D14.1 – Mid-term interim report on the ARIADNEplus knowledge management system

Version 1.0 (final)

27/1/2021

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 - Polo Universitario "Città di Prato"**
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: **Maria Theodoridou**

FORTH

Contributing partners: **Vangelis Kritsotakis, FORTH**

Achille Felicetti, PIN

Douglas Tudhope, USW

Ceri Binding, USW

Edeltraud Aspöck, OEAW

Quality Control **Holly Wright, UoY-ADS**

Document History

- 11.11.2020– Draft Version 0.1
- 10.01.2021 – Draft Version 0.2
- 13.01.2021 – Draft Version 0.3
- 15.01.2021 – Draft Version 0.4
- 27.01.2021 – Final Version

Table of Contents

1 Executive Summary

This deliverable describes the activities carried out during the first half of the ARIADNEplus project within Work Package 14 (WP14 - JRA3 - The ARIADNEplus knowledge management system) by the partners and describes the results achieved by this work package. The report describes progress in:

- Task 14.1 – Monitoring knowledge integration (JRA3.1)
- Task 14.2 – Application profiles (JRA3.2)
- Task 14.3 – Vocabularies and gazetteers (JRA3.3)

In the first half of the project and the interim assessment produced by:

- Task 14.4 – Assessing the CRM extensions (JRA3.4)

The overall objectives of WP14 are presented in Section 2 while Sections 3, 4, 5 and 6 present the detailed work in Tasks 14.1, 14.2, 14.3 and 14.4 respectively. All sections follow the same structure providing an overview, the actual work achieved by the task, the deviation (if any) from the work plan and the plans for the next period. Work in WP14 is closely related to other work packages and more specifically to WP4 where the ARIADNEplus ontology and specific extensions, the application profiles, to specific sub-domains of archaeology and archaeological science, are being implemented. *Deliverable D4.2 - Initial report on ontology implementation* was submitted previously and provides a detailed description of the AO-Cat (the ARIADNE ontology) and the progress on the application profiles. This work is closely related to Task 14.2, so direct references in D4.2 will be provided whenever needed.

2 Introduction and Objectives

The objectives of WP14 - JRA3 - The ARIADNEplus knowledge management system centre on four main directions:

- define the ARIADNEplus semantic framework;
- define and build the ontologies required for ARIADNEplus;
- analyse, adapt and, if necessary, build the required Knowledge Organization Systems;
- assess the ARIADNE ontology extensions, and validate the work done in this work package, and in WP4 where implementation takes place.

This report presents detail about the activities carried out during the first half of the ARIADNEplus project, in each Task of the work package.

Section 3 presents the activities of Task 14.1. The objective of Task 14.1 is to monitor the aggregation process. In this direction, we designed and implemented Activity Dash, a dashboard that will support monitoring the aggregation process. Activity Dash supports a group of users to keep track of the whole aggregation process. It provides information about which resources have been uploaded/mapped/transformed and when. It also provides a collaborative environment allowing multiple users to work within the same workspace simultaneously.

Section 4 presents the activities of Task 14.2. The primary purpose of task 14.2 is the coordination of activities related to the definition of application profiles. All the activities of this task have been carried out in close collaboration with Task 4.4. The results of these activities, including the requirements and the application profiles that have been defined up to this stage of the project, have been presented in detail in the D4.2 report ("Initial report on ontology implementation"). In Section 4 of this report we provide a general overview of the models being developed, referring to the specific sections of D4.2 for more detailed information.

Section 5 presents the activities of Task 14.3 on vocabularies and gazetteers. Work on multilingual vocabularies overlaps to some extent with work for WP5 on vocabulary integration via mappings from Partner vocabularies to the Getty Art and Architecture Thesaurus (AAT). These mappings contribute to a multilingual vocabulary resource compiled from different sources. Section 5 presents the work that has been done to extract, convert and combine multilingual resources from Wikidata, Getty AAT and other relevant sources. Moreover, we present temporal integration via the Perio.do framework for temporal periods. The work on vocabulary integration is detailed in the User Manual for the ARIADNEplus Data Aggregation Pipeline and D5.2 first report.

Finally, section 6 presents the interim assessment of the AO-Cat, the application profiles and the CRM extensions produced by of Task 14.4. All sections follow the same structure providing an overview, the actual work achieved by the task, deviation from the work plan (if any) and the plans for the next period.

3 Monitoring knowledge integration (Activity Dash)

3.1 Overview

Activity Dash is an online web application, accessed through any web browser, to track the several processes (activities) of a workflow that might or might not be executed in certain order. End users of the system can monitor the progress of one or more workflows, and update them if necessary. In general, the application combines:

i) a workflow manager and tracker;
ii) a collaborative environment;
iii) a notification means for stakeholders;
iv) a reference point for stakeholders, to monitor workflow progress.

Activity Dash serves the needs of many different projects, each of which might have many workflows, making it a generic tool suitable for serving different needs and targets. Furthermore, the system can be used by many users simultaneously, working within different or the same workspaces, applying changes and receiving updates from other users.

All workflows within the system are grouped by project, where access is regulated by user groups, defined by some project owner user. Following the same pattern, each workflow within a project has its own user group defined by the workflow owner. Projects and workflows are managed by the respective owners in terms of metadata they hold and user accessibility.

The workspace environment offered by the system, presents the workflow divided into activities, which are further divided into tasks. The workflow owner assigns users responsible for activities, and each activity assignee, assigns other users responsible for tasks. All stakeholder users can simultaneously edit or extend the workflow, without worrying about missing potential changes, being made by others within their shared workspace. This is mainly achieved, due to the system's capability of syncing the shared workspace view, among all users behind different browsers.

Activities within the workflow can also hold associations. This is mainly done when some activity depends on the completeness of other activities. The same can also happen between tasks of a single activity. The system is capable of automatically informing stakeholders, under specific circumstances, mainly about the completeness of some activity (or task respectively) or when a new assignment or dismissed assignment has been implemented.

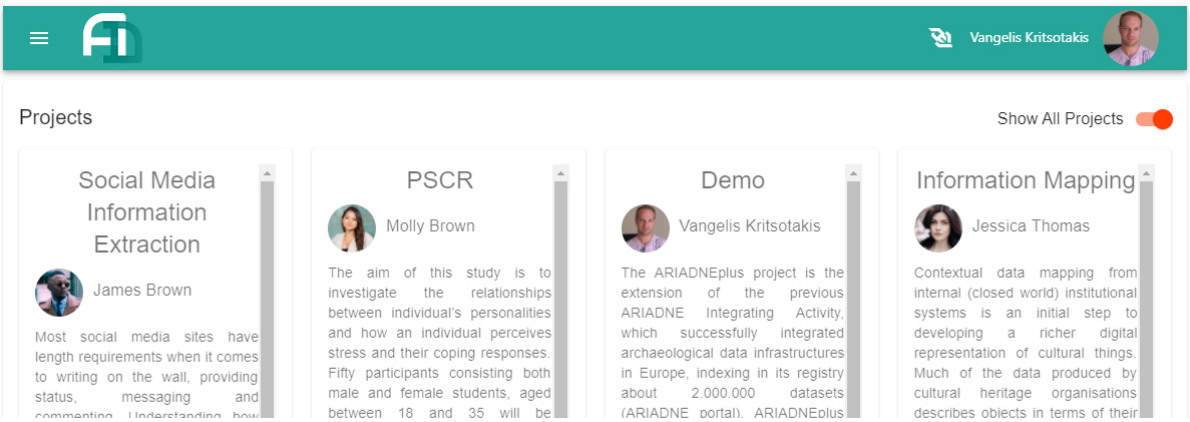

Figure 1: Activity Dash – Projects.

