D5.2 First Report on Data Infrastructure update and extension

Final version

8 June 2020

Grant Agreement number:	**823914**
Project acronym:	**ARIADNEplus**
Project title:	**Advanced Research Infrastructure for Archaeological Dataset Networking in Europe - plus**
Funding Scheme:	**H2020-INFRAIA-2018-1**
Project co-ordinator name, Title and Organisation:	**Prof. Franco Niccolucci, PIN Scrl - Servizi Didattici e Scientifici per l'Università di Firenze**
Tel:	**+39 0574 602578**
E-mail:	**franco.niccolucci@pin.unifi.it**
Project website address:	**www.ariadne-infrastructure.eu**

The research leading to these results has received funding from the European Community's Horizon 2020 Programme (H2020-INFRAIA-2018-1) under grant agreement n° 823914.

Author:	**Julian Richards**
	University of York: Archaeology Data Service
Contributing partners:	**Ceri Binding, University of South Wales**
	Doug Tudhope, University of South Wales
	Alessia Bardi, CNR-ISTI
	Carlo Meghini, CNR-ISTI
	Achille Felicetti, PIN
	Maria Theodoridou, FORTH
Quality control:	**Stephanie Williams, PIN**

Document History

- 20.04.2020 – Draft Version 0.1
- 15.05.2020 – Complete draft with partner contributions v.1.0
- 29.05.2020 – Quality control review
- 02.06.2020 – Final version
- 08.06.2020 - Submission

Table of Contents

1 Executive Summary

The deliverable reports the work done in Tasks 5.1, 5.2, 5.3, 5.4, 5.5 and 5.7 during the initial 18 months, i.e. period 1 of the project, assessing it and planning the related activities for the second period, i.e. months 19-36. Work done under Task 5.6 has previously been reported under D5.1. related activity is reported in D4.1

ARIADNEplus seeks to update and extend the research data infrastructure delivered within the preceding ARIADNE project (2013-17). It extends ARIADNE in several dimensions:

1. Wider geographical coverage with new partners.
2. Wider disciplinary coverage with a greater emphasis on the sub-domains of palaeo-anthropology, bioarchaeology, environmental archaeology, material sciences, dating methods, and on the archaeology of standing structures.
3. The time span considered.
4. The depth of database integration, with a greater degree of item-level integration.
5. Greater integration of texts.
6. Broader audiences.
7. Greater range of services.

This deliverable describes the update procedures, which are being followed by partners, and introduces the steps in the aggregation pipeline. There are two options for aggregation: the standard approach using a suite of tools for the semi-automated aggregation of large data-sets, and a basic approach for the manual upload of small numbers of records. The majority of partners use the standard approach, which has been developed and tested on over 1m records from UoY-ADS, from a range of datasets. Aggregation proceeds according to an agreed priority list. ARIADNE subject types are agreed, and partners choose whether they will upload their data via XML files, or automated harvesting methods, such as OAI-PMH.

Partners following the standard approach must:

1. Describe their data according to the AO-Cat using the 3M tool, usually with one mapping per partner
2. Map subject terms to the Getty AAT using the Vocabulary Matching Tool
3. Define any period terms used so that they are uploaded to Perio.do

Where temporal data needs cleaning to create consistent use of date ranges and periods, partners use an additional tool, Time Spans, to normalise date ranges. They must also ensure that spatial data is compliant with WGS 84.

Partners using Fast Cat instead manually enter their data records in a spreadsheet-like tool, where the column headings already correspond to AO-Cat core mandatory fields, so that there can be a single mapping covering multiple partners.

Data aggregated by both routes is then transformed into the ARIADNE triplestore, and is also used to create the indices used to power Elasticsearch in the ARIADNE portal. Data is initially loaded into a "ghost" portal for checking, before it is published.

Progress has so far been monitored via a shared Googlesheet, which provides an aggregation dashboard, but during the next phase we will use a new software tool, Activity Dash (under implementation in WP14), which will make it easier to monitor the progress across a large number of partners.

During the reporting period we have so far aggregated over 1.5m records covering archaeological sites and monuments and archaeological 'events' (i.e. excavation and other fieldwork activities) from a small number of partners: UoY-ADS, AIAC, HNM, DANS-KNAW, ARUP. During the remainder of 2020 our next priority will be to complete the aggregation of this type of data, and move onto a broader range of data types, including the development of application profiles for data types which extend the subject range of the ARIADNE infrastructure, and take us into item level aggregation. This was an experimental area in ARIADNE, but will be a priority for the development of VREs in ARIADNEplus during the second phase of the project. We will also extend the range of the ARIADNEplus data infrastructure to catalogue information about people, institutions and services. We will throughout ensure compatibility with other catalogues. We have done pilot work with EOSC in the TEXTCROWD application, and going forward we will ensure the visibility of ARIADNE resources within the EOSC Hub as that is implemented.

2 Introduction and Objectives

The deliverable reports the work done within WP5, comprising Tasks 5.1, 5.2, 5.3, 5.4, 5.5 and 5.7 during the initial 18 months, i.e. period 1 of the project, assessing it and planning the related activities for the second period, i.e. months 19-36. Work done under Task 5.6 on the Data Management Plan has previously been reported under D5.1.

The overall aim of WP5 is to maintain and extend the ARIADNE Data infrastructure (ADI). This leads to three objectives:

- Keeping the ARIADNEplus Data Infrastructure up-to-date
- Extending the ARIADNEplus Data Infrastructure to other sectors
- Keeping the liaisons with other Catalogues, including the future EOSC catalogue of services

The ADI was developed during the first ARIADNE infrastructure project, with funding from the European Commission under the Commission's 7th Framework Programme for the period 2013-2017.[1] ARIADNE's goal was to provide open access to Europe's archaeological heritage and to overcome the fragmentation of digital repositories, placed in different countries and compiled in different languages (Niccolucci and Richards 2019).

Integration has been achieved using state-of-the-art ICT and open data standards, creating the ARIADNE Catalogue with advanced search functionalities and services to use and re-use data.[2]Innovative vocabulary mapping techniques have brought interoperability to a huge and heterogeneous collection of texts, drawings, images, videos, 3D models and maps. ARIADNE succeeded in integrating archaeological datasets in its Registry, with more than 1,700,000 datasets recorded and managed according to the FAIR principles (Aloia et al. 2017).

ARIADNEplus builds upon the success of ARIADNE, extending its scope and improving the technology to embed the ARIADNE infrastructure in the European Open Science Cloud (EOSC).[3] It **extends** ARIADNE in a number of dimensions, covering the domains served and the users addressed:

- The **geographic coverage**, which in ARIADNE already reached almost all European regions, by integrating in the ARIADNEplus Infrastructure a greater number of archaeological partners and giving particular attention to areas where the coverage was less intensive (Figure 1).

- The **disciplinary coverage**, which in ARIADNE included mainly excavation data and a few specialist topics such as, for example, dendrochronology, by integrating in the new ARIADNEplus Infrastructure data produced by palaeoanthropology, bioarchaeology, environmental archaeology as well as the results of scientific analyses, such as material sciences, dating and so on, and those related to standing structures, be they small remains of

[1] http://legacy.ariadne-infrastructure.eu/
[2] http://portal.ariadne-infrastructure.eu/
[3] https://www.eosc-portal.eu/

ancient constructions or complex and massive monuments as, for example, Hadrian's Wall in the UK or the Magna Graecia temples in Southern Italy (Figure 2).

- The **time-span** considered, pushing back the earliest datasets included, by incorporating palaeoanthropology, and forwards the end-date until recent times, e.g. including industrial archaeology and Cold War archaeology; in practice covering the full time-span of the human presence on Earth.

- The **depth of database integration**, exploiting the potential of well-structured datasets such as databases, for which the interoperability will be extended to item level (in ARIADNE implemented only experimentally), and archaeological Geographic Information Systems (GIS), for which integration will be achieved through the introduction of dedicated services, going beyond mere digital maps and overcoming incompatible reference systems.

- The **integration of text datasets** by extending the use of Text Mining through Natural Language Processing (NLP) and Named Entity Recognition (NER), previously applied only experimentally.

- The **research community** involved. The ARIADNEplus target is to make contact with the majority of all researchers and professionals (particularly important in this domain where research and heritage management often go hand in hand), and address most if not all the needs of computer-aware archaeologists. From existing expressions of interest, including the USA, Japan, Latin America and Australia, it is also already anticipated that ARIADNEplus will attract the international research community.

- The **service portfolio** offered to users, incorporating more advanced tools for digital analysis and interpretation in ARIADNEplus.

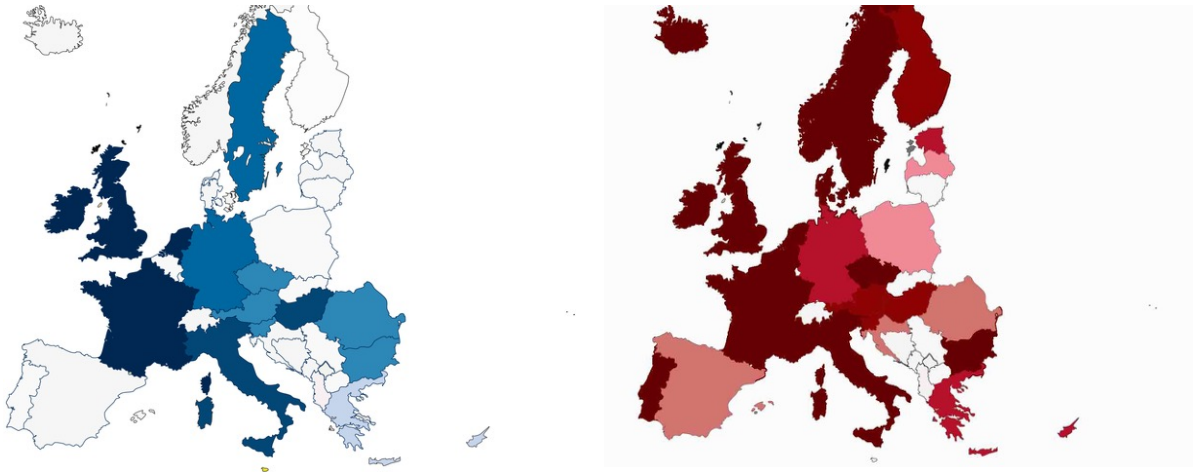


Figure 1: Extension of geographic coverage from ARIADNE to ARIADNEplus

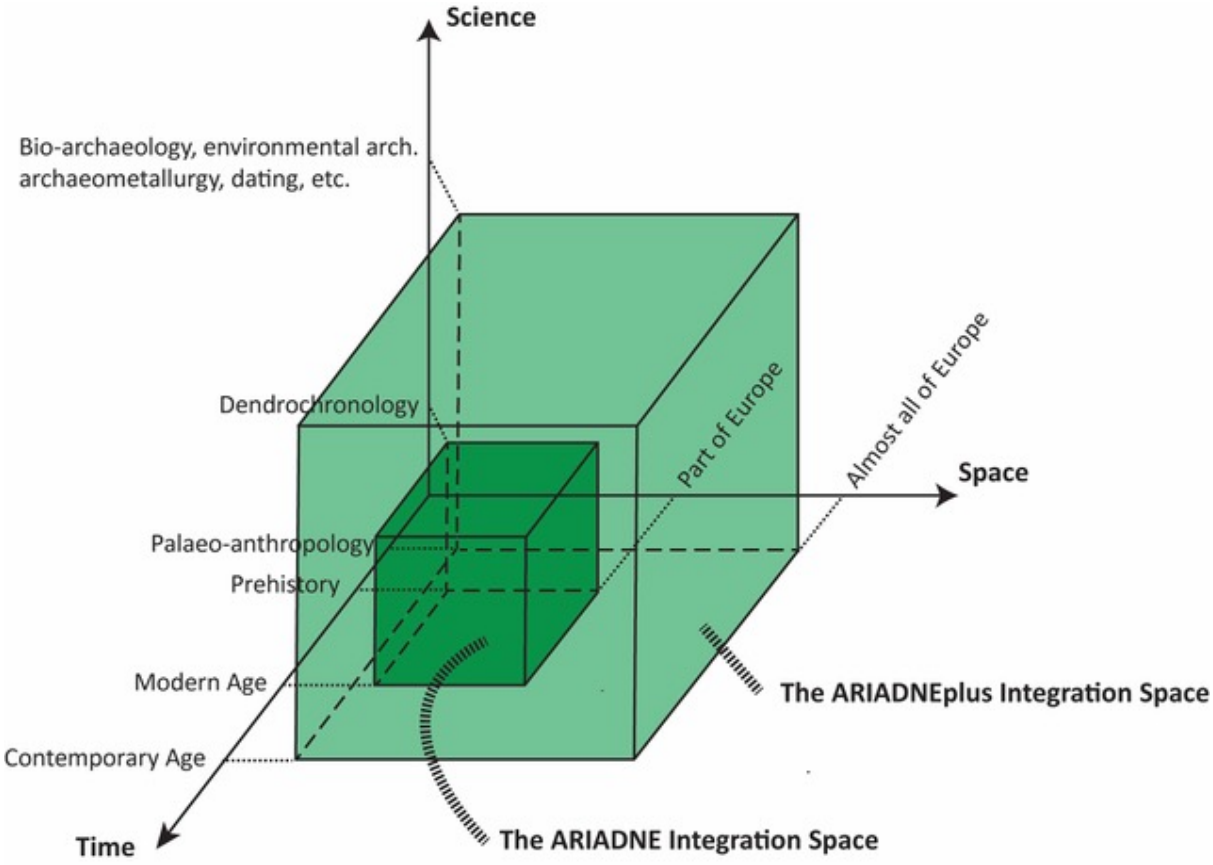


Figure 2. How ARIADNEplus extends its integration scope in time, space and content

This deliverable reports on progress to date in meeting these objectives.

Section 3 defines the ADI update procedure for both old and new partners. Section 4 discusses the data cleansing procedures. Section 5 describes how the current content of the ADI, inherited from ARIADNE, is being updated and how new data sets are being added, using an improved data model, the AO-Cat (described in D4.1). Section 6 reports on plans to extend the scope of the ADI to catalogue people, institutions and services as well as datasets, whilst Section 7 discusses collaboration and cross-discipline interoperability issues with catalogues in other domains, and compatibility with the nascent EOSC catalogue. Finally, Section 8 concludes and summarises priorities for the next phase. Related activities, including A-Cat, and the development and application of the 3M and Fast Cat tools are reported in D4.1.

3 Definition of the update procedure of the ARIADNEplus Data Infrastructure. Task 5.1 (NA4.1) Task leader: UoY-ADS

This task is responsible for defining the most appropriate pipeline for updating the ARIADNEplus Data Infrastructure, i.e. the pipeline of actions to execute. In the Description of Work it was originally envisaged that the pipeline would need to take account of the varying requirements of "old" and "new" partners and that a different procedure would be needed according to whether partners were already providing content to the former ARIADNE Catalogue, whether they were extending their content provision, or whether they were providing content to the ARIADNEplus Data Infrastructure for the first time.

However, at an early stage it was agreed that the ARIADNE Catalogue Data Model (ACDM), which had been used to describe data aggregated within the existing ARIADNE infrastructure, was unnecessarily complex and indeed, that many of its properties had not been implemented in the preceding project and were superfluous to requirements. With the adoption of the improved ARIADNEplus Data Model (AO-Cat) developed in WP4 of ARIADNEplus, data mappings could be simplified, but all the existing data aggregated during ARIADNE would need to be re-processed to reach compliance with AO-Cat (see D4.1). This has the additional advantage that partners are able to improve existing mappings and add to existing datasets at the same time.

Therefore, with the agreement that all partners would need to map their data to the AO-Cat from scratch, we have defined a single pipeline for both existing and new data suppliers, although existing Getty AAT and Perio.do mappings have been re-used where possible (see Section 4).

Aggregation priorities have been defined in Task 4.4 (see D4.1 Section 7) and we have also distinguished between those partners with large data sets who need to undertake a full 3M mapping to the AO-Cat and use the automated tools for aggregation, or those with small datasets which may be manually entered using the Fast Cat tool. These tools are described in more detail in D4.1.

A simplified version of the aggregation procedure is shown in Figure 3. In practice there are also quality control checks at each stage in the pipeline. The D4 Science helpdesk is used for partner support, with tickets corresponding to the technical partner supporting each step of the procedure (see D4.1 Section x). Experience with the test data has shown that it is an iterative rather than entirely linear process and that stages may need to be repeated before they are completed. For example, the 3M mapping process is a useful mechanism in forcing partners to think about their data in relation to the AO-Cat, and may lead to a need for re-supply of the data.

The aggregation procedure is described in more detail, but in an accessible manner, in the Data Aggregation Pipeline User Guide (Bardi et al. 2020), which has been circulated to all partners.

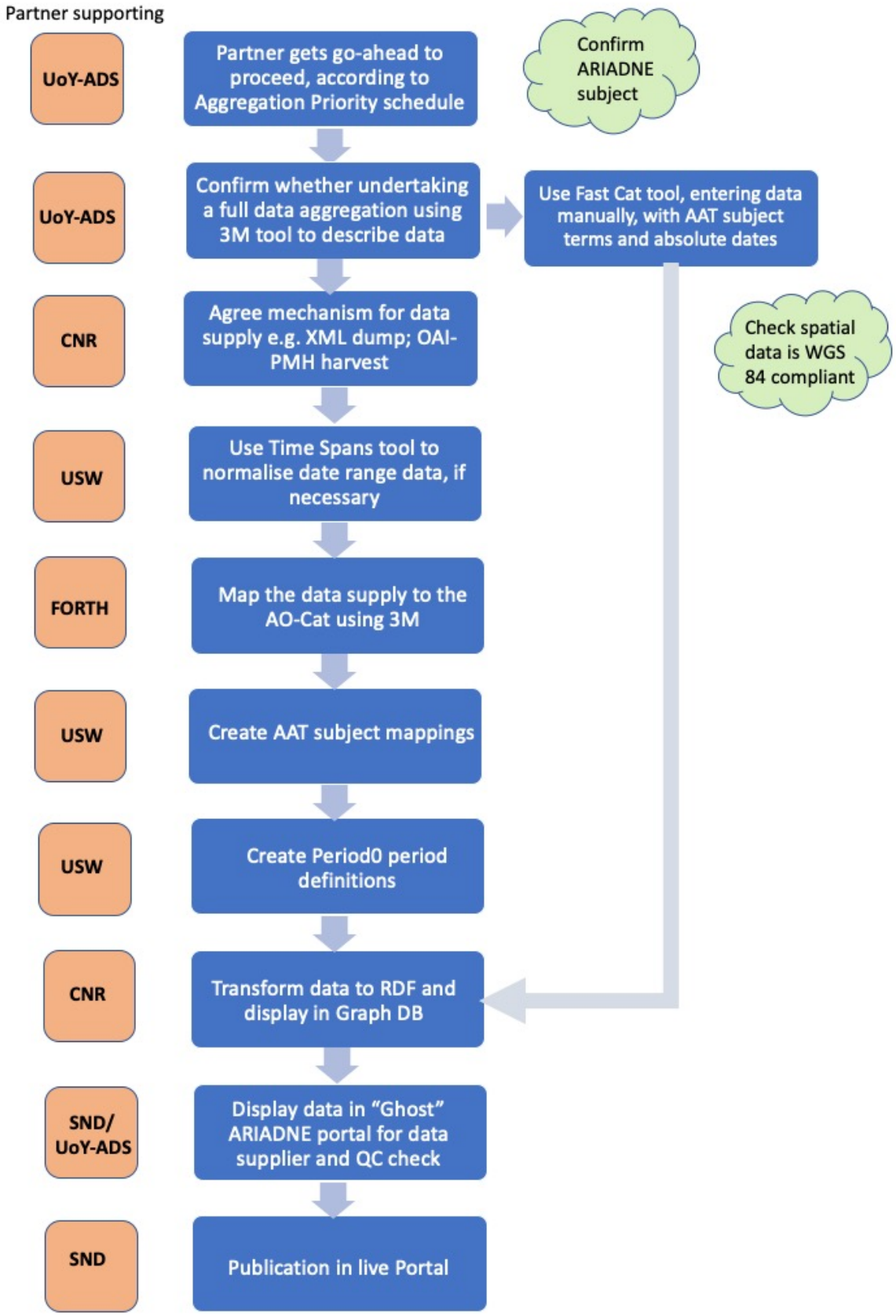
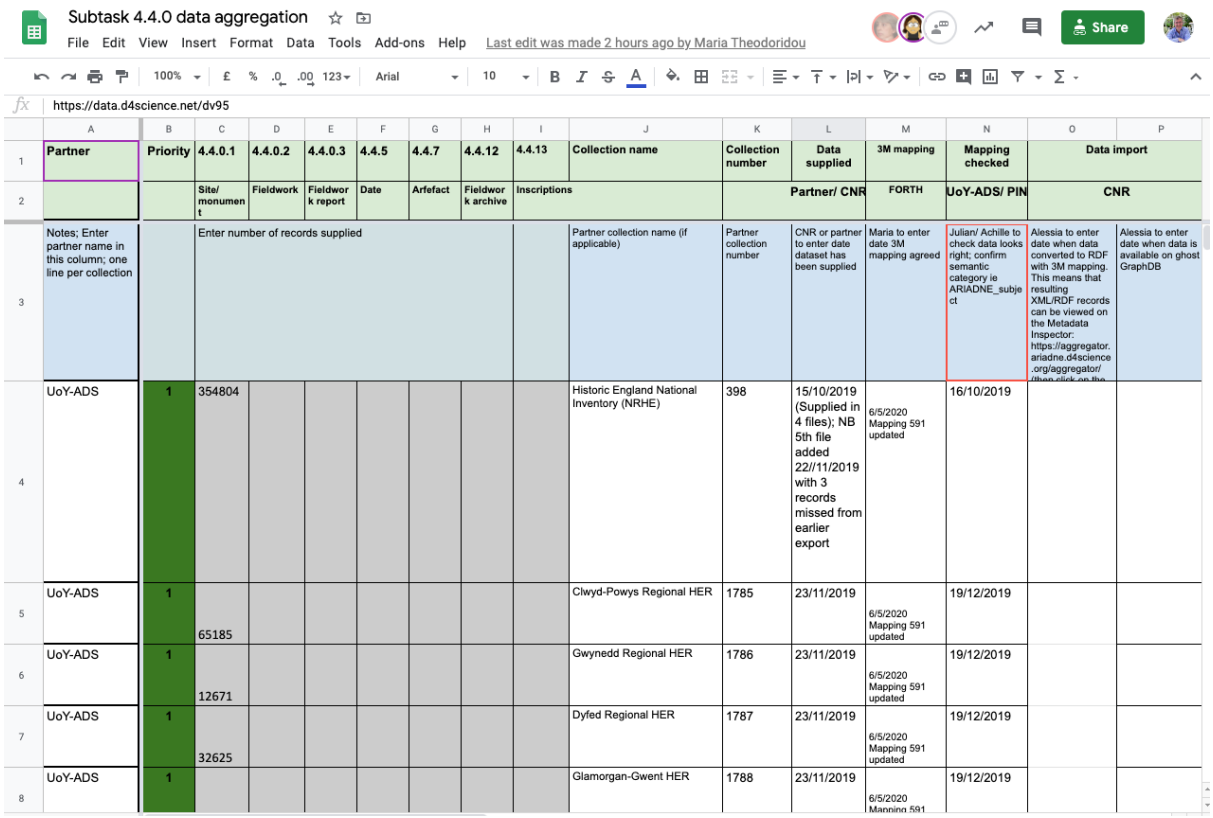


Figure 3: Simplified flowchart showing the update procedure.

Partners have also recorded their preferences for frequency of update and a target during the 2nd phase of ARIADNEplus will be to implement an automated process for regular updating, using the existing mappings.

During the first phase of ARIADNEplus the partners in the aggregation task force used a shared Googlesheet dashboard to define the specific stages and to enter dates as stages were completed (Figure 4).



Subtask 4.4.0 data aggregation

File Edit View Insert Format Data Tools Add-ons Help Last edit was made 2 hours ago by Maria Theodoridou

Share

https://data.d4science.net/dv95

	Partner	Priority	4.4.0.1	4.4.0.2	4.4.0.3	4.4.5	4.4.7	4.4.12	4.4.13	Collection name	Collection number	Data supplied	3M mapping	Mapping checked	Data import	
2			Site/ monument	Fieldwork	Fieldwork report	Date	Arfefact	Fieldwork archive	Inscriptions			Partner/ CNR	FORTH	UoY-ADS/ PIN	CNR	
3	Notes; Enter partner name in this column; one line per collection		Enter number of records supplied							Partner collection name (if applicable)	Partner collection number	CNR or partner to enter date dataset has been supplied	Maria to enter date 3M mapping agreed	Julian/ Achille to check data looks right; confirm semantic category ie ARIADNE_subject	Alessia to enter date when data converted to RDF with 3M mapping. This means that resulting XML/RDF records can be viewed on the Metadata Inspector: https://aggregator.ariadne.d4science.org/aggregator/	Alessia to enter date when data is available on ghost GraphDB
4	UoY-ADS	1	354804							Historic England National Inventory (NRHE)	398	15/10/2019 (Supplied in 4 files); NB 5th file added 22//11/2019 with 3 records missed from earlier export	6/5/2020 Mapping 591 updated	16/10/2019		
5	UoY-ADS	1	65185							Clwyd-Powys Regional HER	1785	23/11/2019	6/5/2020 Mapping 591 updated	19/12/2019		
6	UoY-ADS	1	12671							Gwynedd Regional HER	1786	23/11/2019	6/5/2020 Mapping 591 updated	19/12/2019		
7	UoY-ADS	1	32625							Dyfed Regional HER	1787	23/11/2019	6/5/2020 Mapping 591 updated	19/12/2019		
8	UoY-ADS	1								Glamorgan-Gwent HER	1788	23/11/2019	6/5/2020 Mapping 591	19/12/2019		

Figure 4: Screenshot of a subset of the Aggregation dashboard. Collections are listed one per row, and the columns represent the separate stages in the pipeline. As stages are completed the partner responsible notes the dates and any observations in the appropriate cell.

