation

Major Adopters: British Museum, National Gallery, ARIADNE RI, 4CH, MINGEI, Arches

Encoded in: RDFS, JSON-LD, XML

Official namespace: http://www.cidoc-crm.org/cidoc-crm/

Relevance to ATRIUM: CIDOC CRM can provide a robust framework for the integration and management of cultural data within the ATRIUM project. However, it may require adaptations to align with the specific needs of the ATRIUM project. It is recommended to use CIDOC CRM as a foundation for developing specific ontologies for ATRIUM, ensuring interoperability with other standards.

Model available at: https://www.cidoc-crm.org/versions-of-the-cidoc-crm

Documentation available at: https://www.cidoc-crm.org/use_and_learn

CRMarchaeo

Type: Ontology
Domain: Cultural Heritage, Archaeology, Archaeological excavation

Description: CRMarchaeo is an extension of CIDOC CRM created to support the archaeological excavation process and all the various entities and activities related to it. The model has been created from standards and models already in use by national and international cultural heritage institutions, and has evolved through deep analysis of existing metadata from real archaeological documentation. It has been enriched by continuous collaboration with various communities of archaeologists from different countries and schools. Furthermore, it takes advantage of the concepts provided by CRMsci, from which it inherits most of the geological and stratigraphic principles that govern archaeological stratigraphy, extending these principles. CRMarchaeo is intended to provide all necessary tools to manage and integrate existing documentation in order to formalise knowledge extracted from observations made by archaeologists, recorded in various ways and adopting different standards. In this sense, its purpose is to facilitate the semantic encoding, exchange, interoperability and access of existing archaeological documentation.

Website: https://cidoc-crm.org/crmarchaeo

Maintainer: CIDOC CRM Special Interest Group

Major Adopters: ARIADNE RI, 4CH

Encoded in: RDFS

Official namespace: http://www.cidoc-crm.org/extensions/crmarchaeo/

Relevance to ATRIUM: CRMarchaeo can provide a robust framework for the integration and management of data from archaeological excavations within the ATRIUM project. However, it may require adaptations to align with the specific needs of the ATRIUM project.

Model available at: https://cidoc-crm.org/crmarchaeo/fm_releases

Documentation available at: https://cidoc-crm.org/crmarchaeo/use-learn-crmarchaeo

CRMba

Type: Ontology
Domain: Archaeology, Standing Structures, Cultural Heritage

Description: CRMba is an extension of the CIDOC Conceptual Reference Model (CRM) specifically designed to support the documentation and analysis of archaeological standing structures. It facilitates the recording of a building's structural development, usage evolution, and the relationships between its components and the whole structure. By incorporating concepts from stratigraphic analysis theory, CRMba enables detailed representation of the construction phases, modifications, and reuse of buildings over time.

Website: https://cidoc-crm.org/crmba/

Maintainer: CIDOC CRM Special Interest Group

Major Adopters: Researchers and institutions involved in the documentation and analysis of archaeological buildings and heritage structures.

Encoding format(s): RDFS (under finalisation)

Official namespace: http://www.cidoc-crm.org/crmba/

Relevance to ATRIUM: CRMba is a detailed ontology designed to support the representation of complex architectural elements and their historical transformations, which is essential for creating accurate Heritage Building Information Models (HBIM). This integration facilitates the seamless incorporation of 3D survey data, such as point clouds and photogrammetry, into a structured semantic framework, enabling comprehensive analyses of construction phases, material compositions, and structural alterations over time. Integrating CRMba into ATRIUM's framework would significantly enhance the project's capabilities in documenting and analysing standing structures through advanced 3D-modelling techniques.

Model available at: https://cidoc-crm.org/crmba/fm_releases

Documentation available at: https://cidoc-crm.org/crmba/

CRMdig

Type: Ontology
Domain: Cultural Heritage, Provenance, Digital products

Description: CRM Digital is an ontology and RDF Schema to encode metadata about the steps and methods of production ("provenance") of digitisation products and synthetic

digital representations such as 2D, 3D or even animated models created by various technologies. Its distinct features compared to competitive models is the complete inclusion of the initial physical measurement processes and their parameters. It has been developed as a compatible extension of CIDOC CRM, which allows for querying the most relevant facts and returning complete descriptions encoded in this model by generic ISO 21127 terms without need to refer to its specific properties. In contrast, competitive models cannot be queried by a more general standard and are restricted to the computational provenance only. Data encoded in the major competitive models can be transformed without loss of meaning into a CRM Digital form.

Website: https://cidoc-crm.org/crmdig

Maintainer: CIDOC CRM Special Interest Group

Major Adopters: ARIADNE RI, 4CH

Encoded in: There is an RDFS encoding of an older version

Official namespace: http://www.cidoc-crm.org/extensions/crmdig/

Relevance to ATRIUM: CRMdig can provide a robust framework for the integration and management of digital assets within the ATRIUM project. However, it may require adaptations to align with the specific needs of the ATRIUM project.

Model available at: https://cidoc-crm.org/crmdig/fm_releases

Documentation available at: https://cidoc-crm.org/crmdig/use-learn-crmdig

CRMsci

Type: Ontology
Domain: Cultural Heritage, Provenance, Digital products

Description: The Scientific Observation Model (CRMsci) is a formal ontology intended to be used as a global schema for integrating metadata about scientific observation, measurements and processed data in descriptive and empirical sciences (such as biodiversity, geology, geography, archaeology, and cultural heritage conservation) in research IT environments and research data libraries. Its primary purpose is facilitating the management, integration, mediation, interchange and access to research data by description of semantic relationships, in particular causal ones. It is not primarily a model

to process the data itself in order to produce new research results, even though its representations offer themselves to be used for some kind of processing.
CRMsci uses and extends the CIDOC CRM (ISO 21127) as a general ontology of human activity, things and events happening in spacetime. It uses the same encoding-neutral formalism of knowledge representation ("data model" in the sense of computer science) as the CIDOC CRM, which can be implemented in RDFS, OWL, RDBMS and other forms of encoding. Since the model reuses, wherever appropriate, parts of the CIDOC Conceptual Reference Model, we also provide in this document a comprehensive list of all constructs used from ISO 21127, together with their definitions following the version 6.2 maintained by CIDOC.

Website: https://cidoc-crm.org/crmsci

Maintainer: CIDOC CRM Special Interest Group

Major Adopters: ARIADNE RI, 4CH, PARTHENOS

Encoded in: RDFS

Official namespace: http://www.cidoc-crm.org/extensions/crmsci/

Relevance to ATRIUM: CRMsci can provide a robust framework for integrating metadata about scientific observation, measurements and processed data within the ATRIUM project. However, it may require adaptations to align with the specific needs of the ATRIUM project.

Model available at: https://cidoc-crm.org/crmsci/fm_releases

Documentation available at: https://cidoc-crm.org/crmsci/use-learn-crmsci

CRMtex

Type: Ontology
Domain: Cultural Heritage, Ancient documents, Epigraphy

Description: CRMtex is an extension of CIDOC CRM created to support the study of ancient documents by identifying relevant textual entities and by modelling the scientific process related with the investigation of ancient texts and their features in order to foster integration with other cultural heritage research fields, such as archaeology and history. The concept of "written text", introduced by the extension, is intended to identify

a common entity consisting of a particular feature (i.e., set of glyphs) created (i.e., written) on various kinds of support, having semiotic significance and the declared purpose of conveying a specific message towards a given recipient or group of recipients. The modelling of the scientific autoptic examination of the document, consisting of an accurate analysis of the surface and the signs and prescribing the use of specific tools and procedures, is another key aspect taken into account by the extension. CRMtex is intended to identify and define in a clear and unambiguous way the main entities involved in the study and edition of ancient handwritten texts and then to describe them by means of appropriate ontological instruments in a multidisciplinary perspective. Thus, the extension aims at introducing new classes and properties more responsive to the specific needs of the various disciplines involved, including papyrology, palaeography, codicology and epigraphy.

Website: https://cidoc-crm.org/crmtex

Maintainer: CIDOC CRM Special Interest Group

Major Adopters: ARIADNE RI, 4CH, ItAnt

Encoded in: RDFS

Official namespace: http://www.cidoc-crm.org/extensions/crmtex/

Relevance to ATRIUM: CRMtex can provide a robust framework for the integration and management of epigraphic data within the ATRIUM project. By integrating CRMtex into its infrastructure, ATRIUM could document the lifecycle of ancient texts, encompassing their creation, usage, preservation, and scholarly examination. On top of this, one can systematically capture the relationships between textual content and its physical medium, as well as the various scholarly activities associated with these texts.

Model available at: https://cidoc-crm.org/crmtex/fm_releases

Documentation available at: https://cidoc-crm.org/crmtex/use-learn-crmtex

Europeana Data Model (EDM)

Type: Ontology
Domain: Digital Libraries, Cultural Heritage, Bibliographic Data, Library Systems, Scholarly domain

Description: The Europeana Data Model (EDM) was proposed for structuring the data that Europeana ingests, manages and publishes. EDM is a major improvement on the Europeana Semantic Elements (ESE), the basic data model that Europeana started with. EDM is not built on any particular community standard but rather adopts an open, cross-domain Semantic Web-based framework that can accommodate the range and richness of particular community standards such as LIDO for museums, EAD1 for archives or METS2 for digital libraries.

Website: https://pro.europeana.eu/page/edm-documentation

Maintainer: Europeana

Major Adopters: Europeana**,** national aggregators for cultural heritage data (such as Kulturpool for Austria)

Encoded in: XSD, OWL

Official namespace: http://www.europeana.eu/schemas/edm/

Relevance to ATRIUM: The mapping and encoding in the EDM format could provide seamless integration with the Europeana ecosystem, facilitating the aggregation and harmonisation with the information of many other research and cultural institutions.

Model and documentation available at:
https://pro.europeana.eu/page/edm-documentation

LRMoo

Type: Ontology
Domain: Bibliographic Information, Library Science

Description: LRMoo is an object-oriented formulation of the IFLA Library Reference Model (IFLA LRM), designed to align with the CIDOC Conceptual Reference Model (CIDOC CRM). It succeeds and replaces FRBRoo, providing a streamlined and generalised model while retaining full expressivity. LRMoo facilitates the integration, mediation, and interchange of bibliographic and museum information by capturing and representing the underlying semantics of bibliographic data.

Website: https://cidoc-crm.org/lrmoo

Maintainer: International Federation of Library Associations and Institutions (IFLA), International Council of Museums (ICOM)

Major Adopters: Europeana, national libraries (several national libraries adopt LRMoo to align their bibliographic records with international standards, facilitating data sharing and consistency)

Encoded in: RDFS, OWL

Official namespace: http://iflastandards.info/ns/lrm/lrmoo/

Relevance to ATRIUM: LRMoo offers a comprehensive framework for modelling bibliographic information within an object-oriented paradigm, aligning with the CIDOC CRM. This alignment is particularly relevant to ATRIUM's goal of facilitating access to digital research infrastructures and advancing knowledge across disciplines, languages, and media. By integrating LRMoo, ATRIUM can enhance the interoperability of its data services, ensuring that bibliographic information is consistently and accurately represented across various platforms.

