



## **D15.2 – Final report on ARIADNEplus services**

**Version 1.0 (final)**

**2022-11-28**

<b>Grant Agreement number:</b>	<b>823914</b>
<b>Project acronym:</b>	ARIADNEplus
<b>Project title:</b>	Advanced Research Infrastructure for Archaeological Dataset Networking in Europe - plus
<b>Funding Scheme:</b>	H2020-INFRAIA-2018-1
<b>Project co-ordinator name, Title and Organisation:</b>	Prof. Franco Niccolucci, PIN Scrl - Polo Universitario "Città di Prato"
<b>Tel:</b>	+39 0574 602578
<b>E-mail:</b>	<a href="mailto:franco.niccolucci@pin.unifi.it">franco.niccolucci@pin.unifi.it</a>
<b>Project website address:</b>	<a href="http://www.ariadne-infrastructure.eu">www.ariadne-infrastructure.eu</a>

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:** Johan Fihn Marberg, Swedish National Data Service

**Contributing partners:** Alessia Bardi, Consiglio Nazionale delle Ricerche  
Andreas Vlachidis, University College London  
Carlo Meghini, Consiglio Nazionale delle Ricerche  
Ceri Binding, University of South Wales  
Douglas Tudhope, University of South Wales  
Fabio Sinibaldi, Consiglio Nazionale delle Ricerche  
Federico Ponchio, Consiglio Nazionale delle Ricerche  
Francesco Mangiacrapa, Consiglio Nazionale delle Ricerche  
Ivan Radman-Livaja, Archaeological Museum Zagreb  
Marco Callieri, Consiglio Nazionale delle Ricerche  
Marco Potenziani, Consiglio Nazionale delle Ricerche  
Marion Lamé, Université de Tours  
Massimiliano Assante, Consiglio Nazionale delle Ricerche  
Pasquale Pagano, Consiglio Nazionale delle Ricerche  
Sorin Hermon, The Cyprus Institute  
Valentina Vassallo, The Cyprus Institute

**Quality Control Review:** Holly Wright, University of York  
Paola Ronzino, PIN

This work is licensed under the Creative Commons CC-BY License. To view a copy of the license, visit <https://creativecommons.org/licenses/by/4.0/>

## Table of Contents

<b>1. Executive Summary .....</b>	<b>4</b>
<b>2. Introduction and Objectives .....</b>	<b>6</b>
<b>3. Progress on service development.....</b>	<b>8</b>
<b>3.1 Task 15.1 – Design of ARIADNEplus services features .....</b>	<b>8</b>
<b>3.2 Task 15.2.1 – Visualisation of archaeological imagery .....</b>	<b>13</b>
<b>3.3 Task 15.2.2 – Visual organisation of archaeological data.....</b>	<b>15</b>
<b>3.4 Task 15.2.3 – Visual documentation of an archaeological excavation .....</b>	<b>18</b>
<b>3.5 Task 15.3.1 – Archaeology text annotation .....</b>	<b>23</b>
<b>3.6 Task 15.3.2 – Image annotation.....</b>	<b>24</b>
<b>3.7 Task 15.4 – Implementing the ARIADNEplus text mining and NLP services.....</b>	<b>30</b>
<b>3.8 Task 15.5 – Implementing space-time services .....</b>	<b>42</b>
<b>3.9 Task 15.6 – Query services .....</b>	<b>50</b>
<b>3.10 Task 15.7 – Cloud geoserver .....</b>	<b>53</b>
<b>4. References.....</b>	<b>57</b>
<b>Appendix A: Template for describing services.....</b>	<b>59</b>

# 1. Executive Summary

This deliverable describes the activities carried out within Work Package 15 (WP15) of the ARIADNEplus project by the different partners and describes the results achieved.

The work package consists of several individual tasks and subtasks with the overall goal to develop and provide useful services to archaeologists. This means the work package is by nature heterogeneous with stand-alone tasks and services. Efforts have been made to facilitate collaboration between the individual tasks through joint work package meetings. This has resulted in new cross-task contacts being made, and some sharing of expertise to improve services has been done.

A service design template aligning the ARIADNEplus services with the requirements from European Open Science Cloud (EOSC) has been created. In connection with this, the ARIADNEplus AO-CAT ontology has been adapted to the requirements from EOSC Resource Data Model (Task 15.1).

The Visual Media Service (Task 15.2.1) has had a new format added, allowing for 2D visualisation of LIDAR data in DEM format. In addition, three other standards have been added: glTF, ThreeJS and IIIF, supporting various functionality in the service. The service has also been adapted to support integration with the ARIADNEplus infrastructure in D4Science.

A visual wizard has been defined to guide Visual Media Service users to add hotspots to a 3D scene easily and quickly. This extension, initially implemented in 3DHOP will allow archaeologists to create interactive links from the digital 3D model to the related documentation without writing any source code (Task 15.2.2).

Task 15.2.3 reworked the Online 3D Database System for Endangered architectural and archaeological Heritage in the south Eastern MEditionRanea area (EpHEMERA). EpHEMERA is a service provided by the Cyprus Institute to visualize in 3D archaeological excavations, ancient buildings, and their related documentation. In EpHEMERA, it is possible to visualise, online and through standard web browsers, 3D architectural and archaeological models (classified according to a specific type of risk), query the database system and retrieve metadata attached to each digital object, and extract geometric and morphological information about the Cultural Heritage asset.

The visualisation and annotation tool of the TSS project have been ported to the OpenLime library and integrated into the Visual Media Service (Task 15.2.1). An additional layer of SVG annotations have been developed and added to the service. The Annotation service have been used and improved in three different pilot projects. (Task 15.3.2)

Various strands of work have been done improving services for text mining and Natural Language Processing (Task 15.4). One of these efforts has been building upon the outcomes of the preceding ARIADNE project. A set of archaeological Named Entity Recognition NLP pipelines were reconfigured and deployed for easier use on the General Architecture for Text Engineering (GATE) cloud. Another effort has been on extracting temporal archaeological information using two different parallel approaches, normalisation and named entity recognition. A Python development platform has been used to unify the various services.

A Vocabulary Annotation Tool (Task 15.3.1) was developed using the same platform, as part of Task 15.4. The tool facilitates the locating and tagging of vocabulary terms within free text and outputs suggested subject annotations in a range of formats.

The GeoPortal service (Task 15.5) is a new REST service designed to manage complex spatio-temporal documents defined by metadata profiles. It was released as a component of the gCube framework. A



prototype using the service was deployed and operated to manage archaeological excavation projects (Task 15.7).

Two services for querying the RDF AO-Cat metadata records aggregated by the ARIADNEplus Infrastructure was established (Task 15.6): a full-text index service and a SPARQL endpoint. The full-text index service is based on OpenSearch and supports the needed query functionality of the ARIADNEplus portal. The SPARQL endpoint allows performance of semantic queries on the RDF records within the ARIADNEplus data and knowledge cloud.

## 2. Introduction and Objectives

The overall goal of Work Package 15 is to provide useful services for archaeologists using ARIADNEplus data. This includes integrating existing services, developing new features for existing services, as well as development of brand-new services. All services will be integrated into the ARIADNEplus infrastructure via the D4Science platform.

The work package consists of 10 individual tasks which cover services with a wide range of features useful for researchers. These range from front-office services like visualisation services, various types of annotations and documentation, to more back-office oriented services like the multilingual query service used by the ARIADNEplus Portal for querying the underlying aggregated partner resources. The tasks and subtasks included in this work package are:

- T15.1 Design of ARIADNEplus services features (JRA4.1)
- T15.2 Implementation of ARIADNEplus visual services (JRA4.2)
  - T15.2.1 Visualisation of archaeological imagery (JRA4.2.1)
  - T15.2.2 Visual organisation of archaeological data (JRA4.2.2)
  - T15.2.3 Visual documentation of an archaeological excavation (JRA4.2.3)
- T15.3 Implementing the ARIADNEplus annotation services (JRA4.3)
  - T15.3.1 Archaeological text annotation (JRA4.3.1)
  - T15.3.2 Image annotation (JRA4.3.2)
- T15.4 Implementing the ARIADNEplus text mining and NLP services (JRA4.4)
- T15.5 Implementing space-time services (JRA4.5)
- T15.6 Query services (JRA4.6)
- T15.7 Cloud Geoserver (JRA4.7)

Task 15.1 analyses the existing services available within the former ARIADNE Portal, as well as services already available to partners and others as open source. The task is also responsible for periodically surveying the market to find and possibly integrate other possible candidates services into the ARIADNEplus infrastructure. A service design, in compliance with the cloud design developed in WP12, will also be produced which aligns the ARIADNEplus services with EOSC requirements.

Tasks 15.2 through 15.5 cover the front-end services and Tasks 15.6 and 15.7 are the back-end services, which are developed under the ARIADNEplus umbrella. Activities in Task 15.2 and Task 15.3 were performed in the respective subtasks of each main task and will thus be reported in each subtask.

The Visual Media service (Task 15.2.1), built upon the VisualMedia EOSCPilot Science Demonstrator, provides easy publication and presentation of complex visual media assets on the web. It is an automatic service that allows the uploading of visual media files on a server and transforms them into an efficient web format, making them ready for web-based visualisation. The service enables the display of archaeological information in the form of images and 3D models, and in the ARIADNEplus context, will be further developed and adapted within the ARIADNEplus infrastructure.

In the context of the implementation of innovative visual services for users, Task 15.2.2 aims to build new components addressed to the ARIADNEplus archaeological data management. More specifically, the goal is to design and develop a web service to improve the visual organisation of 3D archaeological datasets. The service, built in the framework of the above-described Visual Media Service, links the 3D model of an artefact or monument to its archaeological documentation, visualising it accordingly. The task revised existing tools and then implement a brand-new service, recently ported in the ARIADNEplus infrastructure.

The Online 3D Database System for Endangered architectural and archaeological Heritage in the south Eastern MEditional area (EpHEMERA). service provided by CYI (Task 15.2.3), is used to visualise archaeological excavations, ancient buildings, and their related documentation in 3D. In EpHEMERA, it is possible to visualise, online and through standard web browsers, 3D architectural and archaeological models (classified according to a specific type of risk), query the database system and retrieve metadata attached to each digital object, and extract geometric and morphological information.

Task 15.3.1 provides support for archaeologists to annotate textual documents that form part of the ARIADNEplus infrastructure, such as archaeological reports, grey literature, archaeological texts in general and possibly textual descriptions forming part of content metadata. The objective of the task is to provide an open-source text annotation tool that is easy to use. The tool will draw on freely available vocabularies including Linked Open Data. Thus, vocabulary concepts with URI identifiers should be output where possible. The tool should provide auto-suggestions that encourage, but do not force users to annotate with standard vocabularies.

Task 15.3.2 concerns the extension of the annotation tool DAP (Digital Autoptic Process). The tool allows the annotation of archaeological images in a CIDOC CRM compliant way. The goal of this task is to develop an online tool for annotating archaeological 2D media: images and relightable images for use with RTI. There is, however, nothing available for the annotation of such media. Standards for image annotations are not widely supported, and the need to connect the metadata annotations with the ARIADNEplus ontology AO-CAT and vocabularies is another challenge. The tool will eventually also be available in the Visual Media Service.

Task 15.4 provides text mining and Natural Language Processing (NLP) services that will assist the integration of textual information into the ARIADNEplus Infrastructure, allowing semantic linking between archaeological reports (e.g. grey literature) and the linking of textual reports to datasets, via the semantic framework. The objective of the task is to produce open source, NLP services that will draw on relevant entities and vocabularies within the ARIADNEplus semantic framework. Thus, the services will support Named Entity Recognition (NER) of key entities, such as Physical Object, Material, etc., with Time being a particular focus. Vocabulary concepts with URI identifiers should be output where possible. The services will provide added value by suggesting additional subject metadata for textual material within the infrastructure with little original subject indexing. The task builds on outcomes from the preceding ARIADNE project, expressing the ARIADNE NLP pipeline utilities as NER services in the cloud environment and extending the languages beyond the coverage in the ARIADNE project.

The management of complex collection of spatial documents requires a technical framework that can be tailored to the specific needs of the community while remaining affordable to the European ARIADNEplus infrastructure. The design and implementation of the service is maintained and operated through extendible Profiles and software handlers that can be customized per VRE. This was developed in Task 15.5, with the definition of a complete framework that using that service can exploit standard SDI services and capabilities is developed in Task 15.7.

To support the necessary query capabilities needed by the ARIADNEplus portal, for querying the underlying aggregated partner resources, various querying services were established in Task 15.6. The services are based on an RDF triplestore containing the entirety of the ingested resources and an OpenSearch service specifically designed to support the functionality of the ARIADNEplus portal. Based on the functionality of OpenSearch, multilingual query capabilities have been added to enable the full use of the Getty Arts and Architecture Thesaurus (AAT) by the portal.

## 3. Progress on service development

### 3.1 Task 15.1 – Design of ARIADNEplus services features

This subtask identified and integrated services into the ARIADNEplus Infrastructure, and produced a service design aligning the ARIADNEplus services with EOSC requirements.

#### Service design

The ARIADNE Catalogue includes information about the resources managed by the ARIADNE infrastructure, including the services provided by the members of the ARIADNE Community. Consequently, the AO-Cat ontology, which provides the terms for the ARIADNE Catalogue, includes terms for the description of services. These terms form a small vocabulary, which was defined during the first ARIADNE project and has been fully adopted by ARIADNEplus. In the meantime, the European infrastructure landscape has undergone an important change, with the introduction of the European Open Science Cloud, EOSC for short.

According to the EOSC Strategic Implementation Roadmap, ‘EOSC will allow for universal access to data and a new level playing field for EU researchers, providing easy access through a universal access point for ALL European researchers’<sup>1</sup>. The role of domain infrastructures like ARIADNE in the realisation of the EOSC vision is fundamental: it will act as conduits between individual researchers and the EOSC. In light of this development, it is crucial that the archaeological services described in the ARIADNE Catalogue be discoverable, along and interoperable with the services of the other communities federated by EOSC.

Discoverability and interoperability with EOSC have therefore become a fundamental requirement for the part of the AO-Cat Ontology that deals with the description of services. In order to implement this requirement, the EOSC Resource Data Model has been analysed in detail and harmonised with the AO-Cat ontology.

Below are the results of this analysis and harmonisation. Appendix A provides a template for the EOSC-compliant service description to be used in ARIADNEplus.

In AO-Cat, the class AO\_Service is introduced to model services, where the adopted notion of service is that of the PARTHENOS Entities Model (PEM) [PEM Specifications 3.1] as ‘an offer by some actor of their willingness and ability to execute an activity or series of activities upon request’. As such, an AO-Cat service is an infrastructural resource (instance of AO\_Resource), which makes it also an AO-Cat entity. The AO-Cat class taxonomy surrounding the class AO\_Service is given in the figure below:

---

<sup>1</sup> European Commission, Directorate-General for Research and Innovation, European Open Science Cloud (EOSC) strategic implementation plan, Jones, S.(editor), Abramatic, J.(editor), Publications Office, 2019, <https://data.europa.eu/doi/10.2777/202370>

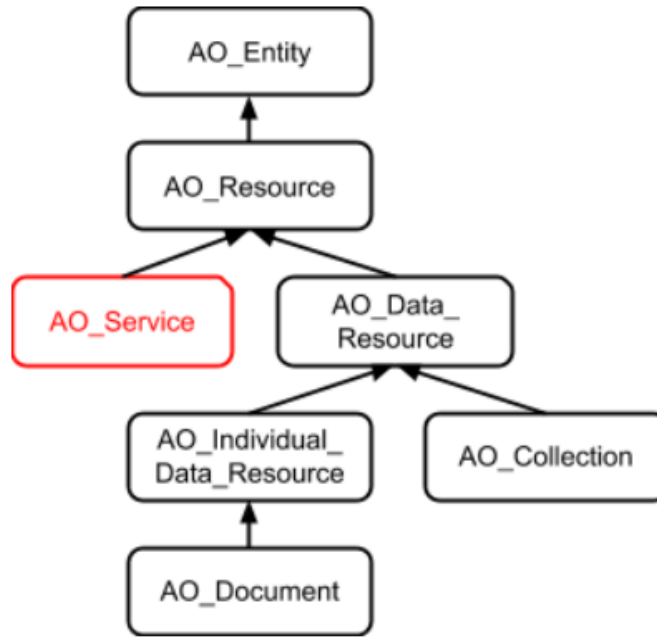


Figure 1: AO-CAT class taxonomy surrounding the class *AO\_Service*.

The properties currently defined for services are (the domain of each property is *AO\_Service*, the range is indicated below):

- *is\_accessible\_at*, associates a service with an IRI where the service is accessible, range is *rdfs:Resource*
- *has\_functionality*, associates a service with its functionality, expressed in some vocabulary, range is *AO\_Concept*
- *has\_consumed\_media*, associates a service with the media type(s) handled by the service, range is *AO\_Concept*
- *has\_produced\_media*, specifies the media types of the objects created by the service, range is *AO\_Concept*
- *has\_consumed\_format*, specifies the MIME type of the objects handled by the service, range is *rdfs:Resource*
- *has\_produced\_format*, specifies the MIME type of the objects created by the service, range is *rdfs:Resource*
- *has\_supported\_language*, specifies the languages supported by the service, encoded according with ISO 639 standard (ISO 639-1:2002). Range is *AO\_Concept* (linked to *lexvo.org*)
- *has\_technical\_support*, specifies the agent offering technical support on the service, if any, range is *AO\_Agent*.