3.2 Abstraction, Templates and Organisation Schema

The Activity Tracking System (ATS) was designed to be as abstract as possible in terms of the activities to be tracked. The "Activities" and "Tasks" constituting the full workflow, have a range of dynamic attributes, but use the same structure and format, following a common generic schema. Moreover, it is important to reuse existing schemas, which is achieved by using existing instances of "Workflows" as templates for the creation of new instances. The metadata fields of each instance activity unit include the following:

Name	Description
Task	
Name	Free text
Description	Free text
UsersAssigned	User responsible for this "Task"
PercentageOfCompleteness	Integer with range from 1 to 100
LastModified	Date it was last modified
CurrentStatusDescription	Free Text
ListOfcompletedTasksRequired	List of task Objects required to be fully complete before it can be worked on
Sequence	Integer defining the sequence index of this task inside the "Activity" (using the same index for more than one tasks, indicates that those tasks can be applied in parallel)
Activity	
Name	Free text
Description	Free text
UsersAssigned	User responsible for this "Activity"
Tasks	List of Task Objects
PercentageOfCompleteness	Integer with range from 1 to 100 to be auto-generated by the list of "Tasks"
ListOfcompletedActivitiesRequired	List of activity Objects required to be fully complete before it can be worked on
Sequence	Integer defining the sequence index of this "Activity" inside the project (using the same index for more than one "Activity", indicates that those processes can be applied in parallel)
Workflow	
Name	Free text
Description	Free text
Owner	User responsible for this "Workflow"
Activities	List of "Activity" Objects, constituting this workflow
PercentageOfCompleteness	Integer with range from 1 to 100 to be auto-generated by the list of "Activities"
Complete	Boolean value denoting that the "Workflow" has been fulfilled
Project	
Name	Free text
Description	Free text
Owner	User responsible for this "Project"

Activity dash organises everything, according to a specific schema that can easily group workflows, activities and tasks respectively.

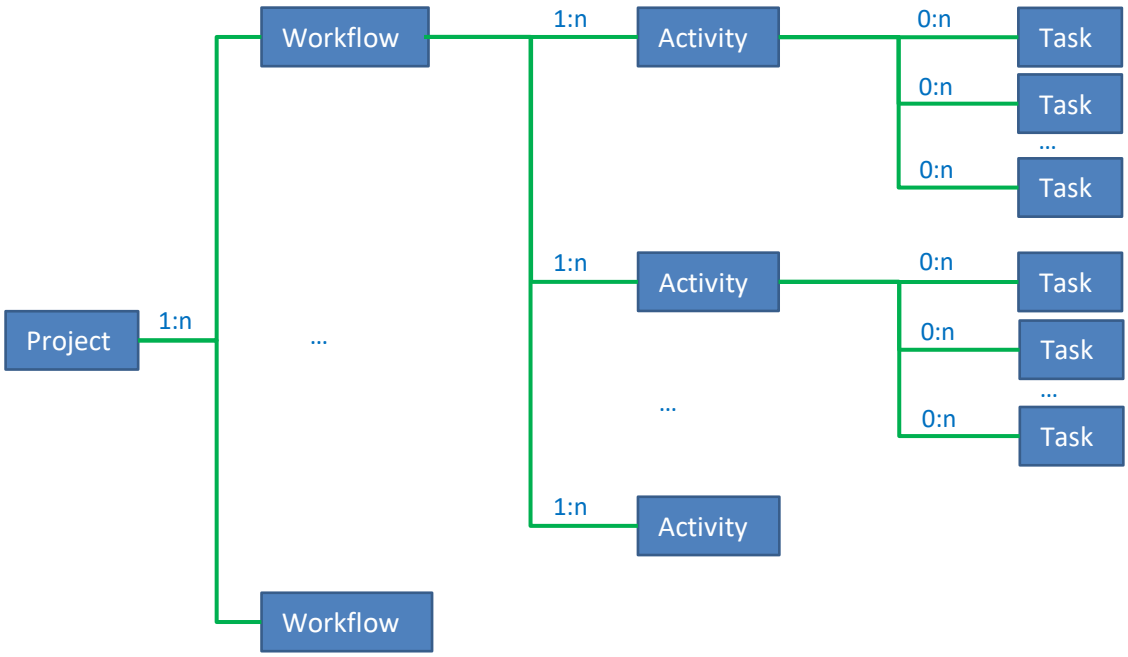

Figure 2: Organizational Schema.

The top root category, named “Project”, is used for grouping a set of “Workflows”, each of which consists of “Activities”. “Activities” are the steps to be accomplished within a “Workflow”. Finally, each “Activity” consists of smaller units, which are the “Tasks”, representing the several operations through which an “Activity” can be divided. The relation between any of them is 0:n, meaning that each “Project” might have zero or more “Workflows” and each “Workflow” might have zero or more “Activities” and each “Activity” might have zero or more “Tasks”.

3.3 Functional model

Activity dash combines an information manager/tracker system, within a multi-user collaborative platform. This chapter describes the functional model of the system.

Security	
S01	The system is secured by a powerful and highly customisable authentication and access-control framework, which is mainly used for authenticating users, and granting access to URLs (services and views), according to their roles.
S02	The JSON Web Token (JWT), is used to provide a compact and self-contained way to securely transmit information between parties, with optional signature and/or encryption.
S03	Role Based Access Control (RBAC) and Data Ownership are combined to provide the required user permissions and regulations.
S04	The system provides a registration mechanism for the new users, which also requires administrator approval before gaining access.

Environment	
F01	The system is accessed through the Web using any modern browser.
F02	Provides a simple, easy and friendly environment to the user.
F03	Provides good performance and makes use of web caching, where appropriate.
F04	Provides a dynamic UI environment, which depends on user roles and ownership.
F05	The system is fully configurable, providing better flexibility.
F06	The system is fault tolerant, which means it is capable of properly handling errors and remaining functional after any error recovery.
F07	The system notifies the user if a functional error occurs.
Administration	
A01	The system allows users to manage their own information.
A02	The system provides full user administration management.
A03	The system provides full workflow administration management.
A04	The system provides full notification messages for administration management.
A05	The system organises the workflows into projects, which are also managed by the administrator.
Workspace	
W01	The system allows end-users to define their own preferences in certain functionality or parts of the system.
W02	The system presents workflows as a set of activities, which may or may not be divided into tasks.
W03	The system allows users to create relations between activities within the same workflow, or tasks within the same activity in the form of dependencies.
W04	The system allows users to simultaneously work within the same workspace, editing or expanding some workflow, which is automatically synced by the system.
W05	The system allows users to clone workflows or tasks.
W06	The system provides a silent mode, which pauses all notifications relating to the completeness of an activity or task, or due to the new task assignment/dismissal (very convenient after cloning and then editing a workflow).
W07	The system always asks for user permission before applying actions that cannot be undone.
W08	The system helps users avoid making mistakes, and provides directions and tooltips guiding end-users towards the proper usage.

Notifications & Messages	
N01	The system provides user notification, which is used either internally by the system or users.
N02	The system provides two different means of notification, via either internal notification messages, or emails (or both).
N03	Users can provide feedback on activities by others, by exchanging short messages or whole conversations.
N04	The system automatically informs stakeholders when dependent activities are 100% complete within a workflow.
N05	The system automatically informs stakeholders when they are assigned or dismissed from an activity or task.
N06	The system provides feedback to the end-user for every action.

3.4 Security and Access Control

Activity Dash provides its own powerful and highly customisable authentication mechanism, which can be used to set up the system as a standalone application, however it is easily configurable to support integration with third party applications that provide their own authentication. In either case, the JSON Web Token (JWT), is used to provide a compact and self-contained way to securely transmit information between parties, with optional signature and/or encryption. This is achieved by encoding data into a JWT and sending it to the Client, which then is attached on every protected request made. The Server will validate that JWT on every such request and return the Response.