Model available at: https://cidoc-crm.org/lrmoo/fm_releases

Documentation available at: https://cidoc-crm.org/lrmoo

Narrative Ontology (NOnt)

Type: Ontology
Domain: Cultural Heritage, Digital Libraries

Description: The Narrative Ontology (NOnt) is a formal framework created to represent narratives in digital contexts, particularly for cultural heritage applications. It conceptualises narratives as structured entities, including events, characters, and their temporal relationships. NOnt is implemented as an extension of three standard models: CIDOC CRM, FRBRoo, and OWL-Time, utilising the Semantic Web Rule Language (SWRL) to express its axioms. This alignment enhances interoperability across digital libraries and Semantic Web applications. By employing models like Allen's time algebra, NOnt effectively captures narrative complexities, facilitating efficient querying and analysis. Its formalisation supports the integration of various narrative components and has been applied in tools like the Narrative Building and Visualising Tool (NBVT), which allows users to create and visualise narratives. This makes NOnt a valuable resource for researchers in literary studies and digital humanities, bridging the gap between narrative theory and

technology while promoting a deeper understanding of storytelling in digital environments.

Website: https://dlnarratives.eu/ontology

Maintainer: CNR Pisa (Italy)

Major Adopters: MINGEI Project

Encoded in: RDFS, SWRL

Official namespace: https://dlnarratives.eu/ontology#

Relevance to ATRIUM: Using NOnt in the ATRIUM project could significantly enhance its ability to support research across the Arts and Humanities by providing a robust framework for the integration and management of narratives related to various cultural artefacts and historical contexts. By formalising narratives as distinct entities, NOnt enables ATRIUM to capture complex relationships between events, characters, and temporal aspects, facilitating effective querying and exploration of these narratives. This integration, once implemented, could support interdisciplinary collaboration among scholars and bridge narrative theory with practical applications in the Arts and Humanities, enriching the digital landscape.

Model available at: https://dlnarratives.eu/ontology/

Documentation available at: https://doi.org/10.3233/SW-200421

Nomisma Ontology

Type: Ontology
Domain: Numismatics, Archaeology, Epigraphy, Cultural Heritage

Description: The Nomisma Ontology is a structured framework developed to standardise the representation of numismatic concepts within the digital realm. It provides a comprehensive set of classes and properties that describe various aspects of numismatic objects, including their typology, material composition, minting authority, and historical context. By adhering to Linked Open Data (LOD) principles, the ontology facilitates interoperability and data sharing among numismatic collections, researchers, and institutions worldwide. This standardisation enhances the accessibility and comparability of numismatic data, supporting scholarly research and public engagement.

Website: https://nomisma.org/ontology

Maintainer: American Numismatic Society

Major Adopters: American Numismatic Society, Online Coins of the Roman Empire (OCRE), Coin Hoards of the Roman Empire (CHRE), ARIADNE RI

Encoded in: OWL, RDF/XML, Turtle

Official namespace: http://nomisma.org/ontology#

Relevance to ATRIUM: The Nomisma Ontology, with its standardised representation of numismatic concepts, serves as a valuable resource to harmonise data descriptions related to numismatic objects, facilitating seamless data aggregation and knowledge discovery of this peculiar information. Moreover, the ontology's adherence to LOD principles aligns with ATRIUM's goal of improving data quality within the research ecosystem, enabling complex workflows and supporting the integration of numismatic data into broader cultural heritage studies.

Model available at: https://nomisma.org/ontology

Documentation available at: https://numismatics.org/

OpenArchaeo Semantic Data Model

Type: Ontology, Application Profile
Domain: Archaeology

Description: OpenArchaeo is a Semantic Web platform for archaeology, devised by the MASA (Mémoire des Archéologues et des Sites Archéologiques) consortium of the research infrastructure Huma-Num. The MASA triplestore is based on the CIDOC CRM ontology and some of its extensions (CRMarchaeo, CRMsci and CRMba). This special selection from the CIDOC CRM ecosystem forms a generic model that covers the basic concepts found in most archaeological corpora (site, operation, structure, feature, wall, burial, stratigraphic unit and artefact).

Website: http://openarchaeo.huma-num.fr/openarchaeo_v2/home

Maintainer: MASA

Major Adopters: MASA

Encoded in: RDFS, SHACL

Relevance to ATRIUM: The OpenArchaeo schema is tailored for the documentation of archaeological heritage assets. It is a concrete working example of Application Profiles devised from the CIDOC CRM ecosystem. This ontology is used both to generate knowledge graphs from legacy datasets and to set up the SPARNATURAL intuitive visual querying system. By integrating the OpenArchaeo ontology, ATRIUM can facilitate the aggregation and interoperability of diverse archaeological datasets and offer researchers access to them via a common federated interface.

Model available at:
http://openarchaeo.huma-num.fr/openarchaeo_v2/home
(+ parts of https://cidoc-crm.org/)

Documentation available at:
http://openarchaeo.huma-num.fr/openarchaeo_v2/home
(+ parts of https://cidoc-crm.org/)

PROV-O

Type: Ontology
Domain: Scholarly domain

Description: The PROV Ontology (PROV-O) is a W3C recommendation that provides a standardised framework for representing provenance information on the web. PROV-O defines a set of classes, properties, and constraints to model the entities, activities, and agents involved in data creation and manipulation. This structured approach facilitates the interchange of provenance information across diverse systems and contexts, enhancing transparency and trustworthiness in data-driven processes.

Website: https://www.w3.org/TR/prov-o/

Maintainer: W3C

Major Adopters: Data integrity and audit trails in research and enterprise

Encoded in: OWL, Turtle

Official namespace: http://www.w3.org/ns/prov#

Relevance to ATRIUM: PROV-O allows the systematic documentation of the provenance of data, detailing the entities, activities, and agents involved in their lifecycle. This comprehensive provenance tracking ensures that data remains accurate and trustable while exchanged. Moreover, implementing PROV-O facilitates transparency and trust in data processes, which is essential for collaborative research environments. The structured provenance information provided by PROV-O could support complex workflows and data integration, enabling researchers to trace the origin and transformation of data assets within the ATRIUM ecosystem.

Model available at: https://www.w3.org/ns/prov-o

Documentation available at: https://www.w3.org/TR/prov-o/

RHDTO

Type: Ontology
Domain: Cultural Heritage, Architecture, Digital Twins

Description: The Reactive Heritage Digital Twin Ontology (RHDTO) is a CIDOC CRM extension for managing and interconnecting the broad range of data forming the informational core of Heritage Digital Twins. The RHDTO model is primarily designed as a tool to capture and encode information both of the tangible and intangible nature of Cultural Heritage and to provide, through specific classes and properties, a mechanism for documenting and dynamically analysing their mutual relationships. It is also designed to model their digital counterparts and the mechanisms that make them interoperable as interconnected elements within the Heritage Digital Twin.

Maintainer: VAST-LAB PIN, FORTH

Major Adopters: 4CH

Encoded in: RDFS

Official namespace: https://hdt/1.0.0/

Relevance to ATRIUM: The RHDTO model has the ability to enhance the documentation and preservation of cultural heritage through a structured semantic framework. Since ATRIUM aims to integrate advanced technologies for cultural heritage management, the

RHDTO ontology provides a robust model that captures both tangible and intangible aspects of heritage assets. By leveraging the principles of digital twins, the HDT ontology facilitates dynamic interactions between real-world objects and their digital counterparts, enabling comprehensive data management and interoperability across various systems. This alignment with ATRIUM's objectives supports enhanced knowledge organisation and the effective sharing of cultural information, ultimately fostering better conservation practices and public engagement with many Arts and Humanities fields.

Model available at: https://is.gd/rhdto

Documentation available at: https://doi.org/10.3390/s24123978

RO-Crate

Type: Ontology
Domain: Research Data

Description: RO-Crate is a community-driven standard for packaging research data and metadata. It is designed to provide a simple and flexible way for researchers to share and reuse data. RO-Crate is used across a wide range of disciplines, including the social sciences, humanities, and natural sciences.

Website: https://www.researchobject.org/ro-crate/

Maintainer: RO-Crate Community

Major Adopters: RoHub, WorkflowHub, Life Monitor (EOSC-Life), Language Data Commons of Australia Program

Encoded in: JSON-LD

Official namespace: https://w3id.org/ro/crate/1.1/context

Relevance to ATRIUM: One of the objectives of ATRIUM is to create demonstrators that serve as exemplary use cases for workflows. Such demonstrators would involve the use of specific data. Since RO-Crate allows for an easy way to package and share research data, it is possible that this ontology will play a role in the preparation of the final package that will represent a demonstrator.

Model available at:
https://www.researchobject.org/ro-crate/specification/1.1/root-data-entity.html#ro-crate-metadata-file-descriptor

Documentation available at: https://www.researchobject.org/ro-crate/specification

Scholarly Ontology

Type: Ontology
Domain: Scholarly domain

Description: Scholarly Ontology (SO) is an ontological framework for representing knowledge in the scholarly domain. It offers a flexible mechanism for modelling scholarly practices in the form of research processes that are carried out by scholars and represent their work. Modelling scholarly work with SO allows for easy access to information, encoded in such a way that can answer questions of the form "who does what", "how", "why", "where", etc. The ontology revolves around the central notion of an activity (i.e. a biological experiment, an archaeological excavation, a medical survey, etc.). Each activity captures various perspectives: namely a) Resource, comprising concepts such as Content Item (i.e. bibliographic references, articles, figures, tables, equations, etc.) and Topic, which model what is produced, used or documents an activity; b) Procedure, comprising the Methods that describe how an activity is conducted; and c) Agency, comprising concepts such as Actors or Goals, which model who is participating in the activities or why they are conducted.

Website: http://isdb.cs.aueb.gr/scholarlyontology/

Maintainer: Vayanos Pertsas, Digital Curation Unit (DCU), Athena Research Centre & Athens University of Economics and Business

Major Adopters: Research Spotlight

Encoded in: RDF/OWL

Official namespace: http://dcu.gr/ontologies/scholarly-ontology/

Relevance to ATRIUM: Scholarly Ontology is integral to ATRIUM, serving as the foundational semantic framework for the Research Spotlight (RS) automated workflow, which is part of the project's inventory. Specifically, RS exploits Scholarly Ontology's core

entities and relations to extract entities from unstructured research article text and metadata and transform the extracted data into structured knowledge graphs.

Model available at:
https://github.com/athenarc/scholarly-ontology/blob/main/ScholarlyOntology_Schema_and_ActyivityTypes_v1.3.owl

Documentation available at: https://github.com/athenarc/scholarly-ontology/

SKOS

Type: Ontology
Domain: Domain-agnostic

Description: The Simple Knowledge Organization System (SKOS) is a standardised framework for representing and managing knowledge organisation systems (KOS) — such as thesauri, classification schemes, subject heading systems, and taxonomies — as machine-readable data. With SKOS, these systems can be exchanged between computer applications, and published on the Web in machine-readable format. The SKOS Specifications are published as W3C Recommendations, while the SKOS data model is formally defined in the specification as an OWL Full ontology, representing knowledge organisation systems as concept schemes made up of individual concepts, each identified by a unique URI. This URI-based identification allows for unambiguous referencing of SKOS concepts across contexts, integrating them seamlessly into the Web.

Website: https://www.w3.org/2004/02/skos

Maintainer: W3C Semantic Web Deployment Working Group (SWDWG)

Major Adopters: Getty Research Institute, British Museum, Europeana

Encoded in: RDF/OWL, RDF/XML

Official namespace: http://www.w3.org/2004/02/skos/core#

Relevance to ATRIUM: SKOS provides a standardised framework to organise and publish controlled vocabularies and concepts associated with research information, thus facilitating data consistency and interoperability. By exploiting SKOS, the ATRIUM project

ensures its resources are well-organised, interoperable, and findable, supporting broader engagement with and reuse of research data across the Web.