The properties inherited from the class *AO\_Resource*, which *AO\_Service* is a sub-class of, are:

- *was\_issued*, the date when the record of the resource was firstly acquired. Range is *xsd:dateTime*
- *was\_modified*, the date when the record of the resource was lastly modified. Range is *xsd:dateTime*

- `has_publisher`, the agent responsible for making the resource publicly accessible. Range is `AO_Agent`.
- `has_contributor`, a contributor of a description of the resource to the AC. Range is `AO_Agent`.
- `has_creator`, a creator of the resource. Range is `AO_Agent`.
- `has_owner`, the owner of the resource. Range is `AO_Agent`.
- `has_responsible`, any person who is scientifically responsible of the resource. Range is `AO_Agent`.

Finally, the properties inherited from class `AO_Entity`, which `AO_Resource` is a sub-class of, are:

- `has_type`, a type of the resource in any classification system. Range is `AO_Concept`.
- `has_title`, a title of the resource. Range is `xsd:string`.
- `has_description`, a textual description of the resource in natural language. Range is `xsd:string`.

### Property adjustments

In order to accommodate the properties needed to describe services in a way that is compliant with the ERDM, some modifications to the AO-Cat ontology are required. These modifications are presented below.

The properties linking a resource to an agent that have been defined so far, are only related to data resources, even though they have been given `AO_Resource` as domain. Indeed, the modelling of services also requires properties of this kind, but these properties are different from those about data resources. To accommodate the new properties, therefore, the currently defined properties:

- `has_publisher`, the agent responsible for making the resource publicly accessible. Range is `AO_Agent`;
- `has_contributor`, a contributor of a description of the resource to the AC. Range is `AO_Agent`;
- `has_creator`, a creator of the resource. Range is `AO_Agent`;
- `has_owner`, the owner of the resource. Range is `AO_Agent`;
- `has_responsible`, any person who is scientifically responsible of the resource. Range is `AO_Agent`;

are given `AO_Data_Resource` as their domain, pushing them down in the taxonomy of resources. This modification has no impact on data already ingested, as these data only concern data resources.

A new property is introduced as the most general property connecting a resource to an agent:

- `has_agent`. The domain of this property is `AO_Resource` and its range is `AO_Agent`.

All the above properties are sub-properties of `has_agent`.

## EOSC vocabularies

The ERDM includes several vocabularies devoted to capture in a controlled and systematic way the values that certain property can assume. These vocabularies are:

- Scientific Domain and relative Subdomain; each domain has a number, a title, a description and several subdomains, each identified by a number and a title;
- Supercategory, Category, Subcategory; each has a number, a title, and categories also have a description;
- Target user; each such category has a number, a title and a description;
- Access Type; each such type has a number, a title and a description;
- Access Mode; each mode has a number, a title and a description;
- Geographical Availability; each such type has a number, a country and a description;
- Language Availability (from ISO 639-1:2002 standard - Codes for the representation of names of languages); each such a code has a number, a title and a code;
- Geographic Location (from ISO 3166 standard – Codes for the representation of names of countries and their subdivisions, except for Greece and the United Kingdom (use EL and UK respectively instead of GR and GB) from Eurostat code lists); each such a code has a number, a country (string) and a code;
- Technology Readiness Level according to Horizon 2020;
- Life Cycle Status;
- Funding Body; each body has a number, a name and a code;
- Funding Program; each program has a number, a name and a code;
- Order Type; each type has a number, a title and a description.

Following the modelling pattern for vocabularies recommended by the CRM, we introduce in AO-Cat a general class:

- EOSC\_Vocabulary, the most general class encompassing every term from any one of the above EOSC vocabularies. EOSC\_Vocabulary is a sub-class of AO\_Concept.

and one specific class for each vocabulary, which is:

- EOSC\_Domain, the class of EOSC domains and subdomains, a sub-class of EOSC\_Vocabulary. Every EOSC domain or sub-domain is an instance of this class, the latter being linked to the former by property P127 has broader term. To capture the salient feature of each domain or subdomain, the following properties are used:
  - has\_title
  - has\_number
  - has\_description

Likewise, we have:

- EOSC\_Category, the class of EOSC supercategories, categories and subcategories, a sub-class of EOSC\_Vocabulary. The relationships between these types of categories are also captured by P127, and the three properties
  - has\_title
  - has\_number
  - has\_description

are also in this case used to capture the features of every category.

Following the same style we introduce the classes:

- EOSC\_Target\_User
- EOSC\_Access\_Type
- EOSC\_Access\_Mode
- EOSC\_Geographical\_Availability
- EOSC\_Language\_Availability
- EOSC\_Geographic\_Location
- EOSC\_Technology\_Readiness\_Level
- EOSC\_Life\_Cycle\_Status
- EOSC\_Funding\_Body
- EOSC\_Funding\_Program
- EOSC\_Order\_Type

Some of these vocabularies, though, overlap with AO-Cat vocabularies. In particular,

- the EOSC\_Geographical\_Availability vocabulary is a list of names of all world's countries taken from the Resource Geographic Location of the EOSC Resource Profile, with the addition of area aggregates, such as 'all countries' or 'all countries of the Eurozone'. In AO-Cat, countries are modelled as spatial regions identified by a standard name, hence they are instances of class AO\_Spatial\_Region\_StdName, a sub-class of AO\_Spatial\_Region. In order to handle this overlapping, countries are also declared instances of class AO\_Concept, and the mapping between such countries and the corresponding terms in the EOSC\_Geographical\_Availability vocabulary is declared as an instance of the skos:exactMatch property.
- Similarly, the EOSC\_Geographic\_Location vocabulary is a list of names of countries and their subdivisions from Eurostat code lists, also modelled in AO-Cat as spatial regions identified by a name. These are declared as instances of AO\_Concept, and the mapping between them and the corresponding terms in the EOSC\_Geographic\_Location vocabulary is declared as an instance of the skos:exactMatch property.
- The EOSC\_Language\_Availability vocabulary is based on the ISO 639-1:2002 standard - Codes for the representation of names of languages. In AO-Cat, names of languages are represented as instances of AO\_Concept, taken from the Lexvo ontology. In order to establish the mapping between these different representations of the same resources, skos:exactMatch statements are asserted in the AC between pairs of IRIs that represent the same language.

### **Service descriptions in the ARIADNEplus portal**

As a part of promoting the various services developed and used by the ARIADNEplus project to the end users, a specific section and search interface has been setup as a part of the ARIADNEplus portal. The services are divided into four different categories: Data management training and tools, Data processing and visualisation services, Vocabulary services, and Other services, tools and resources.



## 3.2 Task 15.2.1 – Visualisation of archaeological imagery

Visual Media Service<sup>2</sup> [Ponchio et.al 2016] provides easy publication and presentation on the web of complex visual media assets. It is an automatic service that allows the uploading of visual media files to a server and transforms them into an efficient web format, making them ready for web-based visualisation.

The visual media supported are high-resolution images, relightable images (RTI), and 3D meshes (in PLY or OBJ format). The user can upload one of these visual media and they are automatically processed and configured for web presentation. Finally, the user can download the optimised models and the configuration files, deploying the result in their own website or directly embedding the viewer produced by the platform.

Visual Media Service exploits 3DHOP<sup>3</sup> [Potenziani et.al 2015] and Nexus<sup>4</sup> for 3D models support. 3DHOP and Nexus feature 3D mesh visualization, streaming, compression, and, most importantly, support multiresolution (adaptive, seamless, and view-dependent rendering of the mesh). Relightable images instead make use of Relight<sup>5</sup>, an efficient online representation of RTI that transforms each image in a web-compliant format (similarly to Google maps, the RTI image at full resolution is regularly divided into chunks and a hierarchy of images at different resolutions is produced from these chunks). High-resolution images also exploit the same technology to be processed and visualized by the Visual Media Service. All of the mentioned software is developed by ISTI-CNR.

The goal of this task is to expand Visual Media Service features, enhance robustness and performance and integrate it as a service within the ARIADNEplus Portal and D4Science platform.

### OpenLIME

This is an online WebGL viewer for a wide variety of 2D data: large images, relightable images (RTI and BRDF), digital elevation models (DEM) and multispectral datasets.

### Large images

Visualisation of large images over the web requires the image to be converted into a pyramidal structure where the image is saved at different resolutions and split into tiles. OpenLIME supports established formats such as Deepzoom, Zoomify, Google and most notably the IIIF ImageAPI. This capability is the building block for using these images as layers for more complex 2D datasets.

### Relightable images

A relightable image makes it possible to interactively change the incoming light direction. This technique can be used for a realistic reproduction of the visual aspect of any surface, and combined with photogrammetry to reveal subtle details on the surface. These datasets are created starting from a large collection of photographs taken from the same viewpoint but different light direction.

Two main approaches are in use: **RTI** (reflectance transformation imaging) where a linear basis (be it a second degree polynomial, spherical harmonics or principal component analysis over radial basis

---

<sup>2</sup> <http://visual.ariadne-infrastructure.eu>

<sup>3</sup> <http://3dhop.net>

<sup>4</sup> <http://vcg.isti.cnr.it/nexus>

<sup>5</sup> <http://vcg.isti.cnr.it/relight>

functions) is fitted on the images. A second approach uses a standard non-linear **BRDF** material characterisation combining albedo, normal, glossiness components.

OpenLIME supports both approaches. Unfortunately RTI has no standard web-friendly format, so we adopted the format generated from the Relight tools. Visual Media Service can also convert standard RTI files into a web-friendly format used for online visualisation. The code for this conversion has been added to the Relight library.

These formats (other than changing the light direction) allows for additional visualisation modes:

- show the normal components in false colours
- display only the albedo removing the influence of the illuminations (shadows, reflections, etc)
- display only the geometry removing the albedo component
- change the material for displaying the geometry: from a pure Lambertian material to a shiny surface where reflections further enhance small geometric details.

Additional visualisation modes are possible and relatively easily implemented in OpenLIME due to its modular design.

## **DEM**

The prototype for DEM online visualisation in the previous deliverable was easily integrated into OpenLIME, with obvious advantages in terms of support and interface consistency.

## **Annotations**

OpenLIME supports SVG vector annotations. The main features are:

- CSS customisation for annotations appearance: colour, stroke width, etc.
- Zoom level dependent appearance and visibility
- Simple annotations navigation over a hierarchical structure
- Online annotation editor.
- Export/import both to standard format (IIIF), and common SVG editors such as Inkscape

Details on these features and related publications can be found in WP 15.2.3.

## **Lens**

Lens-based exploration is an interesting paradigm for user interaction, combining motion and discovery it can be extended with buttons and other features directly on the lens and has multi-user capabilities. OpenLIME has a solid lens implementation.

## **Interface**

OpenLIME comes with a default interface, with functionalities such as layer navigation, zoom-in, zoom-out, fullscreen, annotation visibility, a ruler, etc. The interface is fully customisable and extensible, both using CSS, replacing or adding icons in SVG and some JavaScript, with minimal coding experience.

## Library design

Great care was taken in the design and implementation of these three core properties:

**Modular:** the library is composed of many components: the rendering engine is usable through an API without the interface classes, class organisation allows the loading of only needed parts (RTI for example does not need BRDF code), both use the same core class of cache, loading images etc.

**Extensible:** the library is also designed to serve as a sandbox to test and develop new algorithms. A robust core of common classes: (e.g. loading images and pyramid formatting, shader and canvas support, etc) minimise the amount of work needed. For example DEM support requires just a few shaders.

**Customisable:** while striving for a sensible default, everything can be easily customised, in particular the user interface, the lens, the annotation appearance and user interaction. The mechanism involves use of css and svg for appearance (no coding needed), and call-backs for functionalities.

## International Image Interoperability Framework (IIIF)

IIIF<sup>6</sup> is an increasingly adopted standard allowing communities to share images online. The Relight viewer adopted in Visual Media Service was extended to support the IIIF protocol as noted in the last deliverable. OpenLIME now replaces the Relight viewer improving support also for annotations.

The “IIIF 3D Community Group” is growing and ARIADNEplus continues to participate in its development, working toward defining standards for 3D content sharing on the Web. At the same time ARIADNEplus continues to work to define relightable extensions for the IIIF image API.

## Pilot

The described technologies are also used to support pilot 16.2. An implementation of Visual Media Service supporting the pilot is being defined, and then will be populated with the data identified by the pilot case.

## 3.3 Task 15.2.2 – Visual organisation of archaeological data

In the context of the implementation of innovative visual services for users, this subtask aimed to build new components addressed through ARIADNEplus archaeological metadata management. More specifically, the goal was to design and develop new services for improving the visual organisation of 3D archaeological datasets.

These services, built within the framework of the Visual Media Service platform, redesigned the way of linking the 3D model of an artefact or a monument to its archaeological documentation, in order to visualise it appropriately. During the project, this vision has been expanded, and the redesign process has involved not only the 3D data enrichment stage, but all the other configuration options already provided by the Visual Media Service for customising the 3D presentation.

The work started with a revision of the existing tools, continued with the implementation of new services, which were then ported into the Visual Media Service cloud environment and made available via a new custom interface. The result of this work is a brand-new configurator for 3D models that

---

<sup>6</sup> <https://iiif.io/api/image/3.0/>

provides all the existing configuration options, plus new ones (most of them already introduced in the beta version in the previous deliverable, D15.1).

The configuration section for 3D models has been completely renewed, changing from a simple single-page editor with very few functionalities, to a complete multi-step wizard able to guide the content creator through the publication process.

The current 3D configurator is composed of five different configuration steps, appropriately divided according to the type of customization they provide: Alignment, Material & Light, Navigation, Interface, and finally, Annotation.

## **Alignment**

The first step allows the content creator to fix the model orientation. One of the more common issues with 3D models is the model is often visualised with the wrong orientation. This happens because the 3D model was not oriented when created, or because the different tools composing the 3D working pipeline (data acquisition, editing, publication) use different visualisation systems. In any case, if the model is not in the correct orientation, the content creator can use the alignment wizard to re-orient it using different commands:

- The orientation can be corrected by rotating the model along an axis; there are three different sliders, one for each axis, which can be used to independently rotate the model on a selected axis;
- The 'UP' vector can be defined in two different ways, which can lead to unexpected behaviour in visualisation. It is not possible to know in advance what notation is used for a given 3D model. In the Alignment panel the user can switch from one notation to the other;
- Finally, advanced straightening can be used to orient the model to match predefined views directly interacting with it in the 3D scene, or by using the buttons provided.

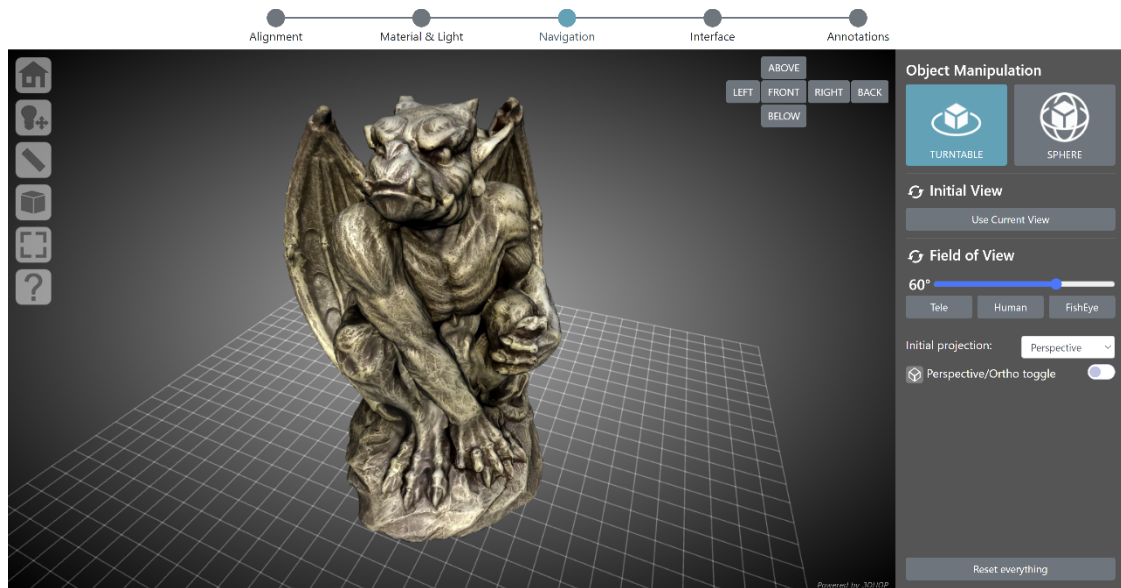
## **Material & Light**

The second step allows the content creator to set up the appearance of the 3D. Specifically:

- the Material wizard provides a set of tools to select the start-up colour of the 3D model (texture or solid colour), define the solid colour, add or remove a toolbar button for switching between texture and solid colour, and finally customise the level of glossiness (dull, low, medium or shiny);
- Lighting provides a set of tools to define the start-up lighting of the scene (enabled for use with 3D model normal maps, or disabled for ambient lighting), adding or removing a toolbar button for switching between enabled and disabled lighting, and finally adding or removing a toolbar button for enabling light direction control.