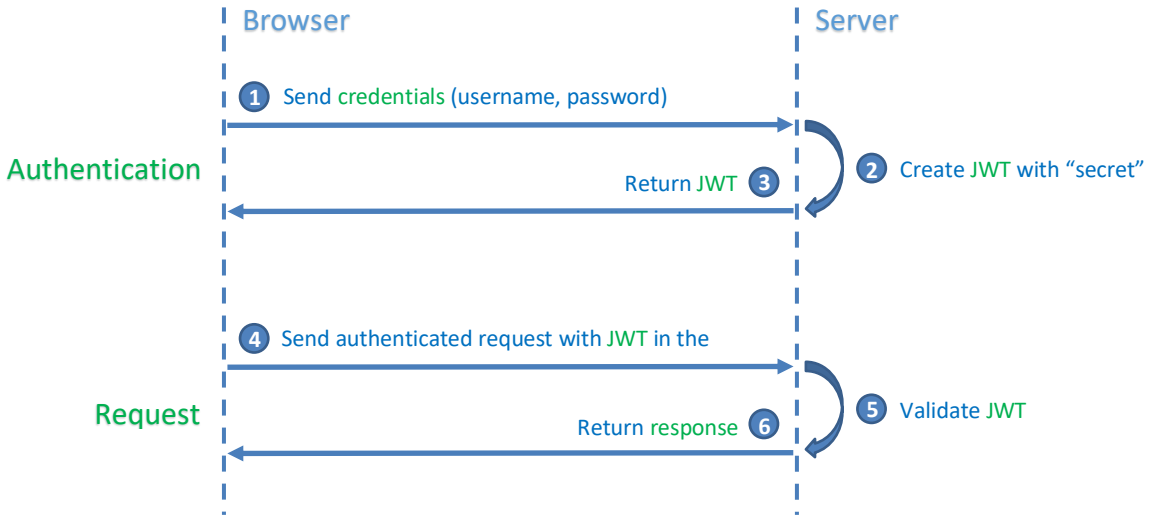

Figure 3: Securely transmitting information between parties with JWT.

In terms of regulating user actions, Activity dash combines both Role Based Access Control (RBAC) and Data Ownership restrictions, to achieve an optimal result. Users are classified in two categories, the administrators and the normal users, according to their roles, but as ownership regulations have their own section within the system, the following set of rules are applied:

- The administrator user is able to manage projects by creating new projects and assigning owners to them, or deleting and editing existing ones

- The administrator user and the project owner is able to edit the project, add workflows within it and assign an owner for each workflow (workflow management within the project)
- The workflow owner is able to edit his/her workflows (not add new or delete)
- The workflow and project owners are able to edit the sorting of activities and add new ones
- The workflow and project owners and also the activity assignee are able to edit the activity
- The workflow and project owners and also the activity and task assignees are able to edit the task
- Users can only see projects where they are members (apart from users with the role of administrator, that can access all projects)

3.5 Architecture & Implementation

Activity Dash is a single page application that follows the classic Client-Server Architecture. It consists of three main components, which are the front end, back end and database. The front end is implemented in React[1], the back end in Spring Boot[2] and the database is a NoSQL MongoDB[3].

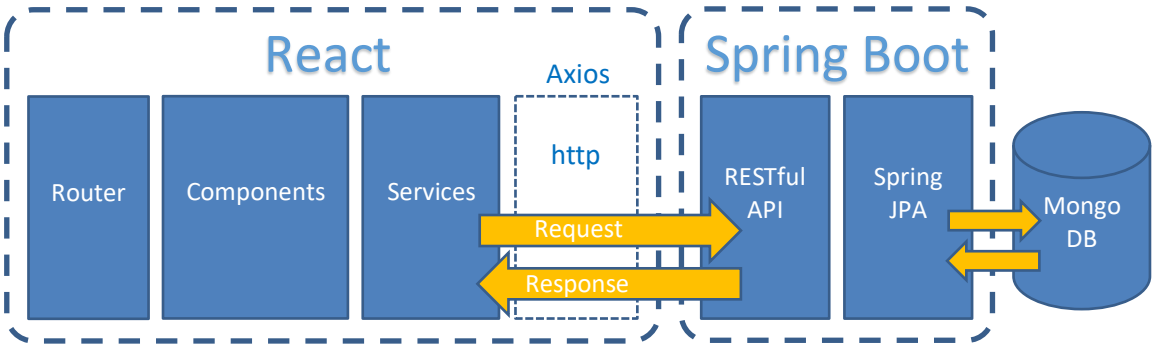

Figure 4: Architectural Design Diagram.

3.5.1 The front end

The front end has a router, which provides the navigation rules. It includes several re-usable components and has a set of services that communicate to the server through "Axios[4]" HTTP requests. Special emphasis is given to the way rendering is controlled through that layer, avoiding re-rendering unnecessary components, and thus boosting performance.

The technology used for the front end is React, which is an open-source JavaScript library, used to build user interfaces for single-page or mobile applications. It is maintained by Facebook and a community of individual developers and companies. The main component library used is Material-UI[5], however many other libraries were used to supplement Material-UI as well.

[1] https://reactjs.org/
[2] https://spring.io/projects/spring-boot
[3] https://www.mongodb.com/
[4] https://github.com/axios/axios
[5] https://material-ui.com/

3.5.2 The back end

At the back end, there is a RESTful API, which consists of several micro services capable of reading and responding to client requests. These micro services use the Spring JPA to interact with the database (retrieve or save data). The server side is also responsible for much more, like encapsulating anything to do with the security of the system (i.e. user authentication, encoding passwords), or storing files within the file system (i.e. user avatar images). Furthermore, the back end is responsible for communicating with third party applications through external RESTful APIs, and achieving proper integration with them. This has already been used to deliver user authentication performed on external platforms.

Finally, the synchronisation of the views of the active users that work within the same workspace is also performed in the back-end. This is achieved by utilising WebSockets[6], that provide a bi-directional, full-duplex, persistent connection between the web browsers and a server. The WebSocket functionality offered by the server uses the Simple Text Oriented Messaging Protocol (STOMP[7]), providing an interoperable wire format between clients and any STOMP message broker.

Spring Boot is used for the development of the back end, which is an open source Java[8] based framework used to create high performing, stand-alone and production ready applications, based on the Spring Framework[9]. This is an application framework and inversion of control container for the Java platform. Spring framework makes development easier, as it promotes good programming practice by enabling a Plain Old Java Object (POJO) based programming model. Finally, Spring Security[10], which is a powerful and highly customisable authentication and access-control framework, was used to secure the application.

3.5.3 Database

The database is the persistence layer of the application, used as data repository to hold information generated or provided by users. Credentials in the database are stored in an encrypted form, by using a one-way encoding algorithm.