Model available at: https://www.w3.org/2004/02/skos/vocabs

Documentation available at:
https://www.w3.org/TR/2009/REC-skos-reference-20090818/

TRIPLE Ontology

Type: Ontology
Domain: Social Sciences and Humanities

Description: It consists of a formal representation of the GoTriple data model. GoTriple entities consist of documents (the most relevant content), projects and authors. The Data Model is largely based on Schema.org. The ontology is also composed of a series of controlled vocabularies that describe some of the document's attributes (disciplines, document types, licences, conditions of access). An effort has been made to make these controlled vocabularies interoperable.
In particular, the "Disciplines" controlled vocabulary is a linked data resource that uses the SKOS ontology to provide a formalisation of the SSH domain in 27 fields of study. This formalisation is one of the outcomes of the TRIPLE project and exploits the vision of a former EU-funded research project called MORESS (Mapping of Research in European Social Sciences and Humanities). In this vocabulary each discipline is linked to one or more corresponding concepts of other widely known classification systems, including the Library of Congress Subject Headings, Wikidata and the Dewey Decimal Classification. This vocabulary is also multilingual, with concepts available in English, Italian, French and German.
"Document type" provides a linked data description of the 20 admitted types by linking them to the corresponding resource types of the COAR Controlled Vocabularies for Repositories.
"Licences" and "Conditions of Access" provide a controlled set of the values the corresponding attributes can assume in a GoTriple document.
Finally, GoTriple document search REST APIs can return results not only in JSON but in JSON-LD as well, by using the Document vocabulary to formalise the output.

Website: https://www.gotriple.eu/ontology/triple

Maintainer: The TRIPLE Consortium/Net7

Major Adopters: GoTriple.eu

Encoded in: RDF/XML, JSON-LD

Relevance to ATRIUM: The enrichment and finalisation of the TRIPLE ontology is one of Net7's contributions to ATRIUM. Part of the work has already been done in this first year of the project. The remaining activities include:

- defining other correspondences with selected ontologies of interest for ATRIUM, CIDOC-CRM in particular. This will facilitate the use of GoTriple's content in integration scenarios with the catalogues of the other partners of the project;
- making the search APIs for projects and authors Semantic Web-compliant, allowing results to be retrieved in JSON-LD format as well;
- expanding the multilinguality of the ontology, by also using the Mondaecus translation tool provided by the University of Coimbra (see ATRIUM task 3.4.1).

Model available at: https://www.gotriple.eu/ontology/triple
https://github.com/atrium-research/triple-ontology

Documentation available at: https://www.gotriple.eu/ontology/triple

5. Vocabularies, Thesauri and Gazetteers

This section provides an overview of various vocabularies, thesauri and gazetteers relevant to the ATRIUM project. As for the previous section, each entry includes specific fields to elucidate the nature and applicability of these resources, such as the type of resource, a brief description, its domain of applicability, maintainer, and major adopters. Where available, the description also includes technical information like namespace and encoding (encompassing both the modelling language being used and the specific formats in which a resource is serialised). Links to access each resource and its accompanying documentation are also provided. Additionally, a 'Relevance to ATRIUM' field explains how each resource aligns with and supports the project's objectives, highlighting its significance within the ATRIUM framework.

CCR Concepts Vocabulary (CLARIN Concept Registry)

Type: Controlled vocabulary
Domain: Language resources

Description: The CLARIN Concept Registry (CCR) forms the basis of the semantic interoperability layer of CLARIN, especially in the context of metadata, i.e., the Component Metadata Infrastructure (CMDI). It does so by offering a collection of concepts relevant to the domain of language resources and identifiable by their persistent identifiers. The registry is SKOS-based (each concept in the registry is defined as a skos:Concept) and implemented as a vocabulary in a Skosmos instance.

Website: https://www.clarin.eu/content/clarin-concept-registry

Maintainer: CLARIN ERIC

Major Adopters: CLARIN ERIC

Encoded in: SKOS, RDF/XML, JSON-LD, Turtle

Relevance to ATRIUM: The concept registry serves as a shared semantic reference vocabulary for the structurally and syntactically highly heterogeneous components and profiles ecosystem of the Component Metadata Infrastructure (CMDI) used in CLARIN. CLARIN is one of the major research infrastructures involved in ATRIUM, and several of its services and catalogues play an important role in the harmonisation process of

(meta)data, the development of workflows, and the enhancement of service interoperability pursued in ATRIUM.

Model available at: https://concepts.clarin.eu

Documentation available at: https://www.clarin.eu/content/clarin-concept-registry

COAR Vocabularies

Type: Controlled vocabulary
Domain: General

Description: The COAR vocabularies include three different sets of vocabularies. The Access Rights vocabulary defines concepts to declare the access status of a resource. The Resource Type vocabulary defines concepts to identify the genre of a resource. The Version Type vocabulary defines concepts to declare the version of a resource.

Website:

- https://vocabularies.coar-repositories.org/
- https://coar-repositories.org/what-we-do/coar-controlled-vocabularies/
- https://github.com/coar-repositories/vocabularies/tree/master (GitHub repository)

Maintainer: COAR Controlled Vocabularies Editorial Board

Major Adopters: GoTriple

Encoded in: SKOS, RDF (N-Triples), Turtle, DSpace XML schema, HTML

Official namespace:

- http://purl.org/coar/access_right/
- http://purl.org/coar/resource_type/
- http://purl.org/coar/version/

Relevance to ATRIUM: In the context of several WPs, metadata regarding catalogues, datasets, and services are collected, compared, and enhanced. The COAR Vocabularies offer the possibility to describe some of the characteristics of these resources by using pre-existing controlled vocabularies that are already widely used and provide a wide range of links to concepts in other authority sources. In this way we can ensure the interoperability of our metadata.

Model available at:

- https://vocabularies.coar-repositories.org/access_rights/
- https://vocabularies.coar-repositories.org/resource_types/
- https://vocabularies.coar-repositories.org/version_types/

Documentation available at:

- https://vocabularies.coar-repositories.org/
- https://coar-repositories.github.io/vocabularies-implementation-guide/

FISH Vocabularies

Type: Controlled vocabulary
Domain: Cultural Heritage, Archaeology, Architecture, Historic Environment

Description: The FISH (Forum on Information Standards in Heritage) Vocabularies are a suite of controlled terminologies developed to standardise the recording and indexing of cultural heritage information within the United Kingdom. These vocabularies encompass a wide range of heritage-related terms, including monument types, building materials, maritime vessels, historic aircraft, cultural periods, and land/seascape characteristics. By providing structured and hierarchical lists of terms, the FISH Vocabularies facilitate consistent data entry, retrieval, and interoperability across various heritage databases and information systems. This standardisation is crucial for effective heritage management, research, and preservation efforts, ensuring that information about heritage assets is accurately categorised and easily accessible.

Website: https://heritage-standards.org.uk/fish-vocabularies/

Maintainer: Forum on Information Standards in Heritage (FISH)

Major Adopters: Historic England, Archaeology Data Service (ADS)

Encoded in: SKOS, CSV

Relevance to ATRIUM: The FISH Vocabularies provide a structured and standardised set of terminologies specifically tailored for the documentation of cultural heritage assets, including monuments, buildings, and archaeological objects. By integrating FISH Vocabularies, ATRIUM can harmonise data descriptions related to archaeological objects, building materials, and monument types, thereby facilitating seamless data aggregation and knowledge discovery. Moreover, the use of established vocabularies like FISH supports the development of interoperable services and complex workflows, aligning

with ATRIUM's goal of improving the interconnection of data within the research ecosystem.

Model available at: https://heritage-standards.org.uk/fish-vocabularies/

Documentation available at: https://heritage-standards.org.uk/

Getty Art and Architecture Thesaurus (AAT)

Type: Thesaurus
Domain: Art History, Subject Classification, Bibliographic Data, Cultural Heritage

Description: The Art & Architecture Thesaurus (AAT) is a structured vocabulary developed by the Getty Research Institute to provide standardised terminology for describing art, architecture, decorative arts, and material culture. It encompasses a wide array of terms, including objects, materials, styles, and techniques, organised hierarchically to reflect relationships among concepts. The AAT serves as a critical tool for cataloguers, researchers, and information professionals, facilitating consistent documentation and improved retrieval of research information across various databases and institutions.

Website: https://www.getty.edu/research/tools/vocabularies/aat/

Maintainer: The Getty Research Institute

Major Adopters: Museums, libraries, archives, and academic institutions worldwide

Encoded in: SKOS, RDF, XML

Official namespace: http://vocab.getty.edu/aat/

Relevance to ATRIUM: The AAT can contribute to the goals by providing a standardised vocabulary for art and architecture, facilitating consistent entity descriptions and enabling seamless data integration. By incorporating the AAT concepts and naming conventions, ATRIUM can improve data discoverability and support complex workflows involving arts and architecture information, thereby advancing its mission to create a cohesive and interoperable research ecosystem.

Model and documentation available at:
https://www.getty.edu/research/tools/vocabularies/aat/

Getty Thesaurus of Geographic Names (TGN)

Type: Controlled vocabulary
Domain: Geography, Cultural Heritage, Art History

Description: The Getty Thesaurus of Geographic Names (TGN) is a comprehensive structured vocabulary that provides standardised information about geographic locations pertinent to art, architecture, and cultural heritage. It encompasses over 1,000,000 names, including vernacular and historical names, coordinates, and place types, focusing on places significant for the study of art and architecture. TGN includes names and associated information about places, such as administrative political entities (e.g., cities and nations) and physical features (e.g., mountains and rivers). Both current and historical places are included, along with information related to history, population, culture, art, and architecture.

Website: https://www.getty.edu/research/tools/vocabularies/tgn/

Maintainer: Getty Research Institute

Major Adopters: Getty Research Institute, Europeana, DARIAH-DE

Encoded in: SKOS, RDF, XML

Official namespace: http://vocab.getty.edu/tgn/

Relevance to ATRIUM: The ATRIUM project aims to enhance the interoperability and composability of services within the Arts and Humanities research infrastructures. TGN's structured vocabulary of geographic names is crucial for achieving semantic interoperability, enabling accurate and unambiguous data exchange across different systems. By integrating TGN, ATRIUM can ensure consistent referencing of geographic entities, facilitating complex workflows and supporting data permeability within the ecosystem. Furthermore, TGN's focus on places relevant to art and cultural heritage aligns with ATRIUM's objectives to improve data interoperability in these domains. Utilising TGN will aid in standardising geographic references, thereby enhancing data aggregation, knowledge discovery, and the federation of information systems across participating research infrastructures.

Model and documentation available at:
https://www.getty.edu/research/tools/vocabularies/tgn/

PACTOLS

Type: Thesaurus
Domain: Archaeology

Description: PACTOLS is a controlled, standardised and multilingual vocabulary for archaeology, from prehistory to the contemporary period, and for the sciences of antiquity in all their aspects. It serves the indexing of the CCI (Catalogue collectif indexé de Frantiq = Indexed collective catalog of Frantiq) and is used by several French resources (academic journals, scientific databases, heritage enhancement programs), including the OpenArchaeo Semantic Web platform. It is partly aligned to the AAT.

Website: https://pactols.frantiq.fr/index.xhtml

Maintainer: Frantiq, MASA

Major Adopters: CCI, OpenArchaeo

Encoded in: SKOS, RDF/XML, JSON, JSON-LD, Turtle

Official namespace: https://ark.frantiq.fr/

Relevance to ATRIUM: The PACTOLS Thesaurus may be used for describing resources related to archaeology, one of the main focus areas of ATRIUM.