## **Navigation**

The third step allows the content creator to define the scene navigation setup. In particular, this wizard provides a set of tools to select the trackball paradigm to be used in the scene for manipulating the 3D model (turntable; allowing tilting and vertical rotating the model, preserving the vertical axis; or sphere; enabling full and free rotation and panning on all axes), define the initial trackball position, customise the field of view angle of the virtual camera (using a slider or three different buttons with predefined values), specify the type of projection used by the camera at the startup (perspective or orthographic), and finally add or remove a toolbar button for switching between perspective and orthographic projection.



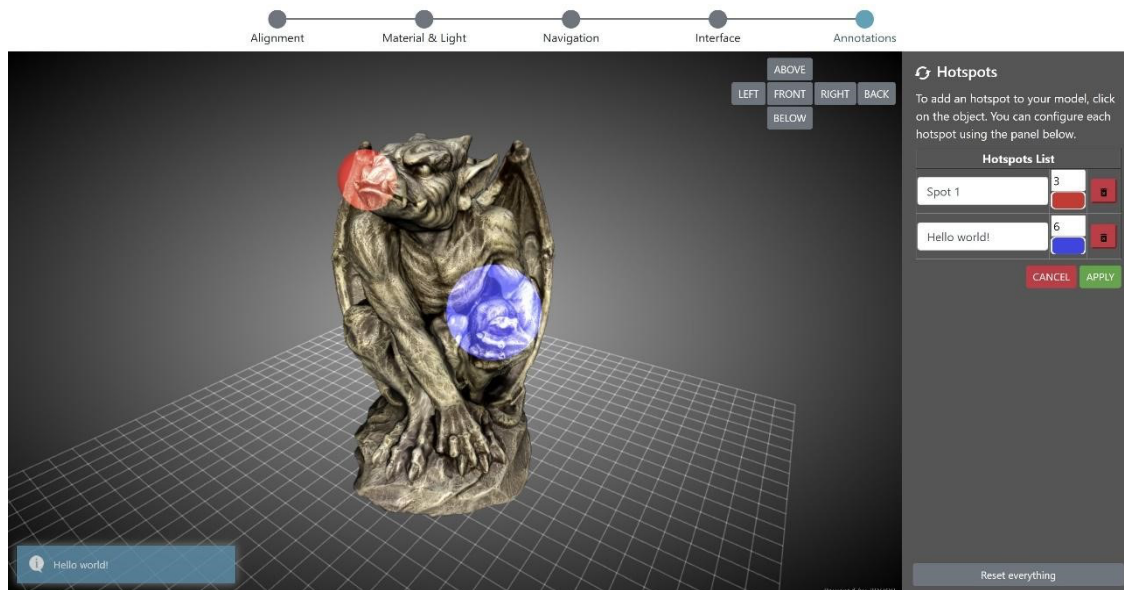
*Figure 2: The navigation step of the 3D configurator wizard. The panel on the right contains the controls for changing the trackball, setting the initial view, modifying the FOV, defining the type of camera projection, and adding/removing the camera projection button in the toolbar. The 3D viewer on the left is updated in real time according to the customisation in the right-hand panel.*

## Interface

The fourth step allows the content creator to customise the viewer interface. This specific wizard provides a set of tools to customise the toolbar skin (six different themes are available: light, dark, minimal light, minimal dark, transparent light, transparent dark), add or remove tool buttons from the toolbar (lighting, light direction, measurement, point picking, sections, solid colour, projection, full screen), define the viewer background (flat colour, linear gradient, radial gradient, predefined images), and finally select which widget provide to the end user (base grid, compass, canonical views, navigation cube).

## Annotation

The fifth step allows the content creator to add hotspots (points of interest highlighted by clickable geometries) to the 3D model and link them to some basic annotation. The annotations wizard provides a set of tools to add or remove hotspots (spherical geometries) to the 3D model, define the radius of each sphere associated with a specific hotspot, customise the colour of the hotspot geometry, and finally edit the text information linked to the hotspot. Once at least one hotspot is defined, the wizard automatically adds a toolbar button to enable or disable the hotspot visibility. If hotspots are visible, moving the mouse cursor over their geometry recalls a popup box showing (in the bottom left of the viewer) the text associated with that specific hotspot.



*Figure 3: The annotation step of the 3D configurator wizard. The panel on the right contains the controls for adding/removing spherical hotspots (in red and blue) to/from the 3D model, modifying the setup of hotspot (radius, colour), and editing the information linked to each hotspot. Moving the mouse cursor over one of the hotspots in the 3D scene recalls the linked information through a popup box visible on the bottom left of the viewer.*

Obviously, this 3D wizard is tailored for use with 3D models, and the same five configuration steps might not make sense with the other visual media supported by Visual Media Service (RTI & High-res 2D Images). However, as the same approach could be replicated for other media, moving the RTI & High-res 2D images configurators from single page editors to a multi-step wizard is one of the future development possibilities for improving the Visual Media Service.

## Pilot

The technologies described were also used to support pilots 16.4 and 16.6. An implementation of Visual Media Service supporting the pilots is being defined, and will be populated with the data identified by the pilot cases.

## 3.4 Task 15.2.3 – Visual documentation of an archaeological excavation

This service builds on the EpHEMERA system (Online 3D Database System for Endangered architectural and archaeological Heritage in the south Eastern MEditionRanea area)<sup>7</sup> developed at Cyl to visualise archaeological excavations, ancient buildings and standing structures and their related documentation in 3D [Abate et al. 2017]. The EpHEMERA platform allows the visualisation online of architectural and archaeological models (classified according to a specific type of risk) of 3D resources through a standard web browser; query the database system and retrieve metadata attached to each digital object; extract geometric and morphological information. The service was conceived as a platform open to multidisciplinary research approaches.

<sup>7</sup> <http://ephemera.cvi.ac.cy>

## Technical characteristics

The EpHEMERA online portal is being developed in adherence with Open Access (OA) guidelines for online access to scientific information, and it has been built on top of the open-source Drupal Content Management Framework (CMF). A key feature of the system is managing high-resolution large and complex 3D models on the web. A dedicated Network Attached Storage (NAS) server installed on the “CyTera” HPC infrastructure of the Cyprus Institute allows the system to handle a large amount of 3D data and other media files. The storage system is scalable and expandable, depending on the availability of new 3D assets. With the main aim of supporting 3D models and their relevant data (e.g., text files, pictures, panoramic images, external links, and resources) and to further map their attributes to existing cultural heritage metadata schemas, a Content Model has been developed. Through the database architecture, three-dimensional models, stored in a separate NAS server, are available online via the open-source WebGL-based Potree viewer. Potree<sup>8</sup> is a point-based rendering solution specifically developed for visualising large point clouds using standard web-based technologies [Schütz 2014; Schütz & Wimmer 2015, Shütz 2016]. It provides a responsible interactive viewer that only requires a traditional web browser (WebGL enable), freeing the user from configuration or specific software issues. A list of command utilities is provided to navigate, interact, and perform geometric analysis of the visualised 3D assets. Two different types of users can access the platform with different privileges assigned:

- Non-authenticated (or Anonymous) users: these users have read-only access, and may not modify the settings or configuration. By default, any user who accesses the EpHEMERA service can browse the content, interact with the 3D models, and perform geometric analysis.
- Registered users: Through username/password credentials, authenticated users can log in to the system to create or edit content through a dedicated user interface.

## Reviewing, updating and porting the EpHEMERA platform

Before the EpHEMERA platform could be incorporated into the ARIADNE service platform, the system needed revision. It was decided that the implementation of the service would be based on user comments. As reported in the D15.1 - *Mid-term interim report on ARIADNEplus services*, a survey was opened to the research community<sup>9</sup> to obtain feedback on the platform and suggestions for additional tools. The the survey results were then analysed, and a list of feedback and possible new components suggested by the users was prepared.

The second phase of work consisted of implementation of the new components requested in the survey and analysing what was already available and feasible in terms of tool and instrument development for the interaction with the content. The new tools are aimed at helping archaeologists and Cultural Heritage professionals, with minimal or no digital skills, interactively visualise online 3D models of CH assets at risk and analyse them. Moreover, the tools allow professionals to extract information and build management plans for the preservation and conservation of Cultural Heritage sites at risk.

As the EpHEMERA platform uses the Potree viewer, the standard Potree Libraries were extended to develop new functionalities. For the development of the tools, the VueJS front-end framework was used, and bundled using Webpack. The user can choose to employ the first or updated version of the Potree visualiser, according to their needs and the type of Cultural Heritage asset being published and analysed online. These new tools allow users to interact online with both 2D and 3D data.

---

<sup>8</sup> <http://potree.org>

<sup>9</sup> The survey is online at the following link: <https://forms.gle/ZSmthfTt3S1jg6ja8>



## Image gallery

Contemporary and historical 2D images can accompany the 3D models, and these images can be visualised within the 3D web environment. Groups of images can be uploaded by the user/administrator through the 'administrator panel' and create an image gallery that enriches the visualisation of the uploaded 3D model (figure 4). Moreover, the user/administrator can align the images on the 3D model according to the angle from which they were taken. This way, once the image is selected by the Anonymous user, it is rendered on top of the model following the same perspective. Beyond the image gallery designed for visualising 2D static images, another element has been developed to visualise images related to the terrain and landscape around the Cultural Heritage asset (e.g., GPR images). In this case, the collection of images includes embedded, real-world geospatial coordinates. Specifically, these 2D images are rendered into 3D space as flat 3D objects, allowing them to be oriented in space and behave as a 3D point cloud of the heritage asset. The Administrator/User can link the image with known GPS points thanks to the panel; each image is associated with parameters for the direction (e.g., Easting and Northing) and the elevation.



*Figure 4: Image gallery and image visualized on the 3D model.*

## Image slider

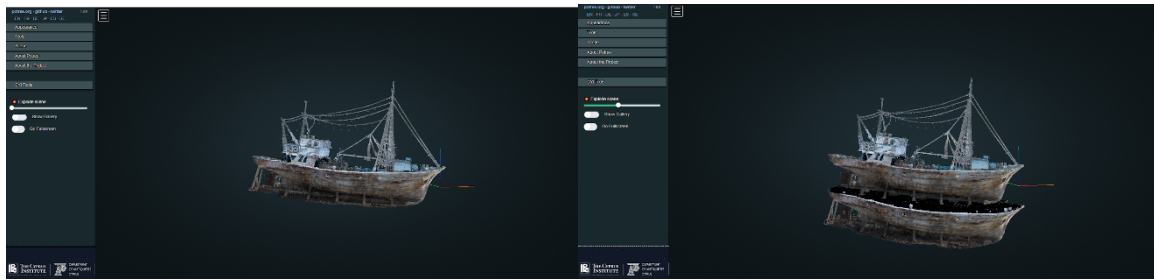
A further tool for the 2D images has been implemented. It is a slider that allows the Anonymous user to change the transparency of the images. Such an instrument allows the user to visualise the image and the 3D model simultaneously. That is particularly useful for archaeologists and cultural heritage professionals to analyse and monitor the state of the asset with regard to conservation. For instance, the comparison between older photos and the 3D representation of the asset's current state allows analysis of the differences that occurred through time and allow the planning of possible interventions.

## Model explosion

A 3D model of an archaeological excavation is composed of several layers. Similarly, a historical building is composed of several architectural components. Often, these assets are digitally visualised as a whole: the layers and the sublevels, usually resulting from different chronological periods superimpositions, are not explicit or not usually visible. For this reason, to understand the composition of a cultural heritage asset, a 3D model explosion tool has been added. When setting up the viewer parameters, the Admin/User can define the axes along which every 3D layer/3D feature can move in the virtual space and the speed of the 'de-composition' movement. Using a slider, the Anonymous user can activate in real time the de-construction of the CH asset 3D point cloud, according to the



directions previously set up by the User/Administrator further away from the centre of the image (0,0,0) (figure 5).



*Figure 5: Visualisation of the cultural heritage asset as a whole (left) and “de-constructed” through the use of the Model explosion tool.*

## Metadata description

Along with the 3D model, the user can add all the metadata information describing the Cultural Heritage asset using an associated web form. More than 45 fields and a dedicated taxonomy connected to the customised Content Model were created. This is linked with the 3D model, providing a data management system for querying the database and retrieving metadata attached to each digital object. The index card of every 3D model is subdivided into four categories describing the physical and technical characteristics of the monument. The ‘physical object’ category describes the main physical characteristics of the monument, including metadata about the monument's name, date, type, dimensions and location. The ‘description’ category gives the user a detailed architectural description of the exterior and interior of the monument (figure 6). Regarding the technical characteristics of the 3D monument, the category ‘Digital Representation’ includes metadata about the digital provenance: the model's creator, the digitisation technique, input and output format, types of equipment used and the survey date. Specifically, the structure of the metadata web form is based on the STARC metadata schema [Ronzino et al. 2012; Vassallo et al. 2016], a proprietary schema further developed and shared by the archaeological community (e.g., CARARE). The user can also upload images of the monuments to create the profile picture of the monument in the collection using the platform design template. Finally, the 3D model and all metadata about the model will appear on the database front-end and be accessible publicly online.

It is worth underlining that a set of metadata concerning the category “Standing structures” under the ARIADNEplus aggregation task and describing the resources currently uploaded and displayed in the EpHEMERA database is currently being mapped to AOCat for aggregation in the ARIADNE portal. As mentioned, the platform publishes several archaeological and cultural heritage sites at risk. This link creates access not only to the metadata integrated into the ARIADNE platform but also to the original data, 3D models in this case, and the specific ARIADNE service to interact with them.

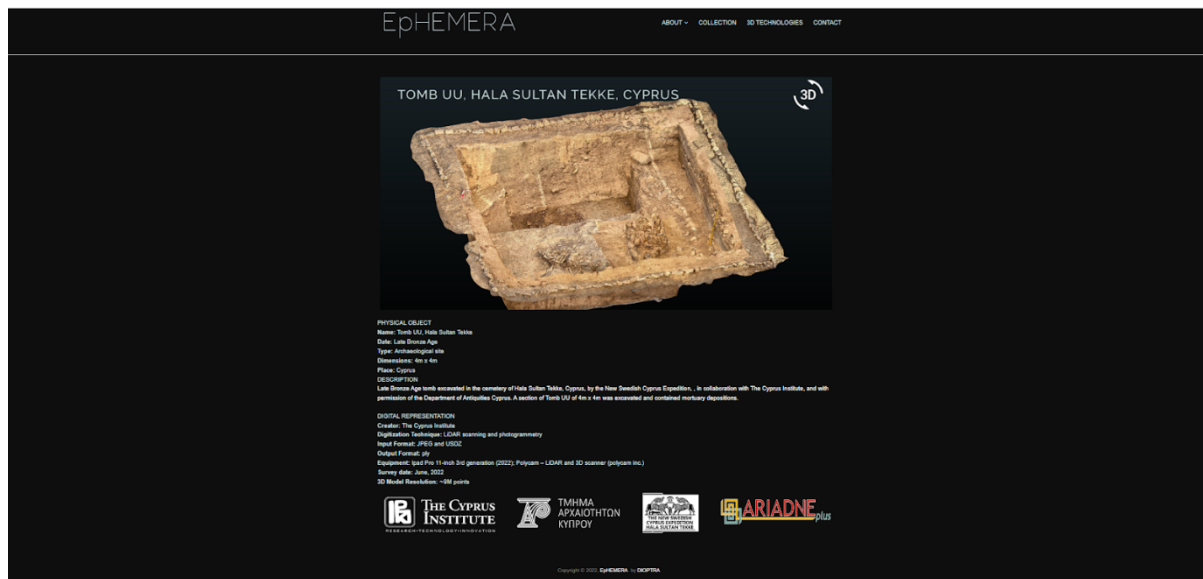


Figure 6: Metadata related to the 3D model of the endangered CH asset.

## Integration of the EpHEMERA platform into the ARIADNE services environment

The next step after the implementation is the integration of the EpHEMERA platform into the ARIADNEplus service environment. Since the EpHEMERA platform is already hosted and integrated in the Cyprus institute infrastructure, an EpHEMERA-Plus instance was created to be hosted by the ARIADNE Infrastructure. While the existing EpHEMERA uses Drupal in a traditional VM environment, a new system was created using an open-source stack consisting of Django, a high-level Python Web framework, Postgres database, Bootstrap, an open-source CSS framework, the Potree viewer and running on virtualised Docker containers.

The installation in the ARIADNE Infrastructure is now being finalised, but in the interim, external users can publish their models using the existing EpHEMERA system. The requirements are listed in the ARIADNE services web-page under 'Services for Humans/ EpHEMERA', including: working with a point cloud (texturised or not); having good Data quality (resolution), being geo-referenced (yes or no), having an Image gallery (yes or no) and provision of metadata. The Administrator makes sure the model(s) are clean and in a logical size and then they are uploaded to the system with all the necessary metadata provided by the creator. A snapshot of the EpHEMERA-Plus version of the system is below (figure 7).