[6] https://developer.mozilla.org/en-US/docs/Web/API/WebSockets_API
[7] https://stomp.github.io/
[8] https://www.java.com/
[9] https://spring.io/
[10] https://spring.io/projects/spring-security

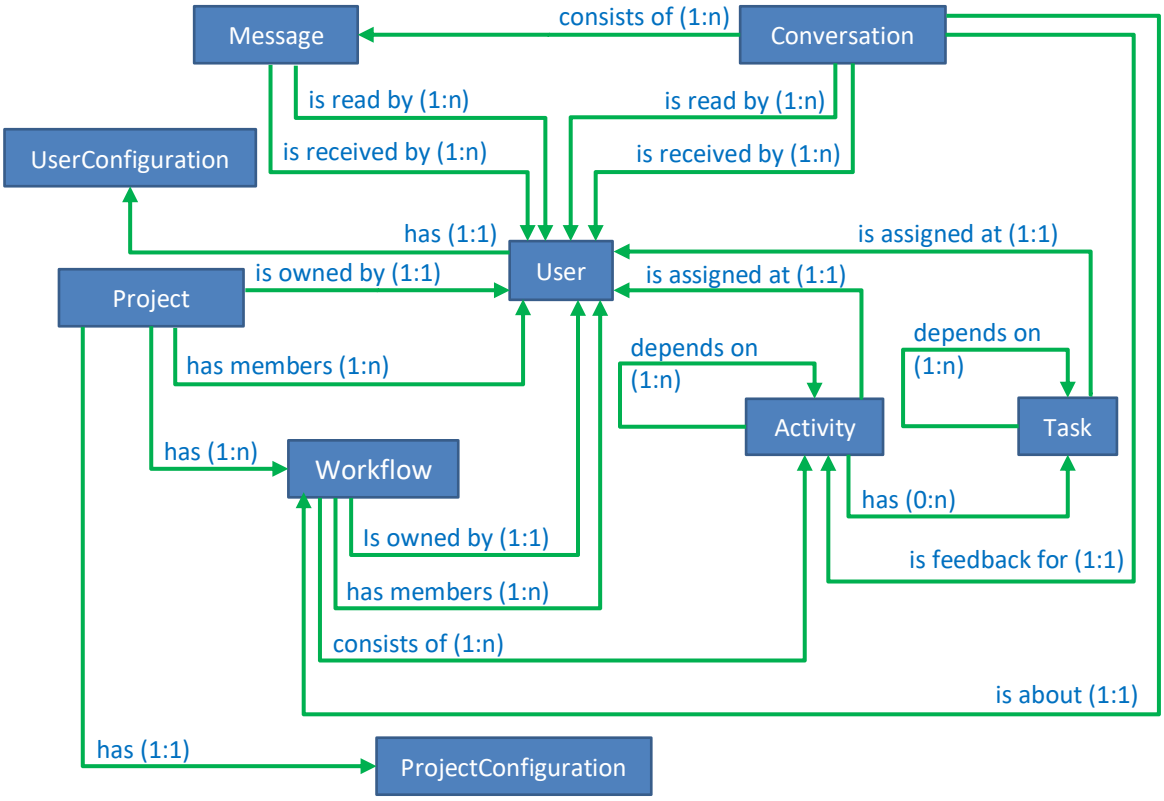

Figure 5: Entity-Relationship Diagram.

Figure 5 presents the database schema of the system. It shows the relationships between entities with respect to their cardinalities. As this is a document database, there are many entity objects natively referenced by other entity objects (including two cases of entities that refer to themselves), the back-end regulates what is to be returned to the client by making use of JSON[11] views, which control serialisation and deserialisation of the objects. In that way, it is easy to exclude properties from the returned object and avoid infinite recursion issues or constructing objects which are too large that are not really needed.

The database used is MongoDB, which is an object-oriented, simple, dynamic, and scalable NoSQL database, based on the NoSQL document store model. The data has the form of JSON objects, and is stored as separate documents inside a collection.

3.6 Integration, testing and validation

The first version of Activity Dash was released on M19 and has been successfully integrated into the D4Science platform. Integration has been implemented in the authentication level, which means that users can login using their D4Science account and have access to Activity Dash without the need for

[11] https://www.json.org/json-en.html

further authentication. Since the functionality of Activity Dash is significantly dependent on end-users (ownership, assignments, memberships, messaging), users that access Activity Dash from within D4Science, are automatically checked. Users that have an account in D4Science but do not exist in the internal database of Activity Dash are automatically added with their D4Science credentials. To accomplish this, Activity Dash makes use of a restful API, provided by D4Science, through which user profiling information is made available. That information is then used to register D4Science users into Activity Dash.

As a first testing and validation step, Activity Dash was made available to a restricted group of users, which spent time testing it by constructing workflows that cover real use case scenarios. During that phase, valuable feedback was gathered and used to optimise the system, either by applying changes to it or by expanding it to provide new functionality. In certain cases during this phase, hidden bugs were uncovered, the resolution of which contributed to making the system more robust and stable.

3.7 Deviation from work plan

There was no deviation from the work plan. However, there was some slight delay in making Activity Dash available to all the end-users. Currently, only a limited group of users have taken part in testing and validation as described in Section 3.6 above. Official release of Activity Dash to all users is planned for February 2021 (M26 of the project).

3.8 Plans for the next period

The main development of Activity Dash is complete and AD v1.0 has been made available to a small group of people for testing and collection of feedback as described in Section 3.6. Testing and debugging continues and further validation will be applied, under real use conditions. An official workshop to present Activity Dash to all interested partners has been scheduled for February 10, 2021. During the workshop, FORTH will present the tool and provide a live demonstration of its usage. Following the workshop, the tool will be used by all involved partners to fully track their activities and tasks. During that phase, FORTH will continue collecting requirements or changes that need to be applied. Debugging, refactoring and user support will be a continuous process in the next period.

Moreover, areas where the system could be improved or optimised have already been identified, along with areas for expansion and new functionality, missing from the current version.

Activity Dash version 2 is expected to be available by the end of the third project year (M36).

4 Application Profiles

4.1 Overview

The primary purpose of Task 14.2 is the coordination of activities related to the definition of application profiles. All the activities of this task have been carried out in close collaboration with task 4.4. Application profiles can be seen as sets of entities used to model the data of specific research domains. The classes and properties that constitute each application profile are usually borrowed from existing ontological models or derived from them, and are typically arranged in coherent structures so as to keep compatibility with the AO-Cat general model and the CIDOC CRM, which is the reference ontology of ARIADNEplus.

The work of creating application profiles has been delegated to 14 special interest groups charged with defining the specific requirements of each sub-domain of task 4.4, designing the classes and properties required for the modelling of the specific entities of each research discipline. The groups were also in charge of surveying, collecting, and managing multilingual domain thesauri and vocabularies necessary for describing domain-specific information.

The results of these activities, including the requirements and the application profiles that have been defined up to this stage of the project, have been presented in detail in the D4.2 report ("Initial report on ontology implementation"). The following sections essentially provide a general overview of the models being developed, referring to the specific sections of D4.2 for more detailed information.

4.2 Heritage Science Application Profile

The development of an ontological model for the encoding of information produced by heritage science was one of the main activities of task 14.2. Heritage science is an interdisciplinary branch of research which uses scientific activities to support conservation, access and interpretation of cultural heritage and derives its methods from a number of fundamental sciences, such as physics, chemistry and biology. Activities in this field typically consist of research and analysis intended to provide a better knowledge of objects, materials, artefacts and artworks of cultural and historical significance, and a deeper understanding of their physical composition, dating and provenance.

To define the necessary user requirements for the construction of the application profile for ARIADNEplus, a transversal scientific data task force was created with partners from both WP4 and WP14, including INFN, CyI, HNM, PIN, OEAW and LNEC. Their activity resulted in the definition of suitable scenarios for the scientific domain and the first draft version of the CRMhs model, which is now available as a well-documented and complete CIDOC CRM and AO-Cat compatible ontology. The document describing the scenarios and needs of the scientific community, produced by the task force, has also been published within ARIADNEplus to be used as a template for gathering requirements and defining scenarios for other application profiles.

CRMhs consists of a set of classes used for distinguishing and defining each of the entities involved in a specific scientific analysis, and of a set of relationships used to link these entities, according to the

specific sequences of events in which they are involved. The model results from the harmonisation of different existing ontologies that already excel in modelling general aspects related to research activities, and provides new entities specifically designed for heritage science, in order to represent the specific aspects of this domain.

A modelling example of a heritage science analysis using CRMhs entities is presented in Figure 6.