Model available at: https://github.com/frantiq/PACTOLS

Documentation available at: https://pactols.frantiq.fr/

PeriodO

Type: Controlled vocabulary, Gazetteer
Domain: History, Archaeology, Art History

Description: PeriodO is a public domain gazetteer that compiles scholarly definitions of historical, art-historical, and archaeological periods. It serves as a centralised resource for period definitions, facilitating the linking and visualisation of data across various datasets that may define periods differently. By aggregating diverse period definitions, PeriodO enables scholars and students to discern overlaps and divergences in temporal classifications, promoting a more nuanced understanding of historical timelines. PeriodO

emphasises ease of use and data accuracy, offering a comprehensive set of tools for time series analysis.

Website: https://perio.do

Maintainer: University of Texas at Austin, University of North Carolina at Chapel Hill

Major Adopters: ARIADNE RI, Pelagios Network

Encoded in: JSON, RDF, Turtle, CSV

Official namespace: http://n2t.net/ark:/99152/

Relevance to ATRIUM: PeriodO provides a comprehensive and structured vocabulary of period definitions, which is essential for accurately indexing and retrieving information about historical periods in the Arts and Humanities. This aligns with ATRIUM's goal of facilitating access to digital research infrastructures and advancing knowledge across disciplines, languages, and media. By integrating PeriodO, ATRIUM can improve interoperability of its data services, ensuring that temporal information is consistently and accurately represented across various platforms. This integration supports scholars in accessing and analysing data related to art, architecture, and cultural heritage according to spatio-temporal criteria.

Model available at: https://client.perio.do

Documentation available at: https://perio.do

Pleiades

Type: Gazetteer
Domain: Ancient Geography, Classics, Archaeology

Description: Pleiades is a community-built gazetteer and graph of ancient places, particularly those from the Greek and Roman worlds. It publishes authoritative information about ancient places and spaces, providing unique services for finding, displaying, and reusing that information under open licence. Pleiades aims to provide scholars, students, and enthusiasts with accessible digital information about the ancient world, facilitating research and exploration of historical geography. The content is open to the public and can be downloaded or accessed via an API for further use.

Website: https://pleiades.stoa.org

Maintainer: Institute for the Study of the Ancient World at New York University, Ancient World Mapping Center at the University of North Carolina at Chapel Hill

Major Adopters: Pelagios Project, Perseus Digital Library

Encoded in: RDF, JSON, KML, CSV

Official namespace: https://pleiades.stoa.org/places/

Relevance to ATRIUM: Pleiades offers a comprehensive and structured gazetteer of ancient places, which is essential for accurately indexing and retrieving information about historical locations in the Arts and Humanities. Thanks to its features, ATRIUM can enhance the interoperability of its data services, ensuring that geographic information is consistently and accurately represented across various platforms.

Model available at: https://pleiades.stoa.org/downloads

Documentation available at: https://pleiades.stoa.org/documentation

SSHOMP vocabularies

Type: Controlled vocabulary
Domain: Social Sciences and Humanities

Description: The SSH Open Marketplace (SSHOMP), a discovery portal for SSH resources, comes with a flexible data model allowing dynamic properties to be adapted to the diversity of information coming from heterogeneous sources and covering a wide variety of resource types (tools, services, training materials, datasets, publications, and workflows). Some of these properties are concept-based and take their values from controlled vocabularies. The most important and relevant SSHOMP vocabularies for the ATRIUM project are "Intended audience", "Invocation type", "SSH Open Marketplace Format", "SSH Open Marketplace Keyword" and "SSK standard".
The "Intended audience" vocabulary allows users to browse audience groupings and intended target audiences for different types of content, providing a structured view of these audiences. This helps researchers and developers identify and classify their intended audience more effectively.
The "Invocation type" vocabulary provides classifications for how tools or services are used ("invoked"), particularly in the context of digital research solutions and the SSH Open

Marketplace. It includes categories for different modes of use, supporting the integration of tools in other systems. It was initially created in the context of the SSH Conversion Hub, a registry of tools capable of converting metadata and data from one format to another, now available and maintained within the SSH Open Marketplace.
The "SSH Open Marketplace Format" vocabulary provides information about the data formats a tool or service can use for in- and output.
The "SSH Open Marketplace Keyword" vocabulary for describing items in the SSHOMP is the only open vocabulary used in the SSHOMP. As a folksonomy, it allows users to suggest new concepts and is subject to regular curation by the SSHOMP team. Concepts in this vocabulary are both user-created as well as ingested from various SSHOMP data sources.
The "SSK Standard" vocabulary is a list of standards referenced within the steps of the initial population of the SSHOMP workflows. It serves as a resource for researchers and professionals to navigate different standards used in digital research and standardisation efforts.

Website: https://vocabs.sshopencloud.eu/browse/en/

Maintainer: SSHOMP (ACDH-CH, DARIAH, PSNC)

Major Adopters: DARIAH, CESSDA, SSHOMP

Encoded in: SKOS, RDF/XML, Turtle

Relevance to ATRIUM: The SSHOMP vocabularies describe a large set of heterogeneous entries, covering multiple SSH domains. Some vocabularies are adapted from legacy data sources and expanded for use in the SSHOMP. Within ATRIUM, further curation of some of these vocabularies, and alignment with other existing vocabularies in use within the community is of utmost importance as it could also enable better sustainability and maintenance of these common vocabularies, supporting initiatives like SSH Vocabulary Commons (Broeder and Ďurčo 2023).

Documentation available at:
https://marketplace.sshopencloud.eu/contribute/metadata-guidelines#vocabularies

TaDiRAH

Type: Taxonomy
Domain: Digital Humanities activities

Description: "Taxonomy of Digital Research Activities in the Humanities" (TaDiRAH) was designed to help community-driven sites and projects structure their digital humanities (DH) content and gain better visibility. TaDiRAH provides terminology for DH research activities as well as scope notes that also explain the methods associated with them. Version 2.0 (linked below) is a complete revision and SKOS-ification of the taxonomy. For more information and additional resources, see https://www.tadirah.info/.

Website: https://vocabs.dariah.eu/tadirah/

Maintainer: Luise Borek, Canan Hastik, Vera Khramova, Jonathan Geiger (creator), TaDiRAH Board (provider)

Major Adopters: SSHOMP, TAPoR, Digital Humanities Course Registry

Encoded in: SKOS

Official namespace: https://vocabs.dariah.eu/tadirah/

Relevance to ATRIUM: TaDiRAH is used to describe the ATRIUM catalogue resources (see Software & Services catalogue and Workflows catalogue). Thanks to its comprehensive coverage of DH activities and its community-based maintenance, TaDiRAH is a good candidate to be used as much as possible in ATRIUM to categorise the activities performed by the workflows and demonstrators (from WP4 'Providing Enhanced Workflows for Frontier Research in the Humanities' and WP5 'Curiosity-driven Demonstrators'). In addition, TaDiRAH can also help in the task of identifying overlapping vocabularies as well as exploring the possibility of aligning or merging different vocabularies (see, for example, slide 7 in Broeder and Odijk 2024).

Documentation available at: https://tadirah.info/

TRIPLE Vocabulary

Type: Controlled vocabulary, Thesaurus
Domain: Social Sciences and Humanities

Description: The TRIPLE Vocabulary is a multilingual and hierarchical set of 3,375 SSH-related concepts, linked to several well-known controlled vocabularies, including LCSH (Library of Congress Subject Headings), WorldCat, and Wikidata. It covers popular SSH aspects, it has been formalised in SKOS and provides translations in 11 languages including Greek, French, Polish, German, Italian, Portuguese, Spanish and Croatian. The

vocabulary is used for the automatic annotation of the documents of the GoTriple platform.

Website: https://www.semantics.gr/authorities/vocabularies/SSH-LCSH/

Maintainer: EKT

Major Adopters: GoTriple

Encoded in: RDF/XML, JSON-LD, N-triples, CSV (Triple)

Official namespace: http://semantics.gr/authorities/SSH-LCSH/

Relevance to ATRIUM: The TRIPLE Vocabulary offers a comprehensive representation of the Social Sciences and Humanities domain, adhering to the 5-star principles of Linked Open Data. Besides its use to enrich GoTriple's content, this vocabulary can be reused in any scenario in which the annotation of SSH-related concepts is required.

Model available at: http://semantics.gr/authorities/vocabularies/SSH-LCSH

Documentation available at:
https://www.semantics.gr/authorities/vocabularies/SSH-LCSH/vocabulary-entries?language=en

6. Additional Resources

6.1 Deliverables from Related Initiatives

This section presents a collection of deliverables from various initiatives and projects that are relevant to the task of achieving semantic interoperability in the Arts and Humanities domain. These deliverables provide insights into standards, procedures, protocols and ontologies that have been developed and implemented in related fields.

Title: 4CH - D4.1 Report on standards, procedures and protocols

> **Abstract:** The main objective of the Deliverable is to present an overview on standards, procedures and protocols for Cultural Heritage (CH) digitisation, based on a critical review of the state of the art in data capturing, data processing and data storage/access within the overall framework of Heritage Documentation.
>
> **Available at:** https://zenodo.org/records/7701529

Title: PARTHENOS D4.2 Report on Standardization

> **Abstract:** The report reflects the second stage of the definition of the Standardization Survival Kit (SSK). This report is based on various user scenarios presented in Deliverable 4.1, where each stage of the research process has been annotated according to the actual standards that are needed in order to fulfil the research task. This includes a systematic review of the activities that have to be carried out to provide support to researchers in using, but also contributing to, these standards.
>
> **Available at:** https://zenodo.org/records/2668414

Title: PARTHENOS D4.4 Report on Standardization – Final

> **Abstract:** The document presents the final findings of WP4 activities in the context of complete research scenarios, created by domain experts and designed to be displayed within the Standardization Survival Kit web application.
>
> **Available at**: https://zenodo.org/records/2607014

Title: ARIADNEplus - D4.4 – Final report on ontology implementation

Abstract: The deliverable describes the activity of ARIADNEplus project's Task 4.4 focused on implementing the ARIADNEplus Ontology (AO-Cat) and its extensions for specific sub-domains of archaeology. The result was 14 potential application profiles, one of which is presented in detail for fieldwork activities. The AO-Cat, a new standard, allowed for the integration of over three million archaeological resources and was efficient in combination with CIDOC CRM ecosystem extensions. Among the new models, the application profile for scientific data (CRMhs) and bio-archaeology were of interest. The definition of a diverse set of ontological models is an achievement for ARIADNEplus, prioritising integration and interoperability. The application profiles are innovative tools for other research domains, and the ARIADNE Ontology enriches the family of CIDOC CRM ontologies.

Available at: https://zenodo.org/records/7636720

Title: SSHOC D3.1 Report on SSHOC (meta)data interoperability problems

Abstract: This report provides an inventory of data formats and metadata standards that are currently used and relevant for the research infrastructures managed by the SSHOC main stakeholders, recommendations of specific formats and standards for increasing interoperability, and prioritisations for providing conversion services and planning solutions.

Available at: https://zenodo.org/doi/10.5281/zenodo.3569867

Title: TRIPLE Deliverable: D2.5 - Report on Data Enrichment

Abstract: In this deliverable, the strategies for data enrichment in TRIPLE are presented. Through the Core Pipeline, named SCRE, metadata regarding publications and projects for the Social Sciences and Humanities are automatically harvested, mapped in the TRIPLE data model, curated, enriched and finally saved in the GoTriple platform's indexes.