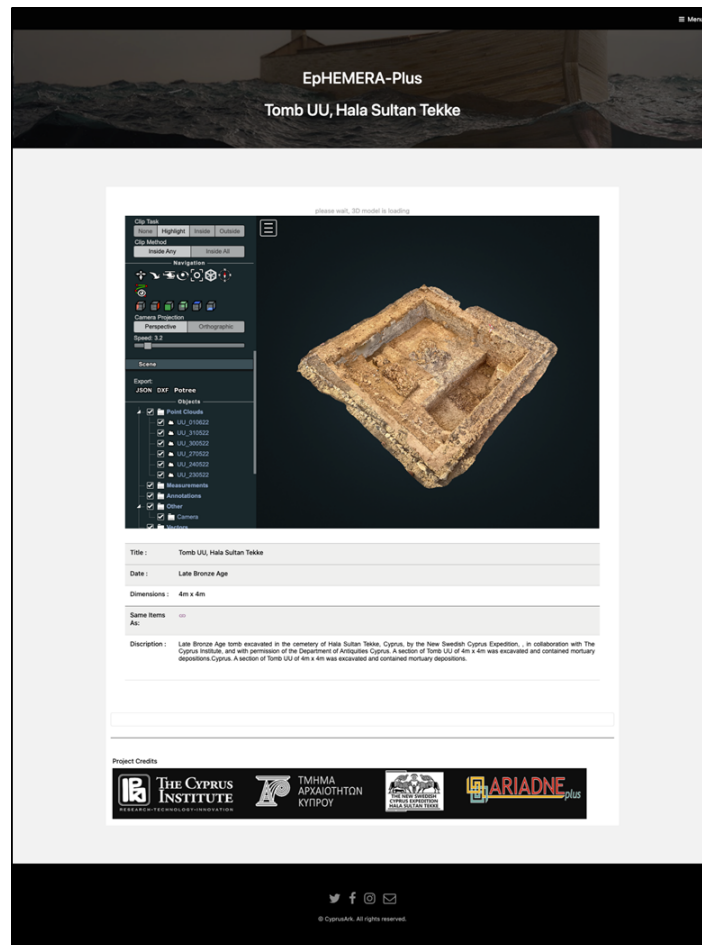


Figure 7: The EpHEMERA-Plus version of the system.

## Pilot

The technologies described were also used to support a pilot developed within Task 16.4. The platform was used to upload data from an external user to test and prove the efficiency of the service for sharing, visualising and analysing archaeological excavation data in 3D as a pilot study. An article to publish the set-up methodology, the survey, its outcomes, the implementations and the final results is currently in preparation. The tutorials created in the previous phase accompany the pilot case.<sup>10</sup> Moreover, a collaboration between the current Task and Task 16.4 included the development of a survey based on task performance dedicated to individuals reviewing the entire pilot procedure in the EpHEMERA platform.

## 3.5 Task 15.3.1 – Archaeology text annotation

As work on this task progressed, it was realised that the task output, a vocabulary-based annotation tool that makes suggestions (based on SKOS vocabularies), overlaps to a large extent with work on temporal NLP (Named Entity Recognition) for Task 15.4. Therefore, it was decided to combine the

<sup>10</sup> A copy of the text documents can be found at the following link: [https://drive.google.com/drive/folders/1gpZ7jZm0QCxHC7JJ9tjlv9ME\\_rLT-VF0?usp=sharing](https://drive.google.com/drive/folders/1gpZ7jZm0QCxHC7JJ9tjlv9ME_rLT-VF0?usp=sharing); video tutorials are available on the YouTube channel: <https://www.youtube.com/channel/UCgkxhBZfM0KFSku2KtQY9tA/featured>

tasks in an integrated development framework. Accordingly work on the annotation tool is reported under Task 15.4 below. This also permits the annotation tool to be documented within the same Jupyter Notebook format employed for T15.4.

### 3.6 Task 15.3.2 – Image annotation

The experience and prototype code and interface developed for the TSS project<sup>11</sup>, i.e. the visualisation and annotation of ancient lead tags has been ported to the OpenLIME library (see section 15.2.1 for details).

On top of the visualisation engine we have developed a Scalable Vector Graphics (SVG) annotation layer support, both for interactive visualisation and online editing. SVG is an image format which is especially suitable for annotations due to its flexibility (CSS selectors can be used to change its properties, such as size, colour and visibility), it is optimal for simple primitives such as rectangles, circles, icons, lines, etc. and because it is vector-based, file sizes are relatively small. In contrast, raster formats for annotations (PNG + transparency) present a number of drawbacks: for example, larger file sizes, and it is particularly difficult if not impossible to efficiently highlight or select single annotations in larger collections. The library extensions to support annotations have been developed during the creation of three pilot projects.

#### Alcazar stone

The **Musée d'Histoire de Marseille** holds several large inscribed stones, discovered in the ancient city harbour. A single very large relightable image of one of these stones has been generated and is now visible using OpenLIME library. Given the difficulty to interpret and display the inscribed stones, the OpenLIME editing tools were used to trace the depictions using the relightable image. An online exhibition has been set up in order to highlight some of the most iconic depictions (boats, griffons, heads).

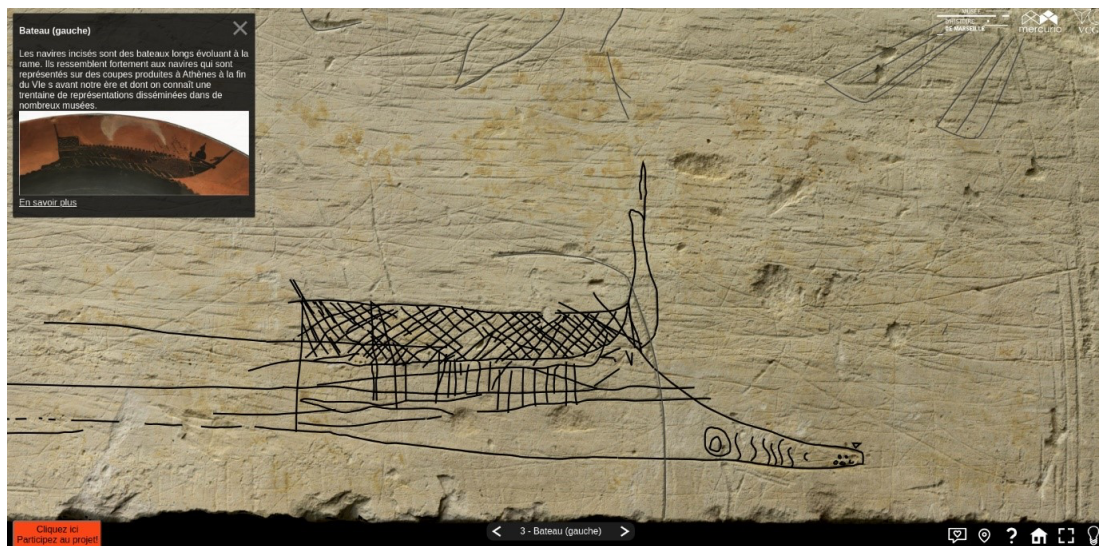


Figure 8: Online viewer for a large inscribed stone found in the port of Marseille, in the depiction a ship.

<sup>11</sup> <http://tss.isti.cnr.it>



This pilot project presented some challenges in terms of customisation and creation of the user interface. The annotations needed to support:

- Easily customisable highlighting: here the annotations get darker and thicker when selected through css selectors.
- Given the scale of the artefacts, annotations needed to have different representations depending on the zoom level. In this case the graphical representation is gradually replaced by a pin when zooming out.
- PHP backend was developed for online editing (drawing, metadata, etc.) of the annotations. Annotations are stored in a custom JSON format, for simplicity.

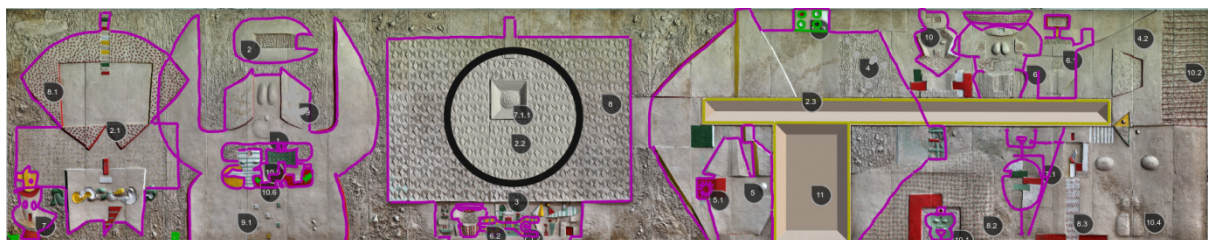
The prototype was hosted and developed in cooperation with <https://mercurioimaging.com/> and is now visible at this link: <https://mercurio3d.fr/alcazar>

## Nivola Sandcast

The viewer was also used for **Nivola's Olivetti Sandcast**: a very large concrete monument, created using the 'sand casting' technique, now hosted in Nivola Museum. It was a large multi-disciplinary project targeting the physical reproduction and virtual documentation and exploration of the Olivetti Sandcast.



*Figure 9: Nicola's Olivetti sandcast on the left, a viewer with lens, annotations and information is on the right.*



*Figure 10: The complete set of annotations on the sandcast.*

We added features to the annotation model and the annotation editor interface allowing users to draw over the model, place pins, and control all rendering, viewing, and lens positioning settings associated with the annotation. Moreover, annotation description and hierarchical annotation ordering have also been defined. SVG annotations could also be created externally using Inkscape or Krita and later imported into the viewer.



Figure 11: OpenLIME annotation editor, customised for the sandcast viewer.

Multilevel annotations are now supported: in the picture below, blue arrows indicate movement at the same abstraction level, while red arrows provide details on demand. The dotted arrows are not authored, but implicitly added.

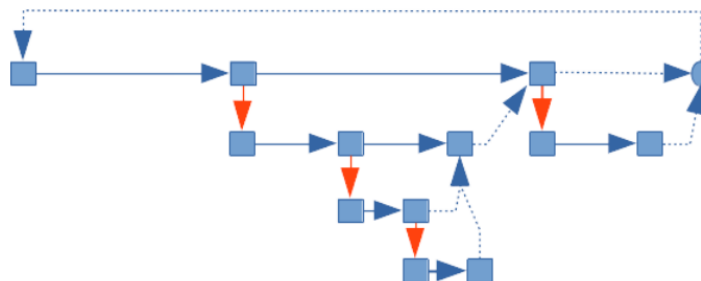


Figure 12: Multilevel annotation structure.

Further details about this work has been published in the 20th Eurographics Workshop on Graphics and Cultural Heritage [Ashan et. al].

## Stele di Nora

**Stele di Nora** is an inscribed stone retrieved in Nora (Sardinia) of particular cultural importance as it includes the first known mention of the name Sardinia in the Phoenician alphabet.

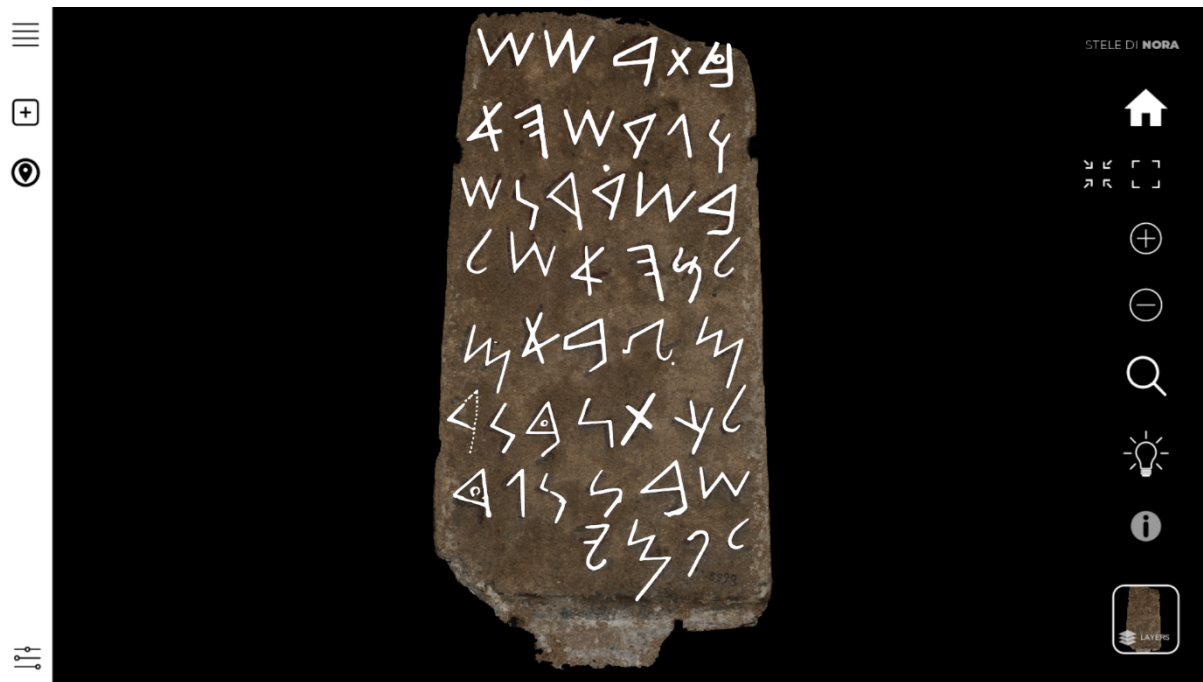


Figure 13: Annotations on the Stele di Nora online application.

In cooperation with Center for Advanced Studies, Research and Development in Sardinia - CRS4 (Moonisa Asham), we worked on a museum exhibition with the goal of faithfully reproducing the visual appearance of the stone using BRDF layers, and annotating the writing in order to highlight the letters and provide an entry point to additional information.

The main contribution for this project was the improvement of the OpenLIME interface code to support additional customisation for icons and tools, and a better annotation editor interface.

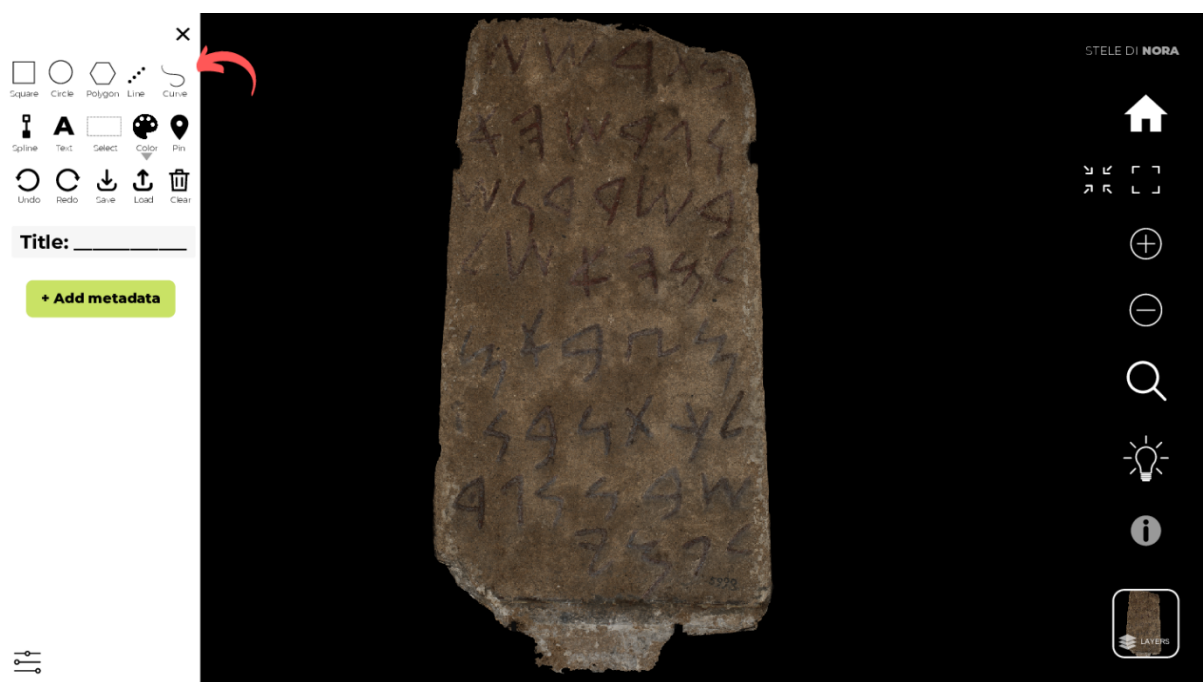


Figure 14: New OpenLIME annotation editor interface.



## Annotations and machine learning

Graphical annotations such as the Stele di Nora example can be used with machine learning for assisted segmentation and then for character recognition and classification. The annotation format has been designed to be easily exported to machine learning friendly datasets.

Manual annotations were traditionally performed on pencil drawings (often derived using the stone rubbing technique). Using analogue drawing can be difficult however, as they were not created in machine learning-friendly formats, and some processing is therefore necessary.

The lead tags of the *Tesserarum Sisciae Sylloge* [Lamé et al., Radman-Livaia] collection provide a good example of handwritten annotations and the challenges involved in using them in machine learning applications:

- Background is not well separated from the letters (shape of the tags and holes).
- Letters are very irregular, and shape is not consistent.
- Text lines are irregular.
- Dotted lines are used for older layers of writings and need to be separated.