However, the Googlesheet dashboard was a short-term solution and is now being replaced with an additional software tool, **Activity Dash**, developed by FORTH as part of T14.1, to be integrated within D4Science so that all partners can access those tasks for which they have permission, using their existing authentication credentials (Figure 5). The Activity Dash tool aims to track several processes (activities) of a workflow, which might or might not be executed in a certain order. It also:

- Provides workflow and activity management
- Offers a collaborative environment
- Provides reliable notification means for stakeholders to be updated of any changes on workflow activities they are interested in
- Is a reference point for stakeholders.

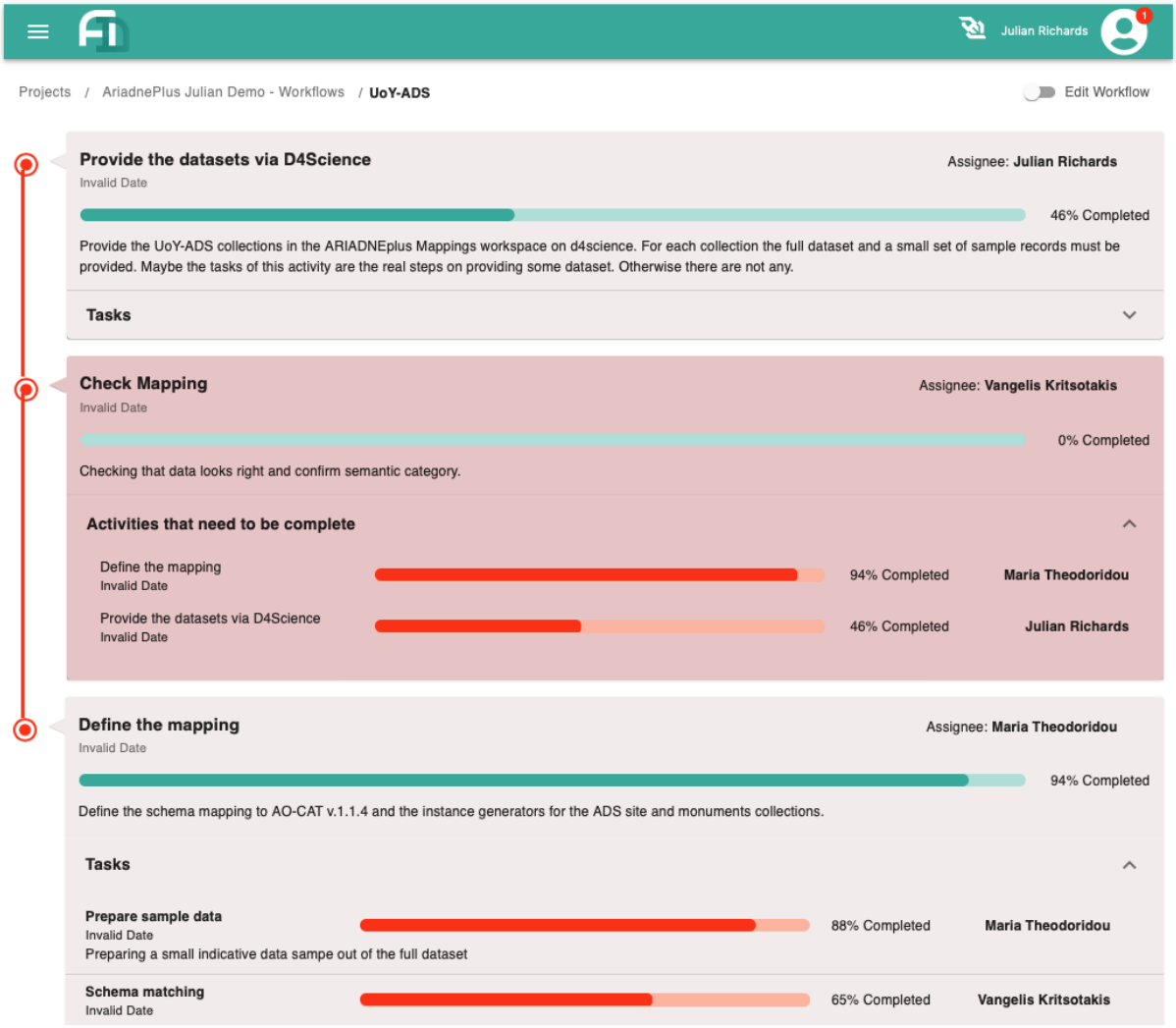


Figure 5: Screenshot of the pre-release version of Activity Dash, May 2020. The screen shows the stages in the workflow for a single UoY-ADS collection, and some stages have been opened up to show individual tasks. The percentage completed for each task is shown in an easily visualised format.

4 Data cleansing. Task 5.2 (NA4.2) Task leader: USW

This task addresses the issue of cleansing the data provided by partners, including mapping of subject terms to the Getty Art and Architecture Thesaurus (AAT). It also covers the use of Perio.do to provide absolute dates for named archaeological periods, as defined by partners, and the Timespan tool to clean period information. The activity is carried out automatically whenever possible, not excluding manual intervention and revision by content providers.

4.1 Subject mapping

The Getty AAT provides identifiers and terms to describe cultural heritage concepts. Getty vocabularies including the AAT are made available as Linked Open Data (LOD), so each concept in the thesaurus has a unique identifier in the form of a URI (e.g. http://vocab.getty.edu/aat/300211545 is the identifier of the concept "penannular brooches"). Continuing with the strategy adopted in the ARIADNE project, ARIADNEplus is using AAT as a central hub for scalable interconnection of local subject vocabularies. Data providers are required to provide a set of mappings from their own local 'native' subject vocabulary (all terms used to describe subjects in their own metadata) to Getty AAT concepts. The aim is to identify subject mappings from source terms/concepts to concepts that are likely to be useful to assist subsequent browsing and searching of the data, improving recall and precision in search, and to enhance the potential for multilingual interoperability and cross search of the aggregated data records.

Vocabulary Matching Tool

During the first ARIADNE project a Vocabulary Matching Tool (VMT) was created by USW to assist data providers with creating consistent mappings between their native subject vocabulary concepts and those of the Getty AAT. This tool has been rewritten and improved in the early stages of ARIADNEplus, the new version retaining compatibility with the output format of the first version thus ensuring we could reuse any of the existing mappings already previously created. The Vocabulary Matching Tool (Figure 6) is a browser-based application with no user installation or configuration requirements, and has been deployed within the ARIADNEplus D4Science web infrastructure. It has a multilingual user interface – at the time of writing the current UI languages supported are German, English, Spanish, French, Italian and Dutch. Users can search and browse the AAT structure, viewing the hierarchical context and scope of each concept to make a far better-informed match than relying on terms alone.

The type of match between the local source concept and the target AAT concept will be one of the possible SKOS mapping properties as defined by the SKOS reference document. There is an associated online help page with further instructions and general usage guidance to assist users. The tool and its operation are also described in the Data Aggregation Pipeline User Guide document (Bardi et al. 2020).

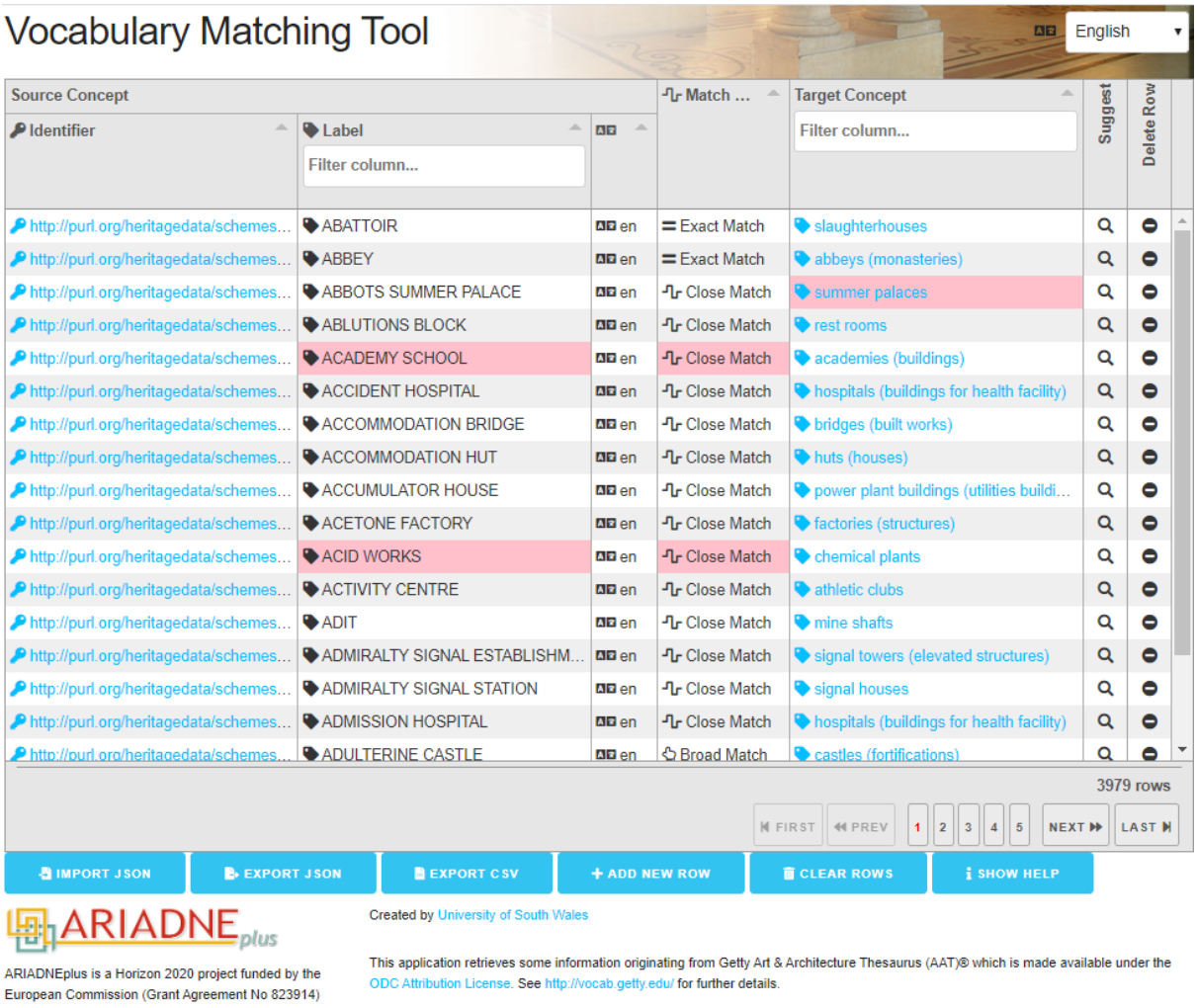


Figure 6: The Vocabulary Matching Tool, illustrating the ability to define Exact, Close, or Board matches between the Getty AAT and target concepts in native vocabularies deployed by partners.

AAT mappings created to date

Mappings are inherently reusable as they describe semantic relationships between persistent concepts relating to the domain but are not restricted in scope to a particular dataset. Therefore in the first instance all 6,400+ existing vocabulary mappings from the preceding ARIADNE I project can be reused. These are then being revised and supplemented as appropriate to account for changes and additions that may have occurred to the source data in the interim period, and new mappings are also being produced by the new ARIADNEplus data providers.

Provider	Description	Mappings
AIAC	(from ARIADNE I) AIAC report that they use Getty AAT terms in their data	129
ARUP	Bilingual – includes mappings for Czech and English terms	2,492
ASU	-	-
AU	-	-
BUP	Mapping facilities are to be built into their new website	-
CARARE	-	-
CENIEH	-	-
CNRS-MASA	-	-
DANS-KNAW	DANS-DCCD (from ARIADNE I) DANS-EASY (from ARIADNE I)	336 114
DGPC	-	-
FI	-	-
HNM	MNM-NOK site types (new, revised)	116
IAA	-	-
IAVP	-	-
INP	-	-
INRAP	PACTOLS (from ARIADNE I)	1,814
KHM-UO	-	-

Provider	Description	Mappings
LNEC	LNEC are creating a bilingual (English/Portuguese) subject vocabulary, which will then have mappings to AAT concepts	-
MIBAC-ICCU	PICO (from ARIADNE I) ICCU (from ARIADNE I)	144 1,304
NARA	-	-
NIAM-BAS	(from ARIADNE I)	238
OEAW	DFMROE (from ARIADNE I) FH (from ARIADNE I) UK-POOL (from ARIADNE I) UK-THUNAU (from ARIADNE I)	7 10 16 4
PP	PP have prepared spreadsheet lists of site types, species, and materials, with mappings to AAT concept IDs.	-
ROCEEH	Interpretation Material Plant remains Category Human remains	13 30 2 18 16
SND	Revised and extended version of previous mappings	431
SNHB	-	-
UH	-	-
UoY-ADS	Building materials (from ARIADNE I) Components (from ARIADNE I) Object types (from ARIADNE I) Maritime craft (from ARIADNE I) Monument types (revised)	12 9 411 24 3,979
ZRC-SAZU	ARKAS (from ARIADNE I) ZBIVA (from ARIADNE I)	93 30
	Total:	**11,792**

4.2 Period mapping

Perio.do is a multilingual gazetteer of scholarly definitions for describing historical and archaeological named periods. It comprises a series of 'collections' (lists of named periods) where each individual period definition is associated with a particular spatial coverage and a start date and end date, then other metadata including a bibliographic reference for the source of the term. Periods and collections have 'permalink' identifiers (URIs) so they can be clearly and unambiguously referenced - e.g. the identifier http://n2t.net/ark:/99152/p0rrjd9gjx9 defines the period "Moyen Âge" (Middle Ages) within the collection http://n2t.net/ark:/99152/p0rrjd9 "INRAP: Chronologie Generale. 2007".

Perio.do can accommodate multiple perspectives regarding the dates for a named period for the same spatial extent as shown in the following table, this is why it is important for ARIADNEplus to make explicit the meaning of named periods where they exist in data records.

Identifier	Label(s)	Source	Spatial Coverage	Start	End
http://n2t.net/ark:/99152/p0gjgrs6qb2	Bronze Age	Portable Antiquities Scheme	UK	2350 BC	801 BC
http://n2t.net/ark:/99152/p0kh9ds7q8m	Bronze Age	Historic England Periods List	UK	2600 BC	700 BC
http://n2t.net/ark:/99152/p0zj6g8rbvk	Bronze Age	ARENA Portal	UK	2500 BC	700 BC

Named periods will therefore be defined with reference to periods existing within Perio.do collections. In the first instance the existing published period definitions used in the first ARIADNE project remain available on Perio.do and can be reused. Period definitions produced by data partners to date are shown in the following table.

Provider (contact)	Description	Periods
AIAC	FASTI - Home. 2004 http://n2t.net/ark:/99152/p06v8w4 AIAC report that they use absolute dates in their underlying data	212
ARUP	ARUP period definitions have been submitted to Perio.do; awaiting publication	-
ASU	-	-
AU	-	-
BUP	-	-
CARARE	-	-
CENIEH	-	-
CNRS-MASA	-	-
DANS-KNAW	Rijksdienst voor het Cultureel Erfgoed. Het Archeologisch Basisregister (ABR). 1992. http://n2t.net/ark:/99152/p0pqptc	53
DGPC	-	-
FI	-	-
HNM	No separate Perio.do collection; Hungarian periods are included in the aggregated collection “ARIADNE Data Collection. 2015.” (With spatial coverage Hungary) http://n2t.net/ark:/99152/p0qhb66	32
IAA	-	-
IAVP	Romanian Dobruja Periodisation http://n2t.net/ark:/99152/p02kbfn	24
INP	Work in progress extracting terms from the 'Treaty of Romanian History'	-

Provider (contact)	Description	Periods
INRAP	INRAP: Chronologie Generale. 2007. http://n2t.net/ark:/99152/p0rrjd9	32
KHM-UO	Norsk arkeologisk leksikon. 2005. http://n2t.net/ark:/99152/p04h98q	28
LNEC	-	-
MIBAC-ICCU	-	-
NARA	-	-
NIAM-BAS	No separate Perio.do collection; Bulgarian periods are included in the aggregated collection “ARIADNE Data Collection. 2015.” (With spatial coverage Bulgaria) http://n2t.net/ark:/99152/p0qhb66	32
OEAW	No separate Perio.do collection; Austrian periods are included in the aggregated collection “ARIADNE Data Collection. 2015.” (With spatial coverage Austria) http://n2t.net/ark:/99152/p0qhb66	31
PP	PP have prepared a draft spreadsheet list of cultural periods with Perio.do references.	-
ROCEEH	Period definitions spreadsheet compiled, submitted to Perio.do team for either manual or bespoke import	-
SND	-	-
SNHB	-	-
UH	-	-
UoY-ADS	Period definitions originating from Historic England Periods Authority File http://n2t.net/ark:/99152/p0kh9ds	42

Provider (contact)	Description	Periods
ZRC-SAZU	No separate Perio.do collection; Slovenian periods are included in the aggregated collection "ARIADNE Data Collection. 2015." (With spatial coverage Slovenia) http://n2t.net/ark:/99152/p0qhb66	31
	Total:	**517**

Time Spans Tool

The purpose of using Perio.do in ARIADNEplus is different to the use of AAT mappings - rather than aligning data to common identifiers, Perio.do will be utilized in a subsequent data enrichment stage to enrich records already indexed with named periods (such as "Bronze Age", "Iron Age" etc.) where the dates associated with these periods are not already made explicit in the input data - determining the intended meaning of named periods in terms of absolute start/end years so that the records can be compared by date and aggregated according to a common timeline. However, in addition to named periods there are occasions where records refer to periods in other forms:

- A span of years (e.g. "*1000 BC to 1785 AD*").
- Year with tolerance (e.g. "*1666*", "*1485+5-10*", "*1540±9")*
- A decade (e.g. "*the 1920s*")
- Ordinal named or numbered centuries (e.g. "*circa C15*", "*Early fifteenth century*")
- A span of centuries (e.g. "*Late 15th-Mid 17th century*")

Terms of this kind do not exist in Perio.do, and where the start/end years are not easily derived in a consistent way this would prevent these records being compared by date. The terms sometimes have additional prefixes and suffixes (e.g. "circa", "early", "AD", "BCE"). A further complication for the parsing of these strings is multilingual variations in the way they may be expressed.

USW have created a bulk data processing tool to resolve this situation - deriving suitable start/end years from such textual strings. Due to the ARIADNEplus workflow and restriction not to make any changes to the existing records provided by the data provider, it was necessary to make a modified version of the tool functionality to import the 'raw' XML data records from the data provider and to supplement any year span values present with new derived start/end year elements, prior to the records being ingested. The example below shows a brief explanatory extract of the output of the tool – a record having an existing "dcterms:temporal" element describing a particular year span has been supplemented with new "minYear" and "maxYear" sibling elements expressing the start/end years as separate xsd:gYear values.

```
<record type="record">

 <dc:title>ST. MARTIN'S CHURCH</dc:title>
 <dc:creator>Exeter City Council</dc:creator>
 <dc:subjectPeriod>
  <dc:subject>CHURCH</dc:subject>
  <dcterms:temporal>850 - 1068</dcterms:temporal>
  <minYear xsi:type="http://www.w3.org/2001/XMLSchema#gYear">0850</minYear>
  <maxYear xsi:type="http://www.w3.org/2001/XMLSchema#gYear">1068</maxYear>
 </dc:subjectPeriod>
 . . .
</record>
```

Following initial testing and adjustment the tool was applied to 'real' input data using 5 XML files originating from UoY-ADS. The files typically contained a mixture of named periods and year spans. In most cases named periods were not processed as they would be present in the associated Perio.do collection anyway, so the dates should subsequently be populated through that route. As this was legacy data that did not necessarily conform to the period controlled vocabulary currently in use in the source organization, some special cases were made to accommodate named period terms which although they did not occur in the associated Perio.do collection, occurred frequently enough to warrant handling – e.g. the term *INTER WAR* was considered to mean the period from 1919 to 1938.

The tool takes only a few seconds to process each data file in turn. The results of the bulk data processing work undertaken are shown in the following table:

Input file name	**Temporal records**	**Processed**	**Not handled**
ariadne_324.xml	20,113	18,982	1,131
ariadne_328.xml	59,278	59,278	0
ariadne_1054.xml	971	971	0
ariadne_1970.xml	20,310	20,310	0
ariadne_1972.xml	16,689	16,689	0
Totals	**117,361**	**116,230**	**1,131**

Investigating the "not handled" records (in ADS collection 324) – the majority were for the named period *COLD WAR* – which is actually present in the associated Perio.do collection in any case, so the appropriate dates will subsequently be populated through that route. The remainder (50 records) were using the term *PRE-WORLD WAR I,* which was not considered specific enough to impose an absolute date span.

5 Feeding content into the ARIADNEplus infrastructure. Tasks 5.3 and 5.4 (NA4.3 and NA4.4) Task leader: CNR-ISTI

This section describes the activities of the tasks 5.3 (NA4.3 - Updating the current content of the ARIADNEplus Data Infrastructure (ADI)) and 5.4 (NA4.4 - Ingesting data into the ARIADNEplus Data Infrastructure) led by CNR-ISTI.

The main objectives of the tasks are:

1. Definition of the aggregation workflow
2. Execution of aggregation activities

5.1 Definition of the aggregation workflow

The current ARIADNE infrastructure aggregated information about archaeological datasets, described according to the ARIADNE Catalogue Data Model (ACDM). With the adoption of the ARIADNEplus Data Model (AO-Cat) developed in WP4 of the ARIADNEplus project, the data aggregated during the ARIADNE project and already publicly available via the ARIADNE portal must be re-processed to reach the compliance to AO-Cat.

Following the decisions taken in task 5.1 – Definition of the update procedure of the ARIADNEplus Data Infrastructure (NA4.1), task 5.3 has been planned to collect from scratch all the records from the ARIADNEplus partners and transform them according to the AO-Cat model by applying new 3M mappings (defined in the context of task 4.2 – Mapping dataset metadata to the ARIADNEplus Data Model (NA3.2)).

The complete re-aggregation of content with the new aggregation infrastructure based on the D-Net framework toolkit (developed and operated in WP12) smoothly addresses the two main issues identified in the Description of the Action:

1. Filling the gaps (sub-task 5.3.1): due to the gap of about two years between ARIADNE and ARIADNEplus, the content available in the ARIADNE infrastructure does not include new and updated records that are available from the data providers.
2. Regular updates (sub-task 5.3.2): the D-Net framework toolkit natively supports the incremental aggregation of records available via online endpoints (provided incremental collection is supported by the endpoints) and the aggregation workflows can be scheduled to run at given intervals in time. Scheduling is configurable with cron expressions[4] that can be different for each partner (e.g. DANS data can be aggregated each Monday morning; UoY-ADS data on the 1st of each month). The aggregation schedule will therefore be agreed with each partner, based on the update frequency of the content on their endpoints.

[4] Cron expression: https://en.wikipedia.org/wiki/Cron#CRON_expression

Thanks to this decision, there was no need to define separate procedures for the aggregation of "old" and "new" content. Therefore, the main objective of both tasks 5.3 and 5.4 in the first project period was to define the aggregation pipeline (or workflow), i.e. the sequence of steps to be performed for the aggregation of metadata records of the ARIADNEplus partners.

For the definition of the aggregation workflow, an iterative approach has been adopted, where refinements to the different steps have been applied based on requirements, suggestions and feedback from WP4 and WP12. The description of the workflow has been defined in the document "Data Aggregation Pipeline: User Guide" (Bardi et al 2020). As of June 2020, the aggregation workflow comprises the flows depicted in Figure 7.