Available at: https://zenodo.org/records/7359654

Title: TRIPLE Deliverable: D2.4 Report on identification and creation of new vocabularies

> **Abstract:** The GoTriple Vocabulary is a multilingual hierarchical set of 3,375 SSH-related concepts. It is a subset of LCSH (Library of Congress Subject Headings), which covers popular SSH subject areas. The English labels are enhanced with labels in Greek, French, Polish, German, Italian, Portuguese, Spanish, Croatian and Ukrainian. The vocabulary conforms to the SKOS data model and is published as Linked Open Data (LOD) under http://semantics.gr/authorities/vocabularies/SSH-LCSH in Semantics.gr, which is a platform developed by EKT for managing and publishing LOD vocabularies, thesauri and authority files of any schema. The vocabulary is used by the annotation service but is also a standalone product, since it is published under an open licence and can be used by the SSH research communities.
>
> **Available at:** https://zenodo.org/records/7539922

Title: Alignment of multilingual vocabularies in the Social Sciences and Humanities (SSH) Working Group

> **Abstract:** The Research Data Alliance (RDA) Working Group intends to address challenges concerning the interoperability and multilingual nature of SSH controlled vocabularies. The Working Group plans to face these issues by providing collectively approved recommendations for creating, extending, updating, and aligning multilingual vocabularies in the SSH. The objective of this RDA WG, currently co-chaired by OPERAS RI (the coordinator of TRIPLE), is to focus on a part of the SSHOC vocabulary roadmap and to collect from a broader community both insights and feedback about a methodology for the alignment of controlled vocabularies in many languages.
>
> **Available at:**
> https://www.rd-alliance.org/app/uploads/2024/05/Alignment-of-multilingual-vocabularies-in-the-Social-Sciences-and-Humanities-SSH-WG-Case-Statement-v.1.0.pdf

Title: D3.9 Report on Ontology and Vocabulary Collection and Publication

> **Abstract:** This deliverable pertains to SSHOC Task 3.1, which was responsible for investigating and providing resources and tools to support the multilingual aspects of the future pan-EU SSH infrastructure. Making data and services

accessible and usable in SSH is very much also a matter of providing relevant translations, translation of metadata concepts, multilingual vocabularies, terminology extraction across languages, and multilingual databases. The deliverable offers a detailed report on the gathering and translation of relevant SSH metadata, ontologies and vocabularies for the use cases indicated in the task's topics: multilingual metadata concepts and vocabularies, the multilingual occupation ontology, with cross-country female occupational titles. In accordance with SSHOC and the EOSC FAIR recommendations and requirements, the metadata vocabularies and ontologies have been published via several different formats and facilities.

Available at: https://doi.org/10.5281/zenodo.5913484

Title: D2.2 An Ontological Approach to the Study of European Popular Culture

Abstract: This document summarises the work conducted by DETECt researchers to design an ontology of European crime narratives, as a methodological contribution to the research on the representation of transcultural identity in European popular culture. The report details how the management of the multiple challenges involved in a complex Digital Humanities project, based on the collaboration among numerous different disciplines, can be facilitated by the adoption of knowledge mapping techniques. Building on the assumption that the development of the Semantic Web has created unique conditions for both the expression and the renewal of specifically humanistic skills, the project aims to propose a replicable model for the integration of research and educational activities in a transcultural/transnational dimension.

Available at:
https://cris.unibo.it/retrieve/e1dcb335-487a-7715-e053-1705fe0a6cc9/D2.2-FINAL-WEB_compressed.pdf

Title: Deliverable D4.1: Ontology and Vocabularies for Olfactory Information

Abstract: This document describes the Odeuropa data model and the olfactory vocabularies developed as part of WP4. It describes the methodology adopted for designing the ontology and provides a number of examples of how to use it. Finally, it shows how this model can be evaluated using competency questions.

Available at:
https://odeuropa.eu/wp-content/uploads/2022/05/D4_1_Ontology_and_Vocabularies_for_Olfactory_Information.pdf

Title: Semantic Data for Humanities and Social Sciences (SDHSS)

Abstract: The extension of CIDOC CRM for semantic data for humanities and social sciences (SDHSS) stems from the need to conceptualise the reality in the world, and more specifically factual information, from the point of view of historical research. The ontological commitment is therefore related to the domain of discourse of history. But insofar as history — as a discipline that studies the life of humans and societies in the past — is interested in all the different aspects of social, economic, political, religious, literary and cultural life, the scope of this extension could be defined as the whole of social and human life, apprehended from the descriptive point of view and global approach to reality that characterises historical research and, more generally, humanities and social sciences.

Available at: https://ontome.net/namespace/11

Title: Deliverable D4.18 SSHOCro v.1.0 beta version

Abstract: The SSHOC Reference Ontology (SSHOCro) proposes an ontological model and RDF Schema to be used as a top-level ontology for organising knowledge and information distributed across various primary sources of information in the Social Sciences and Humanities Open Cloud (SSHOC).

Available at: https://doi.org/10.5281/zenodo.3744925

Title: 5D2.1: Ontology-based knowledge graphs for music objects (V1.0)

Abstract: This deliverable describes the first version of the ontologies and knowledge graphs of music objects and patterns in Polifonia, as well as the software engineered to this end. It contains a description of the ontologies developed for musical objects, patterns, and musical heritage metadata; and a preliminary technical report on a software package that transforms music collections — containing those aforementioned objects, patterns, and metadata — into knowledge graphs that use the developed ontologies as basic semantic schemas to organise their content. The specific point of departure of this deliverable are the existing ontologies and ontology design patterns (ODP) that are

relevant to the domain of European musical cultural heritage, the existing related datasets that were surveyed in — as well as requirements that stemmed from — previous Polifonia work in WP1 Pilots, and their related stories, personas and datasets. These pilots, stories and personas constitute the used modelling domain and main source of requirements.

Available at:
https://polifonia-project.eu/wp-content/uploads/2022/01/Polifonia_D2.1_V1.0.pdf

Title: CLS INFRA D6.3 Standards beyond TEI / Extended Transformation Matrix / Alternative Formats

Abstract: This deliverable builds on and further extends the findings of D6.1 'Inventory of existing data sources and formats' surveying the landscape of literary corpora, as well as D8.1 'Tools for NLP' cataloguing the set of tools in the context of CLS. Focusing on the wealth of formats used when encoding and processing text, it offers a comprehensive overview of common formats for encoding textual data, beyond the "lingua franca", TEI, both in the domain of computational literary studies and computational linguistics, highlighting potential discrepancies in the approach between these two areas of research. The overview reveals a very heterogeneous landscape with a plethora of formats, devised for differing tasks, from philological encoding of historical text material, to computational annotation and processing of text.
Considering interoperability an indispensable key to reusability, the deliverable explores the challenges and approaches converting between formats.
This information compilation is considered as input for further developing the Transformation Matrix, introduced in D6.1, which shall serve as a conceptual framework to consolidate existing solutions for format conversion in the Transformation Toolbox to be delivered by the end of the project (D6.2). The Transformation Matrix shall allow to capture information about specific data structures (features) present in datasets as well as data structures required or produced by tools. This requires a sufficiently expressive formalised description, which is proposed in the CLSCor data model.

Available at: https://doi.org/10.5281/zenodo.10209607

7. Conclusion

In conclusion, this document provides an overview of the data formats, metadata schemas, ontologies, controlled vocabularies, and other resources crucial for achieving (semantic) interoperability between resources available in the ecosystem constituted by the research infrastructures involved in ATRIUM. The document primarily focuses on resources that are widely used in their respective fields and have demonstrated significant benefits in their application. We have outlined the potential benefits of using these resources in ATRIUM, particularly regarding the project's goal to enhance findability, accessibility, interoperability, and sustainability of research and cultural heritage data by improving the quality of the metadata used to describe the diverse data objects.

Building on this document, future work will include providing support to the project's development activities, with particular attention to efficient data encoding for specific cross-domain integration. This includes finding solutions to provide and improve semantic interoperability via mappings between schemas and vocabularies. This also involves the optimal application of models and vocabularies identified in this report to other activities in the ATRIUM project, especially the workflows and demonstrators (WP4 and WP5) as well as the interoperability of services (WP6). Given the dynamic and evolving landscape, the list of resources in this document will remain a living and open collection, which will accommodate additional useful resources that may emerge in the future. To this end, and also to enable reuse of this meta-resource as a registry, the information will be maintained in a structured way and openly published, encoded according to the AO-Cat ontology.

References

Aloia, N., Binding, C., Cuy, S., Doerr, M., Fanini, B., Felicetti, A., Fihn, J., Gavrilis, D., Geser, G., Hollander, H., Meghini, C., Niccolucci, F., Nurra, F., Papatheodorou, C., Richards, J., Ronzino, P., Scopigno, R., Theodoridou, M., Tudhope, D., Vlachidis, A. and Wright, H. 2017 Enabling European Archaeological Research: The ARIADNE E-Infrastructure, Internet Archaeology 43. https://doi.org/10.11141/ia.43.11.

Bertozzi, A. 2024. TRIPLE Ontology: A reusable formalisation of scientific entities in the Social Sciences and Humanities domain [Zenodo]. https://doi.org/10.5281/zenodo.14203413

Broeder, D. and Ďurčo, M. 2023. SSH Vocabulary Commons. In: Use of vocabularies for metadata curation and quality assessment in Social Sciences and Humanities, Session 1. DARIAH-Campus. https://doi.org/10.5281/zenodo.7871204

Broeder, D. and Odijk, J. 2024. Vocabularies in CLARIN: Problems and suggested solutions. CLARIN Annual Conference 2024. https://www.clarin.eu/sites/default/files/03_Broeder-Odijk_CLARIN2024_Metadata.pdf

Castelli, L., Felicetti, A. and Proietti, F. 2021. Heritage Science and Cultural Heritage: standards and tools for establishing cross-domain data interoperability. International Journal of Digital Libraries, 22, 279–287. https://doi.org/10.1007/s00799-019-00275-2.

Felicetti, A. and Murano, F. 2021. Semantic modeling of textual entities: The CRMtex model and the ontological description of ancient texts. Umanistica Digitale, 11, 163–175. https://doi.org/10.6092/issn.2532-8816/13674.

Niccolucci, F. and Felicetti, A. 2018. A CIDOC CRM-based Model for the Documentation of Heritage Sciences. In: Proceedings of the 2018 3rd Digital Heritage International Congress (DigitalHERITAGE) held jointly with 2018 24th International Conference on Virtual Systems & Multimedia (VSMM 2018), San Francisco, CA, USA, 26–30 October 2018; pp. 1–6. https://doi.org/10.1109/DigitalHeritage.2018.8810109.

Niccolucci, F. and Felicetti, A. 2024. Digital Twin Sensors in Cultural Heritage Ontology Applications. Sensors 24, 3978. https://doi.org/10.3390/s24123978.

Niccolucci, F., Felicetti, A. and Hermon, S. 2022. Populating the Data Space for Cultural Heritage with Heritage Digital Twins. Data 7, 105. https://doi.org/10.3390/data7080105.