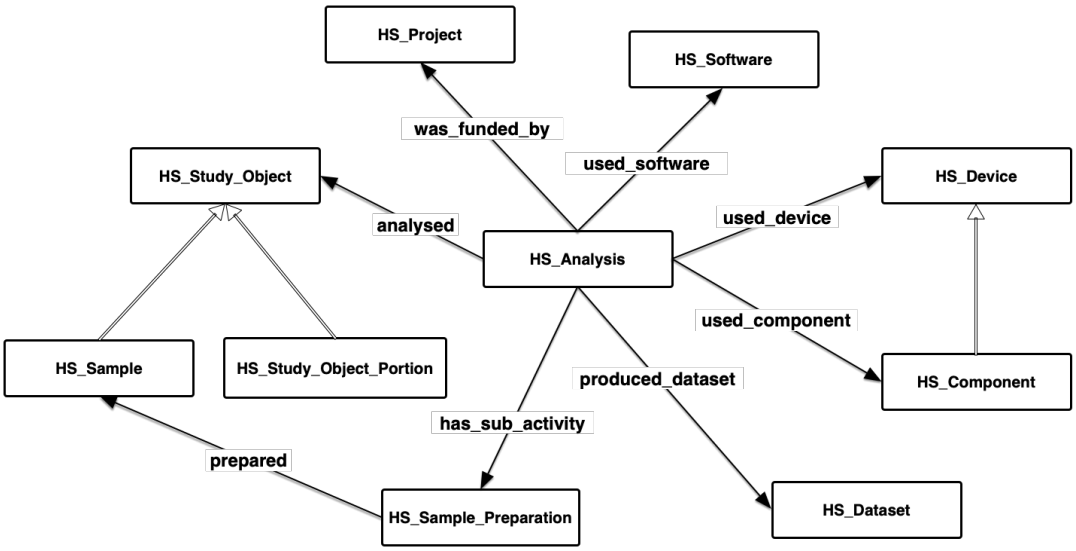

Figure 6: CRMhs description of analysis activities.

An activity of alignment of CRMhs with the model proposed for bioarchaeology and ancient DNA, described in the following section, is planned as part of a complete framework for the description of all data related to scientific activities with which ARIADNEplus aims to integrate.

CRMhs shares the data modelling philosophy of CIDOC CRM and the FAIR principles for accessibility, integration and (re)usability of digital data, which have been used consistantly to both outline the conceptual basis of the model and build its entities. The complete definition of the CRMhs ontology is presented in the deliverable D4.2, Section 5 ("Case Study 1: CRMhs and scientific data") and all its classes and properties are described in detail in the Appendix ("Classes and properties of the CRMhs model") of the same document.

4.3 Application Profile for ancient DNA

To define the necessary user requirements for the construction of the application profile for ancient DNA for ARIADNEplus a model was defined to describe ancient DNA wetlab methodologies. We analysed projects currently running in the aDNA laboratory facilities of FORTH-IMBB. Each project was described as an AO_Collection since it is an aggregation of resources. The properties of AO_Collection were found to be sufficient to describe aDNA projects at a level of high abstraction. In order to describe the aDNA wetlab services in more detail, classes and properties were used from the CIDOC-CRM family of compatible models:

1. CIDOC CRM – The base model, version 6.2.1
2. CRMsci – Scientific observation model, version 1.2.2
3. CRMdig – Model for provenance metadata, version 3.2
4. CRMarchaeo – Excavation model, version 1.4.1
5. CRMpe – Model for Research Infrastructure management, the PARTHENOS Entity Model version 3.1.2

As already mentioned, activities were undertaken in the context of an existing project with responsible actors and a description of its goals and where the use of a controlled vocabulary to classify the projects could be very helpful. The project was modelled as a PE35_Project and supports a number of different activities. A core activity is the acquisition of the samples to be used in the experiments. An acquisition activity (E8_Acquisition) consists of one or more sample taking activity (S2_Sample_Taking) that deals with each distinct sample. For each sample basic information was recorded about the archaeological excavation within which the sample was found. The sample itself is modelled as S13_Sample and also, in the case of aDNA, as E20_Biological_Object.

The archaeological excavation is modelled as an A9_Archaelogical_Excavation class with basic information about the archaeological excavation, including the responsible entity, the broad location where it took place, the time period to which it refers, and the findings of interest from where the sample(s) are taken.

Having acquired the appropriate samples there are a number of steps that are followed in a particular protocol. Each step was modelled as an S4_Observation, along with any associated basic information regarding the person who processed the sample, the place where it was processed, the date, the protocol upon which it is based, any input parameters in use and of course, the results.

A tentative model for bioarchaeology and aDNA application profiles, along with a specific use case, is presented in D4.2, Section 6 (Case Study 2: Bio-archaeology and Ancient DNA).

The aDNA modelling was undertaken in parallel with the development of CRMhs. A full alignment with CRMhs is planned, in order to create a complete framework for the description of all data related to scientific activities in ARIADNEplus. A very first harmonisation attempt was done without encountering any alignment problems and/or incompatibilities, and is presented in D4.2, Section 6 (Case Study 2: Bio-archaeology and Ancient DNA).

4.4 Application Profile for Inscriptions

The definition of an application profile for data relating to inscriptions started with a preliminary analysis of existing models within this domain, investigating in particular the possibility of reusing some of their entities and adapting them to the context of epigraphic data, and other similar information present in ARIADNEplus. This mainly includes inscriptions, stamps, graffiti and other physical features usually found on portable objects, such as coins, vases, weapons, jewels and other artefacts that are coming from archaeological contexts, museums or collections, or found on architectural objects (for example, monuments, buildings or building sections) or in natural contexts such as rocks, caves and caverns.

As a result of this preliminary activity, two models were chosen as the basis for the construction of the application profile for inscriptions and the modelling of its various features:

- CRMtex, an extension of CIDOC CRM specifically created to model ancient texts and other semiotic features appearing on inscriptions, papyri, manuscripts and other similar supports, and to describe the phenomena related to the production, use, conservation, study and interpretation of textual entities[12].
- EPNet, a CIDOC CRM based ontology designed to deal with inscriptions, events and objects connected with the distribution of food in the Roman world[13].

Both these models proved to be widely compatible with the AO-Cat ontology and able to ensure persistent interoperability with the data encoded by means of the other components of the ARIADNEplus Ontology. For instance, elements such as the physical carrier, of particular importance both in CRMtex and in EPNet and represented as the bearers of inscriptions, stamps or graffiti, constitute fundamental contact points with the ARIADNEplus ecosystem, being very often archaeological objects and, as such, suitable to be modelled using AO-Cat and CRMarchaeo for their discovery, provenance, archaeological context and so on.

Errore. L'origine riferimento non è stata trovata. shows an example of possible interoperability between archaeological and textual entities within ARIADNEplus.

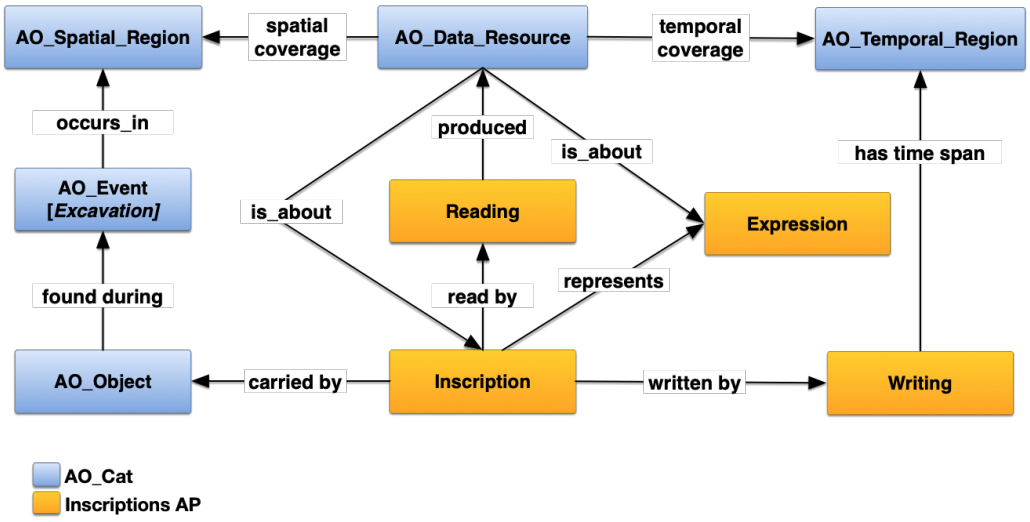

Figure 7: Overview of the interactions between the main classes of the application profile for inscriptions and the AO-Cat model.