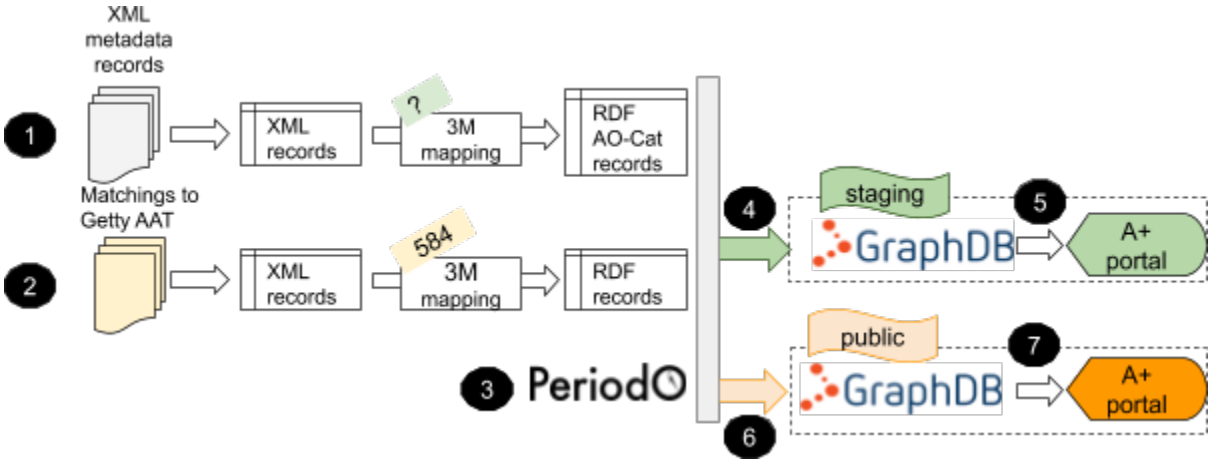


Figure 7: The aggregation workflow

The flows identified as 1, 2, 3, 4 and 6 in figure 7 are intended to feed the knowledge graph (implemented with an instance of GraphDB[5]):

1. *Ingestion of XML records of the provider*. Records are collected, transformed into AO-Cat by applying a dedicated 3M mapping, and fed into the knowledge graph.
2. *Enrichment with Getty AAT subjects*[6]. The subject mapping to Getty AAT (see Section 4.1) is transformed into RDF by applying the 3M mapping 584 and fed into the knowledge graph. As a result, the knowledge graph will contain the correspondences between native subjects and terms of the Getty AAT vocabulary.
3. *Enrichment with PeriodO*. The PeriodO datasets defined by the provider (see Section 4.2) is ingested into the knowledge graph and used to generate explicit aocat:has_period properties.
4. *Feed the staging knowledge graph.* In order to support the providers at checking the content before it is made publicly available, the push on the knowledge graph initially targets a "staging" instance.
5. *Feed the staging A+ portal*. For each set of records ingested with flow 1, a procedure reads the corresponding triples and builds json records to feed the index server that serves the A+

[5] GraphDB: http://graphdb.ontotext.com/
[6] Getty AAT: https://www.getty.edu/research/tools/vocabularies/aat/index.html

portal. The procedure will include in the json records the properties that have been added by the other flows (2-3) and others that can be inferred exploiting the reasoning capability of GraphDB. The goal of this flow is initially to generate records that can be used by the current ARIADNE portal. In a second stage, the procedure will be updated to include properties that were not available in ACDM but that have been added in the AO-Cat.

6. *Feed the public knowledge graph*. This flow is executed if the provider successfully completed the content checking on the staging knowledge graph and portal
7. *Feed the public A+ portal*. The flow applies the same procedure of 5, using the public instances of the knowledge graph and portal instead of the staging instances.

Steps 4 and 5 supports the quality checks of content and will be bypassed once the input format of the records and the 3M mappings of a provider are stable, so that it will be possible to automatically update the knowledge graph and the ARIADNEplus portal with updated and new records without human intervention. If necessary, steps 4 and 5 might be reactivated for a given source (e.g. because the provider upgraded the information system and wants to perform extensive checks before the data supplied by the new system goes public).

An overview of the implementation of the workflow in the aggregation infrastructure is given in D12.1, while technical details will be available in D12.2 (due in M25).

5.2 Execution of aggregation activities

The aggregation activities have been planned to support the work and match the timeline of WP12 (for the development and deployment of the services of the aggregative infrastructure) and WP4 (for the preparation of 3M mappings).

Phase 1: testing the aggregative infrastructure with UoY-ADS records
The content update on the new A+ infrastructure has been tested with records provided by the ADS. This phase informed WP12 about requirements and needed to be fulfilled with the integration of additional tools and services for data processing and curation that were not foreseen in the DoA. This phase was completed in the reporting period.

Phase 2: aggregation of records of ARIADNE partners according to the pipeline defined in T5.1
This phase started at the end of the reporting period with the partners ADS, AIAC, and HNM. For the complete pipeline, please refer to Section 3 and D4.1.

6 Extending the ARIADNEplus Data Infrastructure scope. Task 5.5 (NA4.5) Task leader: CNR-ISTI

Since the ADM enables cataloguing people, institutions and services together with datasets, task 5.5 is responsible for identifying how this is to be carried out and supports the implementation of the Data Infrastructure extension.

In the first 18 months of the project, the Consortium members in charge of Task T5.5 have focussed on endowing the ARIADNE Catalogue ontology (AO-Cat, as it is named within the Consortium) with the vocabulary needed to model people, institutions and services at the Catalogue level. The vocabulary includes classes, properties and the axioms providing the meaning of these in relation with the other classes and properties of the AO-Cat. These classes and properties will be briefly and informally illustrated in the rest of this Section. For a full account including the axioms, the reader is referred to Deliverable D4.1 (Initial report on dataset integration), which gives a full account of the AO-Cat.

6.1 Modelling people and institutions in AO-Cat

Agents play important roles in the ARIADNE information space: they are responsible for making resources available and publicly accessible; they hold various types of responsibilities for those resources; finally, they carry out activities. For these reasons, AO-Cat defines the class AO_Agent to model entities that can act. AO_Agent is a subclass of the CRM class E39 Actor, defined as “people, either individually or in groups, who have the potential to perform intentional actions of kinds for which someone may be held responsible”.

Moreover, from an archaeological perspective, it is important to distinguish between two kinds of agents: the person and the institutions, which in AO-Cat are generally called “organizations”. This distinction is captured in AO-Cat by two subclasses of Agent:

- Person, modelling individual humans, for which AO-Cat provides class AO_Person;
- Group, modelling “gatherings or organizations of agents that act collectively or in a similar way due to any form of unifying relationship” [CRM Spec 6.2], for which AO-Cat provides class AO_Group.

The following knowledge is represented for instances of AO_Agent, regardless whether the agent is a person or a group:

- the name of the agent (property has_name)
- any identifier of an agent outside of the ARIADNE namespace (property has_agent_identifier)
- the email address of the agent (property has_email)
- a web page for the agent (property has_homepage)
- the institution(s) of the agent (property has_insitution). If the agent is itself an institution, this property may be used to record the mother institution, if any, or be left unspecified.

6.2 Modelling services in the AO-Cat

AO-Cat follows the view of a service provided by the PARTHENOS Entity Model (PEM), an extension of the CIDOC CRM focussed on infrastructural entities [PEM Specifications 3.1].[7] According to this view, a service is "an offer by some actor of their willingness and ability to execute an activity or series of activities upon request". Applyinh this definition in the ARIADNE ontology, a service is modelled as a resource in the archaeological domain, and it is therefore represented as an instance of class AO_Service, a subclass of AO_Resource. The complete AO-Cat class taxonomy regarding resources is provided in Figure 8.

For the purposes of the Catalogue, the following information is represented about a service:

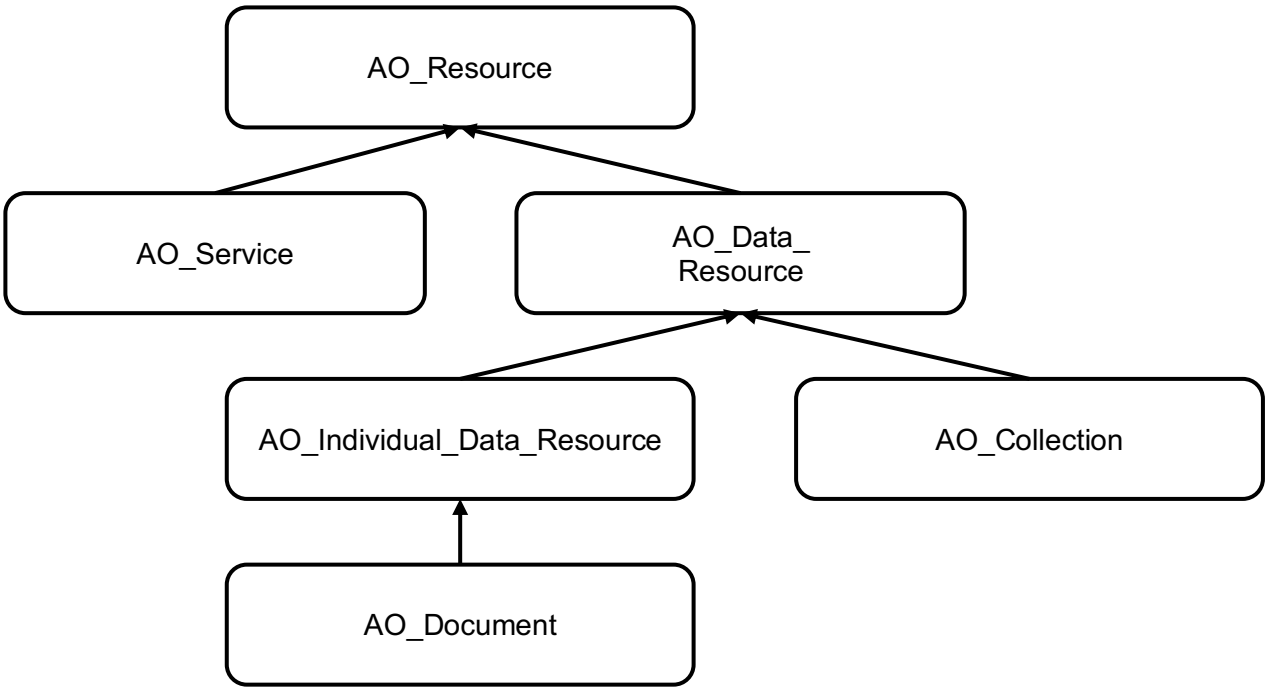


Figure 8. A fragment of the AO-Cat class taxonomy including AO_Service

- an IRI giving the resource making the service accessible (property is_accessible_at). If the service is a web service, this IRI is the technical access point of the service. Otherwise, it is the IRI of a resource describing how the service can be accessed.
- a concept giving the functionality of the service (property has_functionality). We have not yet started collecting this knowledge from the partners, so at the moment there is no fixed vocabulary for service functionality.
- concepts giving the produced and consumed media types and format of the services (properties has_consumed_media, has_produced_media, has_consumed_format and has_produced_format). A natural candidate vocabulary for these properties is the IANA MIME type hierarchy.[8]

[7] https://parthenos.acdh.oeaw.ac.at/resource/?uri=p%3ADocsModel
[8] https://www.iana.org/assignments/media-types/media-types.xhtml

- concepts giving the supported natural languages of the service (property has_supported_language). A natural candidate for this property is the language vocabulary established by Lexvo.org.
- an agent providing technical support for the service (property has_technical_support).

As the project tackles item-level integration, these classes and properties may be extended to capture relevant knowledge about services from the providers.

7 Collaborating with other Catalogues. Task 5.7 (NA4.7) Task leader: PIN

This task is responsible for collaborating and addressing cross-discipline interoperability issues with catalogues in other domains, as well as for ensuring compatibility with the future EOSC catalogue, as well as OpenAIRE.

We have also developed a collaboration with the Pelagios network which includes North American partners and covers the classical world (see D2.2).[9] Pelagios is developing tools such as Recogito[10], an annotation tool that allows the creation of relationships between textual fragments or areas of images, and geographical entities in order to define geographic maps of textual and semantic content. The collaboration with Pelagios includes the definition of a series of case studies spanning over various disciplines, in which archaeology, cultural heritage and digital heritage interact and become interoperable both at the level of data and as regards tools and services. This can be implemented through the interaction of the tools developed by both ARIADNEplus and Pelagios in a network of functionalities capable of querying distributed information (formatted according to the Linked Open Data paradigm) and to create virtual platforms capable of providing answers to complex, interdisciplinary scientific questions.

Activities in EOSC have started with the development of a tool derived from ARIADNE within EOSCpilot as a so-called Science Demonstrator. The tool, renamed TEXTCROWD, was used for text-mining of archaeological documents. The new development concerned texts in Italian, while those in English and Dutch had been addressed in ARIADNE. Also, the tool was embedded in the cloud environment D4Science provided by CNR-ISTI.[11]

The EOSCpilot (www.eoscpilot.eu) duration overlapped with ARIADNEplus only for a short period. However, it was possible to demonstrate the needs and solutions of the archaeological community in the EOSC framework. The ARIADNE community was represented by PIN, supported by CNR.

TEXTCROWD used some open source linguistic libraries necessary for text analysis which were accessible online at the time it was developed. Later, however, such libraries were removed and although they are still available as source, they need reinstalling in a different framework. This has slowed down the implementation of another activity based on the same package.

TEXTCROWD was then included as one of Tasks of the current EOSC-Pillar project (www.eosc-pillar.eu) with the goal of integrating the possibility of linking data from archaeological science, stored in a database, with related texts analysed using TEXTCROWD. The previous version of TEXTCROWD can be used once the issue of the linguistic libraries is resolved, which is currently work in progress. In the

[9] https://pelagios.org
[10] https://recogito.pelagios.org
[11] https://eoscpilot.eu/science-demos/textcrowd

meantime, the scientific database has been implemented in a cloud environment and is currently being populated with data. The combination of both aspects is expected to be set up by Autumn 2020. The result will also be useful for ARIADNEplus, as it can be straightforwardly transferred in the ARIADNEplus framework as one of the services provided by the project. It will be compared with the re-engineered version of the text-mining tool, currently under development in ARIADNEplus. Further developments include plans to "export" the same approach to France and French; for this, the reference EOSC-Pillar partner is CNRS, also an ARIADNEplus partner.

The ARIADNEplus partner in charge of the EOSC-Pillar TEXTCROWD Task is INFN. In general, both CNR and INFN are strongly involved in the global development of the EOSC and are able to keep ARIADNEplus partners informed about its progress. In addition, UoY-ADS, DANS-KNAW and CNR are also partners in SSHOC, where UoY-ADS represents the Heritage Science community and is responsible for a deliverable addressing the challenges of FAIR data in the heritage science and archaeological communities. This also ensures useful cross-fertilisation between ARIADNEplus and EOSC.

Our work on digital libraries (described in D4.1 Section 7) is in collaboration with OpenAIRE, which is one of the services now integrated with EOSC.[12] OpenAIRE's mission is closely linked to the mission of the European Commission: to provide unlimited, barrier free, open access to research outputs financed by public funding in Europe. OpenAIRE fulfils the EOSC vision substantially, as its operations already provide the glue for many of the user and research driven functionalities, including domain disciplined research communities or Research Infrastructures.

ARIADNEplus is also working to make its platform interoperable with the EOSC cloud by means of the D4Science platform, used to implement the Virtual Research Environments in which the ARIADNEplus cloud services operate. D4Science is already largely compatible and interoperable with EOSC at cloud level, and already shares some services and features with it, including authentication and management of federated identities, interconnection of distributed infrastructures, orchestration of distributed services. The interaction between ARIADNEplus and EOSC will be further enhanced in the next period also at data level in order to make ARIADNEplus information and related services available and accessible in different ways from the EOSC side.

[12]https://marketplace.eosc-portal.eu/services/digital-humanities-and-cultural-heritage-openaire-community-gateway/information

8 Conclusions

In conclusion, we have made significant progress in improving the procedures for updating and extending the ARIADNEplus infrastructure from those, which were used during the first ARIADNE project. The development and testing of these tools has occupied much of our first year, in order to ensure they are fit for purpose, but we now have a clear pipeline, and are making excellent progress in aggregating partner datasets. This will be prioritised during the remainder of 2020, and we will begin to extend the infrastructure to accommodate new ARIADNE subject types, demonstrating the powerful benefits for researchers in being able to cross-search multiple datasets, of different data types, from different data providers. During the second phase of the project we will continue to extend the number and range of data sets accessible via the ARIADNE portal, ensuring full coverage of all ARIADNEplus partners, as well as the growing number of Associate partners who wish to use our services to make their data available beyond their national borders. We will extend the experiments undertaken on item level integration in ARIADNE, and use the power of the CIDOC CRM (and the fact that the AO-Cat is fully compliant with it) to demonstrate interoperability and connections between dataset at a variety of scales, and with complex relations with other data types. We will also extend current activity in order to catalogue services, and we will work to ensure that ARIADNE infrastructure is able to take its place within EOSC and the wider network of infrastructure providers.

9 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** 2017, https://doi.org/10.11141/ia.43.11

Bardi, A., Binding, C., Felicetti, A., Meghini, C., Richards, J. & Theodoridou, M. Unpublished report 2020. *Data Aggregation Pipeline: User Guide. Version 1.7. 24/4/2020*. See Appendix 1.

Niccolucci, F. and Richards, J.D. 2019 "ARIADNE and ARIADNEplus" in Richards and Niccolucci (eds.) *The ARIADNE Impact.* Archaeolingua, 7-25

Appendix A

Data Aggregation Pipeline: User Guide

Alessi Bardi, Ceri Binding, Achille Felicetti, Carlo Meghini, Julian Richards, Maria Theodoridou

Version: 1.8

Revision history

Date	Version	Author(s)	Description
25/10/2019	1.1	J. Richards, C. Meghini, A. Bardi, A. Felicetti, M. Theodoridou, C. Binding	First draft
20/12/2019	1.2	J. Richards, C. Meghini, A. Bardi, A. Felicetti, M. Theodoridou, C. Binding	Pre-publication revisions
7/1/2020	1.3	J. Richards	Final copy edit
17/1/2020	1.4	J. Richards & C. Meghini	Addition of Appendix 1
2/2/2020	1.5	C. Meghini	Aligned with modified AO-Cat
9/3/2020	1.6	A.Bardi	Added description of the flow to include date normalization and Timespans
27/04/2020	1.7	J.Richards	Aligned names for ARIADNE_subjects with the agreed terms
18/05/2020	1.8	A.Bardi	Section 4 updated: mapping to PeriodO is not required. Removed option to map normnalized temporal coverages separately: TimeSpans will produce the modified XML records as output.

Table of Contents

1 Introduction

The purpose of this User Guide is to provide a short introduction to the ARIADNEplus data aggregation pipeline. It defines, for the archaeological data providers, the process by which their data should be uploaded to the ARIADNE Content Cloud (AC for short), so that it appears in the ARIADNEplus Catalogue, and can be searched via the ARIADNEplus Portal. It provides a basis for Task 5.1 – Definition of the update procedure of the ARIADNEplus Data Infrastructure.

1.1 Related documents

- The specification of the AO-Cat Ontology, the living document where the ontology of the ARIADNE Catalogue is described in detail, available on the project infrastructure at https://ariadne.d4science.org/group/ariadneplus-gateway/workspace?itemid=75a774d6-da40-43d4-96f3-184bca3c297e&operation=gotofolder
- Data resource template – the properties defined in AO-Cat for an AO_Data_Resource (See also Appendix 1) https://docs.google.com/spreadsheets/d/1Yqv1gIbaHI8rNnhF9AY2zBjcTUSaHWxEwnSI6oaM6HM/edit?usp=sharing

1.2 Structure of the document

The document is structured as follows.

- Section 2 outlines the purpose of the aggregation
- Section 3 introduces the AO-Cat
- Section 4 presents the options for how partners should supply their data to CNR-ISTI, and considers the frequency of supply
- Section 5 presents the 3M Mapping tool, which helps partners align their data to the AO-Cat (The data model used in the AC), and thereby to the CIDOC CRM
- Section 6 defines how data should also be mapped to standardised terms, to allow cross-search of the AC according to What, When and Where criteria
- Section 7 summarises the pipeline

2 Background: Why are we doing this?

This Section briefly outlines the requirements that are expected to be fulfilled by the ARIADNEplus Catalogue and more generally by the ARIADNE Content Cloud (AC for short), which the Catalogue is a part of.

There are two basic requirements:

1. Enabling cross-border/cross-institution resource discovery, i.e. *finding* data.
2. Enabling interoperability–across partners, countries, data types, data schemas, i.e. *enabling research.*

Concerning resource discovery, the ARIADNEplus Catalogue should support:

- *What* searches, searches based on a topic. For topics the Getty AAT should be used as the common conceptual backbone, in addition to local reference resources.
- *When* searches, searches based on a temporal period. For temporal periods, the PeriodO vocabulary should be used as it maps periods to absolute dates on a common time scale. In addition, local vocabularies providing region-related period appellations should be represented in the AC and allowed in searches by cultural period e.g. “Iron Age”
- *Where* searches, searches based on a spatial region. For spatial regions, the World Geodetic System 1984 (commonly abbreviated as WGS84) representation should be used.
- Enhanced map-based searching.
- Enhanced queries, specific to data types.
-

More search types that should be supported are being defined by a user needs study undertaken by the project (Deliverable 2.1)

Concerning enabling research, in addition to the above types of search, the AC is expected to store information about digital objects belonging to a large range of data types, forming a heterogeneous set, as data is diverse in character and content. These data types include, but are not restricted to:

- Databases
- Reports
- Finds
- Images
- GIS
- LiDAR data
- Datasets e.g. excavation archives
- Sub-domains, *e.g.,* scientific data; these should be modelled as CRM Application profiles, that is subsets of the CRM that we use for item level integration (cf ARIADNE coins)
- Linguistic resources, such as ontologies and vocabularies relevant to the archaeological domain.

In addition, different levels of granularity must be accounted for, supporting collection and item level interoperability, *e.g.,* sites and the individual artefacts they contain; individual dates; cemeteries and the individual graves they contain.

Moreover, links to distributed digital and paper resources about the described resources should also be maintained.

Finally, archaeological fieldwork events should be accounted for, categorized following standard vocabularies and properly connected to the relevant archaeological resources, as schematized in Figures 1 and 2, where each event category is annotated with the number of the sub-task of ARIADNEplus Task 4.4.0 addressing it:

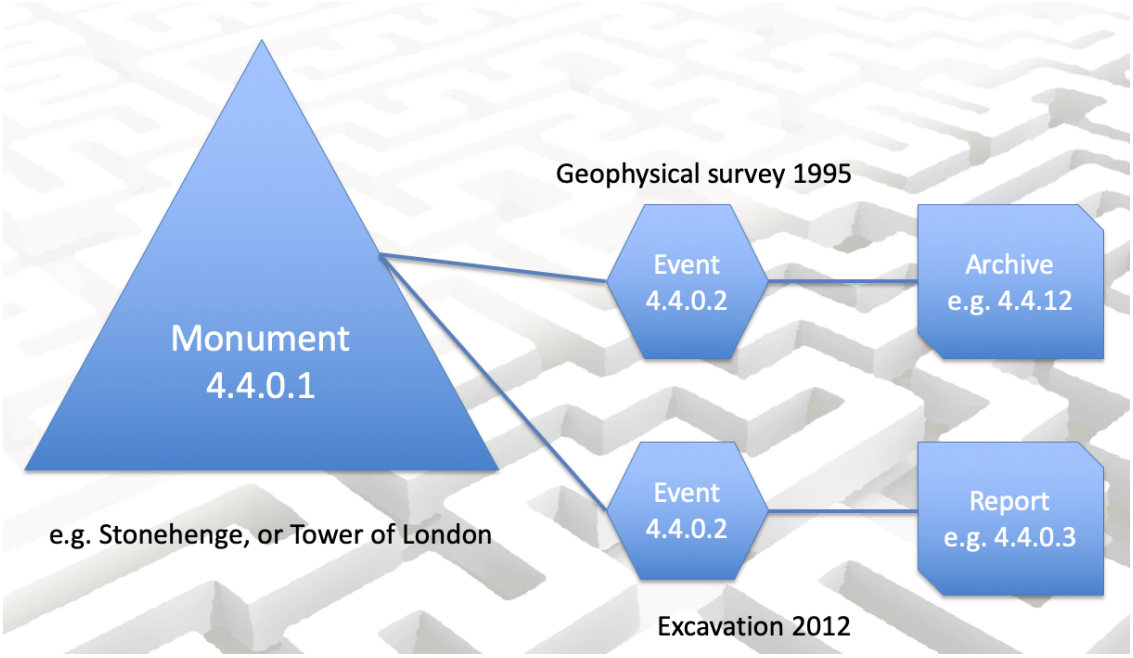


Figure 1. Entities and relationships in the archaeological domain

The diagram follows the UK MIDAS heritage standard terminology[13], whereby a 'Monument' is a physical entity where activity took place in the past, possibly over several periods e.g. Stonehenge; whereas an 'Event' normally refers to an archaeological recording event at that monument, such as an excavation or field survey. e.g. "Excavations at Stonehenge by Richard Atkinson 1950-64". In other cases e.g. the "Tiber Valley Survey", a single archaeological event (the survey) may lead to the discovery of multiple new archaeological monuments, or sites (Figure 2).