Windhouwer, M. and Goosen, T. 2022. Component metadata infrastructure. CLARIN: The infrastructure for language resources, 191-222. https://doi.org/10.1515/9783110767377-008

Online Registries

Basic Register of Thesauri, Ontologies and Classifications (BARTOC): https://bartoc.org/

CLARIN Standards Information System: https://standards.clarin.eu/sis/

FAIRsharing Standard Catalogue: https://fairsharing.org/search?fairsharingRegistry=Standard

Linked Open Data Vocabularies: https://lov.linkeddata.es/dataset/lov/

Linked Open Vocabularies: http://lov.okfn.org/

RDA Metadata Working Group Metadata Registry: https://rdamsc.bath.ac.uk/

Appendix A

This appendix contains a table detailing all the semantic resources collected during the preparation of this deliverable. It derives from the table compiled and maintained in the database tool Baserow, as specified in Section 2 'Methodological Considerations'. It includes not only the semantic resources described in detail in the deliverable, but also any other resources not selected for inclusion in the main body of the deliverable.
This appendix reflects the state of the list at the time of submission of the deliverable; however, the list may be updated in the future and may be published in other forms (in particular, encoded according to the AO-Cat ontology) in the next phases of the project. Basic information about each semantic resource is provided, including the domain, type and provider of the resource, as well as links to websites and documentation.

NAME	DOMAIN	TYPE	PROVIDER	AVAILABLE AT	DOCUMENTATION
AO-Cat	Archaeology	Ontology	ARIADNE RI	https://github.com/ARIADNE-Infrastructure/AO-Cat-Ontology	- The AO-Cat Ontologyhttps://zenodo.org/records/7818375 - ARIADNEplus D4.3 – Final report on dataset integration https://zenodo.org/records/7612672 - ARIADNEplus Data Aggregation Pipeline: User Guide https://zenodo.org/records/8060925
Archaeological Objects Thesaurus	Archaeology	Thesaurus	Forum on Information Standards in Heritage (FISH)	https://heritage-standards.org.uk/fish-vocabularies/#archaeological-objects-thesaurus	

NAME	DOMAIN	TYPE	PROVIDER	AVAILABLE AT	DOCUMENTATION
Archaeological Sciences Thesaurus	Archaeology	Thesaurus	Forum on Information Standards in Heritage (FISH)	https://heritage-standards.org.uk/fish-vocabularies/#archaeological-objects-thesaurus	
ARCO	Cultural Heritage	Metadata schema	ARCO Project	https://www.arco-web.org/publicDocuments/D8/ARCO-D8-1.0-R-251002.pdf	
Backbone Thesaurus	Subject Classification	Thesaurus	DARIAH	https://vocabs.dariah.eu/bbt/ConceptScheme/Backbone_Thesaurus	
BibFrame	Library Systems	Ontology	Library of Congress	https://www.loc.gov/bibframe/docs/index.html	
BIBO (Bibliographic Ontology)	Bibliography	Ontology	W3C	https://bibo.readthedocs.io/en/latest/	https://bibo.readthedocs.io/en/latest/
BIBO Publication Type	Bibliography	Controlled vocabulary		http://purl.org/ontology/bibo/publicationTypeScheme	https://www.dublincore.org/specifications/bibo/bibo/
BiRO	Bibliography	Ontology	David Shotton, Silvio Peroni	https://sparontologies.github.io/biro/current/biro.html	https://sparontologies.github.io/biro/current/biro.html
British Museum Object Names Thesaurus	Museum Collections	Ontology	British Museum	https://terminology.collectionstrust.org.uk/British-Museum-objects/	https://terminology.collectionstrust.org.uk/British-Museum-objects/

NAME	DOMAIN	TYPE	PROVIDER	AVAILABLE AT	DOCUMENTATION
Building Materials Thesaurus	Cultural Heritage	Thesaurus	Forum on Information Standards in Heritage (FISH)	https://heritage-standards.org.uk/fish-vocabularies/#building-materials-thesaurus	
CARARE	Archaeology, Architecture, Cultural Heritage	Metadata schema	CARARE Association	https://www.carare.eu/en/services/carare-aggregation-services/carare-metadata-schema/carare-version-20/	https://www.carare.eu/en/services/carare-aggregation-services/carare-metadata-schema/
Cargo Thesaurus	Library Science	Thesaurus	Forum on Information Standards in Heritage (FISH)	https://heritage-standards.org.uk/fish-vocabularies/#cargo-thesaurus	
CDWA (Categories for the Description of Works of Art)	Art works	Ontology	Getty Vocabulary Program	https://www.getty.edu/research/publications/electronic_publications/cdwa/	https://www.getty.edu/research/publications/electronic_publications/cdwa/
CERL Thesaurus	Book History	Ontology	CERL	https://data.cerl.org/thesaurus/	https://data.cerl.org/thesaurus/
CESSDA	Social Sciences	Ontology	CESSDA	https://zenodo.org/records/4751455	https://zenodo.org/records/4672248
CIDOC CRM	Cultural Heritage, Museums, Historical Studies, Archaeology	Ontology	CIDOC - ICOM International Committee for Documentation	https://www.cidoc-crm.org/versions-of-the-cidoc-crm	https://www.cidoc-crm.org https://www.cidoc-crm.org/use_and_learn

NAME	DOMAIN	TYPE	PROVIDER	AVAILABLE AT	DOCUMENTATION
CiTO, the Citation Typing Ontology	Bibliography	Ontology	David Shotton, Silvio Peroni	https://sparontologies.github.io/cito/current/cito.html	https://sparontologies.github.io/cito/current/cito.html
CLARIN Concept Registry	Linguistics, General	Ontology	CLARIN ERIC	https://concepts.clarin.eu	https://www.clarin.eu/content/clarin-concept-registry
CLARIN/VLO Organisations vocabulary	General	Controlled vocabulary	CLARIN ERIC	https://github.com/clarin-eric/VLO-mapping/blob/master/uniform-maps/OrganisationControlledVocabulary.xml	
CMDI: ISO 3166-1 alpha-2 country codes	General	Controlled vocabulary	CLARIN ERIC	https://catalog.clarin.eu/ds/ComponentRegistry/rest/registry/1.x/components/clarin.eu:cr1:c_1271859438104	https://en.wikipedia.org/wiki/ISO_3166-1_alpha-2
CMDI: ISO 639-1 language codes	General	Controlled vocabulary	CLARIN ERIC	https://catalog.clarin.eu/ds/ComponentRegistry/rest/registry/1.x/components/clarin.eu:cr1:c_1271859438109	https://en.wikipedia.org/wiki/ISO_639
CMDI: ISO 639-3 language codes	General	Controlled vocabulary	CLARIN ERIC	https://catalog.clarin.eu/ds/ComponentRegistry/rest/registry/1.x/components/clarin.eu:cr1:c_1271859438110	https://en.wikipedia.org/wiki/ISO_639
COAR Vocabularies	General	Controlled vocabulary	COAR	https://vocabularies.coar-repositories.org/	https://vocabularies.coar-repositories.org/
COCSH Ontology	3D digitisation of cultural heritage	Ontology	COST Action TD1201	https://heritagesciencejournal.springeropen.com/articles/10.1186/s40494-021-00561-w	

NAME	DOMAIN	TYPE	PROVIDER	AVAILABLE AT	DOCUMENTATION
Component Metadata Infrastructure (CMDI)	Social Sciences and Humanities	Framework / meta-model	CLARIN	https://www.clarin.eu/content/cmdi-specification-version-12 https://catalog.clarin.eu/ds/ComponentRegistry	https://www.clarin.eu/cmdi https://github.com/clarin-eric/awesome-cmdi
CRediT (Contributor Role Taxonomy)	General	Taxonomy	CRediT	https://credit.niso.org/	https://credit.niso.org/
CRMarchaeo	Cultural Heritage, Archaeology, Archaeological excavation	Ontology	CIDOC CRM	https://cidoc-crm.org/crmarchaeo/fm_releases	https://cidoc-crm.org/crmarchaeo https://cidoc-crm.org/crmarchaeo/use-learn-crmarchaeo
CRMba	Archaeology, Standing Structures, Cultural Heritage	Ontology	CIDOC CRM	https://cidoc-crm.org/crmba/fm_releases	https://cidoc-crm.org/crmba/ https://cidoc-crm.org/crmba/use-learn-crmba
CRMdig	Digital Libraries	Ontology	CIDOC CRM	https://cidoc-crm.org/crmdig/fm_releases	https://cidoc-crm.org/crmdig/ https://cidoc-crm.org/crmdig/use-learn-crmdig
CRMinf	Cultural Heritage, Archaeology, General, Scholarly domain	Ontology	CIDOC CRM	https://cidoc-crm.org/crminf	https://cidoc-crm.org/crminf https://cidoc-crm.org/crminf/use-learn-crminf
CRMsci	Cultural Heritage, Provenance, Digital products	Ontology	CIDOC CRM	https://cidoc-crm.org/crmsci/fm_releases	https://cidoc-crm.org/crmsci/ https://cidoc-crm.org/crmsci/use-learn-crmsci

NAME	DOMAIN	TYPE	PROVIDER	AVAILABLE AT	DOCUMENTATION
CRMtex	Cultural Heritage, Ancient documents, Epigraphy	Ontology	CIDOC CRM	https://cidoc-crm.org/crmtex/fm_releases	https://cidoc-crm.org/crmtex/ https://cidoc-crm.org/crmtex/use-learn-crmtex
Crossref relationship types	General	Controlled vocabulary	Crossref	https://www.crossref.org/documentation/schema-library/markup-guide-metadata-segments/relationships/	https://www.crossref.org/documentation/schema-library/markup-guide-metadata-segments/relationships/
Crossref retraction watch	General	Controlled vocabulary	Crossref	https://retractionwatch.com/retraction-watch-database-user-guide/retraction-watch-database-user-guide-appendix-b-reasons/	https://retractionwatch.com/retraction-watch-database-user-guide/retraction-watch-database-user-guide-appendix-b-reasons/
Cultural Objects Name Authority (CONA)	Cultural Heritage	Thesaurus	The Getty Research Institute	https://www.getty.edu/research/tools/vocabularies/cona/index.html	
Data Documentation Initiative (DDI) Codebook	Social, behavioral, economic, and health sciences	Metadata schema	DDI Alliance	https://ddialliance.org/	https://ddialliance.org/Specification/DDI-Codebook/2.5/

NAME	DOMAIN	TYPE	PROVIDER	AVAILABLE AT	DOCUMENTATION
Data Documentation Initiative (DDI) Lifecycle	Social, behavioral, economic, and health sciences	Metadata schema	DDI Alliance	https://ddialliance.org/	https://ddialliance.org/Specification/DDI-Lifecycle/3.3/
DataCite schema	Research datasets and citation	Metadata schema	DataCite	http://schema.datacite.org/	https://datacite-metadata-schema.readthedocs.io
DCAT	Data catalogs and datasets	Metadata schema	W3C	https://www.w3.org/ns/dcat.rdf	https://www.w3.org/TR/vocab-dcat/
DEFC Thesaurus	General	Thesaurus	ACDH-CH	https://vocabs.acdh.oeaw.ac.at/defc_thesaurus/en/	
Dewey Decimal Classification (DDC)	Library Classification	Ontology	OCLC	https://docs.datadrivencooking.com/docs/started-introduction.html	https://docs.datadrivencooking.com/docs/started-introduction.html
Dublin Core Metadata Initiative (DCMI)	Web Resources, Digital Libraries	Ontology	Dublin Core Metadata Initiative		

NAME	DOMAIN	TYPE	PROVIDER	AVAILABLE AT	DOCUMENTATION
DYAS Humanities Thesaurus	Anthropology and ethnology, Art and history of art, Byzantine archaeology, Classical archaeology, Classical studies, Epigraphy, History and philosophy of sciences, Linguistics, Literary studies, Modern and contemporary history, Musicology, Numismatics, Prehistoric archaeology, Theatre studies, Ancient History	Thesaurus		https://humanitiesthesaurus.academyofathens.gr/HUMANITIES-THESAURUS/en/	
EAD (Encoded Archival Description)	Archaeology, Cultural Heritage, Library Systems, Library Science	Ontology	Society of American Archivists	https://www.uscis.gov/green-card/green-card-processes-and-procedures/employment-authorization-document	https://www.uscis.gov/green-card/green-card-processes-and-procedures/employment-authorization-document