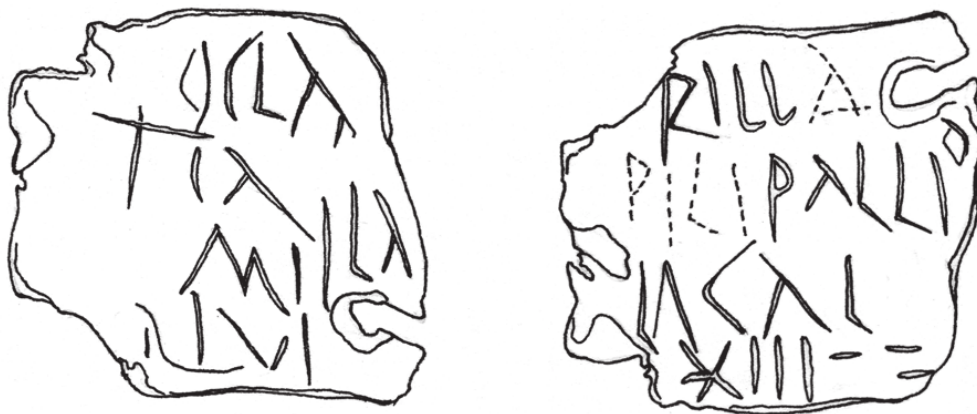


Figure 15: Drawing of a lead tag inscription.

## Tokenizer

To expedite the processing of these unstructured annotations, a (Qt based) application was developed to allow assisted segmentation of the one thousand lead tags in the TSS collection.

The goal of Tokenizer is to convert the existing annotations into an open annotation format, connect the single letters with the existing annotations, and prepare a dataset suitable for machine learning.

The input consists of the digitised drawings of the label and the text of the existing transcriptions, Tokenizer identifies the connected components, discards the tag contours and very small markings, assigns a box to each component and matches them to the existing transcription. This is the starting point for the user. The user can then:

- remove unwanted parts (select + canc)
- group letter parts in a single letter (select + ctrl-G),
- separate the letters from the background deleting some pixels and ungrouping the resulting connected components.
- reorder the letters to match the transcription
- correct the transcription and mark each letter as uncertain or damaged





*Figure 16: Initial automatic segmentation on the left, final work on the right.*

The result can be easily exported in PDF format, and the letters exported separately in PNG format, suitable for machine learning training and classification.

01-40

- NOTE: letter U (V) is uncertain and probably damaged, how to mark it



*Figure 17: PDF export of the segmentation.*

From the user point of view, i.e. as far as epigraphists and palaeographers are concerned, an annotation tool which allows the extraction and documentation of letters, more precisely their shapes, is a tremendous asset when one needs to analyse different handwriting on documents or to study the development of letter forms over a certain period (depending on the chronological time frame of the inscribed documents). This tool does not currently exist however, and any kind of palaeographical statistical analysis related to letter forms would rely on visual observation as well as time consuming manual annotation of observable differences in the shapes of letters belonging to documents discovered within the same archaeological context (such as is the case with Siscia's lead tags or, for instance, Egyptian papyri and/or ostraka originating from the same sites).

## 3.7 Task 15.4 – Implementing the ARIADNEplus text mining and NLP services

There are various strands to this work which are outlined below.

### General archaeological NLP services

This task builds on the NLP outcomes from the preceding ARIADNE project (eg Binding et al. 2019; Vlachidis & Tudhope 2022) and a Cooperation Agreement with ARIADNEplus Associate Partner University College London. A set of archaeological Named Entity Recognition NLP pipelines developed within the GATE<sup>12</sup> framework (Cunningham et al. 2013) for ARIADNE, were reconfigured and deployed for easier use on the GATE cloud<sup>13</sup>. The same set of services and pipelines is available for the European Language Grid<sup>14</sup>. Services are provided in Dutch, English and Swedish for general archaeological Named Entity Recognition and some services for dendrochronological entities. All services follow a KOS-based information extraction technique driven by a combination of hand-crafted rules and input from domain vocabulary, either via SKOSified thesauri (as with the English and Dutch pipelines) or made available internally to the project (as with the Swedish pipelines). The services are work in progress and as with all NLP applications, there may be false positives and missed entities (e.g. less common patterns or variant spellings) and the services should be considered as suggestions or recommendations. The open-source software is available via a source code repository<sup>15</sup> and requests for further refinement continue to be open.

The services are currently available for English, Dutch and Swedish languages. The English NLP pipeline delivers named entities of archaeological context, physical object, material, time appellation and structure. The named entities are linked to concept labels of the National Cultural Heritage Thesauri (UK)<sup>16</sup>. The Dutch NLP pipeline delivers named entities of archaeological context, physical object, material, time appellation and structure. The named entities are linked to concept labels of the Rijksdienst voor het Cultureel Erfgoed (RCE) Thesauri (NL)<sup>17</sup>. The Swedish pipeline delivers named entities: archaeological context, physical object, material, time appellation and structure. The Dendrochronology pipelines deliver named entities of archaeological elements, wood material, sample, and date in the same three languages. They also annotate phrases with different weights based on the number of entities types that are distinctly mentioned, in order to (optionally) highlight rich sections of the document for this purpose. The named entities apart from the Date, are linked to concepts of the Getty AAT Thesaurus<sup>18</sup>.

Example screenshots from the six NER services are given below (Figures 18-23). The entity legend on the right shows the entities selected for visual highlight according to the colour key shown.

---

<sup>12</sup> <https://gate.ac.uk>

<sup>13</sup> <https://cloud.gate.ac.uk/shopfront#tagged=Archaeology>

<sup>14</sup> <https://live.european-language-grid.eu/>

<sup>15</sup> <https://github.com/avlachid/Multilingual-NLP-for-Archaeological-Reports-Ariadne-Infrastructure>

<sup>16</sup> <https://www.heritagedata.org/>

<sup>17</sup> <https://thesaurus.cultureelerfgoed.nl/>

<sup>18</sup> <https://www.getty.edu/research/tools/vocabularies/aat/>



Figure 18: Dutch Archaeology NER example screenshot.

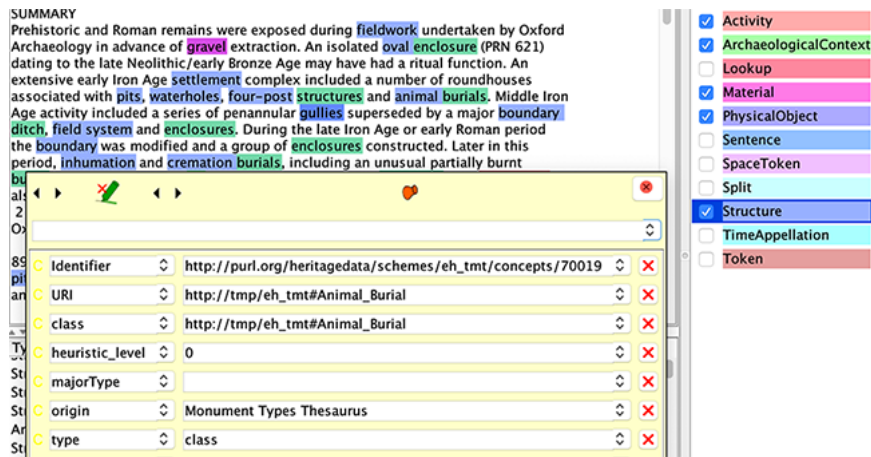


Figure 19: English Archaeology NER example screenshot.

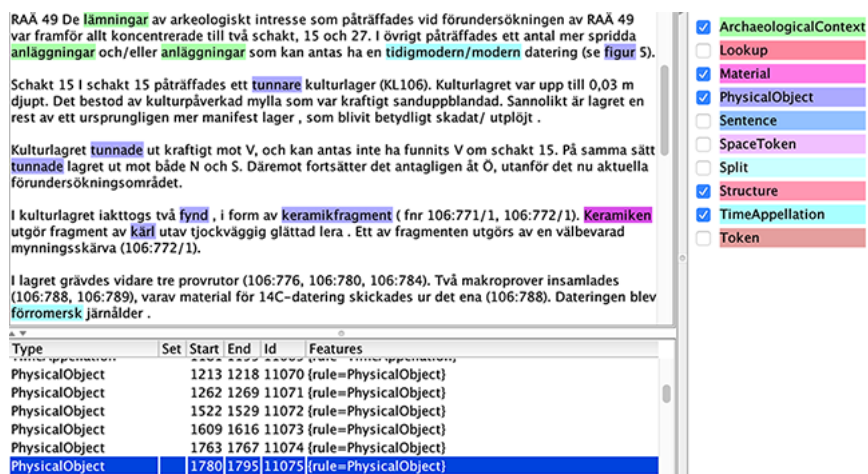


Figure 20: Swedish Archaeology NER example screenshot.



## Temporal services for archaeological data and reports

A second strand of work on automatic extraction of temporal archaeological textual information builds on previous work at USW and overlaps to some extent with ARIADNEplus contributions to WP5 on tools for the aggregation pipeline. There have been two parallel approaches (normalisation and NER).

### Temporal normalisation services

One family of services performs normalisation based on the wide set of date patterns and variant expressions for time spans commonly used in archaeology. Archaeological dataset records often use a textual expression of dating rather than absolute numeric years for the dating of artefacts. These textual data values can be in a variety of formats and different languages. There can be prefixes present such as *Circa*, *Early*, *Mid*, *Late* - and suffixes such as *A.D.*, *B.C.*, *C.E.*, *B.C.E.*, *B.P.* that may influence the dates intended. This situation presents potential problems for temporal comparison of records in a single dataset, but also introduces wider data integration issues, as illustrated in Table 1.

Category	Language	Expression
Ordinal century	Dutch	Begin 11e eeuw voor Christus
	English	Circa Second Century BC
	French	Début du 11e siècle avant JC
	German	Frühes elfte Jahrhundert v
	Italian	XV secolo d.C.
	Norwegian	Tidlig ellevte århundre e.Kr.
	Spanish	Principios del siglo XI d.C.
	Swedish	Tidigt elfte århundrade f.Kr.
Year span	Welsh	Canol y 15fed ganrif
Year span	English	1450-1460
	English	1485-86
Single year (with tolerance)	English	C. 1485
	English	1540±9
	English	AD400+
	English	400 AD
Decade	English	Circa 1860s
	Italian	intorno al decennio 1910

Category	Language	Expression
Century span	Welsh	1930au
	English	5th – 6th century AD
	Italian	VIII-VII secolo a.C.
	Welsh	5ed 6ed ganrif
Month and year	English	July 1855
	Italian	Luglio 1855
	Welsh	Gorffennaf 1855
	English	Summer 1855
Season and year	English	Summer 1855
	Italian	Estate 1855
	Welsh	Haf 1855
	English	Summer 1855
Named periods (from lookup)	English	Georgian
	English	Victorian

*Table 1: Different language temporal expressions by category.*

If normalised date spans can be generated then the data integration problem, resulting from different expressions for the same date span, can be addressed. A few examples in English, French and Italian illustrate various categories of temporal expression with the date span:

1450-1460	1450	1460
1485-86	1485	1486
vers les années 1860	1860	1869
intorno al decennio 1910	1910	1919
Early 2nd Century	101	140
début du IIe s. apr. J.-C.	101	140
Circa Second Century BC	-200	-101
vers le IIe siècle av. J.-C.	-200	-101
XV secolo d.C.	1401	1500
intorno a VI sec. d.C.	501	600
Victorian	1837	1901
Rinascimento	1350	1550

The (Python implementation) services are based on regular expression pattern matches and lookups from standard authority lists, with output as start/end years (in ISO 8601 format), relative to Common Era (CE). Centuries are defined to range from year 1 to 100. The resulting normalised numeric date spans for temporal expressions are offered as additional attributes so as not to preclude application use of the original temporal expression. The languages supported are Dutch, English, French, German, Italian, Norwegian, Spanish, Swedish, Welsh.



The open-source software is available via a source code repository<sup>19</sup>. A suite of tests using the Python *unit test* framework are also provided, covering the various categories of year span textual expressions in each supported language.

A version of this service (available from the same Github repository) is employed in the ARIADNEplus aggregation workflow and this use is outlined in D4.3 (Section 4.4) and described in detail in D5.2 (Section 4.2). It is employed where temporal expressions do not exist in PeriodO, examples are:

- 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”)

The modified version of Year spans imports XML data records from a data provider and adds new derived start/end year numeric attributes to any year span values encountered in the XML data. The result is a new version of the XML data augmented with the additional start/end attributes. This utility is used in combination with the PeriodO elements of the aggregation workflow.

A case study was also conducted for Norwegian language temporal expressions using an early version of the service. A meeting was held with ARIADNEplus partners KHM at the Museum of Cultural History (Oslo) to discuss requirements and Norwegian temporal expressions. In collaboration with KHM, USW elaborated a set of Norwegian textual patterns for time periods and date spans for use in a multilingual date span matching utility. Examples include:

eldre steinalder	9500 BC	4000 BC
tidlig metalltid	2000 BC	300 BC
mellommesolitikum	8000 BC	6000 BC
310-340 e.Kr	310 AD	340 AD
500 f.Kr-500 e.Kr	500 BC	500 AD
Tidlig på 1100-tallet e.Kr.	1001 AD	10040 AD
1950-tallet	1950 AD	1959 AD

## Temporal NLP/NER services

Another set of services provides NLP and in particular Named Entity Recognition (NER) of temporal expressions in text documents. As outlined in T15.3.1, both the annotation tool (T15.3.1) and the temporal NER services are developed in a unified (open source) Python development framework based on the spaCy platform<sup>20</sup>.

The facilities developed are available via a source code repository<sup>21</sup> for long term sustainability. It comprises an open-source library and associated tools for performing rule-based Named Entity Recognition on abstracts and texts relating to archaeological investigations in different languages. These tools utilize the spaCy NLP pipeline architecture as a framework for performing multilingual NER on free text. A suite of spaCy ‘patterns’ has been developed as Python modules together with a series of specialised NER pipeline components to identify and tag various types of temporal entity within passages of free text. Bulk processing scripts and Python notebooks are included to demonstrate usage and to highlight aspects of the available functionality.

<sup>19</sup> <https://github.com/cbinding/years spans>

<sup>20</sup> <https://spacy.io>

<sup>21</sup> <https://github.com/cbinding/rematch2>

## Pipeline components

The components developed are implemented as specialisations of the existing spaCy EntityRuler NER pipeline component (Figure 24). The specialised components perform case-insensitive multilingual matching and the languages currently supported are German, English, Spanish, French, Italian, Dutch, Norwegian, Swedish. Each component consults a language-specific set of spaCy 'patterns' to determine the locations of matching words or phrases in the specified text.

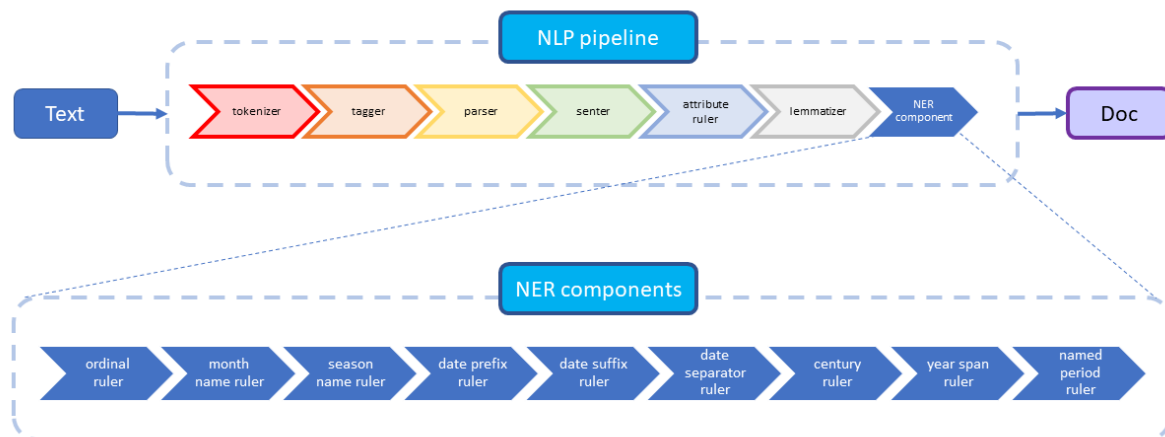


Figure 24: NLP/NER Pipeline in spaCy.

The individual components are described below:

**ordinal\_ruler** e.g. *1st, first, THIRD, 3<sup>RD</sup>* etc. - tagged as *ORDINAL*

Identifies ordinal expressions in text e.g. **15th, nineteenth**. Used in combination with other rulers to identify pattern of ordinal followed by century e.g. **15th century, nineteenth century**. (As an alternative spaCy does have its own built in NER functionality which includes identification of ordinals; this component and its associated patterns were developed prior to adopting spaCy.)

**dayname\_ruler** e.g. *Mon., TUES, Wednesday* etc. - tagged as *DAYNAME*

Identifies day names or their abbreviations in text. Not currently used but remains present and usable as a concrete example showing how to implement a custom multilingual entity recognition pattern ruler.

**monthname\_ruler** e.g. *Jan., FEB, March* - tagged as *MONTHNAME*

Identifies month names or their common abbreviations in text. Used in combination with other rulers to identify pattern of month followed by year e.g. **June 1867**.

**seasonname\_ruler** e.g. *Spring, SUMMER, Autumn, WINTER, Fall* etc. - tagged as *SEASONNAME*

Identifies season names in text. Used in combination with other rulers to identify pattern of season followed by numeric year e.g. **Spring 1867**.

**dateprefix\_ruler** e.g. *Circa, Early, earlier, mid, MIDDLE, Late, later* etc. - tagged as *DATEPREFIX*

Identifies typical dating prefixes commonly associated with years, spans and centuries in text. Used in combination with other rulers to identify patterns of prefixes followed by century or year e.g. **early 1867, circa mid 19th century**.