[13] http://www.heritage-standards.org.uk/midas-heritage/

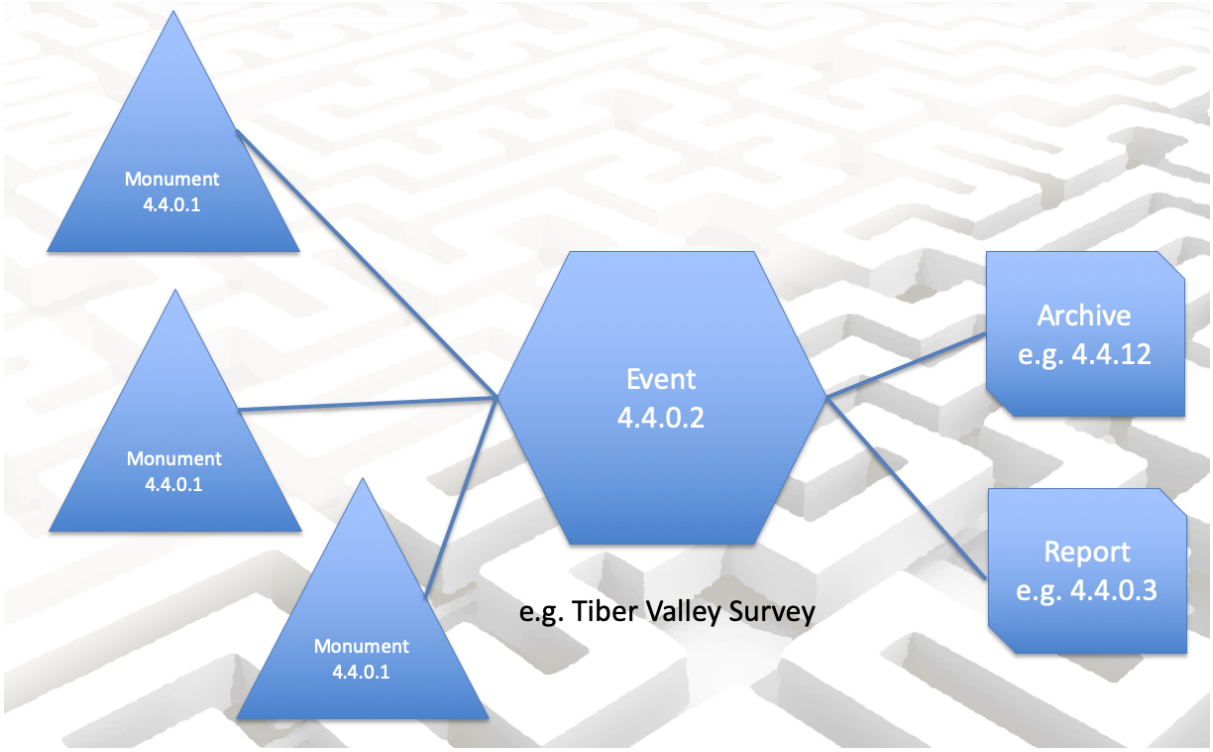


Figure 2. Entities and relationships in the archaeological domain - example of where a single recording event leads to the discovery of several new monuments

The services described above should be offered both to humans, via the ARIADNEplus Portal and supporting multilingualism, and to artificial agents, via APIs. Those APIs should provide access to the whole AC (including the Catalogue) as Linked Open Data, thereby allowing other organisations to implement their own portal or service.

3 Matching the requirements: introduction to the AO-Cat

This Section provides a simplified explanation of how the requirements presented in Section 2 have been taken into account to define the AO-Cat. For further information see the AO-Cat Ontology.

All entities are identified by an IRI (Internationalized Resource Identifier, the identifier of resources on the web, previously known as URI, Uniform Resource Identifier) in the ARIADNE namespace.

The most general class of AO-Cat is AO_Entity. As such, AO_Entity has as instances all resources that have any role in the ARIADNE infrastructure. There are a few properties that have AO_Entity as domain, and therefore apply to all resources:

- has_type, a type of the resource in any classification system.
- has_title, a title of the resource.
- has_description, a textual description of the resource in natural language.

In addition, AO_Entity is the range of property is_about, having AO_Data_Resource (see below) as domain, to express the fact that a data resource may document any entity in ARIADNE.

The AO-Cat is designed to catalogue several types of digital infrastructural resources. The most general infrastructural resource class is AO_Resource, representing all the digital resources the ARIADNE research infrastructure deals with. Any ARIADNE Resource has a type, which can be specified in two different ways.

- One way is via the property has_type, which associates an instance of AO_Entity with a type represented as an AO_Concept.
- The other way is via the sub-classes of AO_Resource, also introduced next.

In addition to the properties inherited from AO_Entity, a number of descriptive properties are defined for an ARIADNE resource:

- was_issued, the date when the record of the resource was firstly acquired.
- was_modified, the date when the record of the resource was lastly modified.
- has_publisher, the agent responsible for making the resource publicly accessible.
- has_contributor, a contributor of a description of the resource to the AC.
- has_creator, a creator of the resource.
- has_owner, the owner of the resource.
- has_responsible, any person who is scientifically responsible of the resource.

These properties have been chosen based on the experience gained in the first ARIADNE project.

AO-Cat provides two subclasses of class AO_Resource to represent the main types of digital resources: AO_Data_Resource for data resources and AO_Service for services. In this document we are simply concerned with Data Resources (AO_Data_Resource).

Data resources are also characterized in terms of the entity they are about. The connection between a data resource and the entity the resource is about is captured by property refers_to.

- refers_to associates a data resource with an entity that the resource refers to, by making assertions, whether implicitly or explicitly and regardless of the format, about that entity.

As a special case of reference, there is also the property is_about.

- is_about associates a data resource with an entity that the data resource documents, including of course events. Notice that the documenting data resource may be an external resource, identified by a IRI, or DOI, or any other identifier, to capture the requirement that the ARIADNE AC should have links to distributed digital and paper resources in the outer world.

In turn, is_about is categorized into three sub-properties making it possible to trace the provenance of the association between the data resource and entity it is about:

- has_native_subject models the association directly imported from the provider of the data.
- has_derived_subject models the association computed by the ARIADNE aggregator by mapping the native subject to an ATT term, to match the requirement that “the Getty AAT should be used as the common conceptual backbone, in addition to local reference resources”.
- has_ARIADNE_subject models the association between the data resource and one of the fundamental ARIADNE categories (see Section 3.4).

The taxonomy of aboutness properties in AO-Cat is given in Figure 3.

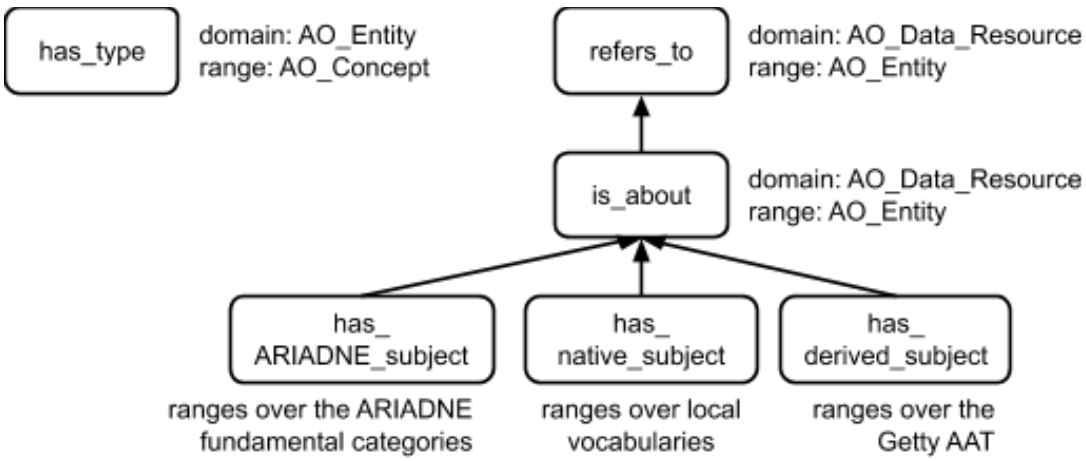


Figure 3. Taxonomy of aboutness properties in AO-Cat

In addition to topicality, required to support *What* searches, the requirements highlight two important aspects of a data resource:

- the spatial coverage, required to support *Where* searches;
- the temporal coverage, required to support *When* searches.

To account for these aspects, AO-Cat introduces the following properties:

- has_spatial_coverage, associating an instance of AO_Data_Resource with a spatial region;
- has_temporal_coverage, associating an instance of AO_Data_Resource with a temporal region.

See Sections 3.2 and 3.3 below for more information on these important properties.

Several descriptive properties are also defined for data resources:

- has_language, the language of the resource.
- was_created_on, the date in which the resource was created.
- has_landing_page, the original landing page of the AO_Data_Resource, if any.
- has_access_policy, a statement of access policy for the resource.
- has_access_rights, a statement of access rights on the resource.
- has_extent, the size of the resource.

The requirements highlight the importance of accounting for the different levels of granularity of the data space, therefore AO-Cat distinguishes data resources into two basic kinds (the complete resource class taxonomy is given in Figure 4):

- individual data resources, which are atomic in the sense that they are not further decomposable in other resources as far as the ARIADNE infrastructure is concerned; these are represented as instances of class AO_Individual_Data_Resource, a sub-class of AO_Data_Resource;
- collections, which are wholes composed of parts that are data resources themselves, and as such they may be collections; collections are represented as instances of class AO_Collection.

Archaeological records provided to ARIADNE by its members are modelled as individual data resources, while property is_about is used to associate each record with the entities (including events) it primarily carries information about. A special kind of individual data resources are documents, to accommodate which class AO_Document is created as a subclass of AO_Individual_Data_Resource. In particular, AO_Document is the class of all the documents ARIADNE needs to deal with, such as for instance the grey literature, published journal articles, radiocarbon dating resource, and possibly others, whose content may possibly be part of the AC.

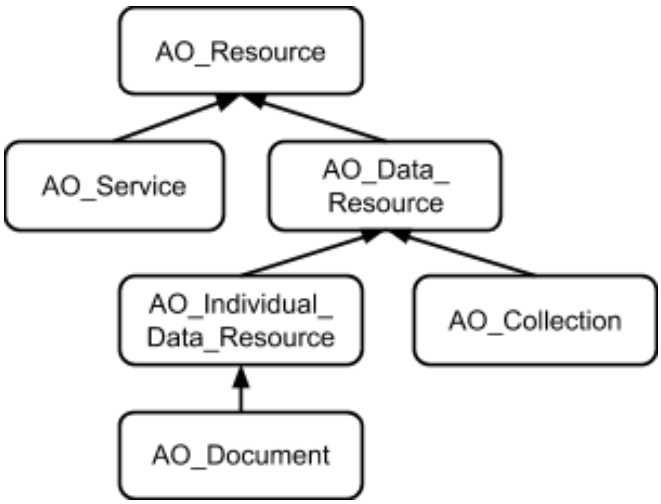


Figure 4. complete resource class taxonomy

The AO-Cat property has_part is introduced to associate a collection with its members.

3.1 Objects

Objects play an important role in the archaeological domain: much of the activity, *e.g.*, in an excavation, concerns objects: discovery, analysis, classification, dating, and so on. As such, objects form an important category for contextualizing the digital resources of the ARIADNE infrastructure and therefore this category needs a place in the AO-Cat.

Class AO_Object is defined for the purpose of classifying in the ARIADNE AC all physical objects that are relevant to ARIADNE. Three properties are defined for it:

- has_time_interval, to relate an object to a temporal region relevant to the object, such as the period when the object was created.
- has_space_region, to relate an object to a spatial region relevant to the object, such as the place where the object is located.
- was_present_at, to relate an object to an archaeological event relevant to the object, such as the excavation event that led to a discovery of the object.

3.2 Defining Space

The generic AO-Cat class for spatial regions is AO_Spatial_Region. Two properties are defined for the instances of this class:

- has_coordinate_system
- has_place_name

Based on the experience of the many content providers in the A+ Consortium and of the previous ARIADNE project, four main representations of a spatial region are provided by AO-Cat, each assigned to a specific sub-class of AO_Spatial_Region. These subclasses are (see Figure below):

- AO_Spatial_Region_Point, having as instances regions that are points. The properties defined on points are:
    - has_latitude
    - has_longitude
- AO_Spatial_Region_Polygon, having as instances regions that are polygons, represented in some format typically managed by a GIS. No commitment is made by AO-Cat on a specific notation for these regions, they are treated simply as abstract objects with a specific property:
    - has_polygonal_representation

giving the XML document representing the polygonal region.

- AO_Spatial_Region_BBox, having as instances regions that are bounded boxes, represented by four points giving the vertices of the box. The properties defined on points are:
    - has_bounding_box_min_lat
    - has_bounding_box_min_lon
    - has_bounding_box_max_lat

    - has_bounding_box_max_lon
- AO_Spatial_Region_StdName having as instances regions that are simply identified by a standard name in some vocabulary. The standard name is given by the property:
    - has_place_IRI

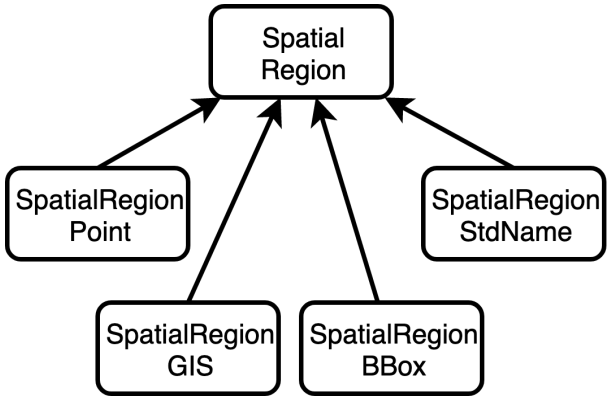


Figure 5 Sub-classes of AO_Spatial_Region

3.3 Defining time

Time is represented and named in many different ways in the archaeological domain, often inextricably associated with space. Fortunately, reference resources have been created in the last decade, which make the task of accounting for time much easier for a research infrastructure such as ARIADNE.

According to the requirements, AO-Cat should support both time points, defined as absolute dates with respect to a reference system, and time intervals, defined as temporal extents having a beginning, an end and a non-zero duration. In addition, time intervals identified by names (e.g., *"Neolithic"*) should be supported, whether these names are drawn from a standard or a local vocabulary.

For the representation of time points, AO-Cat relies on the "date" data type of the XML Schema type system, written as xsd:date.

For the representation of time intervals and their names, AO-Cat relies on the PeriodO gazetteer service[14], which allows to define a temporal interval as a web resource, associated to a label and to

[14] https://perio.do/en/

two absolute dates giving the earliest start and the latest stop of the interval. The service also allows to cluster period resources in collections, thus facilitating the exploration of the gazetteer data space[15].

The AO-Cat class AO_Temporal_Region has time intervals as instances. Each such instance can be described by means of the following four properties:

- has_period, giving the PeriodO time interval as an IRI.
- has_native_period, giving the local identifier of the period, as an instance of AO_Concept.
- from, giving the beginning of the interval as an xsd:date.
- until, giving the end of the interval as an xsd:date.

For more information on mapping to space and time see Section 6.

3.4 The ARIADNE fundamental categories - archaeological data types

There are many concept spaces that can be used to describe the topics a data or a service resource is about. Generally speaking, these topics are grouped in three main categories in the ARIADNE AC.

- the ARIADNE fundamental categories
- the terms of the ATT Thesaurus
- any other term used in the data of a provider.

In the ARIADNE portal the ARIADNE fundamental categories are defined as a very high level classification and used as a mechanism to filter searches according to types of data sets. The initial ARIADNEplus fundamental categories - ARIADNE_subject - (and current corresponding A+ sub-tasks) are:

- Site/monument (4.4.0.1)
- Fieldwork (Sub-task 4.4.0.2)
- Fieldwork report (sub-task 4.4.0.3)
- Scientific analysis (Sub-task 4.4.4, plus 4.4.1, 4.4.2, & 4.4.3)
- Date (4.4.5)
- Artefact (4.4.7)
- Fieldwork archive (4.4.12)
- Inscription (4.4.13)
- Burial (4.4.14)

[15] including about six thousand period definitions as of June 14, 2019.

These categories emerged as *ad hoc* pragmatically defined data clusters in the first ARIADNE project and it is anticipated that they will evolve and increase as the WP4 subtasks gather information about data categories commonly held by A+ partners.

3.5 Events and activities

Much of the archaeological research activity concerns fieldwork. Field work results in digital data resources being collected or created for the purpose of documenting the research activity, presenting its results, or arguing in various ways, for instance expressing hypotheses or refuting other scholars' hypotheses. Digital data resources may also be the result of analyses that are conducted in a research laboratory.

In AO-Cat, these research activities are generally modelled as activities, in the sense of the "actions intentionally carried out by actors that result in changes of state in the cultural, social, or physical systems documented. This notion includes complex, composite and long-lasting actions such as the building of a settlement or a war, as well as simple, short-lived actions such as the opening of a door" [CIDOC CRM Specs 6.2]. AO-Cat follows the conceptual structure of the CRM and models activities as special kinds of events, where an event is a "change of states in cultural, social or physical systems, regardless of scale, brought about by a series or group of coherent physical, cultural, technological or legal phenomena" [CIDOC CRM Specs 6.2].

4 Supplying Data to CNR

As a content provider in ARIADNEplus, partners are expected to make their metadata available so that it can be collected, transformed into the ARIADNEplus data model (AO-Cat), stored on a triplestore based on the GraphDB technology, and enriched with links to PeriodO and Getty AAT.

This section describes what partners need to provide so that the ARIADNEplus aggregation system (managed by ISTI-CNR) can be configured to apply the workflow (collect, transform, enrich) to their records. It is strongly recommended that they also check the results before their content is made publicly available.

Figure 6 shows the data flow that will result in the availability of enriched, AO-Cat compliant metadata records in the ARIADNEplus portal.

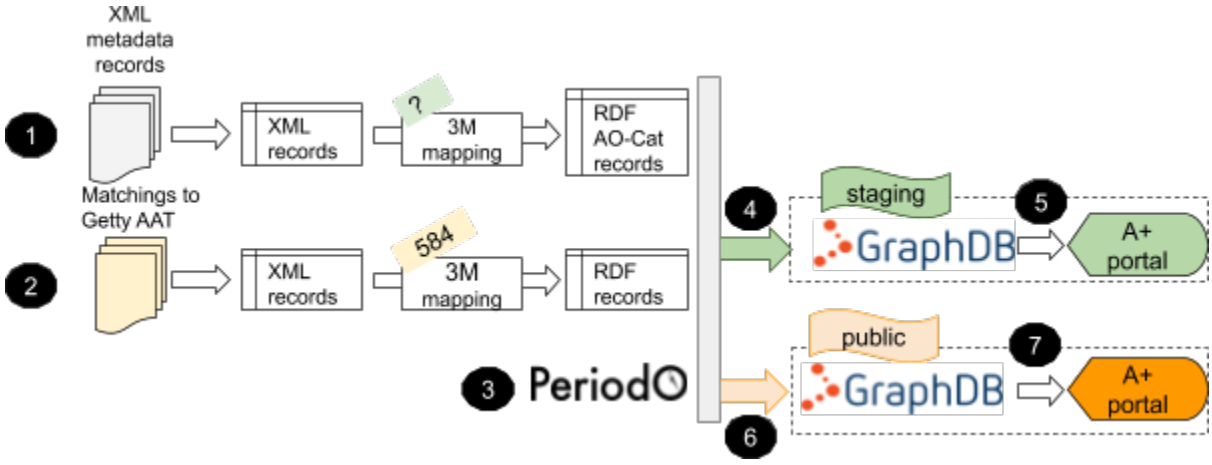


Figure 6. The aggregation workflow applied to each set of records

4.1 In a nutshell

- Make your XML records available (see section 4.2 “Metadata Ingestion”)
- Inform the WP4 leader if you cannot provide normalised dates: he will give you instructions on how to use a tool that normalises them and how provide the output of the tool along with your XML records (if you cannot modify the records themselves to include the generated normalised dates)
- Inform the WP4 leader which is the mapping for your metadata sets and when it is ready (see section 4.3 “Metadata Transformation”)
- Inform the WP4 leader when you are ready with your matchings to Getty AAT and your PeriodO dataset (see section 4.4 “Metadata Enrichment”)

- You'll be notified when the records are ready and you will be able to preview them in a "ghost" portal, only accessible to the members of the consortium (see section 4.6 "Check your content!")
- If everything is ok, give the WP4 leader the green light for publishing and the records will appear in the public ARIADNEplus portal

4.2 Metadata ingestion

The ARIADNEplus aggregation system is based on the D-Net software toolkit (implemented and maintained by ISTI-CNR), which provides built-in features for collecting metadata records in different ways. The main options are[16]:

- **OAI-PMH:** OAI-PMH is a standard protocol for metadata exchange. The aggregator can collect the whole content or only from one or more specific OAI sets. Incremental collection of records is supported if your OAI-PMH publisher supports it.
- **SFTP**: there should be one XML file per resource (i.e. not one single, big file). Incremental collection is supported. You are responsible for the SFTP server. Public key or username/password authentication modes are supported.
- **FTP(S)**: there should be one XML file per resource (i.e. not one single, big file). Incremental collection is supported. You are responsible for the FTP server.
- **Dumps**: you can upload the records in your folder here on the d4science workspace (Workspace > VRE Folders > ARIADNEplus_Mappings > Metadata Ingestion - login required). If you have problems in accessing the folder or login, please open a ticket to the helpdesk.

IMPORTANT: XML files will be transformed with the 3M mapping you have defined. Please ensure the export format is the same as the one you used when defining the mapping. In case of OAI-PMH, each oai:record will be passed to the transformation.

IMPORTANT: Each XML file should have one XML field containing a unique identifier. In case of OAI-PMH, it can be the oai:identifier. The availability of a unique identifier per resource is also a requirement of the 3M mapping.

IMPORTANT: if you have several collections that must be transformed with different 3M mappings, then you should provide the relative records in different sets/folders/dumps.

[16] If you are not sure what to choose or if none of the options suits, please open a ticket on the helpdesk.

4.3 Metadata Transformation

The aggregator will automatically transform the collected records to the ARIADNEplus data model by applying the 3M mapping you have defined with the X3ML mapping tool (see Section 5).

If your XML records already contain normalised dates, then your 3M mapping must properly map the values into AO-Cat temporal coverages.

If instead your dates are not normalised, Ariadne+ is offering a tool, called Timespans, that automatically performs normalization based on a wide set of date patterns commonly used in archaeology. The tool will produce as output a modified version of your XML records, including the normalised dates. You might need to update your 3M mapping to properly map the normalised dates into AO-Cat temporal coverages ("from" and "until" properties).

4.4 Metadata Enrichment

You have created the matching between your local terms and Getty AAT with the Vocabulary Matching Tool.

Save the output on your folder here on the d4science workspace (Workspace > VRE Folders > ARIADNEplus_Mappings > Matched Vocabularies - login required). If you have problems in accessing the folder or login, please open a ticket to the helpdesk.

You can also have multiple mapping files, if it is easier for you to manage. See what UoY-ADS has done in their folder.

Please note that the workspace supports versioning: if you have an updated version of the mapping, upload it with the same name of the previous version. The old version will still be accessible by clicking on the file with the right button of the mouse and then select "Versions".

The files will be collected by the ARIADNEplus aggregator, converted into RDF by applying the 3M mapping 584.

See also Section 6.1.

To enrich your collection with the PeriodO URIs (e.g. fill the values in aocat:has_period), you just need to communicate the URL to your PeriodO dataset.

See also Section 6.2.

IMPORTANT: if you are already using Getty AAT and PeriodO URIs in your records, then you can directly map them into has_derived_subject and aocat:has_period, respectively, with your 3M mapping.

4.5 Check your content!

When the process is completed, you'll be asked to access the ARIADNEplus "ghost" (test) portal where you will be able to preview how the records will look in the ARIADNEplus portal.

If you are an RDF expert, we can also give you access to the ghost GraphDB, so that you can explore the RDF graphs and run the SPARQL queries you prefer. To request access to the ghost GraphDB, please open a ticket on the helpdesk and specify your IP address (or address range).

We will publish your records on the GraphDB serving the public SPARQL and the public ARIADNEplus portal only when you give us your green light!

4.6 Frequency of updates

In the initial phase, we will perform several rounds of collection, transformation, and enrichment in order to tune all the mappings. Once we reach stability, we can agree on the frequency of execution. The ARIADNEplus aggregator can be configured to aggregate the records automatically according to a given schedule: once a week, once a month, or whatever fits best your use case. Examples of options might include:

- I have an OAI-PMH publisher that supports incremental harvesting and deleted records. Records are inserted/deleted/updated daily. I would like to see changes in the ARIADNEplus portal after one week → let's schedule an automatic incremental aggregation once a week
- I have an FTP server where records are automatically published once every 6 months (e.g. January and June), existing records are overwritten every time, records to be removed are deleted from the FTP server → let's schedule an automatic complete re-aggregation every 6 months (e.g. February and July).
- I manually upload the records on the d4science workspace when I know there are relevant records for ARIADNEplus → No schedule: inform WP4 leader of the new records when you are ready.

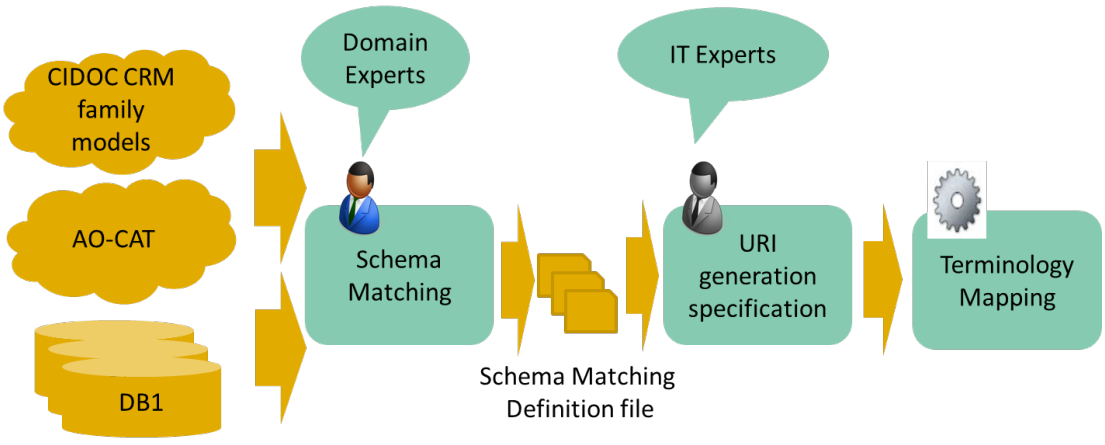
5 Using the 3M Mapping tool

The X3ML Toolkit is a set of small, open source, microservices that follow the SYNERGY Reference Model[17] of data provision and aggregation. They are designed with open interfaces and they can be easily customized and adapted to complex environments. The key components of the toolkit are:

- Mapping Memory Manager - 3M is a tool for managing mapping definition files. It provides a number of administrative actions that assist the data supplying partners manage their mapping definition files.
- 3M Editor - is the interface tool that allows the domain experts to build mappings. It provides:
  - Source and target agnostic mapping facility
  - Guided mapping according to deployed ontology's logic
  - Comment and justification facility
  - Mapping storage
  - Separated instance generation practice for IT professionals
- X3ML Engine - The X3ML Engine realizes the transformation of the source records to the target format. The engine takes as input the source data (currently in the form of an XML document), the description of the mappings in the X3ML mapping definition file and the URI generation policy file and is responsible for transforming the source document into a valid RDF document which corresponds to the input XML file, with respect to the given mappings and policy.
- RDF visualizer – The visualizer allows a fast way to inspect the transformed records.