NAME	DOMAIN	TYPE	PROVIDER	AVAILABLE AT	DOCUMENTATION
EOSC Resource Category, Subcategory (and Supercategory)	General	Controlled vocabulary	SSHOC	https://vocabs.sshopencloud.eu/vocabularies/eosc-resource-category/eoscResourceCategoryScheme	
EOSC Resource Geographical Availability	General	Controlled vocabulary	SSHOC	https://vocabs.sshopencloud.eu/vocabularies/eosc-geographical-availability/eoscGeographicalAvailabilityScheme	
EOSC Resource Life Cycle Status	General	Controlled vocabulary	SSHOC	https://vocabs.sshopencloud.eu/vocabularies/eosc-life-cycle-status/eoscLifeCycleStatusScheme	
Epigraphic Ontology	Epigraphy	Ontology		https://zenodo.org/records/4639508	
Europeana Data Model (EDM)	Digital Libraries, Cultural Heritage, Bibliographic Data, Library Systems, Scholarly domain	Ontology	Europeana	https://pro.europeana.eu/page/edm-documentation	https://pro.europeana.eu/page/edm-documentation
EuroVoc	EU Legislation	Ontology	European Parliament	https://op.europa.eu/en/web/eurovoc	https://op.europa.eu/en/web/eurovoc
Event Thesaurus	Cultural Heritage	Thesaurus	Forum on Information Standards in Heritage (FISH)	https://heritage-standards.org.uk/fish-vocabularies/#event-thesaurus	

NAME	DOMAIN	TYPE	PROVIDER	AVAILABLE AT	DOCUMENTATION
Evidence Thesaurus	Archaeology, Cultural Heritage, Cultural Heritage and Archaeology	Thesaurus	Forum on Information Standards in Heritage (FISH)	https://heritage-standards.org.uk/fish-vocabularies/#evidence-thesaurus	
FaBiO	Bibliography	Ontology	David Shotton, Silvio Peroni	https://sparontologies.github.io/fabio/current/fabio.html	https://sparontologies.github.io/fabio/current/fabio.html
Farmstead Thesaurus	Archaeology, Cultural Heritage	Thesaurus	Forum on Information Standards in Heritage (FISH)		
FAST (Faceted Application of Subject Terminology)	Subject Indexing	Ontology	OCLC	https://fast.eriksmistad.no/	https://fast.eriksmistad.no/
First World War Thesaurus	Modern and contemporary history	Thesaurus		https://heritage-standards.org.uk/fish-vocabularies/#first-world-war-thesaurus	
FOAF	Social networks and people	Ontology	FOAF Project	http://xmlns.com/foaf/spec/	http://xmlns.com/foaf/spec/
FRBR /LRMoo	Library Science, Bibliographic Data	Ontology	IFLA	https://www.oclc.org/research/activities/frbr.html	

NAME	DOMAIN	TYPE	PROVIDER	AVAILABLE AT	DOCUMENTATION
GEMET (General Multilingual Environmental Thesaurus)	Environmental Information	Ontology	European Environment Agency	https://www.eea.europa.eu/themes/data/what-is-gemet	https://www.eea.europa.eu/themes/data/what-is-gemet
Getty Art and Architecture Thesaurus (AAT)	Art History, Subject Classification, Bibliographic Data, Cultural Heritage	Thesaurus	The Getty Research Institute	https://www.getty.edu/research/tools/vocabularies/aat/	https://www.getty.edu/research/tools/vocabularies/aat/
Getty Thesaurus of Geographic Names (TGN)	Geography, Cultural Heritage, Art History	Thesaurus	The Getty Research Institute	https://www.getty.edu/research/tools/vocabularies/tgn/	https://www.getty.edu/research/tools/vocabularies/tgn/
Getty Thesaurus of Graphic Materials (TGM)	Visual Materials	Ontology	The Getty Research Institute	https://www.getty.edu/research/tools/vocabularies/tgn/	https://www.getty.edu/research/tools/vocabularies/tgn/
Heritage Crime Thesaurus	Cultural Heritage	Thesaurus		https://heritage-standards.org.uk/fish-vocabularies/#heritage-crime-thesaurus	
HeritageData (FISH UK) vocabularies	Cultural Heritage and Archaeology	Thesaurus, Controlled vocabulary	HE, HES, RCAHMW	https://www.heritagedata.org/blog/vocabularies-provided/	https://www.heritage-standards.org.uk/fish-vocabularies/
HERO - Ontology of Heraldry	Heraldry	Ontology	National Archives of Finland	http://www.yso.fi/onto/hero/	https://finto.fi/hero/en/

NAME	DOMAIN	TYPE	PROVIDER	AVAILABLE AT	DOCUMENTATION
Historic Aircraft Thesaurus	History and philosophy of sciences	Thesaurus		https://heritage-standards.org.uk/fish-vocabularies/#historic-aircraft-thesaurus	
IA - Getty Iconography Authority	Art Description	Ontology	Getty Vocabulary Program	https://www.getty.edu/research/tools/vocabularies/ia_in_depth.pdf	https://www.getty.edu/research/tools/vocabularies/ia_in_depth.pdf
IANA Media Types	General	Controlled vocabulary		https://vocabs.sshopencloud.eu/vocabularies/media-type/mediaTypeScheme	
Iconclass	Art History, Museum Studies	Thesaurus		https://vocabs.acdh.oeaw.ac.at/iconclass/en/	

NAME	DOMAIN	TYPE	PROVIDER	AVAILABLE AT	DOCUMENTATION
iDAI.world Thesaurus	Ancient History, Archaeology, Archaeology of Late Antiquity, Archaeology of Non-European Cultures, Archaeology of the Iberian Peninsula, Building Research, Classical Archaeology, Classical Studies, Conservation of Cultural Heritage, European Archaeology, Humanities, Prehistoric Archaeology, Provincial Roman Archaeology	Thesaurus		https://vocabs.dariah.eu/DAI/en/	
IIIF	Digital Libraries, Cultural Heritage	Ontology	IIIF Consortium		

NAME	DOMAIN	TYPE	PROVIDER	AVAILABLE AT	DOCUMENTATION
INDIGO Graffiti Thesaurus	Archaeology	Thesaurus	ACDH-CH	https://vocabs.acdh.oeaw.ac.at/indigo_browse/en/	
Intended audience	General	Controlled vocabulary	SSHOC	https://vocabs.sshopencloud.eu/vocabularies/sshoc-audience/audienceScheme	
Invocation type	General	Controlled vocabulary	SSHOC	https://vocabs.sshopencloud.eu/vocabularies/invocation-type/invocationTypeScheme	
ISO 639-3 Language Codes	General	Controlled vocabulary		https://vocabs.acdh.oeaw.ac.at/iso6393/Schema	

NAME	DOMAIN	TYPE	PROVIDER	AVAILABLE AT	DOCUMENTATION
KTO - Ontology for Linguistics	Linguistics	Ontology	PUBLISHER: National Library of Finland CREATOR: Institute for the Languages of Finland, National Library of Finland, Semantic Computing Research Group (SeCo0, The Finnish Terminology Centre TSK	https://finto.fi/kto/en/	https://finto.fi/kto/en/
Lexvo - language names and ISO 639-3 language codes	Linguistics, Computational Linguistics	Ontology	Lexvo.org	http://lexvo.org/	https://finto.fi/lexvo/en/
Library of Congress Subject Headings (LCSH)	Subject Classification	Ontology	Library of Congress	https://www.loc.gov/aba/publications/FreeLCSH/freelcsh.html	https://www.loc.gov/aba/publications/FreeLCSH/freelcsh.html

NAME	DOMAIN	TYPE	PROVIDER	AVAILABLE AT	DOCUMENTATION
LIDO	Museums, Cultural Heritage	Metadata schema	ICOM CIDOC	https://cidoc.mini.icom.museum/working-groups/lido/	
LRMoo	Bibliographic Information, Library Science	Ontology	International Federation of Library Associations and Institutions (IFLA), International Council of Museums (ICOM)	https://cidoc-crm.org/lrmoo/fm_releases	https://cidoc-crm.org/lrmoo
MADS/RDF Primer	Bibliographic Metadata, Bibliographic Metadata, Library Systems	Ontology	Library of Congress	https://www.loc.gov/standards/mads/rdf/	
MARC (Machine-Readable Cataloging)	Bibliographic Cataloguing	Ontology	Library of Congress	https://www.loc.gov/marc/marcdocz.html	https://www.loc.gov/marc/marcdocz.html
MEI (Music Encoding Initiative)	Musicology	Ontology	Music Encoding Initiative Consortium	https://music-encoding.org/	https://music-encoding.org/
METS	Digital libraries, Preservation	Metadata schema	Library of Congress	https://www.loc.gov/standards/premis/ontology/owl-version3.html	
MIDAS Heritage	Cultural Heritage	Ontology	The National Trust	https://historicengland.org.uk/images-books/publications/midas-heritage/	https://historicengland.org.uk/images-books/publications/midas-heritage/

NAME	DOMAIN	TYPE	PROVIDER	AVAILABLE AT	DOCUMENTATION
MODS	Digital Libraries	Ontology	Library of Congress	https://www.loc.gov/standards/mods/	https://www.loc.gov/standards/mods/
MORESS (Mapping of Research in European Social Sciences and Humanities)	Social Sciences and Humanities	Controlled vocabulary	European University Association		https://cordis.europa.eu/project/id/HPSE-CT-2002-60060
NeDiMAH Ontology	General	Ontology		https://github.com/athenarc/nemo	
Nomisma Ontology	Numismatics, Cultural Heritage, Epigraphy, Archaeology	Ontology	American Numismatic Society	https://nomisma.org/ontology	https://nomisma.org/
NOnt	Cultural Heritage, Digital Libraries	Ontology	CNR Pisa (Italy)	https://dlnarratives.eu/ontology/	https://doi.org/10.3233/SW-200421
OAI-ORE	General	Ontology	Open Archives Initiative	https://www.openarchives.org/ore/	https://www.openarchives.org/ore/
OeAI Thesaurus - Cultural Time Periods	Archaeology	Thesaurus	ACDH-CH	https://vocabs.acdh.oeaw.ac.at/oeai-cp/en/	
OeAI Thesaurus - Materials	Archaeology	Thesaurus	ACDH-CH	https://vocabs.acdh.oeaw.ac.at/oeai-materials_browse/en/	

NAME	DOMAIN	TYPE	PROVIDER	AVAILABLE AT	DOCUMENTATION
ÖFOS 2012. Austrian Fields of Science and Technology Classification 2012	General	Controlled vocabulary	ACDH-CH	https://vocabs.acdh.oeaw.ac.at/oefosdisciplines/Schema	http://www.statistik.at/web_de/statistiken/index.html
OMA - Ontology for Media Art	Media Art	Ontology	PUBLISHER: Kansalliskirjasto CREATOR: M-Cult	http://www.yso.fi/onto/oma/	
OMS	General	Ontology	Open Metadata Initiative	https://www.ogc.org/standard/om/	https://www.ogc.org/standard/om/
OpenAIRE guidelines	Scholarly domain	Metadata schema	OpenAIRE AMKE	https://guidelines.openaire.eu	https://guidelines.openaire.eu
OpenArchaeo Semantic Data Model	Archaeology	Ontology	Huma-Num consortium MASA	http://openarchaeo.huma-num.fr/openarchaeo_v2/home	http://openarchaeo.huma-num.fr/openarchaeo_v2/home
OpenCitations Ontology	Bibliographic and Citation Data	Ontology	OpenCitations.net	https://opencitations.github.io/ontology/current/ontology.html	
PACTOLS	Archaeology	Thesaurus	FRANTIQ	https://github.com/frantiq/PACTOLS	https://pactols.frantiq.fr/opentheso/
PBCore	Digital Libraries	Ontology	Public Broadcasting Metadata Group	https://pbcore.org/	https://pbcore.org/