**datesuffix\_ruler** e.g. *A.D., AD, B.C., BC, B.P., BP etc.* - tagged as *DATESUFFIX*

Identifies typical dating suffixes commonly associated with years, spans and centuries in text. Used in combination with other rulers to identify patterns of year or century followed by suffix e.g. **early 1867 AD, 5th century BC.**

**dateseparator\_ruler** e.g. *TO, /* - tagged as *DATESEPARATOR*

Identifies typical separator characters commonly used to separate years, spans and centuries in text. Used in combination with other rulers to identify patterns of spans of year or century e.g. *15<sup>th</sup> to 18<sup>th</sup> century, 1811 - 1912*

**century\_ruler** e.g. *early 15th century BC to late fifteenth century AD* - tagged as *CENTURY*

Identifies typical expressions of centuries or spans of centuries in text. Note some component dependencies exist e.g. the *century\_ruler* component includes an *ordinal\_ruler* to first identify ordinals (e.g. *"fourteenth", "15<sup>th</sup>"* etc.) in the text. These dependencies are the result of a modular approach; the lower level components can be used on their own but exist to be used in combination allowing for identification of larger entity combinations e.g. *"circa late fourth to mid 5<sup>th</sup> century AD"*.

**yearspan\_ruler** e.g. *early 1300 to late 1350 AD* - tagged as *YEARSPAN*

Identifies typical expressions of years or spans of years in text. Utilises other rulers (prefixes and/or suffixes) to identify more complex patterns e.g. **late 1712 to early 1714 AD.**

**namedperiod\_ruler** e.g. *Bronze Age, Early Medieval, Victorian* - tagged as *NAMEDPERIOD*

This component utilises the PeriodO time period dataset. PeriodO has been adopted by ARIADNEplus as an authority and repository for defining partner named time periods (see ARADNEplus D14.2). When configured with a valid PeriodO authority identifier e.g. *"p0xxt6t"* (Scottish Archaeological Periods & Ages, ScAPA<sup>22</sup>) the component will match against the labels of periods contained within the specified authority. e.g. **Chalcolithic, Early Bronze Age, Antonine.**

**Python (Jupyter) Notebooks** Jupyter Notebooks combine interactive code and descriptive text and are a useful way to demonstrate usage and functionality and to facilitate experimentation. Notebooks describing key elements of the functionality are included with the source code repository (and will continue to be augmented as use cases emerge). The screenshot in Figure 25 shows an extract of a Notebook demonstrating usage of one of the pipeline components. A copy of the Notebooks (based on the underlying spaCy NER pipelines) is available via the ARIADNEplus Lab VRE.

---

<sup>22</sup> <http://n2t.net/ark:/99152/p0xxt6t>

```

# Using a rematch2 component
import spacy
from spacy import displacy
from rematch2 import components

# use a predefined spaCy pipeline, disabling the default NER component
nlp = spacy.load("en_core_web_sm", disable = ['ner'])
# add rematch2 component(s) to the end of the pipeline
nlp.add_pipe("namedperiod_ruler", last=True, config={
    "periodo_authority_id": "p0kh9ds"})
# process example text using the modified pipeline
# (text from https://www.archaeologydataservice.ac.uk/archives/view/1004810/)
txt = '''
Eight trenches were excavated across the 0.93ha Site. Previous archaeological works immediately to the north of
Site in the early 1990s, including field walking and targeted excavation, recorded worked flint and a pit possib
the Late Iron Age, while other investigations in the wider area have identified remains of Late Bronze Age, Earl
Saxon date. Evidence for truncation of the deposit sequence was noted across the development area, with the Site
having been levelled and drained prior to the establishment of turf pitches, with topsoil present only as a very
sterile subsoil and brickearth deposits. The evaluation identified a single pit of Late Bronze Age to Early Iron
archaeological features were located during the works.
'''

doc = nlp(txt)
# highlight identified entities in the text
displacy.render(doc, style="ent")

```

Eight trenches were excavated across the 0.93ha Site. Previous archaeological works immediately to the north of the Site in the early 1990s, including field walking and targeted excavation, recorded worked flint and a pit possibly dated to the **Late Iron Age** , while other investigations in the wider area have identified remains of **Late Bronze Age** , **Early Iron Age** and Saxon date. Evidence for truncation of the deposit sequence was noted across the development area, with the Site area potentially having been levelled and drained prior to the establishment of turf pitches, with topsoil present only as a very thin turf horizon overlying sterile subsoil and brickearth deposits. The evaluation identified a single pit of **Late Bronze Age** to **Early Iron Age** date. No other archaeological features were located during the works.

Figure 25: Illustrative screenshot from Jupyter Notebook with NER pipeline examples.

A preliminary version of the NER service was applied to a corpus of French language grey literature reports received from ARIADNEplus partners INRAP as part of the pilot study for WP16 (Task 8). The data consisted of 11 ZIP archive files containing a total of 481 French language reports from preventative archaeology (2018 to 2021, across the whole French territory covering all periods) in PDF format, plus an accompanying XML metadata file describing the reports, containing rich summaries worded differently to the equivalent report abstracts. It was decided to perform NER on these textual summaries in addition to the work on the PDF report abstracts as this might potentially yield additional information.

Patterns to match are expressed in a spaCy compatible form. An example pattern is shown below:

```

{
  "label": "CENTURYSPAN",
  "pattern": [
    { "LOWER": { "REGEX": "^[mdclxvi]+|\d+)(er?|[eè]me)$" }},
    { "OP": "?", "LOWER": { "REGEX": "^s(\.|iècle)?$" }},
    { "LOWER": { "REGEX": "^(au|--)]$" }},
    { "LOWER": { "REGEX": "^[mdclxvi]+|\d+)(er?|[eè]me)$" }},
    { "OP": "?", "LOWER": { "REGEX": "^s(\.|iècles)?$" }},
  ]
}

```

This example will match a 'span' of centuries. The pattern locates ordinals expressed either as Roman numerals e.g., "XVIIIe" (18th) or as ordinal numbers e.g. "5ème" (5th) optionally accompanied by the word "siècle" (century) or its frequently encountered abbreviation "s.", then separated by either a hyphen or by the word "au" (to). Incorporating regular expressions within the patterns allowed us to cater for a number of different valid free-text variants without resulting in a combinatorial explosion of possible patterns, e.g., "VIe-IXe siècles" (6th-9th centuries), or "VIe siècle au IXe siècle" (6th century to 9th century). This pattern also enforces lowercase text comparison for more flexible (case insensitive) matching.

Not all reports contained abstracts and some reports consisted predominately of images. A total of 309 text abstracts were successfully extracted from the PDF reports and processed. In addition, a total of 352 abstracts were processed from the XML metadata file. All abstracts were processed using the previously described patterns to locate temporal terms and phrases within the text. Results were written to individual CSV files (per report) and to a composite JSON file (covering all reports). Results were also rendered in a HTML format for each report to visually markup and show the identified entities in their correct context within the abstract (Figure 26).

L'intervention s'est scindée en deux parties qui se sont déroulées en parallèle : l'étude du parement extérieur de l'aile sud du boulevard d'artillerie et l'étude du rempart square Hyacinthe Lorette. A l'instar de ce qui avait été observé de 2013 à 2016 **YEARS** lors de la fouille du boulevard, il apparaît que l'ouvrage est parfaitement homogène. En revanche, le rempart plus au sud a fait l'objet de plusieurs campagnes de travaux dont la plus ancienne ne remonte pas avant 1458 **YEAR** d'après les datations radiocarbone. Par ailleurs, l'étude documentaire se poursuit. Elle s'est focalisée cette année sur la fin du **DATEPREFIX** Moyen-âge **PERIOD** autour du châtelet et a apporté de nombreuses informations inédites qui permettent de préciser les hypothèses émises lors des différentes campagnes. En outre, de nouvelles données sur les liens entre les fortifications et les logis à l'arrière des remparts du XIIIe siècle au XVIIe siècle **CENTURYS** ont été mises en exergue, notamment en soulignant la présence d'une maison des Templiers dont il reste aujourd'hui les baies géminées sur la courtine à l'ouest du châtelet. Le fait d'annexer une partie de l'hôtel de Trélo au milieu du **DATEPREFIX** XVe siècle **CENTURY** pour pallier l'absence d'équipements au sein du châtelet lors de l'arrivée du gouverneur est à souligner et confirme l'hypothèse émise en 2013 **YEARS** d'un corps de logis détaché, disparu depuis. 8 Inrap · Fouille programmée Rennes, Ille-et-Vilaine, Les portes Mordelaises

Figure 26: Visual representation of NER output from example INRAP PDF report abstract.

The example in Figure 26 is the result of processing the abstract of a PDF report. The displayed annotated text is generated by the spaCy library's named entity visualizer "displaCy", and shows identified custom entities *years*, *year*, *dateprefix*, *period*, *centurys* and *century* in context within the text. Named periods drew on INRAP's ARIADNEplus PeriodO collection supplemented by other material. The example in Figure 27 illustrates entities identified in the text of the metadata summary field of an XML record:

au **Bas-Empire PERIOD** , ont été relevées sur le site. La période la mieux documentée étant le **deuxième âge du**  
**Fer PERIOD** , période qui fait l'objet de très nombreuses découvertes dans le quartier de la plaine Saint-Roch, depuis  
 plus d'un siècle. Les traces d'occupations les plus anciennes sont datées du **Bronze final PERIOD** IIIb où une aire  
 d'habitat semble s'être développée sur ce secteur, et s'inscrivent dans un recouvrement limoneux, dans lequel des  
 artefacts datant du **Chalcolithique PERIOD** ont été recueillis. La période qui succède à ces premiers niveaux  
 d'occupations est la **fin du DATEPREFIX** **second âge du Fer PERIOD** , rompant ainsi un hiatus chronologique  
 s'échelonnant du **VIIIe s. CENTURYSpan** au **milieu du DATEPREFIX** **Ile s. CENTURYSpan** **av. notre ère**  
**DATESUFFIX** . Le **début de DATEPREFIX** cette occupation gauloise a été daté du **milieu du DATEPREFIX** **Ile s.**  
**CENTURYSpan** pour perdurer sans interruption jusqu'au premier tiers du **Ier s. CENTURYSpan** **av. n. ère**  
**DATESUFFIX** . Trois phases caractérisées par des réaménagements du site ont été distinguées. Les deux premières  
 semblent regrouper des activités domestiques assez similaires au sein d'un îlot « d'habitats » délimité par un fossé. Les  
 puits font leur apparition à la transition entre le **Ile et le Ier s. CENTURYSpan** **av. n. ère DATESUFFIX** . Un vaste  
 épandage de morceaux d'amphores mêlé à des galets, constituant ainsi un niveau de circulation, caractérise la fin des  
 occupations gauloises sur le site et semble répondre à de nouveaux besoins. Ce type d'aménagement, comme  
 l'interruption de l'occupation constatée pour une grande partie du **Ier s. CENTURYSpan** **av. n.è DATESUFFIX** .

Figure 27: Visual representation of NER output from example INRAP XML summary.

Note here particularly how multiple variations in the way AD is expressed in the free text have been successfully matched (e.g. “*av. notre ère*”, “*av. n. ère*”, “*av. n.è*”).

Patterns were initially written to take advantage of the part of speech (POS) tagging pipeline component that executes immediately following tokenisation of the text (the “tagger” component in Figure 24 pipeline diagram). An example was to find the word “*période*” followed by an adjective (e.g. “*période gauloise*” – Gallic period). Although this worked the approach proved too restrictive in certain cases because the spaCy POS tagger would sometimes tag the parts of speech incorrectly leading to seemingly valid period expressions being incorrectly excluded from the results. Writing specific patterns for known periods resolved this but then led to a proliferation of very similar patterns just to accommodate the potential for POS tagging errors. These pattern POS rules were relaxed for the WP16 pilot. However, the potential exists in future applications for the parts of speech to be utilised in patterns where appropriate.

Some adjustments were made to patterns to accommodate the possibility (regardless of correctness) of differences in the usage of diacritic marks e.g., *époque* vs. *epoque*. These patterns might be further elaborated in future work to consider the possibility of distinguishing patterns that should be flagged with attributes for grammatical and/or spelling errors. Another fruitful avenue for future work (with domain expert, native language speakers) would be to further elaborate contextual patterns for lone year values that distinguish dates from other lone year measurements.

## Vocabulary Annotation Tool

The previously described facilities concerned temporal annotation. As outlined in Task 15.3.1, a tool has also been developed using the same platform and similar techniques to extract vocabulary-based subject annotation suggestions from archaeological texts. The Vocabulary Annotation Tool (VAT) is a Python application to locate and tag vocabulary terms within free text and to output suggested subject annotations in a range of formats. The tool (currently) incorporates selected vocabularies of the FISH Terminology Working Group including:

- FISH Archaeological Objects Thesaurus<sup>23</sup>
- FISH Archaeological Sciences Thesaurus<sup>24</sup>
- FISH Building Materials Thesaurus<sup>25</sup>
- FISH Monument Types Thesaurus<sup>26</sup>

spaCy compatible patterns were generated representing each of the terms or phrases present in the vocabularies. These patterns make use of NLP features in the spaCy pipeline, such as parts of speech (POS) tagging and lemmatization. POS tagging allows the patterns to match the grammatical sense of the words (e.g., to match “lodging” as a noun rather than a verb, or “ground” as a noun rather than a verb or adjective). Lemmatization introduces flexibility for matching free text where the words may not be specified in precisely the same way as in the controlled vocabulary (e.g., singular/plural forms). Currently the vocabularies used are English language.

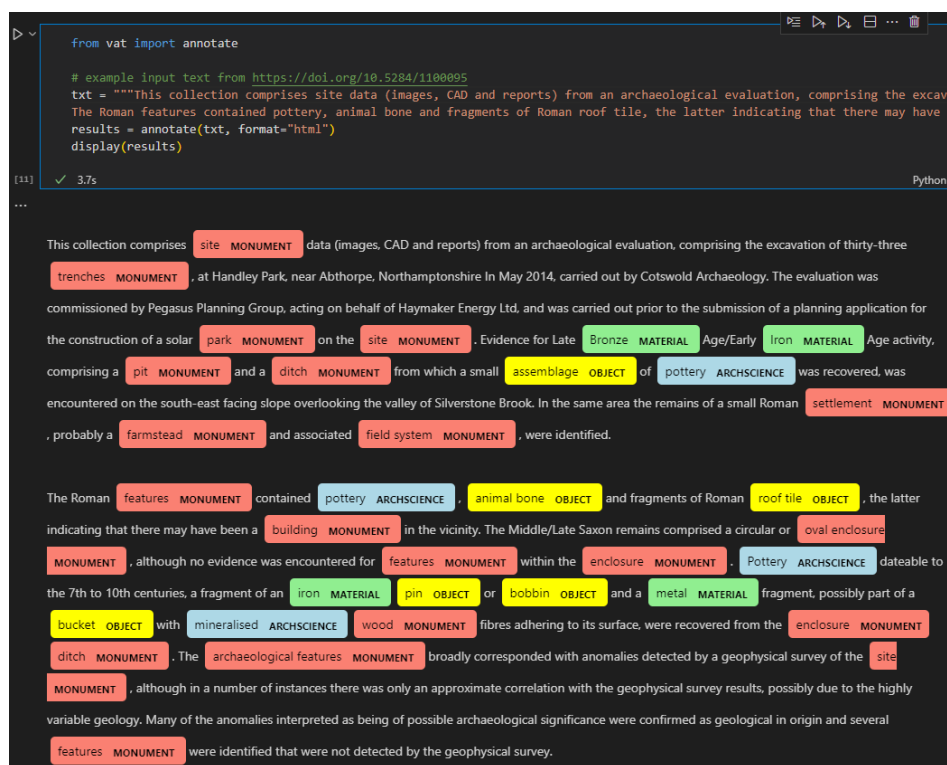


Figure 28: Example usage of the Vocabulary Annotation Tool via a Python notebook.

In addition to the HTML representation shown in Figure 28, the tool can also output results in other formats such as delimited data (CSV/TSV), RDF and JSON. As the output subject suggestions originate

<sup>23</sup> [https://heritagedata.org/live/schemes/mda\\_obj.html](https://heritagedata.org/live/schemes/mda_obj.html)

<sup>24</sup> <https://heritagedata.org/live/schemes/560.html>

<sup>25</sup> [http://purl.org/heritagedata/schemes/eh\\_tbm](http://purl.org/heritagedata/schemes/eh_tbm)

<sup>26</sup> [https://heritagedata.org/live/schemes/eh\\_tmt2.html](https://heritagedata.org/live/schemes/eh_tmt2.html)

from Linked Open Data (LOD) vocabularies they are associated with URI identifiers allowing for output of subject indexing that facilitates cross searching via vocabulary concepts.

### 3.8 Task 15.5 – Implementing space-time services

GeoPortal Service is a new REST service designed to manage complex space-temporal Documents. It has been released as a component of the gCube [Assante M. et al. 2018] framework, and SmartGears<sup>27</sup> powers it.

It has been designed to manage the publication lifecycle of complex space-temporal documents called Projects, supporting their materialisation and indexing in different platforms (Databases, Catalogues, OGC Services) while maximising reusability.

Project management is based on Use Case Descriptor (UCD) documents, which define for each collection:

- Projects metadata schema
- Role-based access to projects (and metadata sections) and operations
- Configuration of involved Project handlers (Lifecycle Management, Event handlers, GUIs..)