The X3ML Toolkit is deployed for ARIADNEplus on d4science at:

https://ariadne.d4science.org/group/ariadneplus_mappings/mapping-tool



[17] http://www.cidoc-crm.org/Resources/the-synergy-reference-model-of-data-provision-and-aggregation

5.1 Quick Start Guide

1. Start a new mapping: Copy **Mapping 594** which acts as a template and has the basic necessary files already loaded.

2. Open the newly created mapping file and in the **Info tab** select EDIT:
   a. Change the Title of the mapping. It is suggested that you use the provider name in the mapping title.
   b. In the Sample data and Generator policy section, Source record, load a file with a few sample records.
   c. Fill in as much metadata as possible.

3. **Matching Table Tab**: Proceed to the matching of every source node to the respective AO-CAT node. Details on the matching are available in Section 5.4.

4. **Generators Tab**: Define the Instance and Label Generation functions. The manual with details on the existing functions is available at: https://mapping.d4science.org/3M/Manuals/en/X3ML_Generators_Manual.pdf
   It is mandatory to give ONE instance generator to each class and ONE label generator. Label generators can be more than one.

5. **Transformation Tab**: Run the transformation on the sample records.

6. Check the console for errors and go back to steps 3 and 4 to fix them.

7. View transformed records in order to ensure that the transformation did what you had in your mind.

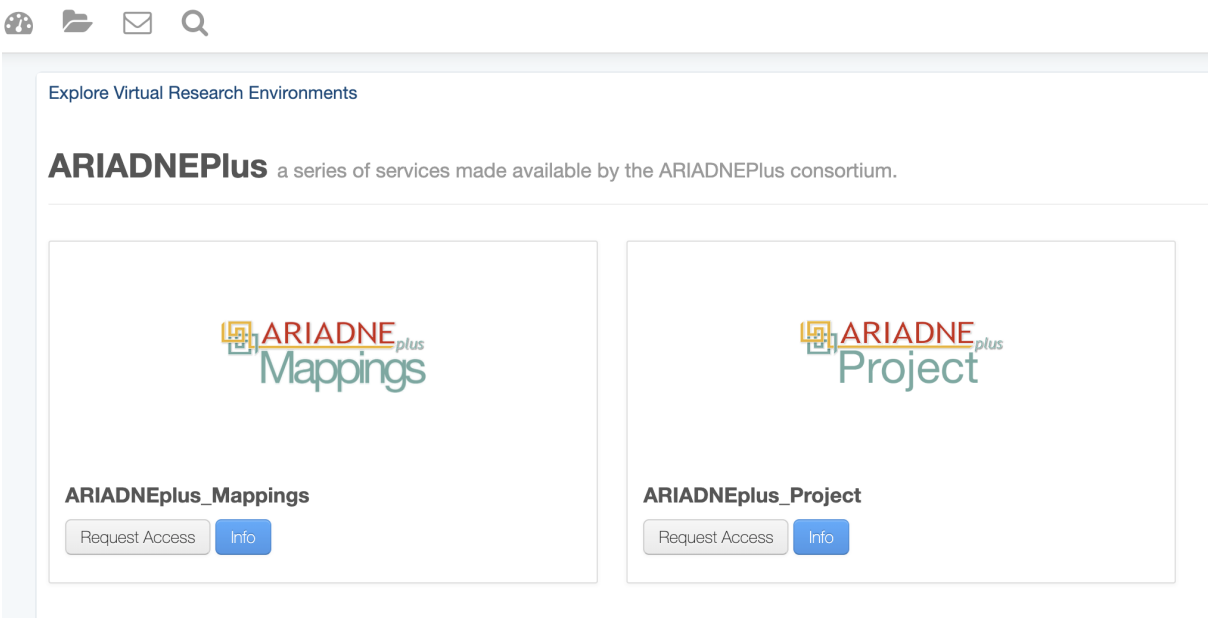
5.2 Basic Interface Operation

5.2.1 Login

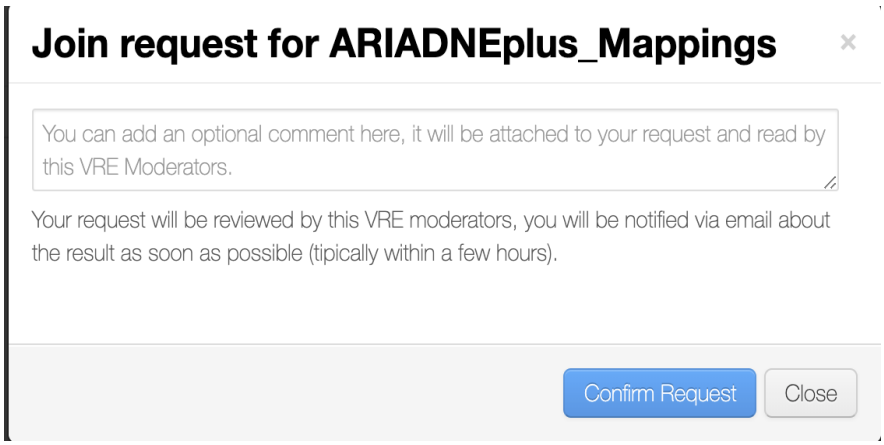
Login using your d4science credentials at:
https://ariadne.d4science.org/group/ariadneplus_mappings/mapping-tool

If you are a new user you first need to register to the ARIADNEplus Mappings VRE on d4science:

- Go here https://ariadne.d4science.org/explore. You will see:

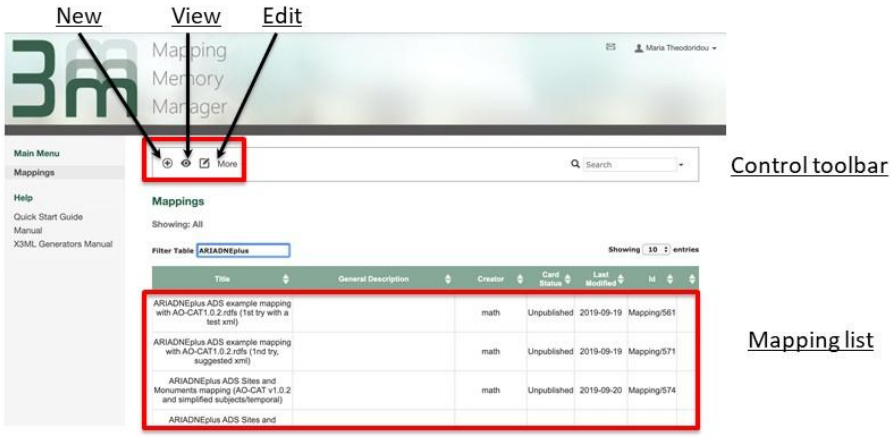


- Click on “Request Access” for the ARIADNEplus_Mappings
- You’ll be asked to login. If it is your first time using D4Science, you’ll be prompted with the terms of use that you have to read and accept (Click on the “I agree” button on the bottom left).
- Once logged in you’ll have to confirm the request by clicking on the “Confirm Request” button on the window that will be prompted.



- You will receive a notification email when the moderators accept your request. As soon as you receive the email you can access the environment and use the 3M mapping tool at https://ariadne.d4science.org/group/ariadneplus_mappings/mapping-tool

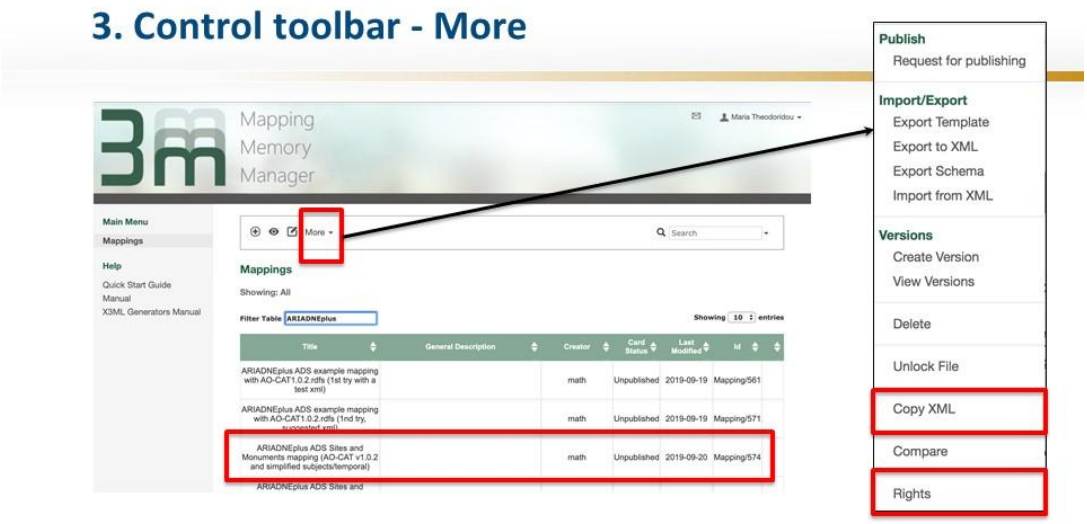
5.2.2 Control toolbar and Mapping list



- Control toolbar for working with mappings.
- To view or edit a map, first select it in the Mapping list, then click the appropriate icon.

14

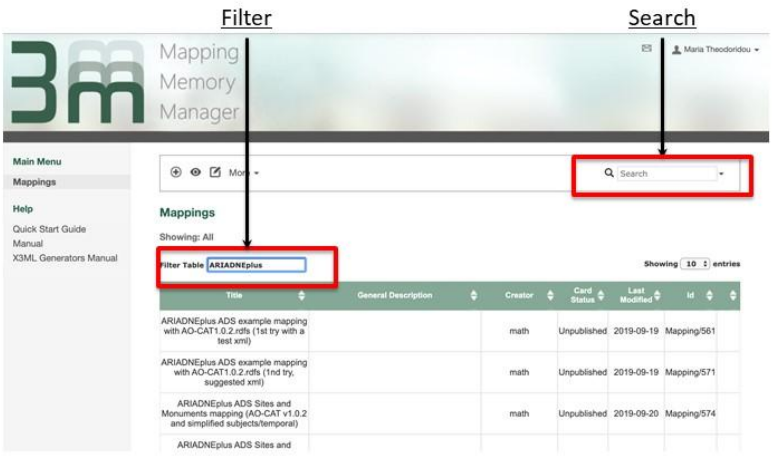
5.2.3 Control toolbar – More/Copy XML/Rights



- ***More*** dropdown menu gives many options for working on a mapping.
- ***Copy XML*** allows you to take a full copy of the selected mapping (useful when not sure of changes).
- ***Rights*** allows you to share edit rights to other users.

15

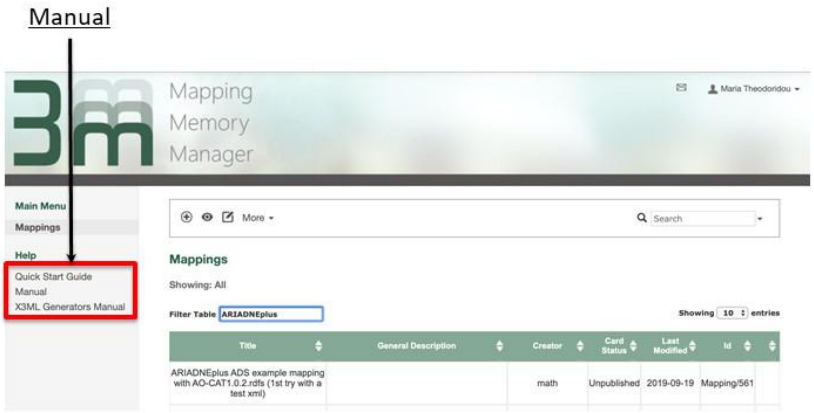
5.2.4 Search and Filter



- There is a search feature allowing you to find mappings.
- There is a filter feature to narrow down within a viewed search set.

16

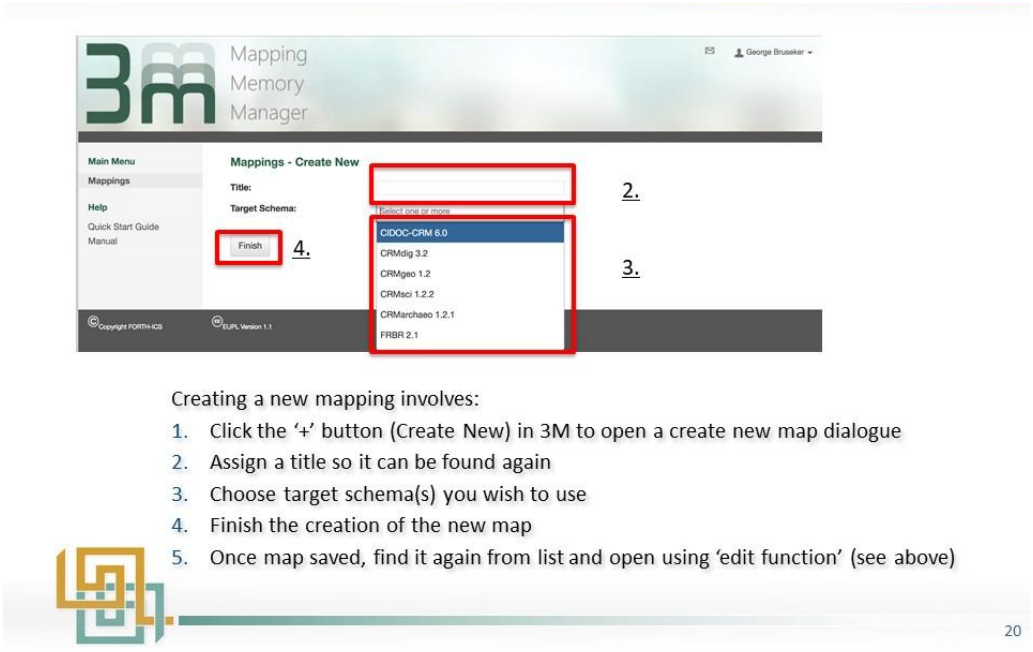
5.2.5 Manual



- A Quick Start Guide and the complete documentation for 3M including general mapping instructions and the generators manual are always accessible directly from the sidebar.

5.3 Setting up a New Mapping

5.3.1 Give Title and choose Target Schemas

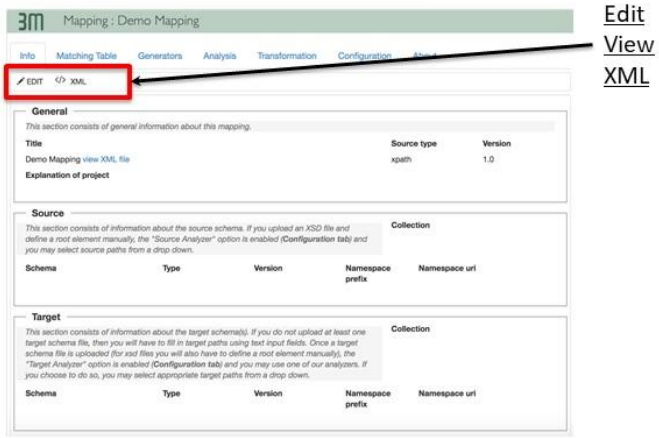


Creating a new mapping involves:

1. Click the '+' button (Create New) in 3M to open a create new map dialogue
2. Assign a title so it can be found again
3. Choose target schema(s) you wish to use
4. Finish the creation of the new map
5. Once map saved, find it again from list and open using 'edit function' (see above)

20

5.3.2 Edit/view Info Tab



- To edit the mapping, you will need to enter edit mode. This can be done by choosing the 'Edit' button.
- To leave edit mode, you can choose the opposite 'View' button. Experts can also edit the XML directly by clicking the 'XML' button.

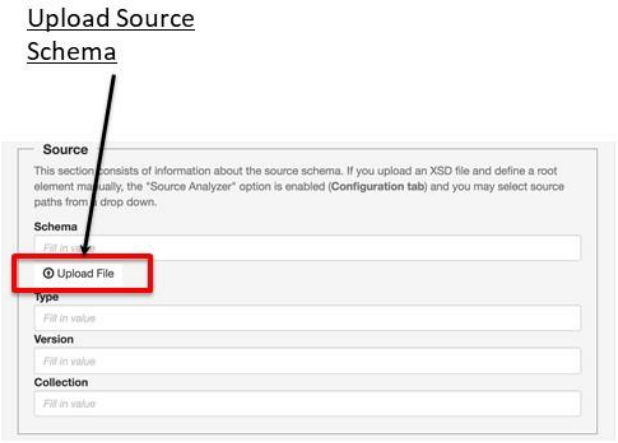
21

5.3.3 Add Generic Mapping Metadata

Mapping Metadata provides crucial provenance information to the mapping so that it can be re-used in the future.

Who did this mapping? How can they be contacted?

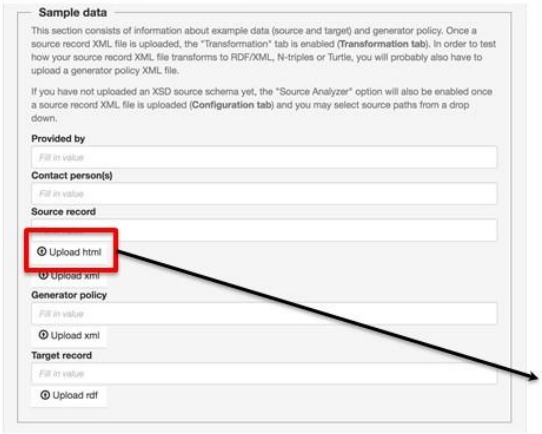
5.3.4 Adding Source Schema



Adding source schema allows 3M to help guide mapping by only allowing valid paths to be selected from source.

- Upload a source schema using 'upload button'
- Source can be in XML or XSD.
- Additional metadata can be added for
  - the type of schema
  - the version of the schema
  - what collection schema used for

5.3.5 Adding Sample Data



- Upload Source Data by choosing 'Upload xml' button
- Add provenance information of who gave data and how they can be contacted.

Upload Sample Source Data

Adding source sample data allows testing of the mapping with real data using the 3M transformation tool.

Sample source record

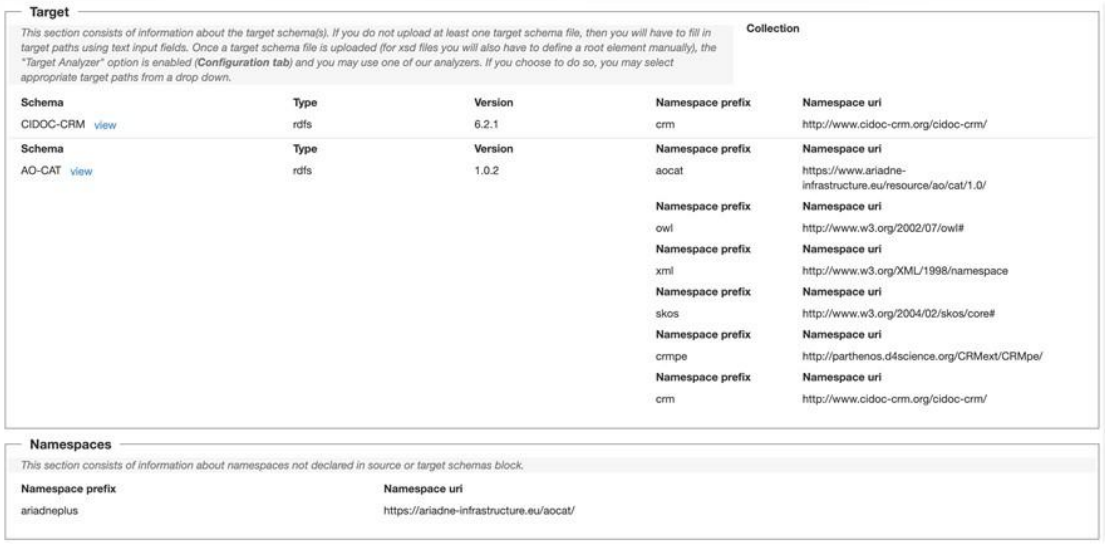
```
<?xml version='1.0' encoding='UTF-8'?>
<Collections xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
      xmlns:dc="http://purl.org/dc/elements/1.1/"
      xmlns:dcterms="http://purl.org/dc/terms/"
      xmlns:ads="https://archaeologydataservice.ac.uk/">

      <collection>
            <meta>
                  <dc:title>Historic England National Record of the Historic Environment (NRHE) </dc:title>
                  <dc:identifier>1000398</dc:identifier>
                  <dc:source>Historic England</dc:source>
                  <dc:description>The Historic England National Record of the Historic Environment (NRHE) consists of core elements of the record of England&apos;s
                  archaeological and architectural sites held by the NRHE.  It encompasses the historic environment in its widest sense and includes archaeological,
                  architectural and historical sites from earliest times to the present day, covering England and its territorial waters (the 12 mile limit).  Each
                  record is linked to its corresponding entry in Historic England&apos;s Pastscape web site.  </dc:description>
                  <dc:publisher>Archaeology Data Service</dc:publisher>
                  <dc:creator>Historic England</dc:creator>
                  <dc:date>15 Sept 2004</dc:date>
                  <dc:format>Dataset Collection</dc:format>
                  <dc:language>eng</dc:language>
                  <dc:rights>http://archaeologydataservice.ac.uk/advice/termsOfUseAndAccess</dc:rights>
            </meta>
            <records>
                  <record>
                        <dc:title>Part of a cropmark with...</dc:title>
                        <dc:creator>Historic England</dc:creator>
                        <dc:subjectPeriod>
                              <dc:subject>CROPMARK</dc:subject>
                              <dcterms:temporal>PALAEOLITHIC</dcterms:temporal>
                        </dc:subjectPeriod>
                        <dc:subjectPeriod>
                              <dc:subject>HOUSE</dc:subject>
                        </dc:subjectPeriod>
                        <dc:description>Part of a cropmark with internal markings seen on air photograph. Field investigation in 1957 found the marks to be due to
                        natural undulations and vegetational changes, and no traces of antiquity were observed.</dc:description>
                        <dc:identifier>Depositor ID: NT 70 NE 1</dc:identifier>
                        <dc:identifier>NMR_NATINV-3</dc:identifier>
                        <dc:source>https://archaeologydataservice.ac.uk/archsearch/record?titleId=967971</dc:source>
                        <dc:language xsi:type="dcterms:ISO639-2">eng</dc:language>
                        <dcterms:spatial>ALWINTON, ALNWICK, NORTHUMBERLAND, ENGLAND</dcterms:spatial>
                        <dcterms:spatial xsi:type="dcterms:POINT">-2.758696, 55.945225</dcterms:spatial>
                        <dc:rights>http://archaeologydataservice.ac.uk/advice/termsOfUseAndAccess</dc:rights>
                  </record>
                  <record>
                        <dc:title>Square mark seen on air...</dc:title>
                        <dc:creator>Historic England</dc:creator>
                        <dc:subjectPeriod>
                              <dc:subject>CROPMARK</dc:subject>
                              <dcterms:temporal>UNCERTAIN</dcterms:temporal>
```

5.3.6 Adding Target Schemas

- Click '+Target' to add new target
- Upload target schema using 'upload' button
- For each target schema specify
  - it's namespace prefix e.g.: 'aocat'
  - It's full namespace URI
- Additional metadata can be added:
  - name of the schema
  - type of schema
  - version of the schema
  - collection used for

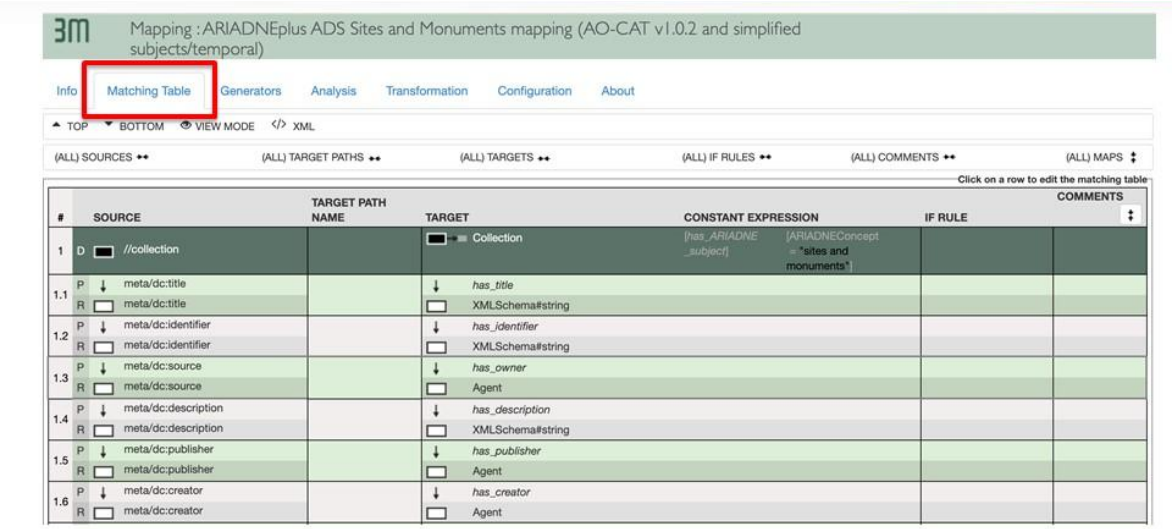
Adding target schema allows 3M to help guide mapping by only allowing valid paths to be selected from target.