[12] CRMtex documentation available at http://www.cidoc-crm.org/crmtex/

[13] More information on EPNet is available at http://romanopendata.eu/sparql/doc/index.html

The work of selecting and adapting the CRMtex and EPNet entities to the ARIADNEplus context is already in process. The requirements, scenarios and the first draft of the possible classes and properties to be developed are fully documented and presented as a case study in D4.2, Section 7 ("Case Study 3: Inscriptions"). In the next period it is planned to finalise this activity in order to produce an application profile for inscriptions that is complete and fully integrated within the ARIADNEplus Ontology.

4.5 Application Profile for mortuary data

Mortuary archaeology consists of a series of research activities and analyses carried out either directly on archaeological evidence containing human remains or on contexts that are interpreted to relate to the disposal of the dead, along with documentation and finds from such contexts. The aim is to acquire information concerning many aspects of past societies, such as disposal of the corpse, past funerary practices, identities, migration, and social composition. The first step to develop an Application Profile for Mortuary Data was to analyse partner data. Mortuary data related to three different stages of the research process have been identified:

1. datasets generated in the field with no or only limited post-excavation analysis;
2. datasets generated as results of workflows based on fieldwork documentation (digital datasets, or, in previous times analogue documentation) and analysis of physical objects (human remains and objects found with them, as well as samples taken from both), typically concerning data from one cemetery;
3. datasets that synthesise/aggregate mortuary data, either presenting an overview of a certain period and region, or dedicated to answering specific research questions.

Ad 1: Field data from the mortuary domain can be treated like a fieldwork archive. They contain mainly collection-level metadata and a description by AO-Cat classes and properties is sufficient.

Ad 2 and 3: On the collection level, i.e. for each site/cemetery, data will be mapped like site and monument data. This will cover information about the geographical extent of a cemetery or other mortuary-related monument such as a mound, its data range and some categorisation by using the AAT (e.g. cremation cemetery). Some of the item-level records may also be described in AO-Cat, for example, a grave with its micro-coordinates and specific data range.

However, for a more detailed description of mortuary deposits (data on human remains, position of the human remains, finds and grave structures) at the item-level, a mortuary data AP will have to be developed. Mortuary data has links with data from other APs, most importantly bioarchaeology, finds and scientific dating. For the development of the mortuary data AP classes and properties were used from CIDOC-CRM and the following extensions:

1. CIDOC CRM – The base model, version 6.2.1
2. CRMsci – Scientific observation model, version 1.2.2
3. CRMarchaeo – Excavation model, version 1.4.1

For more information and a list of classes and properties that were necessary for the development of the mortuary data AP see Section 6., Assessing the CRM extensions, below.

4.6 Other Application Profiles

The other application profiles listed in 4.4 subtasks are still under definition as the activities of the prior period mainly concerned the collection of requirements, and the definition of application scenarios and domain research questions, a very complex and delicate phase which constituted the bulk of the development work. A series of surveys was also conducted in many subtasks to improve knowledge about the availability of data, and to appropriately define related users' needs.

Specific working groups, composed of ontology experts and scholars from the various domains, were each set a subtask to design the most appropriate tools to manage the complexity of the data for each discipline, while maintaining compatibility with AO-Cat and the other components of the ARIADNE Ontology.

As in the case of the models presented above, the details of each application profile under preparation are reported in D4.2, Section 4 ("Application profiles and other compatible models"), therefore only a quick overview of the models being prepared and some general considerations about them is provided in this section.

For some disciplines, application profile development was deemed essential to describe domain information, considering AO-Cat, whose purpose is to describe data aggregated at high level, and is not suitable for describing the granularity of their information. This is particularly true for palaeo-anthropology (subtask 4.4.1) that needs to model complex entities related to human biological and cultural evidences, together with environmental data, structured in different formats including photos (TIFF, JPEG, etc.), Excel tables, free text, etc. Environmental archaeology (subtask 4.4.3) also deals with complex information, including the description of insects, pollen, seeds, as well as scientific dating information, and needs a sophisticated application profile for the semantic description of its data. The model, once ready, can also take advantage of CRMhs for modelling the laboratory analyses of this domain since this ontology is already equipped to deal with scientific procedures and protocols of this type.

In other research domains, the data seems to need no specific models since their data can be sufficiently described using existing ontologies or application profiles that offer common solutions to similar problems. The AO-Cat model seems to be an ideal tool in this regard. In particular, the working groups dealing with airborne LiDAR, satellite data and aerial imagery has proposed to manage aerial photography using GISs, and has concluded that the AO-Cat model is sufficient to adopt as the ontology for their subtask (4.4.8). The same was agreed for metal detector surveys and archaeological finds made by the general public (subtask 4.4.7), maritime and underwater archaeology (subtask 4.4.11) and standing structure data (subtask 4.4.9) since the aggregation in their case mainly concerns collection level information which AO-Cat is perfectly capable of handling.

CIDOC CRM and its extensions provide excellent coverage for item level information description. Standing structure data, for instance, additionally provides examples of information about buildings and physical infrastructures which might be suitable for item level integration: specific extensions of CIDOC CRM and CRMba are going to be defined in this case to provide semantic significance for these entities. CIDOC CRM was also chosen for the harmonisation of field survey documentation and protocols (subtask 4.4.6). A general purpose extension for this discipline is also planned to be

developed under its framework, to also confer the adequate degree of reusability required for this kind of data. Existing extensions of the CIDOC CRM, including CRMarcheo and CRMdig, will provide the basis for an application profile for archaeological fieldwork data (subtask 4.4.12), intended especially for modelling information in databases and spreadsheets of layers, digital photographs and plans of earlier excavations and CAD.

All working groups agreed that fine-grained vocabularies are essential to achieve interoperability and most of them already identified the relevant technical terms and periods for mapping to Getty AAT and PeriodO. It was also observed that full interoperability will be actually achieved by means of extensions necessary to include terms specific to each discipline especially into the Getty AAT. Environmental archaeology presents a very peculiar case of terminology integration in this sense, due to the great variety of non-comprehensive and overlapping vocabularies available in this discipline. An intense work of harmonisation is required and will be carried out within the next period to accomplish this task.

4.7 Deviation from work plan

The definition of some application profiles was slightly delayed due to internal organisational problems of some partners during the first phase of the project (staff recruitment, infrastructure preparation and so on).

A series of activities, including meetings and workshops, aimed at speeding up the completion of the application profiles still pending and the definition of the vocabularies necessary for a detailed description of the various information was planned for 2020, but it has been put on hold due to COVID-19. Planning will resume in 2021.

4.8 Plans for the next period

The work plan prepared for the second part of ARIADNEplus ensures that the definition of user needs and application profiles will be completed at the beginning of the next period, despite the obstacles posed by the COVID-19 pandemic.

The testing phase will begin as soon as the new application profiles are released and will allow their validation and integration as part of the ARIADNE Ontology.

5 Vocabularies and gazetteers

5.1 Overview

Work on multilingual vocabularies overlaps to some extent with work for WP5 on vocabulary integration via mappings from partner vocabularies to the Getty Art and Architecture Thesaurus (AAT) and via the Perio.do framework for temporal periods, as detailed in the User Manual for the ARIADNEplus Data Aggregation Pipeline and D5.2.