The service's main features are

- Documents Publication Management:
  - CRUD (Create, Read, Update, Delete) operations, publication lifecycle, UCD - based validation
  - Role-based access to resources and metadata sections
  - Cooperative and extensible workflows
- Automatic Indexing & aggregation based on UCD annotations:
  - GIS indexes (both centroids & multi-polygon), with dynamic aggregation based on zoom level
  - Text indexes
- Extensible Document model:
  - Generic meta-model
  - Custom extensions of the model
- Extensible behaviour
  - Core engine based on a meta-model
  - Publication management implemented by an extendable set of document Handlers
  - Workflow configured in UCD by declaring LifeCycle Managers and involved Handlers
  - Consumption of documents configured in UCD
  - Configurable exploitation of gCube facilities via dedicated Handlers (e.g. SDI Materialisation and Indexing)

Users can define their collection as a UCD in the VRE and then manage their Projects publication lifecycle both via gateway's GUIs and by interacting with the service's API (both JAVA and/or REST).

The service provides a set of Handlers which can be declared and configured in the UCD (defaults apply) to be invoked in the different phases of the publication lifecycle.

Handlers are implemented as plugins, thus dynamically loaded by the service allowing for

- Support of custom implementation of Handlers

---

<sup>27</sup> <https://wiki.gcube-system.org/gcube/SmartGears>



- Service capabilities defined at deployment time (allowing for different solutions for different situations)

The provided java client library allows for dynamic model Serialisation / Deserialisation of custom-provided Java model classes, allowing for easy integration of different systems.

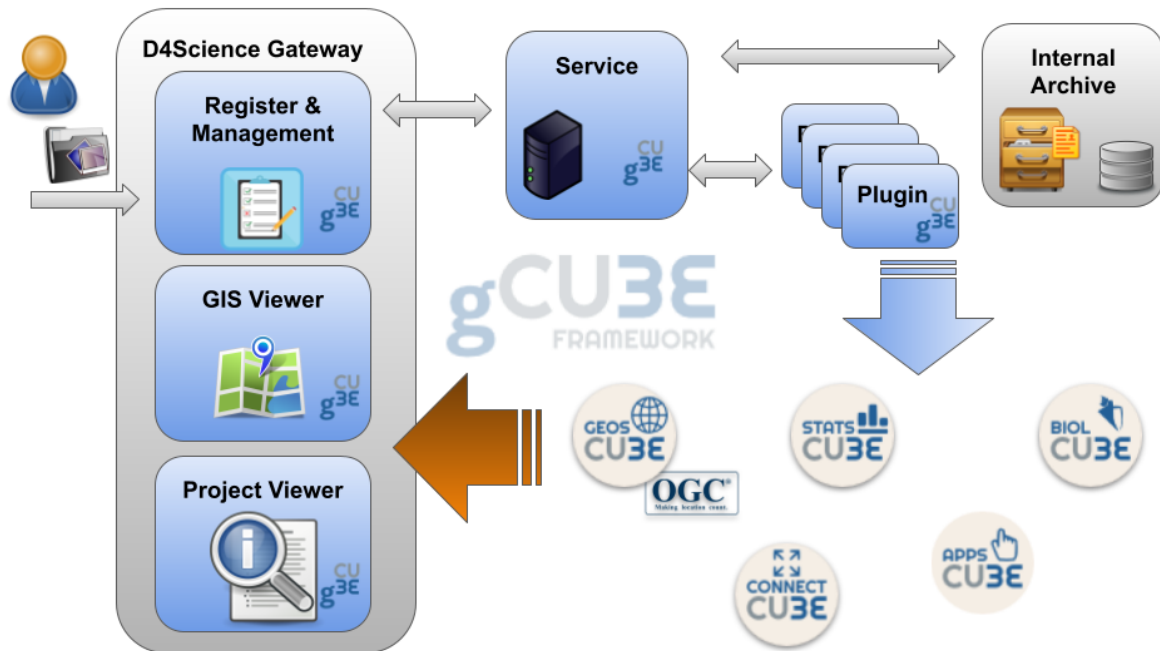


Figure 29: GeoPortal integration with gCube facilities.

### Extensible model and Use Case Descriptors

Projects are the main entities managed by the application. They represent complex documents consisting of both metadata and datasets provided by the user. Metadata schemata are defined in Use Case Descriptors (UCD), and every managed collection of Projects refers to a specific UCD.

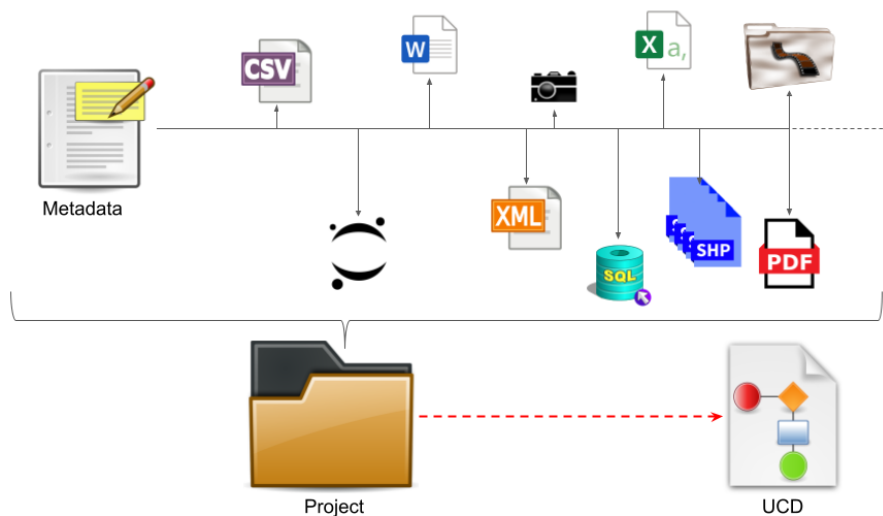


Figure 30: Projects and UCD.

This approach allows for the definition of tailored Use Cases to manage a heterogeneous collection of documents. Moreover, it supports evolution by intrinsically enabling the extensions of any UCD.

Projects Metadata contains the following information:

- Basic:
  - ID, version, UCD id
- Publication/accounting info:
  - creation and last editing information
  - Access type info
- Lifecycle Management:
  - Current PHASE
  - Last operation STATUS
  - Last operation report messages
- Identification References:
  - Spatial reference (e.g. BBOX, Centroid)
  - Temporal reference
  - Catalogues reference
- Relationships towards other Projects
- Custom document
  - Custom Metadata fields
  - Associated FileSets
    - gCube Storage references
    - Fileset Materializations (e.g. SDI GIS, DBMS..)

Projects attachments have been implemented as FileSets to support:

- Isolated parts of the same resource (e.g. zips)
- Multi-file resources (e.g. layers comprising of .shp and .shx files)

Projects Metadata are stored on a Document database (powered by MongoDB) allowing for dynamic querying and schemata, while FileSets are stored in gCube StorageHub and then materialized in the right engine (e.g. SDI GeoServer) in order to let clients (both users and applications) consume them properly.

Use Case Descriptors define collections of Projects in a VRE, specifying:

- Document structure
  - Fields (cardinality, type, constraint, defaults, indexing, GUI declaration)
- Role Based rights to Projects
- Project Relationships definition
- Handlers declaration and configuration

Project Management is performed by the service which relies on the available Handlers to implement Lifecycle Management STEPS and EVENTS. In particular:

- CRUD operations:
  - Performed by CORE engine
  - Lifecycle Management supports Events (e.g. ON\_DELETE, ON\_CREATE)
- Lifecycle Management STEPS and EVENTS: operations depending on the configured Lifecycle (E.g. PUBLISH, SUBMIT\_FOR\_REVIEW)



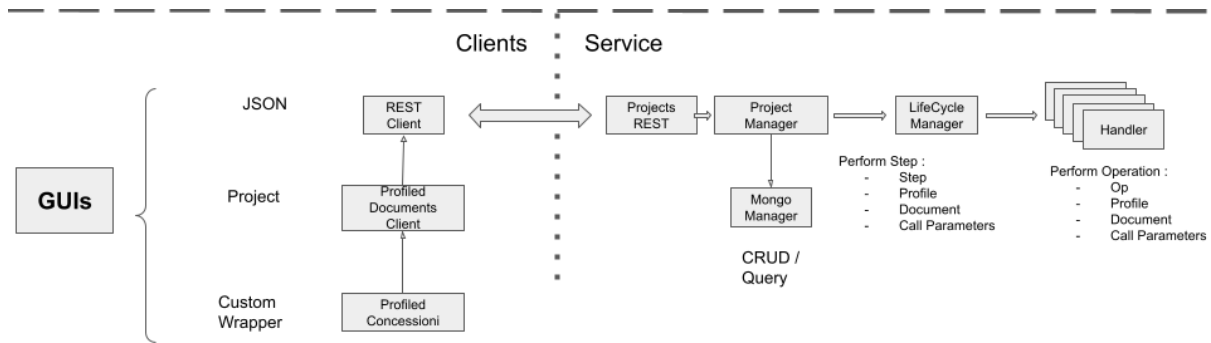


Figure 31: Operations support by Lifecycle Managers.

Concurrency is granted by implementing the following Lock Management



Figure 32: Lock support in STEP invocation.

## Handlers

Handler refers to components that can be optionally configured in UCDs, controlling the behaviour of the use case in a certain scenario. They are designed to manage Projects to:

- perform a specific task (e.g. materialisation of GIS Filesets)
- support / extend the basic functionalities (e.g. Lifecycle management)
- consume Projects (e.g. GIS Collection Viewer)
- offer user-oriented functionalities on top of the system (e.g. Data entry and management GUI)

Handlers can be implemented in different ways, varying from service-side plugins to portlets and web components.

Service capabilities depend on the Handler(s) plugin(s) deployed along with the service itself, allowing for dynamic and tailored behaviour in different VREs.

The service exposes a REST capabilities interface that describes the available implementations along with supported operations and expected parameters. In this section, we present the Default Handler set implemented to satisfy community needs.

In this section, we present the Handlers developed during the project and provided along the geoportal service as a default configuration.

## Lifecycle Management

Lifecycle Management plugins provide support for STEPS and EVENTS. Each Lifecycle can be defined as an oriented graph of various PHASES, in which the Project can be at any given time (the default initial PHASE is DRAFT). We call STEP the Lifecycle operation that can be invoked on a document, optionally changing its PHASE. We then call EVENT an operation that is triggered upon invoking a STEP.

The following Lifecycle Management handlers are provided by defaults. Role-based rights to invoke the supported STEPS are configured in the collection UCD.

### 2-PHASE Lifecycle

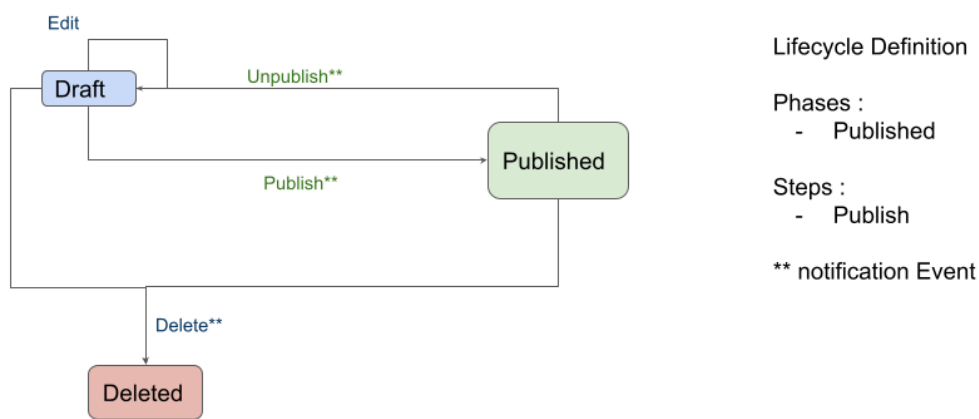
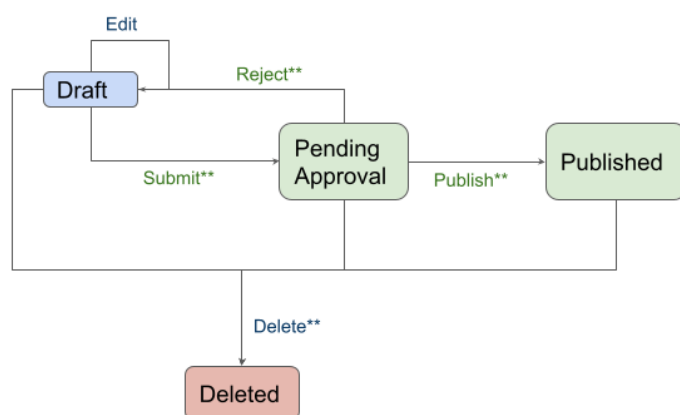


Figure 33: 2 PHASE Lifecycle allowing for basic project publication management.

In a 2-PHASE Lifecycle, Projects are:

- created as DRAFT and optionally EDITED
- a DRAFT can be Published (triggering configured materializers and indexers)
- a PUBLISHED project can be UNPUBLISHED
- Projects can be deleted

## 3-PHASE Lifecycle



Lifecycle Definition

Phases :

- Pending Approval
- Published

Steps :

- Submit
- Reject
- Publish

\*\* notification Event

Figure 34: 3 PHASE Lifecycle allowing for moderated publication management.

The 3-PHASE Lifecycle handler is built as an extension of the 2-PHASE Lifecycle, reducing code management costs.

In a 3-PHASE Lifecycle, Projects are:

- created as DRAFT and optionally EDITED
- a DRAFT can be SUBMITTED for approval (triggering configured materializers and confidential indexers)
- Projects PENDING for APPROVAL can be Rejected (becoming DRAFT) or PUBLISHED (triggering public indexing)
- Projects can be deleted any time

### SDI Materialisation

SDI Materialisation plugin allows for the publication of GIS layers in gCube SDI. Currently, the following formats are supported:

- Esri SHP files (with optionally included SHX)
- GeoTIFF.

It just needs to be configured with the FileSet JSON paths to be checked for GIS datasets, and it automatically:

- evaluates the best dataset format between available files in the Fileset;
- pushes the selected files into gCube's SDI GeoServer in the VRE exploiting gCube Data Transfer service;
- publishes the related layers exploiting GeoServer REST interface;
- enhances the Project metadata with the generated materialization information and OGC endpoints exposed by GeoServer.

## SDI Indexing

SDI Indexer plugin allows for the management of special GIS layers published in gCube's SDI. These special layers represent all registered Projects in a particular lifecycle PHASE. They allow the browsing of projects using GIS operations (e.g. spatial filtering, rendering). Each feature in the layer represents a Project and contains:

- the project ID;
- the geometry (currently supported format is the Project's centroid) that can be configured to be derived from:
  - Fileset SDI Materializations' bounding box;
  - specific coordinates declared in project's metadata;
- a set of additionally configurable fields evaluated from project's metadata.

## Geoportal GUIs

The Geoportal GUIs allow users to create and manage new Projects in the Geoportal Service (Data Entry GUI) and consume them via Web-Map Viewer (GIS Collection Viewer).

### Data Entry GUI

The Data Entry GUI discovers the Use Case Descriptors defined in the VRE. Each Use Case Descriptor discovered contains a series of handlers defining the behaviour and managed by the Data-Entry UI manager, they are:

- [DATA\\_ENTRY\\_GUI](#) for the definition of the Document as a configuration containing the list of sections (1..N) of the Document. Each section consists of the (pointer to) *gCube Metadata Profile* specifying the fields of the section, its cardinality (whether it can be replicated or not), and how the section must be transformed in the JSON Document result that must be published in the Geoportal Service;
- [DATA\\_LIST\\_GUI](#) specifying the fields of the Document that the *UI Data List management* has to display to the end user in the Data List facility, and what fields must be used for searching and ordering facilities;
- [WORKFLOW\\_ACTION\\_LIST\\_GUI](#) specifying the list of STEPS that the *UI workflow management* can call on a Project and they must be provided to the end user in the *Workflow Action List UI* in any phase of the Workflow.

The Data Entry GUI provides the facilities for:

- submitting via web forms created dynamically according to UCD [geoportal-data-entry](#) definition a new Project in the Geoportal Service via Data-Entry UI;
- navigating the Project created, listing, searching and ordering them according to UCD [geoportal-data-list](#) handler definition, via Data-List UI;
- managing the Workflow actions according to UCD [workflow-action-list](#) handler definition, via Data-Management UI;
- creating and navigating the relationships among the Projects, according to UCD [relationshipDefinitions](#) handler definition, via Data-Management UI. For example, type of (qualitative) relations can be: *follows*, or *precedes*. They are configurable via the UCD.

## gCube Metadata Profile

The Document (Section) Structure is defined via gCube Metadata Profile. A gCube Metadata Profile is an XML-based schema for defining a 'generic item' (e.g. a Spatio-temporal dataset) and its attributes.

Each “generic item” defined by the Profile represents a ‘type’ consisting of (metadata) fields that can be descriptive, spatial, temporal and so on.