In this example CIDOC-CRM and AO-CAT are loaded.

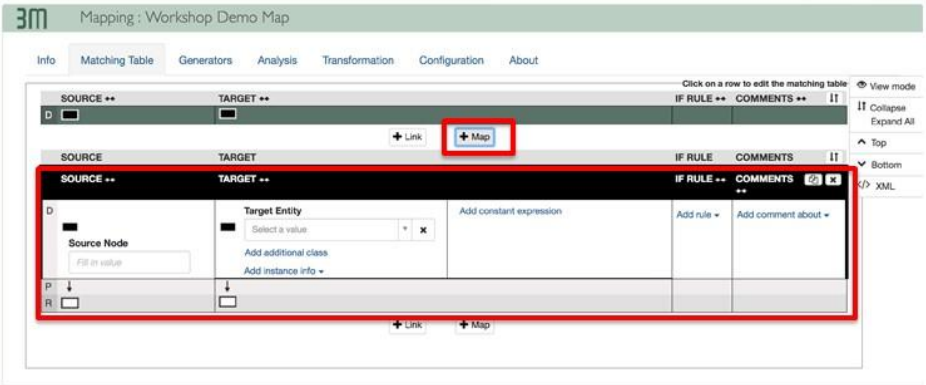
5.4 Matching Table Operation

5.4.1 Accessing the Matching Table



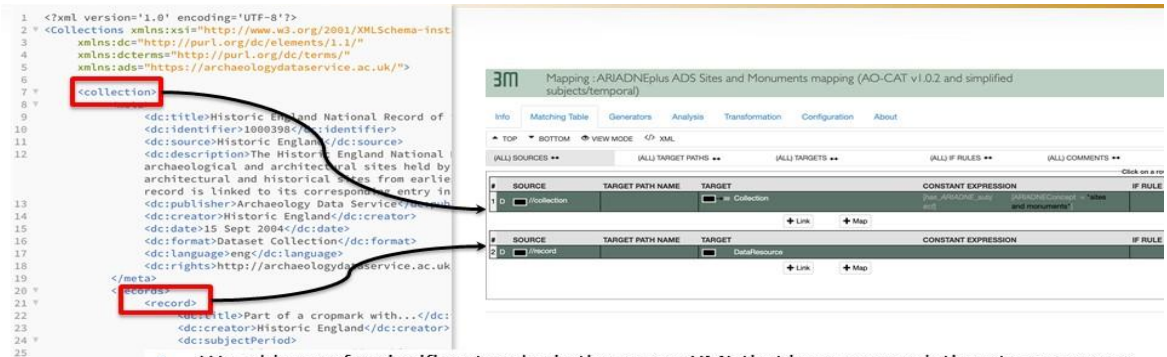
- From within your mapping, click on 'matching table' tab

5.4.2 Adding a New Map



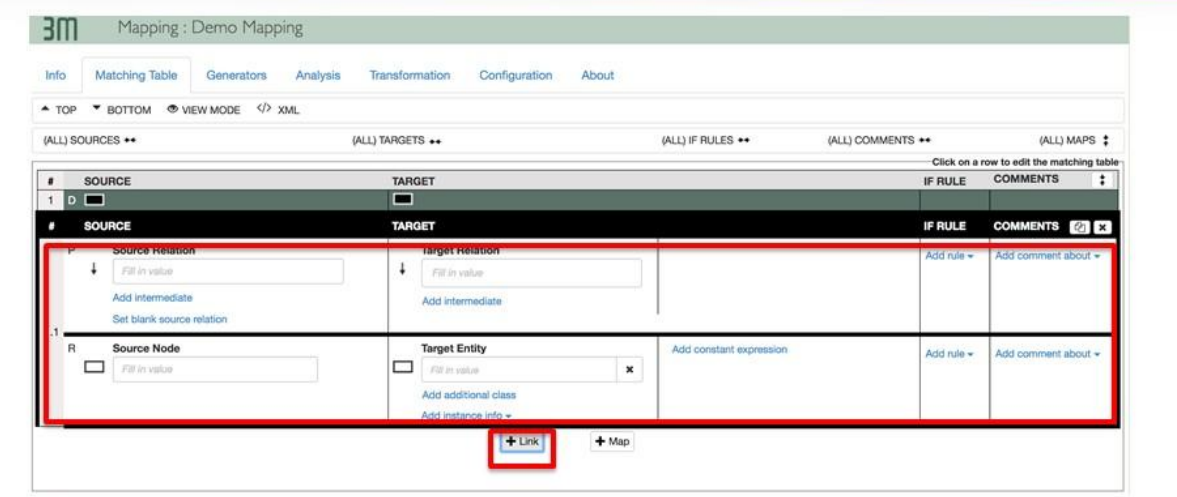
- Maps are used for mapping a domain (root node) from the source to a specific class in the target ontology
- If your data structure is complex (not flat) you can use multiple maps to map multiple domains (root nodes) in the source data structure
- Use '+ Map' button to add additional maps

Adding Map Examples



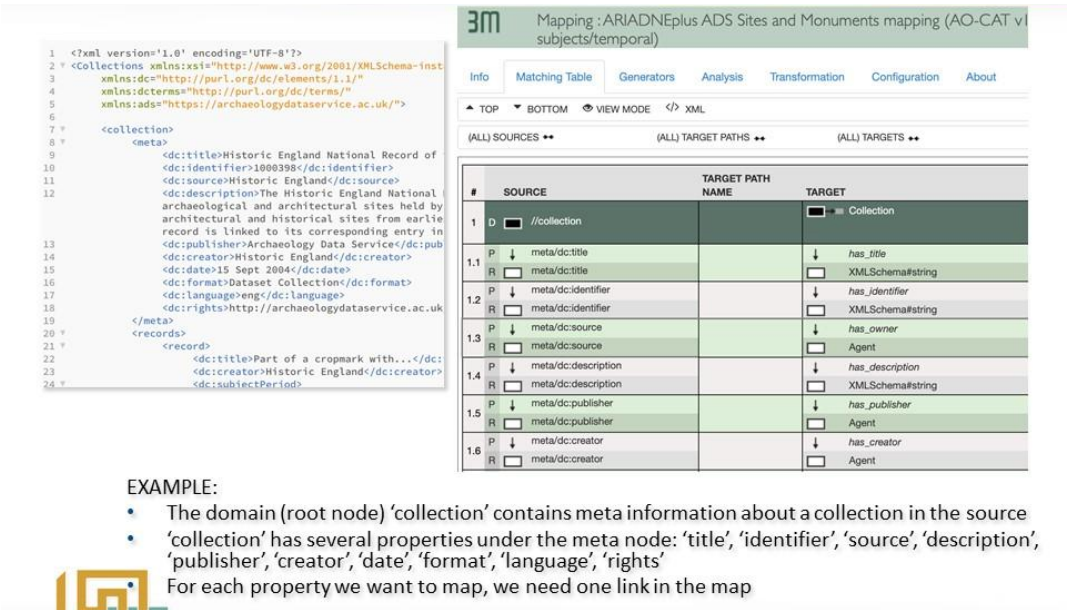
- We add maps for significant nodes in the source XML that have many relations to re-express in the target ontology. These are primary foci of interest.
- In this example '**Collections**' is the root node under which individual '**collection**' records in this schema are documented. Each '**collection**' includes
- general meta information, attributing title, identifier etc. grouped under the '**meta**' node and
- a list of '**record**' nodes grouped under the '**records**' node.
- Our focus of interest are **collection** and **record** and each becomes the source domain of a map. The two maps will be linked eventually.
- The splitting in maps simplifies mapping and makes the mapping definition easier to read.

5.4.3 Adding a New Link



- A link is used for mapping one field/node under a domain (root node) from the source to a semantic path in the target ontology
- For each field/node in the source that should be mapped, there should be at least one link in the map
- Use '+ Link' button to add additional links

Example

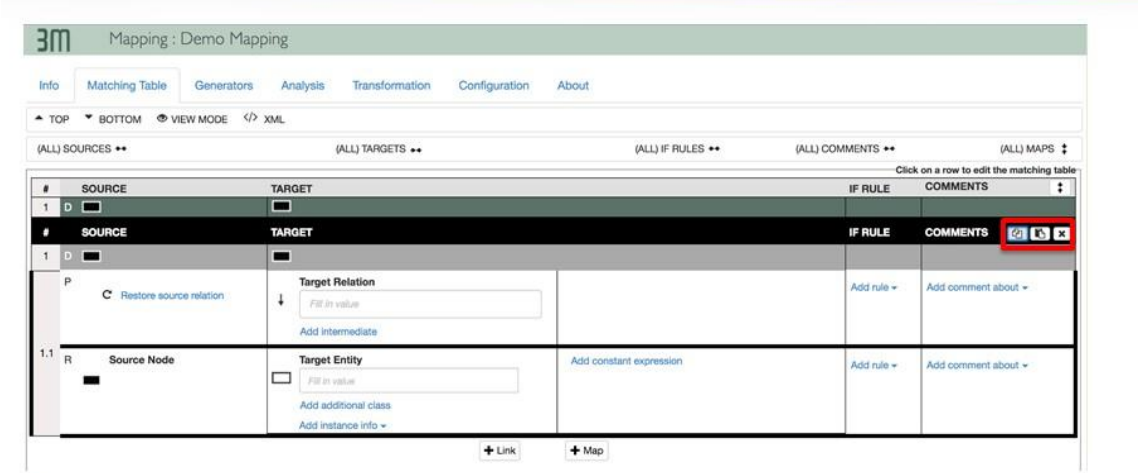


EXAMPLE:

- The domain (root node) 'collection' contains meta information about a collection in the source
- 'collection' has several properties under the meta node: 'title', 'identifier', 'source', 'description', 'publisher', 'creator', 'date', 'format', 'language', 'rights'

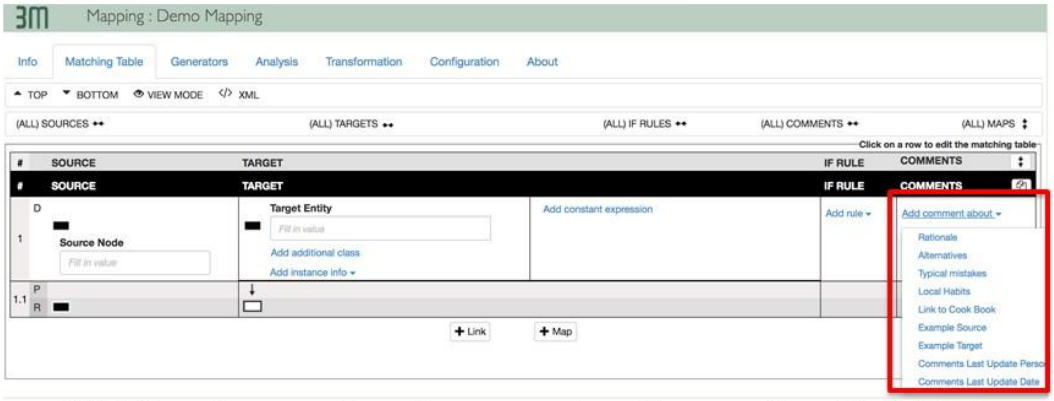
For each property we want to map, we need one link in the map

5.4.4 Copy and Deleting Maps and Links



- Longer maps/links can take time to generate. The copying feature allows one to quickly duplicate existing maps/links.
- To copy a map or link click the 'copy' icon, followed by the 'paste' icon
- To delete a map/link, select the appropriate map/link and then click the 'x'

5.4.5 How to comment

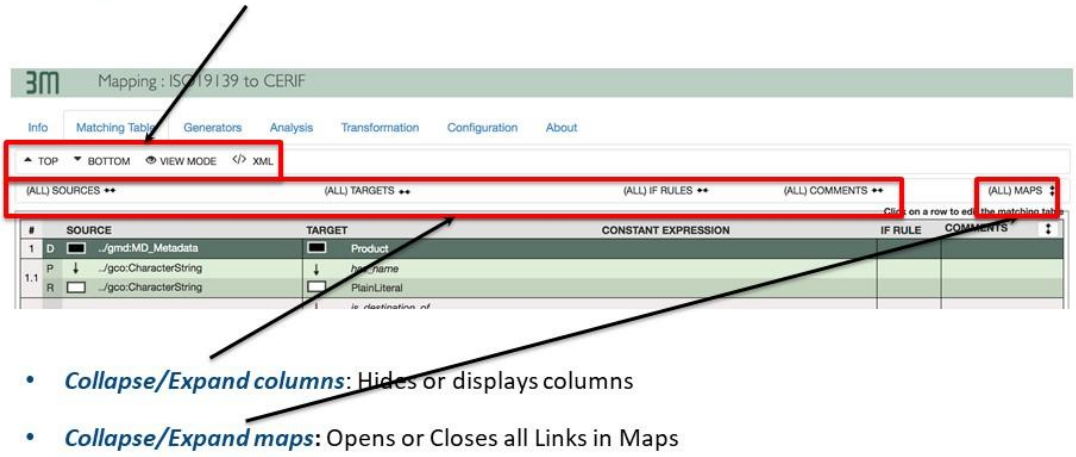


- 3MEditor allows you to make comments on the mapping of individual fields/nodes in links. This is a useful feature to explain/remember mapping decisions
- To add comment click 'Add comment about'

5.4.6 View Controls

3MEditor also offers a series of control tools to help navigate and work with your mapping

- ***Top/Bottom***: Provides scrolling over long maps
- ***View Mode***: toggles map to read only
- ***</> XML***: Allows direct editing of the underlying X3ML definition of the mapping



- ***Collapse/Expand columns***: Hides or displays columns
- ***Collapse/Expand maps***: Opens or Closes all Links in Maps

5.5 Mapping Patterns

5.5.1 X3ML Constructs

X3ML supports **1:N mappings** and uses the following special constructs:

- **intermediate nodes** used to represent the mapping of a simple source path to a complex target path.
- **constant expression nodes** used to assign constant attributes to an entity.
- **conditional statements** within the target node and target relation support checks for existence and equality of values and can be combined into Boolean expressions.
- **"Same as" variable** used to identify a specific node instance for a given input record that is generated once but is used in a number of locations in the mapping.
- **Join operator (==)** used in the source path to denote relational database joins
- **info and comment blocks** throughout the mapping specification bridge the gap between human author and machine executor.

5.5.2 Mapping Source Root to Target Domain

- Click top row in a map
- On Source Side, choose appropriate domain (a node) from source
- On Target Side, choose appropriate class in target ontology

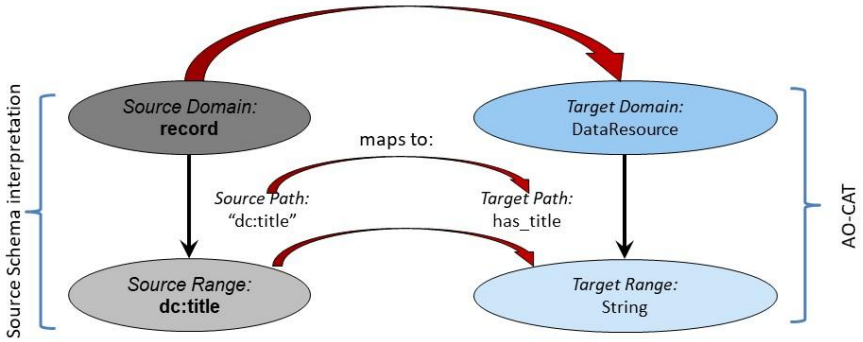
This now says:

"for each instance of source domain (node) generate one instance of target class"

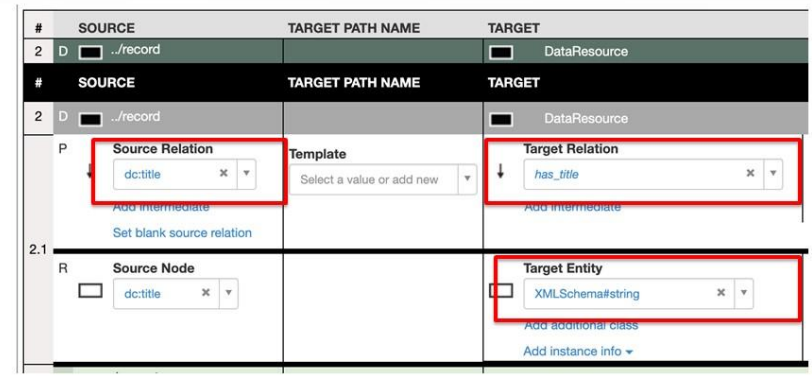
So this mapping instruction says,

"for each instance of 'record' in source, create an RDF instance of type 'DataResource' of the AO-CAT".

Simple Field Mapping: Creating and Equivalent Proposition

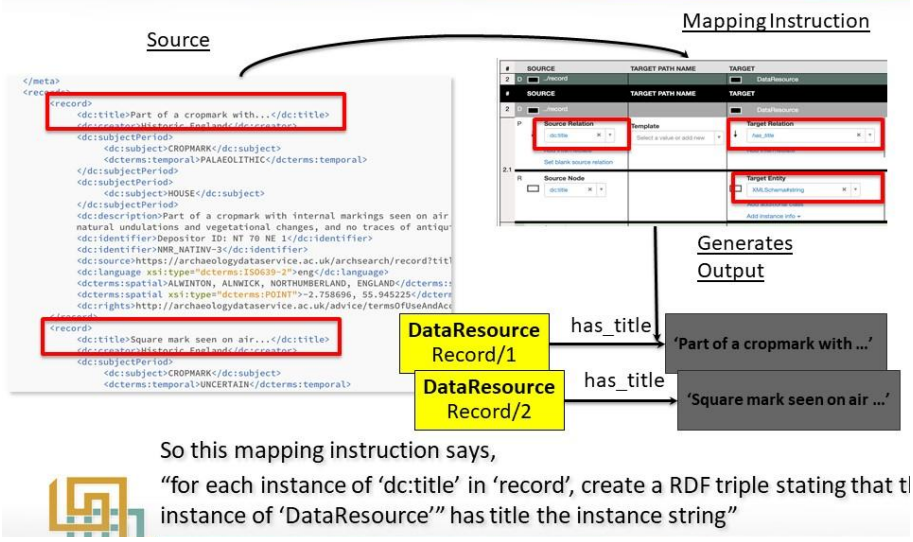


5.5.3 Simple Field Mapping (One to One)



- Choose field for mapping from Source
- Add one relation and one target class in target
- This creates a simple triple statement in RDF of form S-P-O

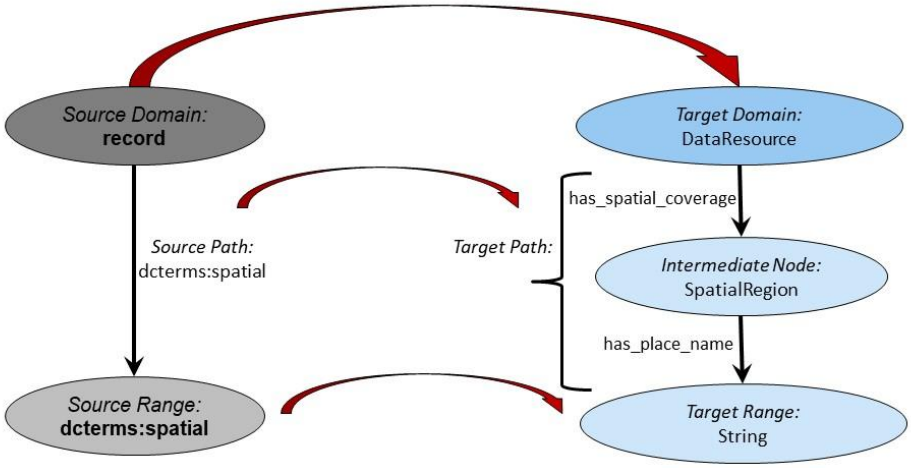
Example

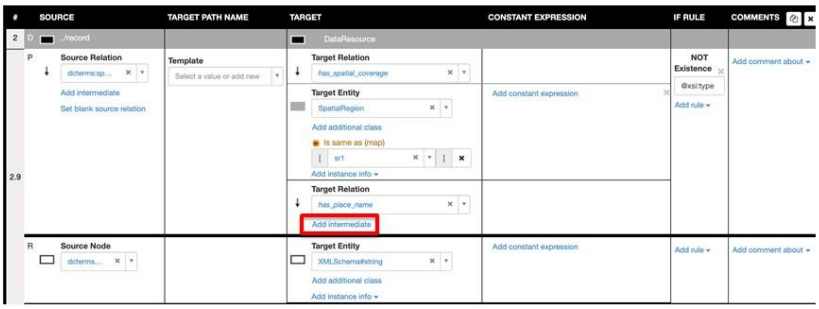


So this mapping instruction says,

"for each instance of 'dc:title' in 'record', create a RDF triple stating that t
instance of 'DataResource'" has title the instance string"

5.5.4 Mapping to Paths: Introducing intermediate nodes



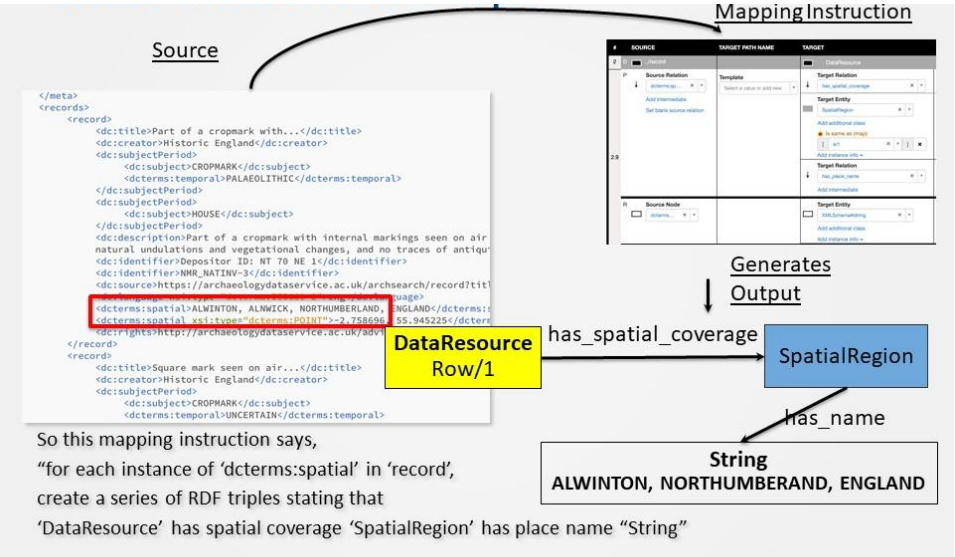


For more complex paths, intermediate nodes are necessary.

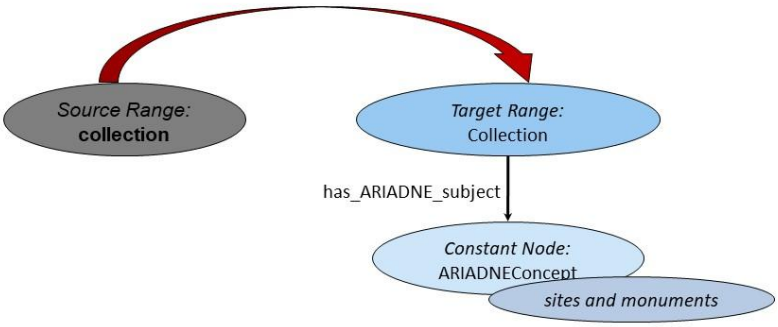
- 'Click 'Add intermediate' for as many intermediate relations as you may need to build.

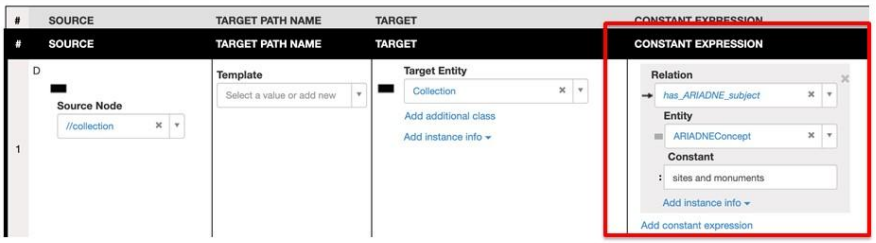
In this example, the place name of the SpatialRegion is expressed.

Example



5.5.5 Mapping to Paths: Introducing constant expressions



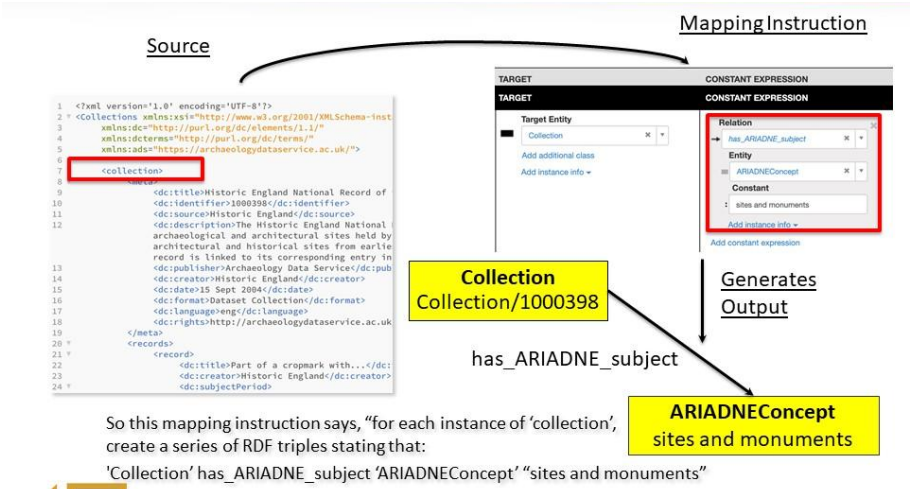


Sometimes, we want to add additional constant information to our mapping.