The mappings from partner native vocabularies in themselves contribute to a multilingual vocabulary resource compiled from different sources. Work has been done to extract, convert and combine multilingual resources from Wikidata, AAT and other relevant sources, such as Fasti terminology. And the partners at the University of South Wales (USW) have extracted 598,000 multilingual terms from the Wikidata knowledge base, by identifying concepts already pre-mapped in Wikidata to AAT concepts, and extracting all the multilingual terms associated with them. Wikidata concepts sometimes have mappings to an AAT concept ID, and multilingual labels. Via SPARQL queries, the relevant URI, label, language and AAT ID were extracted from Wikidata. The extracted AAT mappings were converted to same format as VMT output for a consistent approach. This yields a large entry vocabulary of multilingual terms mapped to AAT concepts, albeit they must be considered more tentative mappings than partner mappings specifically done for ARIADNEplus. It is intended that these terms will further expand the ARIADNEplus entry vocabulary for the development of enhanced multilingual search facilities as planned for Task 15.6 (see D15.1).

Getty Art and Architecture Thesaurus

As an illustration of the potential synergies, consider a partner mapping of the local concept “CEMETERY” to the AAT --- skos:exactMatch aat:300266755 “cemeteries”. This will make available other partner mappings to the same AAT concept:

BARROW CEMETERY, CHOLERA BURIAL GROUND, FRIENDS BURIAL GROUND, INHUMATION CEMETERY, JEWISH CEMETERY, MUSLIM CEMETERY, NONCONFORMIST CEMETERY, PLAGUE CEMETERY, ROMAN CATHOLIC CEMETERY, WALLED CEMETERY

These are all in the same language. However, the AAT has other terms associated with aat:300266755:

Preferred labels: "cemeteries"@en, "campos santos"@es, "campi santi"@it, "cimetières"@fr, "begraafplaatsen"@nl, "Friedhof"@de

Alternate labels: "cemetery"@en, "campos santos (cemeteries)"@en, "campo santo (cemetery)"@en, "campo santo"@es, "campo santo"@it, "cimetière"@fr, "cœmeterium (cemeteries)"@la, "camposanto (cemetery)"@en, "camposanto"@it, "begraafplaats"@nl, "Friedhöfe"@de

Further, we can utilise the AAT (poly)hierarchical structure and include narrower AAT concepts (more specific) than cemeteries:

catacombs, columbaria (cemeteries), graveyards, lawn cemeteries, memorial parks, necropolises, Reihengräberfelder, churchyards, cineraria (cemeteries), military cemeteries (veteran cemeteries), national cemeteries, pet cemeteries, potter's fields, war cemeteries

In addition, each of these concepts has multilingual preferred / alternate terms.

Now consider Wikidata. wikidata:Q39614 is already directly mapped to aat:300266755 ("*cemeteries*") and has many multilingual labels:

Cemetery, graveyard, burial ground, cemeteries, churchyard, cimetière, champ de repos, boulevard des allongés, champ du repos, Friedhof, Totenacker, Begräbnisplatz, Gottesacker, Kirchhof, Leichenhof, Begraafplaas, Asie, Fosal, Fosar, Zimenterio, Corralón, Fusal, Sagrero, Fosal d'os moros, Cimenterio, Fonsal, 墳場, cmentarz, cemitério, pokopališče, гробље etc.

Thus, one local mapping of a partner vocabulary concept to AAT has the potential to bring in many alternative terms/concepts from Wikidata. This can improve multilingual query experience and expand potential results via a semantically expanded multilingual search.

Perio.do

Perio.do is an international collaborative effort to create a freely available, multilingual, linked data reference gazetteer to describe historical and archaeological named periods. It has been adopted by ARIADNEplus to provide a framework for the project's space-time gazetteer and periodisation system. Partners upload their relevant standard sets of named periods into a Perio.do 'collection'. Every individual period definition associates the period name with a particular spatial coverage and with a start date and end date, plus other metadata including a bibliographic reference for the source of the definition. Periods and collections have 'permalink' identifiers (URIs) so they can be clearly and unambiguously referenced. For example, 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".

It is possible for different reputable sources to associate different date ranges with the same spatial extent for the same period name. It is important for data integration in ARIADNEplus that temporal references are made to a specific period and collection within the Perio.do framework. Some examples of different definitions for the same period are given in the following table.

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

The starting point is the published period definitions used in the preceding ARIADNE project, which have been revised and supplemented where necessary. This has been augmented by new period collections from various ARIADNEplus partners with assistance in use of Perio.do from USW. Work is ongoing. The following (partial) examples illustrate the breadth of coverage to date:

AIAC http://n2t.net/ark:/99152/p06v8w4 (Fasti)
DANS-KNAW http://n2t.net/ark:/99152/p0pqptc
IAVP http://n2t.net/ark:/99152/p02kbfn
INRAP http://n2t.net/ark:/99152/p0rrjd9
KHM-UO http://n2t.net/ark:/99152/p04h98q
UoY-ADS http://n2t.net/ark:/99152/p0kh9ds
ROCEEH http://n2t.net/ark:/99152/p0s2rwk (ROCEEH Out of Africa Database)
ARUP http://n2t.net/ark:/99152/p0wctqt
PAS http://n2t.net/ark:/99152/p0gjgrs (Portable Antiquities Scheme)

5.2 Deviation from work plan

No deviation from the work plan.

5.3 Plans for the next period

- USW will continue supporting partners when entering their local collections of time periods into Perio.do. USW will liaise with WP15 portal development and multilingual search issues concerning Perio.do

- USW will continue supporting partners with AAT mappings and AAT issues. This will contribute to the general ARIADNEplus multilingual resource.

- USW will work to support partners with specialised archaeological vocabulary development for the data integration phase.

- USW will continue to develop and provide multilingual vocabulary resources for WP15.6 portal development and multilingual search services. It is anticipated this will involve further exploration of the synergies between Getty AAT and Wikidata, together with partners' AAT mappings, as outlined in Section 5.1.

6 Assessing the CRM extensions

6.1 Overview

In the first 23 months of the project, testing of CRM extensions happened via the implementation of the ontology for the creation of application profiles for the archaeological subdomains. The task was mainly carried out in WP4, T4.4 (but also in WP2, 5, 12 and 14) and the results were reported in detail in deliverable in D4.2 (Initial report on ontology implementation) and partly summarised above in Section 4 (Application Profiles). For a description of the ARIADNE ontology see D4.2, Section 3: The AO-Cat ontology for the ARIADNE catalogue.

6.2 Summary: testing AO-Cat with data from archaeological sub-domains in WP 4 T4.4

It was found that for some archaeological subdomains the AO-Cat provides a suitable application profile. This includes data from sites and monument records and excavation reports (sub-task 4.4.0), as well full digital excavation archives (subtask 4.4.12).

For several other datatypes, the AO-Cat suffices to describe information on the site-level, but more specific application profiles will be required where data will be described in more detail, for example when it comes to item-level integration:

4.4.3 environmental archaeology may need a more specific application profile for analysis;

4.4.7 archaeological finds made by the public may need more fine-grained vocabularies to achieve interoperability at the finds level, which they want to achieve via an extension of the AAT;

4.4.8 remote sensing (mainly reports, which are adequately covered by AO Cat) but LiDAR data may need specific extension;

4.4.9 standing structures where some parts such as material analysis have different requirements;

4.4.10 spatial-temporal data was found to cross-cut most subdomains, hence requirements of individual sub-domains apply;

4.4.14 mortuary data.