NAME	DOMAIN	TYPE	PROVIDER	AVAILABLE AT	DOCUMENTATION
PeriodO	History, Archaeology, Art History	Controlled vocabulary, Gazetteer	PeriodO	https://client.perio.do	https://perio.do
Pleiades	Ancient geography	Gazetteer	Pleiades	https://pleiades.stoa.org/	https://pleiades.stoa.org/
PREMIS	Digital Preservation	Ontology	Library of Congress	https://www.loc.gov/standards/premis/ontology/owl-version3.html	
PROV-O	Scholarly domain	Ontology	W3C	https://www.w3.org/ns/prov-o	https://www.w3.org/TR/prov-o/
RAMEAU (Répertoire d'autorité pour les matières)	Subject Classification	Ontology	Bibliothèque nationale de France	https://rameau.bnf.fr/informations/rameauenbref	https://rameau.bnf.fr/informations/rameauenbref
Re3Data	Digital Libraries	Metadata schema		https://www.re3data.org/	
RHDTO	Cultural Heritage, Architecture, Digital Twins	Ontology	VAST-LAB PIN / ARIADNE, FORTH	https://is.gd/rhdto	https://doi.org/10.3390/s24123978; https://doi.org/10.48550/arXiv.2302.07138
RIC-O	Archives	Ontology	International Council on Archives	https://www.ica.org/standards/RiC/ontology	https://www.ica.org/standards/RiC/RiC-O_1-0-2.html
RO-Crate	Research Data	Ontology	RO-Crate Community	https://www.researchobject.org/ro-crate/specification/1.1/root-data-entity.html#ro-crate-metadata-file-descriptor	https://www.researchobject.org/ro-crate/ https://www.researchobject.org/ro-crate/specification

NAME	DOMAIN	TYPE	PROVIDER	AVAILABLE AT	DOCUMENTATION
Schema.org	Structured data on the web	Metadata schema	Schema.org	https://schema.org/docs/developers.html https://github.com/schemaorg/schemaorg	https://schema.org/ https://schema.org/docs/documents.html
Scholarly Ontology	Scholarly domain, research process	Ontology	ATHENA R.C.	https://github.com/athenarc/scholarly-ontology	http://isdb.cs.aueb.gr/scholarlyontology
SKOS	Domain-agnostic	Ontology	W3C	https://www.w3.org/2004/02/skos/vocabs	https://www.w3.org/TR/2009/REC-skos-reference-20090818/
Smithsonian 3D Metadata Model	3D digitisation, Museum and cultural heritage collections	Metadata schema	Smithsonian Institution	https://dpo.si.edu/index.php/resource/smithsonian-institution-3d-metadata-model-overview-v06	
SPDX Software License	Licensing	Controlled vocabulary	https://sshoc.poolparty.biz/user/swc	https://vocabs.sshopencloud.eu/browse/software-license/	https://spdx.org/licenses/
SSH Open Marketplace Data Model	Scholarly domain	Ontology	SSHOC		https://doi.org/10.5281/zenodo.5749465
SSH Open Marketplace Format	General	Controlled vocabulary	SSHOC	https://vocabs.sshopencloud.eu/browse/sshomp-format/en/	

NAME	DOMAIN	TYPE	PROVIDER	AVAILABLE AT	DOCUMENTATION
SSH Open Marketplace Keyword	Social Sciences and Humanities	Controlled vocabulary	SSHOC	https://vocabs.sshopencloud.eu/vocabularies/sshomp-keyword/	
SSHOC Reference Ontology	Social Science Data	Ontology	SSHOC	https://sshopencloud.eu/	https://sshopencloud.eu/
SSK Standard	General	Controlled vocabulary	SSHOC	https://vocabs.sshopencloud.eu/vocabularies/standard/standardScheme	
STAR model ('STructured Assertion Record')	Humanities, History	Metadata schema		https://ojs.elte.hu/hsce/article/view/8212/6754	https://ojs.elte.hu/hsce/article/view/8212/6754
STARC Metadata schema	Cultural Heritage	Metadata model	Cyprus Institute	https://public.cyi.ac.cy/	https://www.ocs.soton.ac.uk/index.php/CAA/2012/paper/view/649
TaDiRAH	DH activities	Taxonomy	TaDiRAH Board	https://vocabs.dariah.eu/tadirah/	https://tadirah.info/

NAME	DOMAIN	TYPE	PROVIDER	AVAILABLE AT	DOCUMENTATION
TEI	Social Sciences and Humanities	Metadata schema	Text Encoding Initiative Consortium	https://tei-c.org/	https://tei-c.org/release/doc/tei-p5-doc/en/html/index.html
The EAGLE Vocabularies	Epigraphy	Thesaurus	EAGLE (Europeana network of Ancient Greek and Latin Epigraphy)	https://www.eagle-network.eu/resources/vocabularies/	
Thesaurus of Monument Types	Cultural Heritage	Thesaurus	Forum on Information Standards in Heritage (FISH)		
TRIPLE Ontology	Social Sciences and Humanities	Ontology	The TRIPLE Consortium/Net7	https://www.gotriple.eu/ontology/triple	https://www.gotriple.eu/ontology/triple
TRIPLE Vocabulary	Social Sciences and Humanities	Controlled vocabulary, Thesaurus	EKT	https://www.semantics.gr/authorities/vocabularies/SSH-LCSH/	https://www.semantics.gr/authorities/vocabularies/SSH-LCSH/vocabulary-entries?language=en
ULAN -Union List of Artist Names Online	Artists and their works	Thesaurus	The Getty Research Institute	https://www.getty.edu/research/tools/vocabularies/ulan/	https://www.getty.edu/research/tools/vocabularies/ulan/
UNESCO Thesaurus	Archaeology, General	Thesaurus	UNESCO	https://vocabularies.unesco.org/browser/thesaurus/en/	
VIAF	Bibliographic data	Thesaurus	OCLC	https://viaf.org/	https://viaf.org/

NAME	DOMAIN	TYPE	PROVIDER	AVAILABLE AT	DOCUMENTATION
VICAV Taxonomy of Arabic Dialects	Archaeology	Thesaurus	ACDH-CH	https://vocabs.acdh.oeaw.ac.at/vicav_dialects_taxonomy/en/	
Vocabulary of Basic Terms for Cataloguing Costume	Cultural Heritage and Fashion Documentation	Thesaurus		https://terminology.collectionstrust.org.uk/ICOM-costume/	
VoID (Vocabulary of Interlinked Datasets)	Web Resources, Semantic Web, Linked Data	Ontology		https://www.w3.org/TR/void/	https://www.w3.org/TR/void/
VRA Core (Visual Resources Association Core)	Art History, Cultural Heritage, Digital Libraries	Ontology	Visual Resources Association	https://www.loc.gov/standards/vracore/schemas.html	https://www.loc.gov/standards/vracore/schemas.html
World Historical Gazetteer	History, Historical Geography Archaeology, Digital Humanities	Gazetteer	World History Center, University of Pittsburgh	https://github.com/WorldHistoricalGazetteer/whgazetteer	https://whgazetteer.org/documentation/

Consortium

Research Infrastructures

DARIAH-EU
Digital Research Infrastructure for the Arts and Humanities

ARIADNE
Research Infrastructure AISBL

CLARIN

OPERAS
open scholarly communication in the european research area for social sciences and humanities

Beneficiaries

ads Archaeology Data Service

ATHENA
Research & Innovation Information Technologies

ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ
ATHENS UNIVERSITY OF ECONOMICS AND BUSINESS

BELGRADE CENTER FOR DIGITAL HUMANITIES

Consiglio Nazionale delle Ricerche

CHARLES UNIVERSITY

foxcub.

IBL

Inria

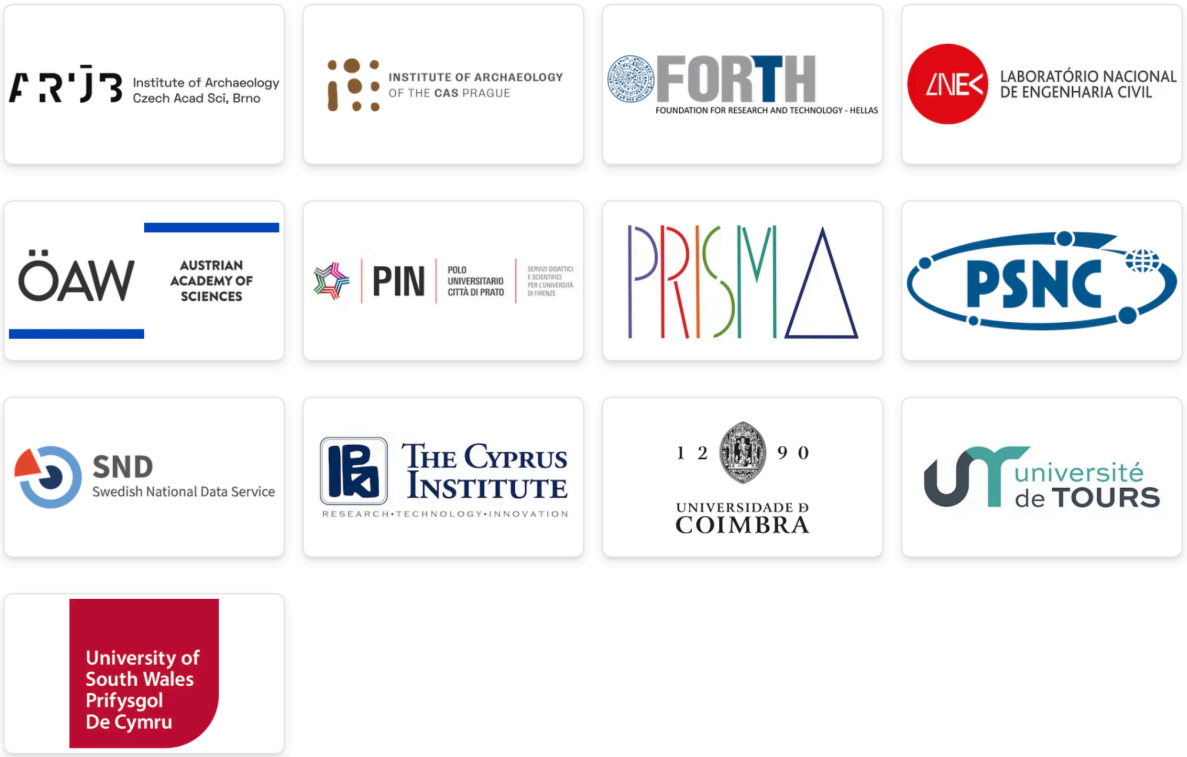
LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

net7 create develop amplify

Radboud Universiteit

University of Sheffield

Affiliated Entities



Disclaimer

This document does not represent the opinion of the European Community, and the European Community is not responsible for any use that might be made of its content. Funded by the European Union. Grant Agreement number 101132163. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Executive Agency. Neither the European Union nor the granting authority can be held responsible for them.