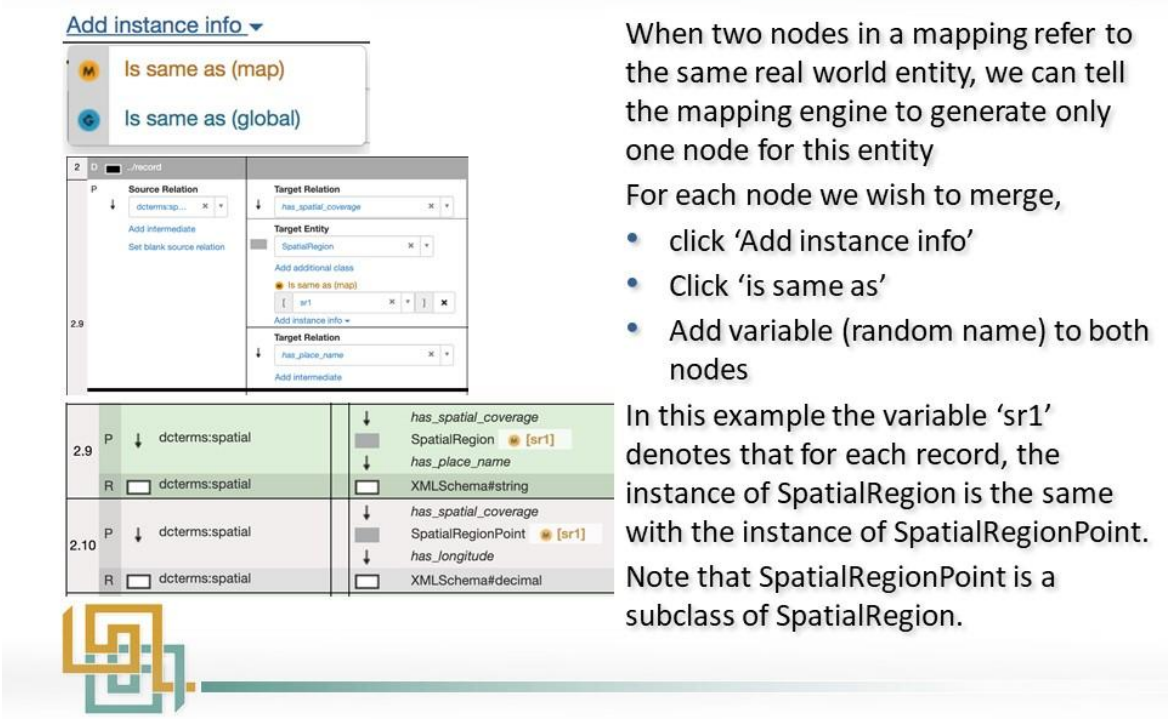
- Click 'Add Constant Expression' (to the right of target)
- Add appropriate links and constant info

In this example, we want to express for each node representing a collection in the target graph has ARIADNE subject 'sites and monuments''.

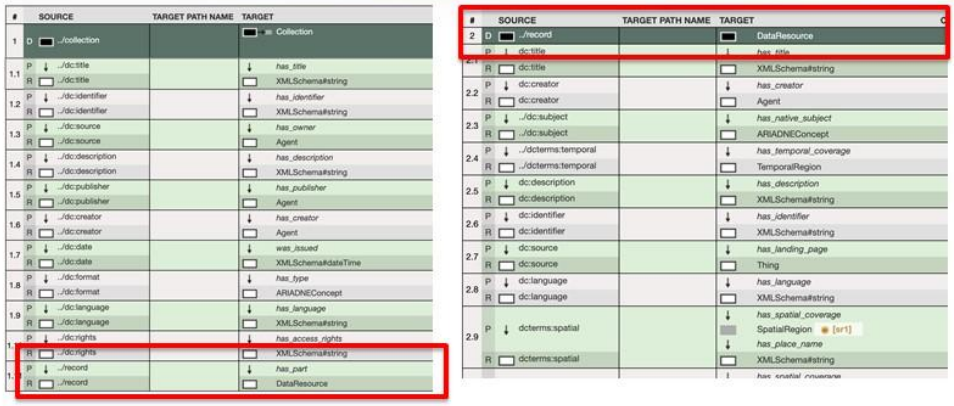
Example



5.5.6 Using variables



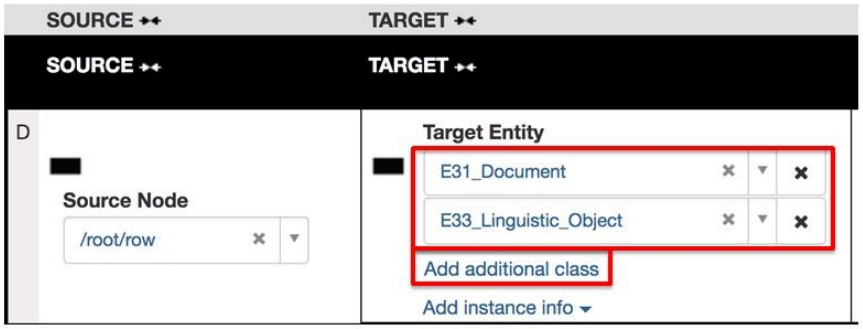
5.5.7 Joining Maps



Two maps can be joined, if the first maps a node which is the domain of the second map. This is often done for convenience purposes, to make mapping easier and simple to read.

E.g.: in the ADS example, 'collection' has a nested list of 'record' nodes. Each 'record' node has many fields describing the record. In map 1, collection references the 'record'. In map 2, we map the fields for 'record' node.

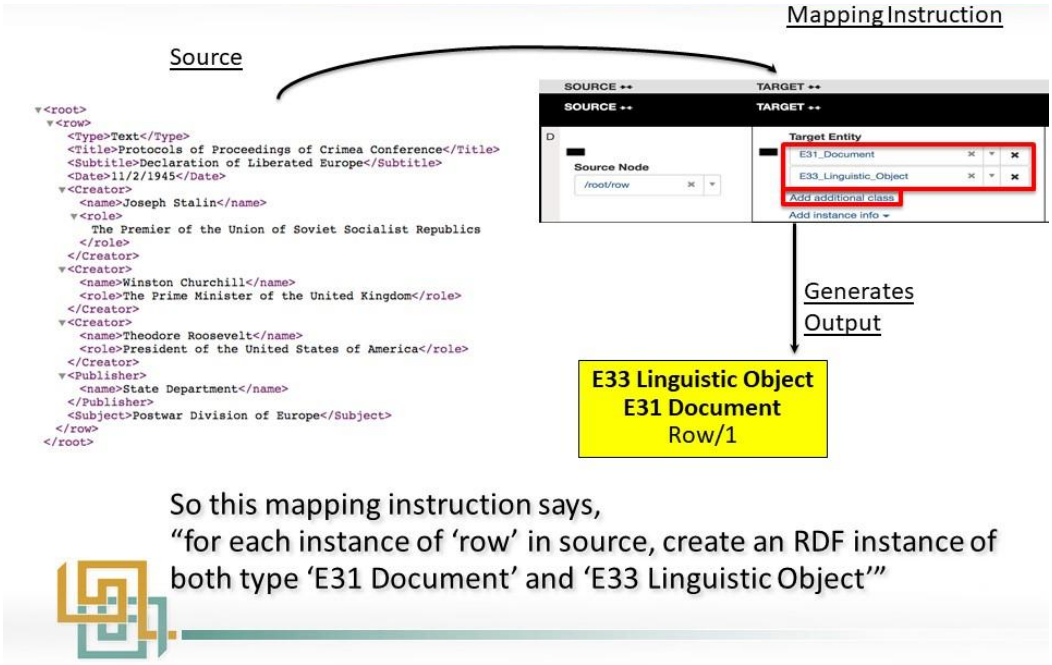
5.5.8 Multiple instantiation



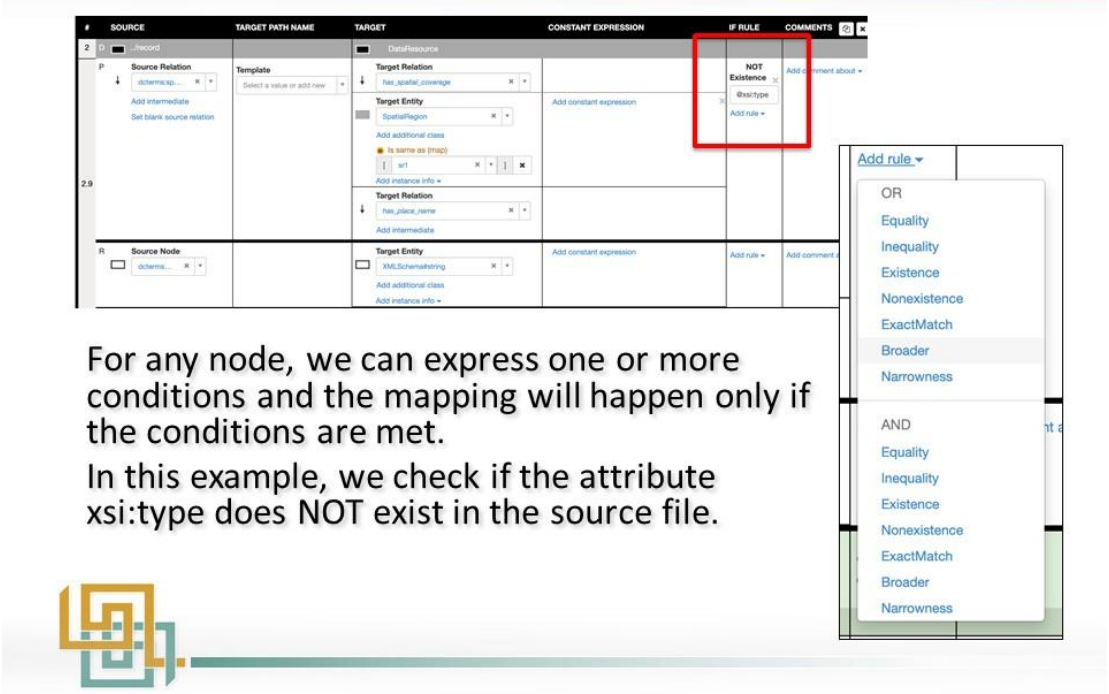
For any node, if it is an instance of two or more classes, we can express this in 3M by using the 'add additional class' feature.

In this example, we say each 'row' from source both is an E31 Document and an E33 Linguistic Object. This allows the use of relations from both classes.

Example



5.5.9 Mapping under conditions - If rules



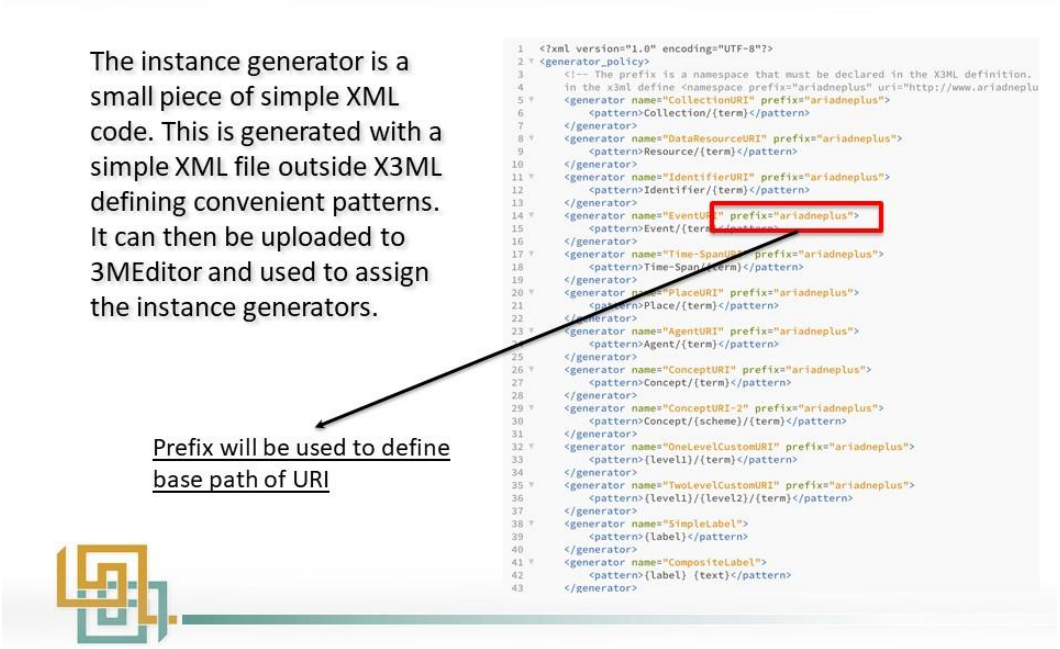
5.6 Instance and label Generators

5.6.1 Why Instance and Label Generators, what do they do?

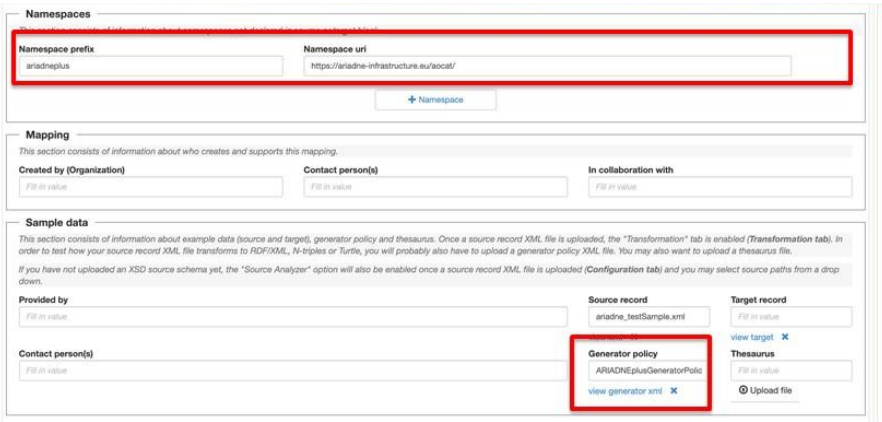
The mapping table allows you to make a translation between a source schema and an RDFS encoded schema like CIDOC CRM. Each node specified in the target will become a separate data entity in the semantic graph that is created through the X3ML transformation engine. This separate data entity will need a unique identifier by which it can be found in the system (like a unique key in a relational database). The instance generators allow specifying patterns for building unique identifiers for instances called 'URIs'.

Because a URI is often unreadable, it is highly recommended that for each node a label (usually the actual data value from your source schema) be added to each node as well. This is also done through the instance generator.

5.6.2 Defining instance generation functions/patterns (offline)



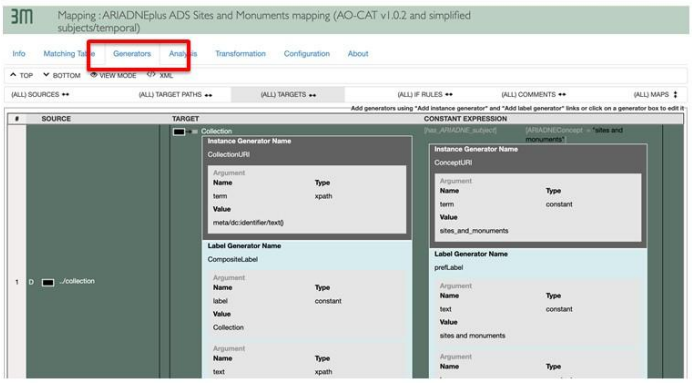
5.6.3 Adding generator file (info tab)



1. Add a new namespace, and specify the same namespace prefix as in your generator file. Use 'namespace uri' field to specify your base namespace.
2. Upload the generator file.

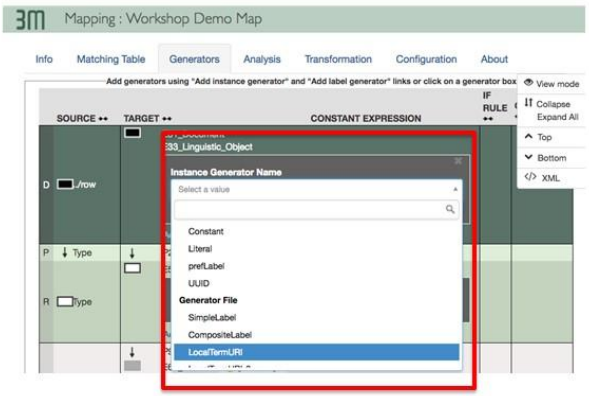
61

5.6.4 Opening up the generator editor

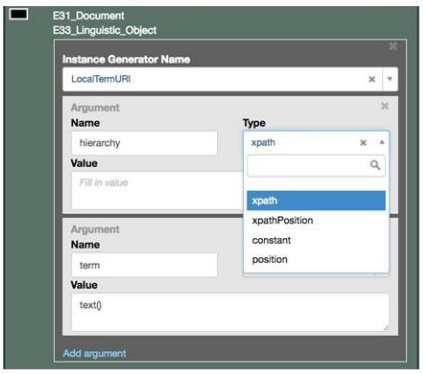


- Opening generator editor is easy, just click to the generators tab.
- The tool adds by default the UUID generator to all entities for fast testing.

5.6.5 Specifying Instance Generators

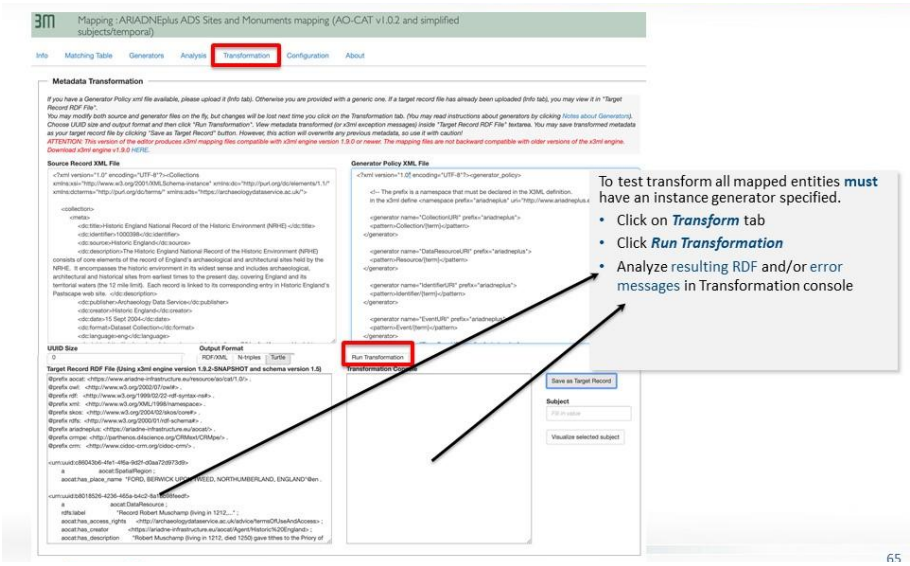


1. Click Add Instance Generator
2. Select Instance Generator

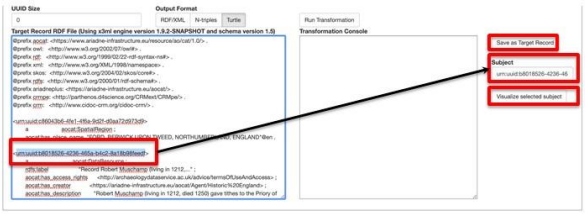


1. Each argument in a generator specifies a piece of the URI string
2. Each argument in the URI can be of different types, chiefly 'Constant' (a value you specify) or 'xpath' (a value from the source)
3. The unique portion of the URI should usually be drawn from the source data. It is accessed from a node using the command 'text()'

5.6.6 Testing Transforms

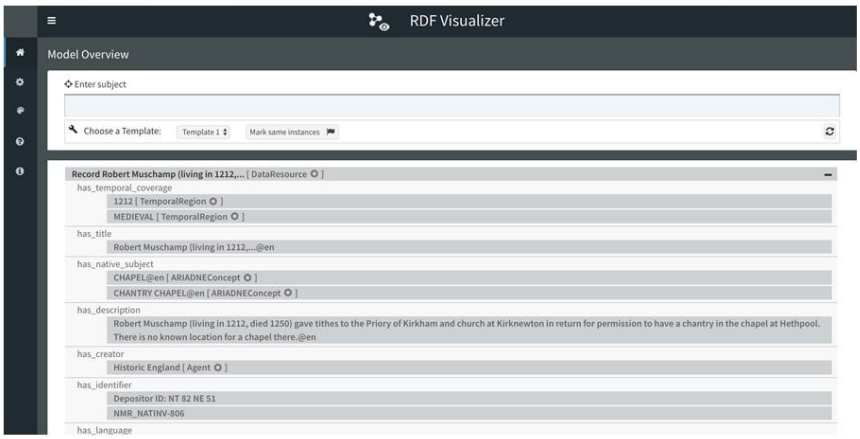


5.6.7 Visualize transformed records



- To visualize a transformed record, you must first ***Save as target record.***
- Select the URI of a record, it automatically copies in the ***Subject*** text box
- Click ***Visualize selected subject***

5.6.8 RDF visualizer



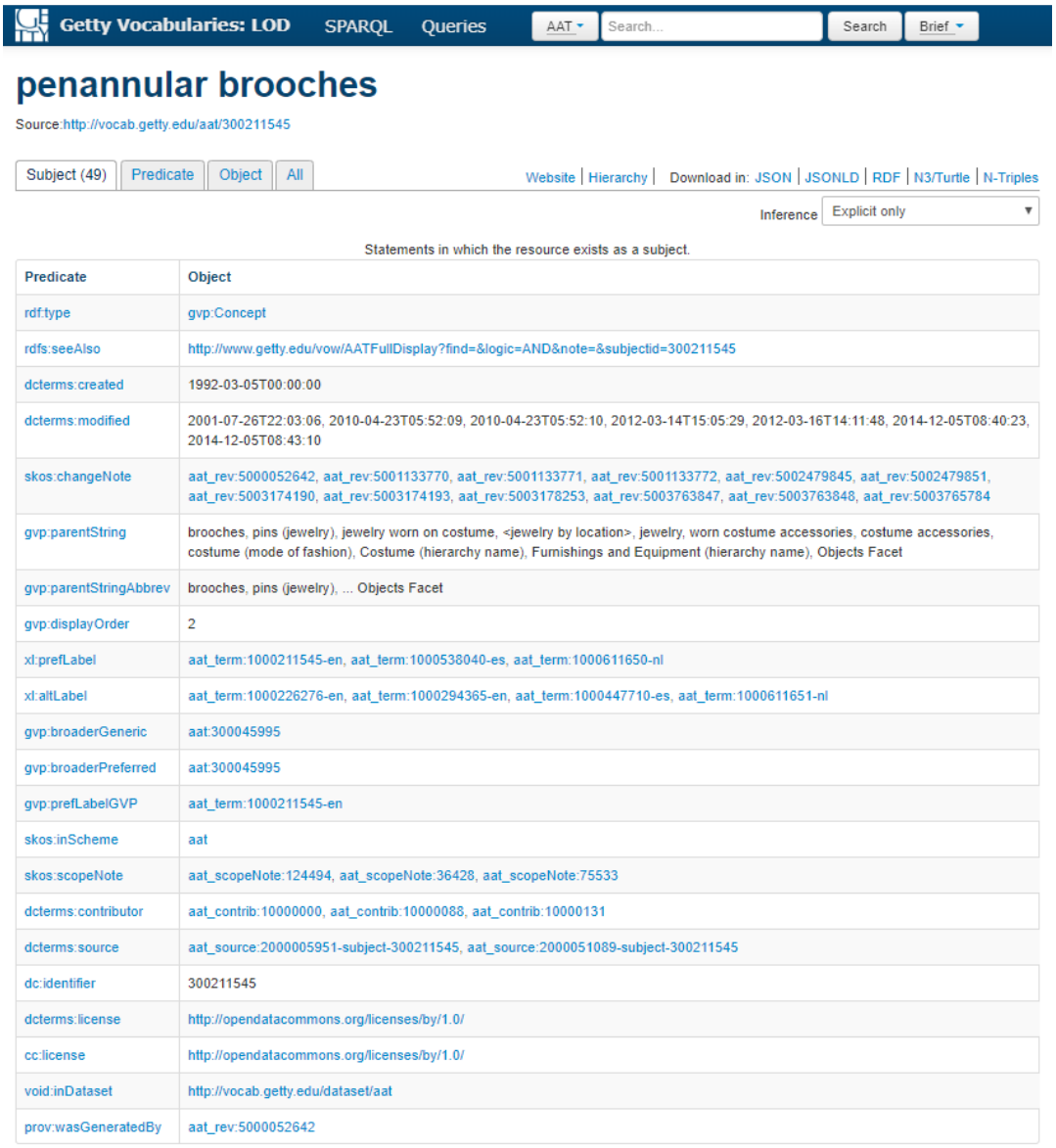
6 Data Mapping

The integrated metadata will be searchable via the ARIADNEplus portal according to three main facets: Where (place) / When (time) / What (subject). We need to support effective cross-search of metadata originating from many different countries; this requires some shared common understanding of the meaning of the metadata. The where and when facets may be communicated using common data types and comparable values (e.g. spatial coordinates relative to a known location on the earth, date ranges relative to a known epoch). The what (subject) aspect can be more difficult to define in a commonly understood and comparable format.