6.2.1 Creation of a CRM extension for Heritage Science Data - CRMhs

For the documentation of archaeological science data and heritage science in general (4.4.4/4.4.5) existing AO-Cat classes and properties were found to be insufficient. This led to the creation of the CRMhs model, which is compatible to CIDOC CRM and AO-Cat. The creation of the model, its conceptual background and entities have been reported in detail in D4.2, Section 5. Case study 1: CRMhs and scientific data. The model may be adapted to cover other laboratory-based sub-domains such as Bio-archaeology and Ancient DNA (4.4.2) - see below, Environmental Archaeology (4.4.3), Inorganic Materials study (4.4.4), and Dating (4.4.5).

6.2.2 Bioarchaeology and ancient DNA

The AO-Cat was found sufficient for the high abstraction level description of aDNA wetlab services (4.4.2 bioarchaeology and ancient aDNA subdomain). For a more detailed description of the services they have used CIDOC CRM classes and properties. The case study has been presented in detail in D4.2, Section 6. Case study 2: Bioarchaeology and ancient DNA.

6.2.3 Inscriptions

A model for the description of the inscriptions, their investigation and the documentation of investigation was created (4.4.13). The model was reported in detail in D4.2, Section 6. Case study 3: Inscriptions.

6.3 Testing the CRM with mortuary data

Archaeological data types from the mortuary realm were described above (Section 4.5). For item-level integration data types 2 and 3 are relevant. As a use case for the more detailed mapping of mortuary data for item-level integration data collection with a database was used on early medieval reopened graves from the repository of ARIADNEplus partner DANS.[14] The database holds information about eleven cemeteries, including basic information on all context types (inhumation and cremation graves, animal graves, pits, ditches, stray finds), skeletal remains, grave goods. It consists of seven main tables and 29 reference tables. The aim of the project was to collect data on the disturbances of the graves for a statistical analysis and a historical-archaeological interpretation of the results.

The data on the individual cemetery sites (table ‘cemetery’) was found not to contain information on the cemeteries (which would typically be their location, date range, number of graves etc.), as one would expect, but it contained data providing information on the reliability of the data from each site. Other than site name and internal identifier this includes: information on the excavation of the

[14] Haperen, M.C. van (Leiden University) (2017): In Touch with the Dead: Early Medieval Grave Reopenings in the Low Countries. DANS. https://doi.org/10.17026/dans-x6b-bvgj

cemetery (period of excavation, quality of excavation, if completely excavated or not), whether it was published and a rating of the quality of the excavation, the soil type and the geographical region, as well as space for remarks. The aim of collecting this data was to exclude that any of these factors, in particular quality of either excavation or publication, or soil type and region would influence the results of the statistical analysis. Hence, we decided to map the 'cemetery' table as an 'Excavation' (A1_Excavation_Process_Unit), because that is what the data is about. Therefore, in this case, also for the very general data AO-Cat did not provide sufficient possibilities for description but needed support from CRMarchaeo.

This was even more the case for the remaining data on contexts, skeletal remains and grave goods. The following CRM extensions were additionally used to cater for the specific requirements of mortuary data (those used frequently, and which are expected to be highly relevant for description of evidence from other archaeological domains too are marked bold):

CIDOC CRM – The base model, version 6.2.1

- **P140i_was_attributed_by -> E13_Attribute_Assignment**
  to express classification/interpretation of archaeological evidence, e.g. if a grave was identified as reopened.

- **P43_has_dimension -> E54_Dimension -> P90_has_value**
  to express the length or other measurements of an archaeological feature such as a grave pit, or number of items belonging to an object, or percentage of a find that was present.

- P45_consists_of -> E57_Material
  to say which material a find from a grave was made of

- P46_is_composed_of -> AO_Object
  to say if a find is composed of several items and in combination with P43 - E54 - P90 how many items there were

- **P124i_was_transformed_by -> E81_Transformation**
  to express if something was intentionally damaged

- **P44_has_condition -> E3_Condition_State**
  to express if something was complete

- **P53_has_former_or_current_location -> E53_Place**
  to say that a find has been moved during human intrusion in a grave

- P46i_forms_part_of -> E22_Man-made_Object
  to express that an object was a fragment/part of another object, possibly found in another context)

- P41i_was_classified_by -> E17_Type_Assignment
  to express classifications, e.g. the dating of a find as 'antique' much older, e.g. prehistoric in its respective context

CRMsci – Scientific observation model, version 1.2.2

- O8i_was_observed_by -> S19_Encounter_Event
  for example to say that fragments of an object were found - that support the inference that once this object was there and complete

- P34i_was_assessed_by -> E14_Condition_Assessment
  for example to say that an inference was made that objects are missing

- O8i_was_observed_by ->S4_Observation
  for example to say that traces of corrosion from another object were observed on a grave find

CRMarchaeo – Excavation model, model, version 1.4.1

- **AP15_is_or_contains_remains_of -> A1_Excavation_Process_Unit**
  to say that this was an archaeological context (e.g. grave) that was removed in the course of an excavation

- **AP15_is_or_contains_remains_of -> A3_Stratigraphic_Interface**
  to describe the grave pit

- **AP18i_is_embedded -> A7_Embedding -> AP19_is_embedding_in -> A2_Stratigraphic_Volume_Unit**
  to say a find was found in a stratigraphic unit

6.4 Deviation from work plan

There has been no deviation from the workplan. The activities of this task started during the second project year, as the assessment of CRM extensions was only possible once AO-Cat had been developed.

6.5 Plans for the next period

Assessing the CRM extensions is a process that will continue in the second half (years 3-4) of the ARIADNEplus project. It will be part of several activities including as a specific task in WP14; in WP 4, Task 4.4 for those archaeological subdomains where the definition of application profiles have not yet been completed; in WP16 where the Virtual Research Environments (VREs) are being developed and in WP12, Task 12.4 where item-level data integration of archaeological subdomains will require more detailed mappings.

For the subdomain of field surveys (4.4.6) partners set up a special interest group to develop a CIDOC CRM extension for the domain. They argue that although the development of a CIDOC CRM extension is a slow process, but global archaeology will benefit from such a solution as it will increase comparability of this data type. This work will be carried out in the second half of the project using the required concepts developed in the first half of the project.

7 Conclusions

Overall the activities of WP14 were carried out successfully and with very good cooperation with other interlinked work packages as WP4, 5, 12, and 16.

Activity Dash version 1.0 for monitoring the aggregation process, has been released and integrated into the projects' infrastructure. The tool was tested and validated by a small group of users and is now available for use by all involved partners. It is expected that the tool will facilitate the aggregation process performed in WP4 and will allow better monitoring.

The activities of task 14.2 and the development of the application profiles was carried out regularly with some small delay due to small organisational problems now solved. The creation of working groups for specific tasks has greatly speeded up the activities of defining the requirements and retrieving information and vocabularies. The close collaboration between the partners of 14.2 and 4.4 led to the release of the first stable versions of some fundamental application profiles, such as those for inscriptions and for scientific and mortuary data. No particular delays are expected for the completion of the activities of these tasks within the next period, which will focus on the release of the remaining application profiles and the testing of their functionality on real data for the item level integration activities planned by the project.

Close collaboration with WP5 led to several mappings from partners' native vocabularies to Getty AAT contributing to a multilingual vocabulary resource compiled from different sources such as Wikidata, and other relevant sources. It is intended that these terms will further expand the ARIADNEplus entry vocabulary for the development of enhanced multilingual search facilities as planned for Task 15.6 (see report D15.1).

The first half of the project showed the AO-Cat is too general to cater for the specific requirements of a more fine-grained description of data from the broad range of archaeological subdomains such as is required for applications profiles and item-level integration. Hence, in WP4 several subdomains have started to make use of the CIDOC CRM and its extensions to develop their application profiles, or, have developed their own extension (heritage science) or started to develop them (field survey). In this report the additional requirements for an application profile for mortuary data have been described as a use case. It has been shown that for this subdomain in particular CIDOC CRM and CRMarchaeo classes and properties are needed.