Clean and consistent metadata is a prerequisite for effective integration and reuse, as we can neither predict nor dictate the way in which it may be reused in the future. Inconsistent or erroneous metadata causes problems in accurately identifying co-references or mappings. Embedded errors may resurface and be magnified where the metadata is subsequently reused in a different context, causing difficulties in performing logical reasoning – and leading to potentially inaccurate assertions or conclusions in future studies.

6.1 Getty AAT

The Getty Art & Architecture Thesaurus (AAT) provides concepts and terms to describe cultural heritage concepts. AAT is made available as Linked Open Data (LOD), so each concept in the thesaurus has a unique identifier in the form of a URI (e.g. http://vocab.getty.edu/aat/300211545 is the identifier of the concept "*penannular brooches*"). ARIADNEplus is using AAT as a central hub for scalable interconnection of local subject vocabularies. Data providers are required to provide a set of mappings from their own local subject vocabulary (all terms used to describe subjects in their own metadata) to the AAT. The aim of the subject mapping exercise is to improve recall and precision when later browsing and searching the ARIADNEplus integrated data by subject (time and space are being handled separately).



Getty Vocabularies: LOD SPARQL Queries AAT Search... Search Brief

penannular brooches

Source:http://vocab.getty.edu/aat/300211545

Subject (49) Predicate Object All Website | Hierarchy | Download in: JSON | JSONLD | RDF | N3/Turtle | N-Triples

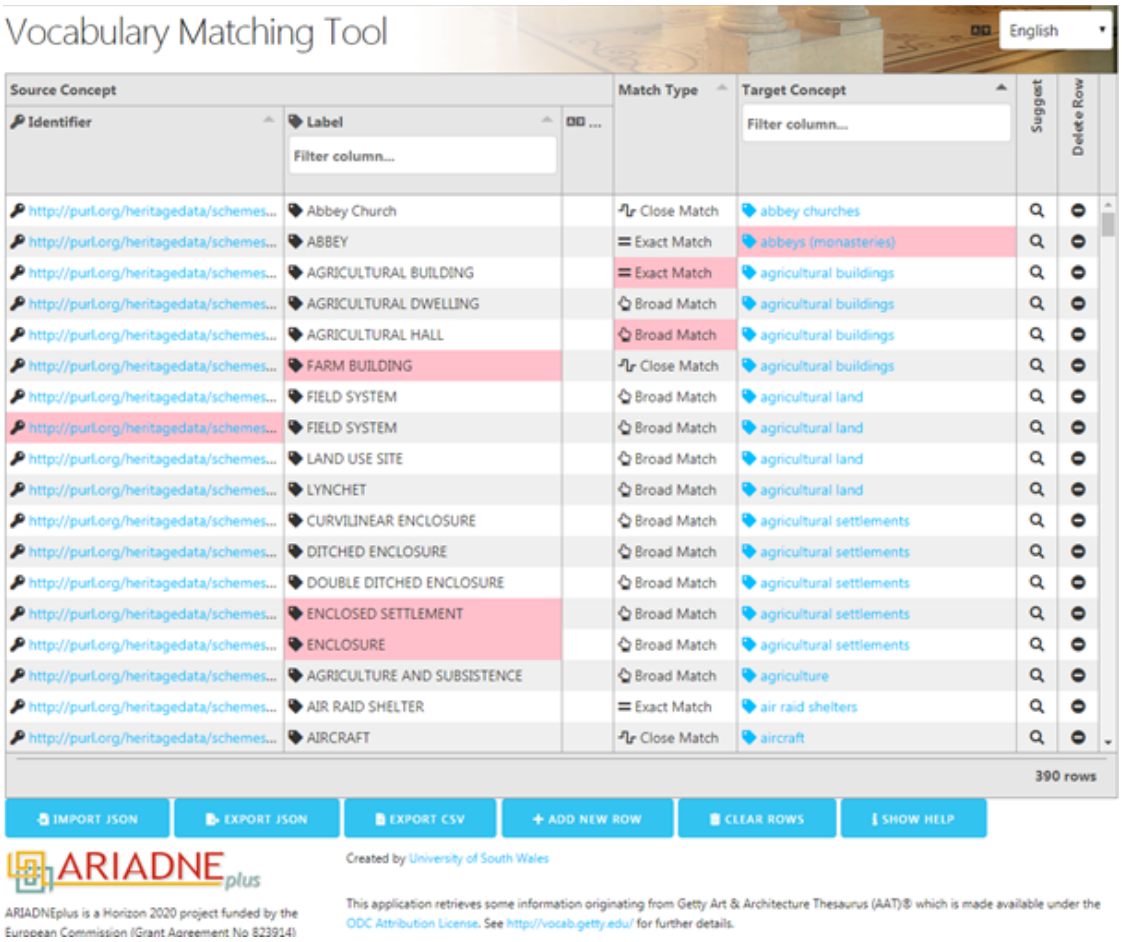
Inference Explicit only

Statements in which the resource exists as a subject.

Predicate	Object
rdf:type	gvp:Concept
rdfs:seeAlso	http://www.getty.edu/vow/AATFullDisplay?find=&logic=AND&note=&subjectid=300211545
dcterms:created	1992-03-05T00:00:00
dcterms:modified	2001-07-26T22:03:06, 2010-04-23T05:52:09, 2010-04-23T05:52:10, 2012-03-14T15:05:29, 2012-03-16T14:11:48, 2014-12-05T08:40:23, 2014-12-05T08:43:10
skos:changeNote	aat_rev:5000052642, aat_rev:5001133770, aat_rev:5001133771, aat_rev:5001133772, aat_rev:5002479845, aat_rev:5002479851, aat_rev:5003174190, aat_rev:5003174193, aat_rev:5003178253, aat_rev:5003763847, aat_rev:5003763848, aat_rev:5003765784
gvp:parentString	brooches, pins (jewelry), jewelry worn on costume, <jewelry by location>, jewelry, worn costume accessories, costume accessories, costume (mode of fashion), Costume (hierarchy name), Furnishings and Equipment (hierarchy name), Objects Facet
gvp:parentStringAbbrev	brooches, pins (jewelry), ... Objects Facet
gvp:displayOrder	2
xl:prefLabel	aat_term:1000211545-en, aat_term:1000538040-es, aat_term:1000611650-nl
xl:altLabel	aat_term:1000226276-en, aat_term:1000294365-en, aat_term:1000447710-es, aat_term:1000611651-nl
gvp:broaderGeneric	aat:300045995
gvp:broaderPreferred	aat:300045995
gvp:prefLabelGVP	aat_term:1000211545-en
skos:inScheme	aat
skos:scopeNote	aat_scopeNote:124494, aat_scopeNote:36428, aat_scopeNote:75533
dcterms:contributor	aat_contrib:10000000, aat_contrib:10000088, aat_contrib:10000131
dcterms:source	aat_source:2000005951-subject-300211545, aat_source:2000051089-subject-300211545
dc:identifier	300211545
dcterms:license	http://opendatacommons.org/licenses/by/1.0/
cc:license	http://opendatacommons.org/licenses/by/1.0/
void:inDataset	http://vocab.getty.edu/dataset/aat
prov:wasGeneratedBy	aat_rev:5000052642

Mappings are inherently reusable as they describe semantic relationships between persistent concepts relating to the domain, but are not restricted in scope to any particular dataset. Therefore if your organization participated in the first ARIADNE project it is likely that a set of subject mappings were produced already, in this case it may be just a matter of locating, reviewing and revising them as appropriate rather than starting again from scratch.

The mappings can be produced using the Vocabulary Matching Tool (VMT). The VMT is a browser-based application with no installation or configuration requirements. It has a multilingual user interface – at the time of writing the current UI languages supported are German, English, Spanish, French, Italian and Dutch. Users can search and browse the AAT, viewing the hierarchical context and scope of concepts to make a better-informed match. Further instructions and general guidance are available online.



Usually you will just make one match from any given local concept to an AAT concept - having mapped to an AAT target concept there is no need for additional mappings to narrower AAT concepts. An exception to this may be the situation where the local source concept is a composite or pre-coordinated concept related to multiple genuinely different AAT concepts e.g. “bronze spearheads”. In this case, two mappings could be made – firstly to aat:300010957 “bronze (metal)” and secondly to aat:300037118 “spearheads”.

The type of match between the source concept and the target concept will be one of the possible SKOS mapping properties as defined by the SKOS reference document, listed here in order of preference:

- Exact Match - this match type indicates that there is "a high degree of confidence that the concepts can be used interchangeably across a wide range of information retrieval applications".
- Close Match - can be used to link concepts "that are sufficiently similar that they can be used interchangeably in some information retrieval applications". It is a more approximate type of relationship between the source and target concepts.
- Broad Match - expresses a hierarchical generic relationship between concepts. If a source concept is more specific in scope than any AAT available concept then you can make a Broad Match link to the AAT concept. This is useful for cases when a source vocabulary contains

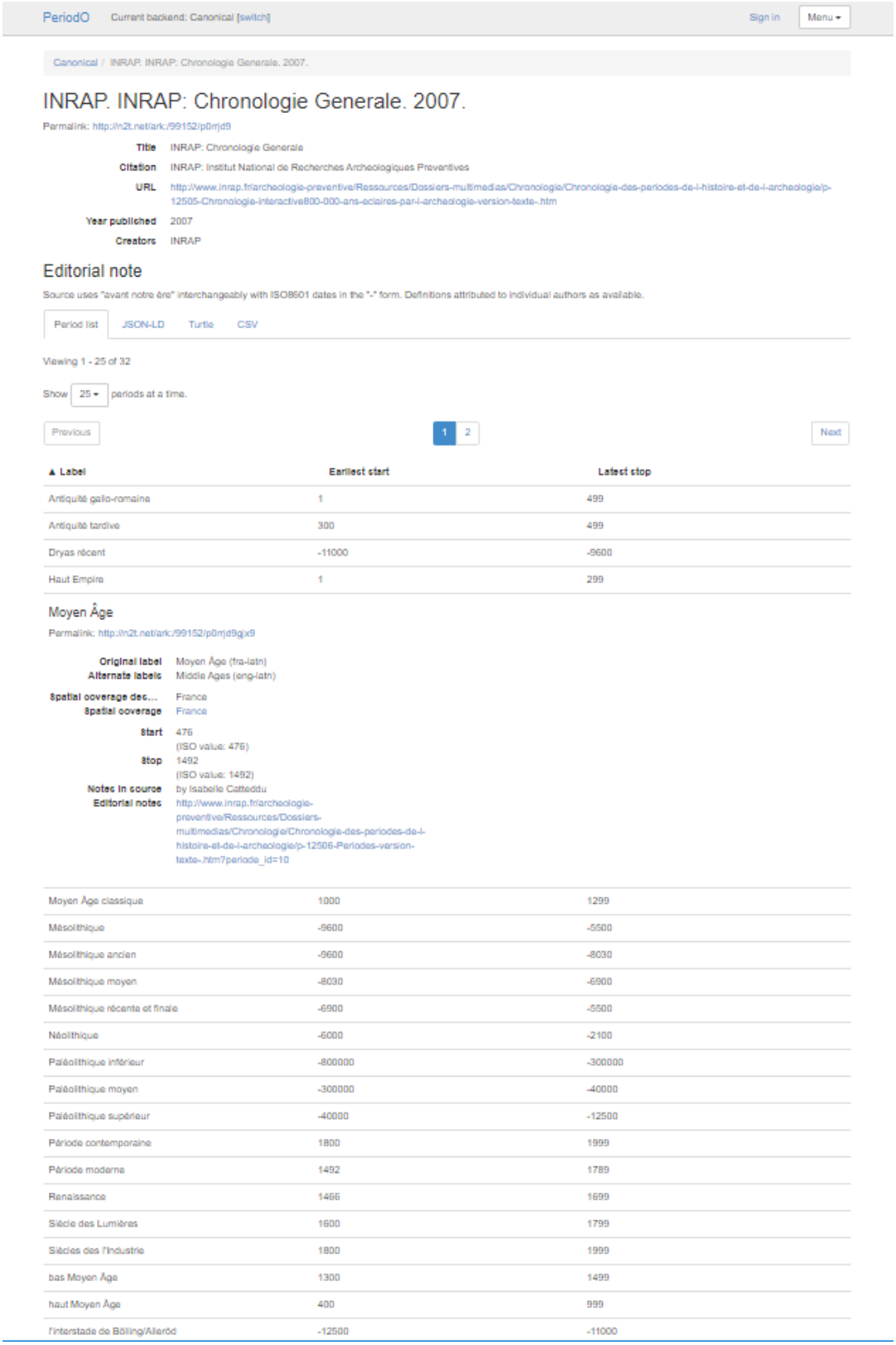
concepts that are more detailed. In this case the "all/some" rule should hold true e.g. "coffee cups" Broad Match "cups" - ALL coffee cups are cups; SOME cups are coffee cups.

- Narrow Match - expresses a hierarchical generic relationship between concepts. It is not expected that you would make much use of Narrow Match relationships for ARIADNEplus vocabularies. The "some/all" rule should hold true e.g. "cups" Narrow Match "coffee cups" - SOME cups are coffee cups; ALL coffee cups are cups.
- Related Match - expresses an associative relationship between concepts. The exact nature of the relationship is not explicitly spelled out, only that there is some see also connection between them. E.g. bullets Related Match guns. Bullets are associated with guns, though you can see that this relationship is different to the other relationship types listed. Preferably you would aim to create a more direct concept mapping wherever the target AAT vocabulary supports it (in this case to aat:300201694 “bullets”).

Once created the sets of mappings are to be exported from the VMT and deposited on D4Science under “VRE Folders > ARIADNEplus_Mappings > Matched Vocabularies”. There is a separate folder per data provider – the mappings will be subsequently picked up from here and used in the data enrichment process.

6.2 Perio.do

Perio.do is a multilingual gazetteer of period definitions for describing historical and archaeological named periods. It comprises a series of 'collections' (lists of named periods) where each individual period definition is associated with a particular geographical area and a start date and end date. Periods and collections have 'permalink' identifiers (URIs) so they can be clearly and unambiguously referenced - e.g. the identifier http://n2t.net/ark:/99152/p0rrjd9gjx9 defines the period "Moyen Âge" (Middle Ages) within the collection http://n2t.net/ark:/99152/p0rrjd9 "INRAP: Chronologie Generale. 2007". All named periods in ARIADNEplus will be defined with reference to periods within Perio.do. This may necessitate creating and submitting a new collection via the Perio.do client application, however If your organization participated in the first ARIADNE project it is likely that a Perio.do collection for the periods defined in your data already exists. In this case it may be just a matter of reviewing and revising them as appropriate rather than starting from scratch. Data providers will inform the data ingestion team of the URI of the Perio.do collection to be used in interpreting named periods in the context of their dataset.



The purpose of using Perio.do in ARIADNEplus is different to the use of AAT data enrichment - rather than aligning data to common identifiers, Perio.do will be utilized in a subsequent data enrichment stage to enrich records already indexed with named periods (such as *“Bronze Age”*, *“Iron Age”* etc.)

where the dates associated with these periods are not already made explicit in the input data - providing start/end dates to make these records comparable and searchable in the same way as other date information. If your data already contains suitable start/end dates then you need do nothing more on this. If your data contains named period terms without any further explanation then read on.

Establish the reference authority in use for the named periods present in your data. This should be an authoritative source - such as an existing published list of period names with associated dates, or a book or other citable publication exhibiting usage of the period terms and describing the dates associated with them. A suitable reference authority may already exist within Perio.do, so the first recommended course of action would be to examine the 'canonical' period list within Perio.do (i.e. the full set of curated period records) for an existing authority list that your data might agree with. If an appropriate entry is found you will need to identify the URI (identifier) of the authority list (the "*permalink*" as mentioned previously). To introduce a slight complication there are (currently) two versions of the Perio.do client available - "*Period Collection*" in the older client application (hosted at https://test.perio.do/) is synonymous with "*Authority*" in the newer client application (hosted at https://client.perio.do/). If using the older client application then select the "*Collection*" tab and click the title of the required "*Period Collection*" – the permalink URI is then displayed in the metadata of the selected collection. If using the newer client application then select the required "*Authority*" to filter the canonical list to terms from that authority only – selecting any of the resultant list of terms will then display the Authority permalink (among other data fields). As an example, the URI of the aggregated list of named periods as used in the first ARIADNE project is http://n2t.net/ark:/99152/p0qhb66. Once the URI has been identified, it will need to be communicated to the data ingestion/aggregation team, as it will be utilized in a subsequent data enrichment stage.

If no suitable authority list for your data is currently present in the Perio.do canonical list then a new list may be created and submitted. Contributing new period information to the Perio.do gazetteer is achieved via the Perio.do client application. If you do not already have one you will need to first set up an ORciD account to identify yourself on Perio.do. The online user guide provides detailed instructions on how to create or edit your own set of named periods, and how to contribute your data to the canonical period list. The new data is submitted to Perio.do using the client application in the form of a 'patch' that will need to be approved by Perio.do administrators prior to merging it into the canonical dataset, so there will be a delay from submission until your period data becomes visible in Perio.do. Once the data is visible you will be able to identify the Period Collection / Authority permalink (URI) that uniquely identifies your data, as described previously.

6.3 WGS84

The World Geodetic System 1984 (WGS84, aka EPSG:4326) is a standard geographic coordinate system (GCS), comprising a global horizontal and vertical datum and a coordinate system used (particularly by GPS systems) to express global positioning on the surface of the Earth. For ARIADNEplus purposes all spatial coordinates are to be expressed using WGS84. Other GCS also exist so coordinates in local datasets may require normalisation/transformation prior to aggregation. Use

of a common GCS to express spatial information will improve opportunities for cross searching the integrated datasets.

WGS84 coordinates are expressed in degrees – the horizontal position is the longitude (having values between -180° and +180°, relative to a fixed datum of 0° at Greenwich, UK) and the vertical position is the latitude (having values between -90° and +90°, relative to a fixed datum of 0° at the equator). As an example, the global location of the Leaning Tower of Pisa (Torre di Pisa) is given by WGS84 coordinates latitude 43.723056 (43° 43′ 23″ N), and longitude 10.396389 (10° 23′ 47″ E).

WGS84 coordinates may be serialized in various machine-readable formats. In terms of the AO-Cat ontological model the coordinate values will be properties of class ao:SpatialRegion and its associated sub-classes (ao:SpatialRegionPoint, ao:SpatialRegionBBox, ao:SpatialRegionPolygon). The AO-Cat properties has_latitude and has_longitude are specified with a range data type of xsd:decimal – so the example (in Turtle RDF format) would be expressed as:

```
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .

@prefix ao: <http://ariadne-infrastructure.eu/ns/> .

@prefix : <http://ariadne-infrastructure.eu/data/> . #this example only


:x a ao:SpatialRegionPoint ;

   rdfs:label "Location of the Leaning Tower of Pisa"@en ;

   ao:has_latitude 43.723056 ;

   ao:has_longitude 10.396389 .
```

7 Aggregation Pipeline Summary

In summary, the pipeline proceeds through the following stages:

1. Get the go-ahead to proceed, according to the workplan
2. Confirm the collections to be aggregated and confirm which ARIADNE fundamental data category they belong to (see Section 3.3)
3. Agree the mechanism for data supply and frequency of update (Section 4)
4. Use the Timespans tool to normalise the dates in your records, if needed (Section 4.3)
5. Map the data to the AO-Cat using the 3M mapping tool (Section 5)
6. Create AAT subject mappings (Section 6.1)
7. Create Perio.do period definitions (section 6.2)
8. Check spatial coordinates comply to WGS84 (Section 6.3)
9. Check that the data displays OK in the test portal
10. Give the green light to publish records on the Ariadne+ public portal
11. Sign it off and have a well-deserved break

All support is managed by the ARIADNEplus helpdesk in D4Science: (https://ariadne.d4science.org/group/ariadneplus_project and click on the tab “HelpDesk”).

Queries will be allocated to the appropriate specialist according to the category of ticket selected.

Appendix 1: Properties of AO_Data_Resource

Domain	Property	Scope note	Range	Optional/ mandatory	One/ many
AO_Enity	has_identifier	an identifier for the resource in some namespace other than the provider's namespace	xsd:string	Optional	one
	has_type	a term from a controlled vocabulary	AO_Concept	Mandatory	many
	has_title	one title identifying the resource in a specific language	xsd:string	Mandatory	one per languag e
	has_description	a free text description of the resource	xsd:string	Optional	one
	collection or individual item?	the nature of the resource, whether it's a collection with multiple items or an individual data resource	yes/no	Mandatory	one
AO_Data_R esource	has_original_id	the local identifier of the resource supplied by the content provider	xsd:string	Mandatory	one
	has_ARIADNE_subject	one the fundamental archaeological categories defined by ARIADNE	AO_Concept	Mandatory	many
	has_native_subject	an original subject in the providing institution	AO_Concept	Mandatory	many
	has_language	the language(s) of the resource	AO_Concept (from lexvo.org)	Mandatory	many
	was_created_on	the date when the AO_Data_Resource was first made available online by the Publisher	xsd:dateTime	Mandatory	one
	has_landing_page	the original landing page	rdfs:Resource	Optional	one

Domain	Property	Scope note	Range	Optional/ mandatory	One/ many
	has_access_policy	the statement of policy (typically, on an organization's website) for the data resource	rdfs:Resource	Optional	one
	has_access_rights	information about who can access a data resource the resource or an indication of the data resource's security status	xsd:string	Mandatory	many
	has_extent	the size of the AO_Data_Resource (i.e., number of members in a collection, number of records in a dataset, etc.)	xsd:string	Optional	one
	has_temporal_coverage	the temporal region covered by the content of the resource	AO_Temporal_Region	Optional	one
	has_spatial_coverage	the spatial region covered by the content of the resource	AO_Spatial_Region	Mandatory	one
	refers_to	the AO_Entity(ies) the resource refers to	AO_Entity	Mandatory	many
AO_Resource	was_issued	the date of formal issuance (e.g., publication) of the resource by the publisher	xsd:dateTime	Mandatory	one
	was_modified	the most recent date on which the resource was modified by the publisher	xsd:dateTime	Mandatory	one
	has_publisher	an agent responsible for making the resource publicly accessible	AO_Agent	Mandatory	one
	has_contributor	the agent primarily responsible for contributing the description of the resource to the ARIADNE Content Cloud	AO_Agent	Mandatory	one
	has_creator	the agent primarily responsible for creating the resource	AO_Agent	Mandatory	one

Domain	Property	Scope note	Range	Optional/ mandatory	One/ many
	has_owner	the legal owner of the resource, who holds the legal responsibility	AO_Agent	Mandatory	one
	has_responsible	the agent holding the scientific responsibility of the resource	AO_Agent	Mandatory	one

Agent

Domain	Property	Scope note	range	Optional/ mandatory	one/ many
Agent	An agent can be a person or a group. It has the following properties				
	has_name	the name of the agent	xsd:string	Mandatory	one
	has_agent_identifier	the identifier of the agent	xsd:string	Optional	many
	has_email	the email of the agent	xsd:string	Desirable	many
	has_homepage	the homepage of the agent	rdfs:Resource	Optional	one
	In addition, if it is a person it also has a mandatory institution				
	has_institution	the institution where a person agent belongs	xsd:string	Optional	one

Spatial Region

Domain	Property	Scope note	range	Optional/ mandatory	one/ many
AO_Spatial_Region	every spatial region has the following properties				
	has_place_name	the name of the region	xsd:string	Mandatory	many

Domain	Property	Scope note	range	Optional/ mandatory	one/ many
	has_coordinate_system	the coordinate system	xsd:string	Optional	one
	and is one of:				
	a point on the surface of the Earth identified by latitude and longitude;				
	has_latitude	latitude	xsd:decimal	Mandatory	one
	has_longitude	longitude	xsd:decimal	Mandatory	one
	a polygon as represented by GIS systems;				
	has_polygonal_representation	polygon as an XML data structure	rdfs:XMLLiteral	Mandatory	one
	a rectangular region on the surface of the Earth identified by its four vertices;				
	has_bounding_box_min_lat	minimal latitude	xsd:decimal	Mandatory	one
	has_bounding_box_min_lon	minimal longitude	xsd:decimal	Mandatory	one
	has_bounding_box_max_lat	maximal latitude	xsd:decimal	Mandatory	one
	has_bounding_box_max_lon	maximal longitude	xsd:decimal	Mandatory	one
	any region identified by a IRI in a standard gazetteer, such as Geonames for modern places, Pleiades for ancient places.				
	has_place_IRI	the IRI of the place	rdfs:Resource	Mandatory	one

Temporal Region

Domain	Property	Scope note	range	Optional/ mandatory	one/ many
AO_Temporal_Region	A temporal region is one of:				
	a temporal period defined in periodO and identified by a IRI				
	has_period	IRI of the periodo in periodO	rdfs:Resource	Mandatory	one
	a concept representing a period in some local vocabulary of the provider				
	has_native_period	concept	AO_Concept	Mandatory	one
	An interval defined by a beginning and an ending time point, both xsd:dateTime				
	from	starting time point of the interval	xsd:dateTime	Mandatory	one
	until	ending time point of the interval	xsd:dateTime	Mandatory	one