A gCube Metadata Profile is an XML model having the format:

```
<metadataformat type="YOUR TYPE HERE">
  <metadatafield>
    <fieldName>Name of Metadata Field</fieldName>
    <mandatory>true|false</mandatory>
    <dataType>String|Time|Time_Interval|Times_ListOf|Text|Boolean|Number|GeoJSON|File</dataType>
    <maxOccurs>N|*</maxOccurs>
    <defaultValue>default value</defaultValue>
    <note>[the note is shown as a suggestion in the insert/update metadata form provided by Publisher Widget]
  </note>
  <vocabulary isMultiSelection="true|false">
    <vocabularyField>field1</vocabularyField>
    <vocabularyField>field2</vocabularyField>
    <vocabularyField>field3</vocabularyField>
  </vocabulary>
  <validator>
    <regularExpression>a regular expression for validating values</regularExpression>
  </validator>
</metadatafield>
</metadataformat>
```

- *Spatial* data - can be specified by using the *dataType* values:
  - (i) *GeoJSON*<sup>28</sup>;
  - (ii) *File* - by uploading a spatial File (e.g. a shapefile).
- *Temporal* data - can be specified by using the *dataType* values (based on ISO-8601<sup>29</sup>):
  - *Time*;
  - *Time\_Interval*;
  - *Times\_ListOf*.

## GIS Collection Viewer

The GIS Collection Viewer uses the Use Case Descriptors and configurations defined in the VRE. Each Use Case Descriptor with its configuration contains the indexes as a layer with the centroids for querying the published Documents collections via web-map GIS. For each collection (of the layer of centroids) added to the map, it is possible to query the centroid referring to a Project published and Open the Project details. The Project details are displayed by UI into sections separated as defined in the UCD, and each one can allow: (i) discovery of metadata, (ii) access to GIS-layers and their features by Web-Maps, (iii) export of the layers into PNG/JPEG format; (iv) access to images via Gallery facility; (v) access to files. Each section is accessible as OPEN/RESTRICTED according to the access policy defined in the UCD.

Moreover, the GIS Collection Viewer provides the facilities also for:

- searching the Project created (and ordering the results), according to UCD `geoportal-min-data-list` handler definition, via Data-List UI, provided into GIS Viewer;
- navigating the relationships created among the Projects.

---

<sup>28</sup> GeoJSON: <http://geojson.org/geojson-spec.html>

<sup>29</sup> ISO-8601: <https://www.iso.org/iso-8601-date-and-time-format.html>

## Evolution

In the first half of the project, the basic core functionalities and prototypal implementations of lifecycle management were developed to support a basic operational version of Cloud Geoserver described in T15.7. This allowed users to operate the logic and provide important feedback for the further development of the application.

In the second half of the project, the feedback was used to re-engineer the logic and components, allowing us to:

- Maximise reusability through modularity, supporting a larger number of case studies through configurations and extensions, including:
  - Improved client libraries for easy extension of behaviour and model
  - Improved Project model for a more generic support of indexing, materialisation, access, reporting
  - Refactored GUIs
  - Refactored developed document lifecycle into 2 modular implementation that can be easily extended
  - Refactored SDI support into configurable modules for materialisation and indexing
  - Refactored Validation and default evaluation
  - Provided common test cases, extensible logic for faster extension development
- Improve performance and robustness:
  - Improved plugin framework, request management, execution management, implementation management support
  - Improved Project model support to lifecycle management, locking, indexing, relations, accounting, access

## 3.9 Task 15.6 – Query services

To support the necessary query capabilities needed by the ARIADNE Portal, for querying the underlying aggregated partner resources, various querying services has been established.

### Query Services

T3.1 was delivered in collaboration with WP12, including two services for querying the ARIADNEplus Knowledge Base: a full-text index service and a SPARQL endpoint.

The full-text index server is based on OpenSearch<sup>30</sup> version 1.3.2, and is only accessible by the ARIADNE Portal. In the first part of the project, the service was implemented with Elasticsearch<sup>31</sup> v7.4.0. Elasticsearch was initially used for persistence of the metadata of the resources. Elasticsearch was a fit-for-purpose database with a built-in search engine where all metadata is stored as JSON documents. Elasticsearch was developed as a scalable database especially focused on searching large amounts of data in near real-time.

The decision was taken to migrate to OpenSearch due to the licensing of the recent version of Elasticsearch<sup>32</sup>, which is not considered an Open license by the Open Source Initiative<sup>33</sup>. OpenSearch started as a fork from the latest open-source version of Elasticsearch and the migration was therefore

---

<sup>30</sup> <https://opensearch.org/>

<sup>31</sup> <https://www.elastic.co/products/elasticsearch>

<sup>32</sup> <https://www.elastic.co/pricing/faq/licensing>

<sup>33</sup> <https://opensource.org/licenses/>

very low impact. There was no loss of features within the ARIADNE query server and no major adaptations of the clients were needed.

A mapping document<sup>34</sup>, which contains rules of the structure of the documents, has been developed to ensure that all documents in OpenSearch have the same structure and are interpreted in the same way. The mapping document also provides various structures, which facilitate the various search options, described below, which the Portal offers.

The SPARQL endpoint allows performing of semantic queries on the RDF records and is based on a GraphDB<sup>35</sup> server (version 9, free edition). It is publicly available via the ARIADNEplus Lab Virtual Research Environment (VRE) at [https://ariadne.d4science.org/group/ariadneplus\\_lab](https://ariadne.d4science.org/group/ariadneplus_lab) and programmatically accessible as described in the relative documentation prepared by WP12 and linked from the VRE as shown in figure 1.

For the communication between GraphDB and the OpenSearch index server, a publisher component was designed and developed. One instance was deployed in a staging environment and one in a production environment.

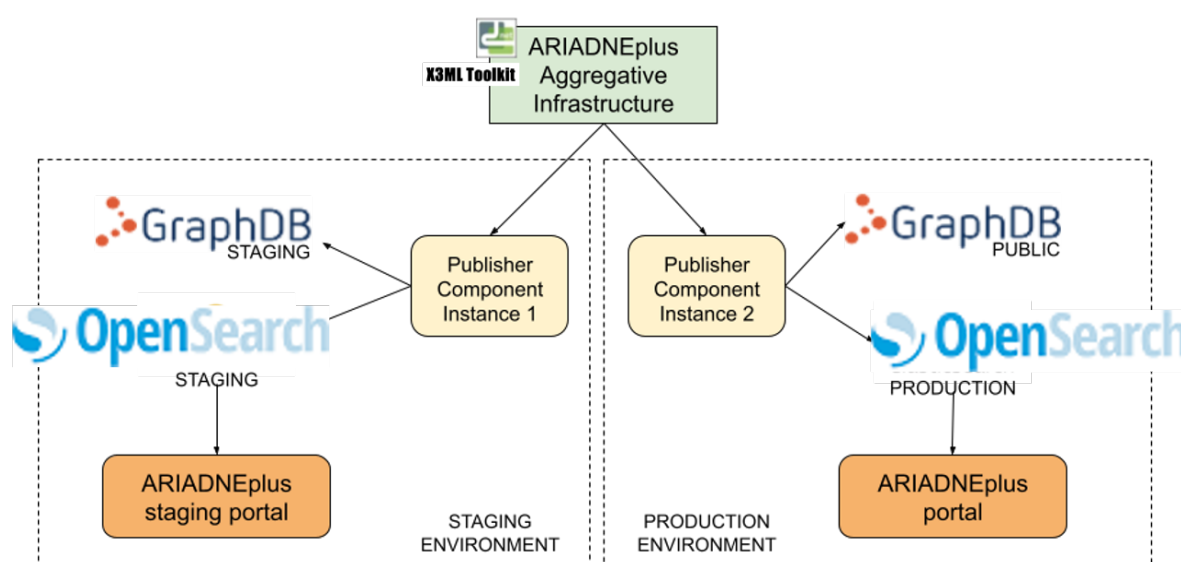


Figure 35: Staging and production environments.

During the second period of the project, the task monitored the performance of the GraphDB servers and the publisher components, increasing the hardware resources and fine tuning their configuration when needed. GraphDB was updated from version 9.0 to 9.8.0. Monitoring is performed continuously by the monitoring infrastructure deployed at CNR premises, based on Prometheus and Grafana.

<sup>34</sup> The mapping document is available here: <https://github.com/ARIADNE-Infrastructure/Index>

<sup>35</sup> <https://graphdb.ontotext.com/>

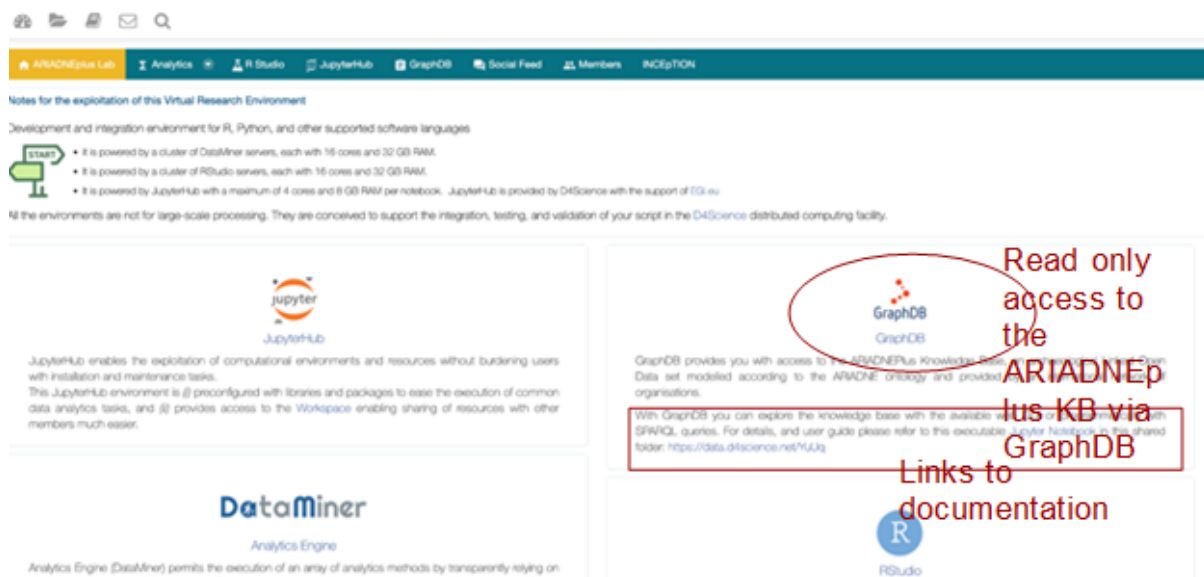


Figure 36: SPARQL endpoint via GraphDB integrated into the ARIADNEplus Lab VRE.

## Multilingual search and query expansion

Based on the functionality of OpenSearch, multilingual query capabilities have been added to enable the full use of the Getty Arts and Architecture Thesaurus (Getty AAT) contents by the portal. With the use of a separate OpenSearch index, containing the full Getty AAT thesaurus and term relations, a multilingual search with suggestions, when selecting Getty AAT Subject as the search parameter has been developed. This feature enables the user to search for resources using any language where Getty AAT has a translation of the term.

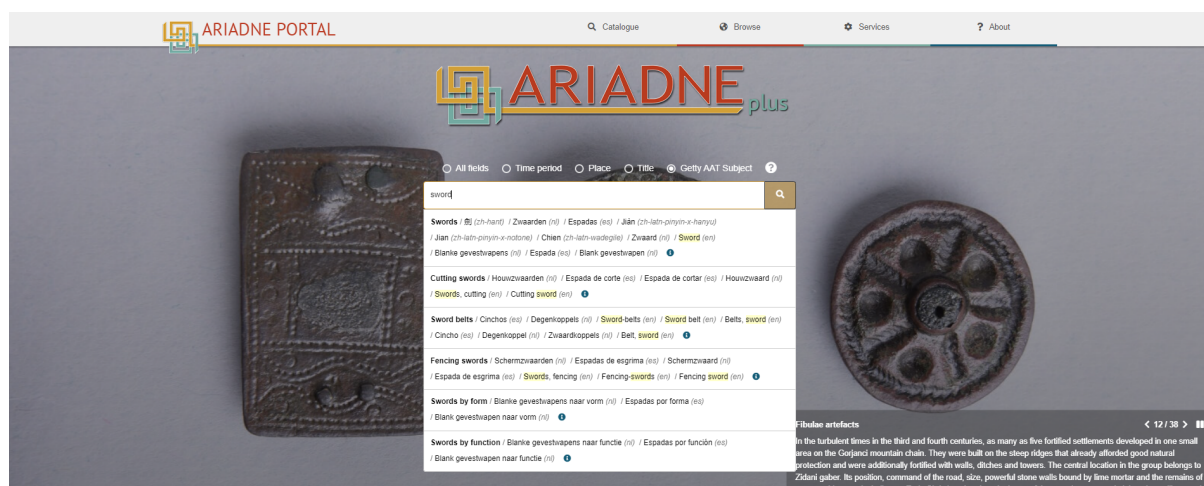


Figure 37: Front page of the portal featuring multilingual suggestions from Getty AAT

Additionally, Getty AAT has been used to enable a hierarchical query expansion of the search query which will include all descendant terms of a found term in the thesaurus. Resources in the portal are tagged with terms from the thesaurus but lack information about terms lower in the hierarchical tree. With the hierarchical query expansion, a search for 'weapon' will return resources directly tagged with



the term 'weapon' and enrich the search result with those resources tagged with terms lower in the hierarchy such as 'bows' or 'clubs'.

### 3.10 Task 15.7 – Cloud geoserver

Cloud GeoServer is a gCube application enabling the publication, access and management of GIS projects consisting of multiple payloads (as documents, images, and datasets), each described with tailored metadata. It is built exploiting the various components developed in 15.5 and configuring them to offer the following features:

- Support for publication lifecycle:
  - Archive complex Data (Meta + Payloads)
  - Enable versioning, workflows, access policies
  - Support several materialisations (GIS, Databases, ...)
  - Manage indexes (Meta catalogues, Index GIS layers)
- Maximise reusability:
  - By exploiting space-time services from Task 15.5
    - Configurable behaviour
    - Generic model
    - Configurable GUIs (Management grid, Insert/Edit Form, Data Viewers)
- External Data Integration:
  - Exploit OGC standards

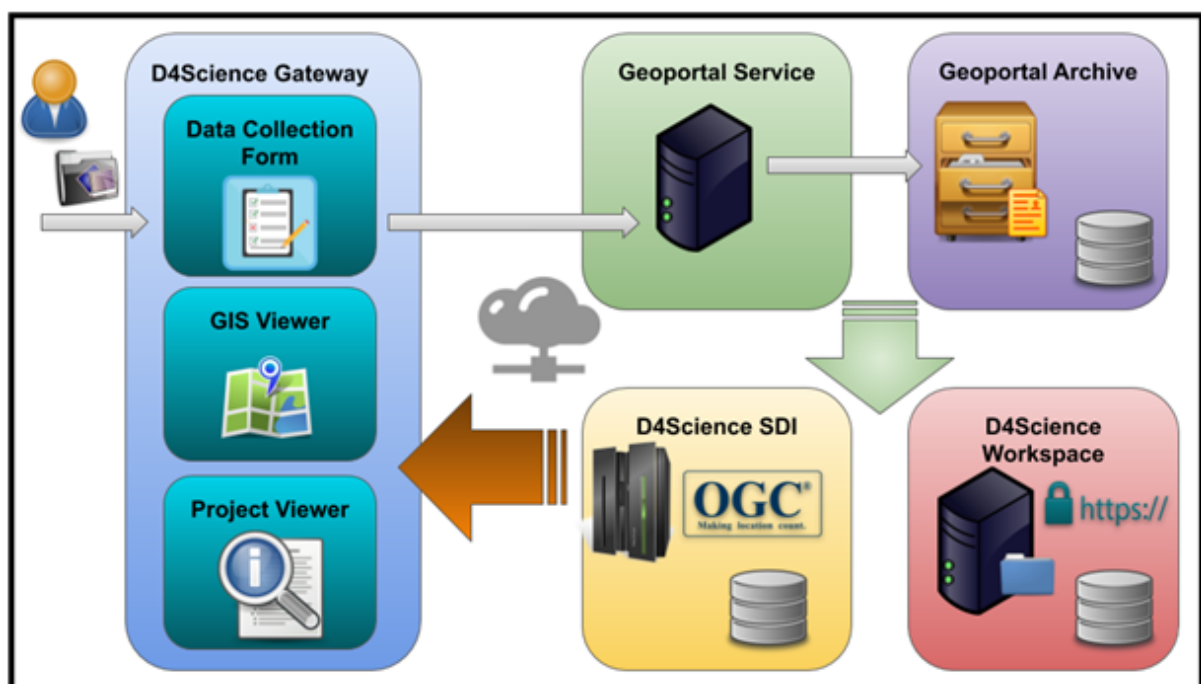


Figure 38: Cloud Geoserver Framework.

The framework provides the following Dynamic GUIs for the VREs hosted by the ARIADNEplus gateway, which use the Use Case Descriptor metadata definition to enable tailored:

- *Data Collection Form* assisting users in publishing GIS projects
- *GIS Viewer* allowing any user to visualise projects on a map
- *Project Viewer* assisting users in accessing information, documents, images, and datasets associated with the GIS project

Dynamic GUIs rely on java client libraries to communicate to the GeoPortal service, managing validation and management of GIS projects.

The service relies on the D4Science Workspace to store, and access attached documents and on the D4Science SDI (Spatial Data Infrastructure) to offer OGC Compliant Services (e.g. WMS, WFS, WCS, etc.). It also uses an internal archive of Projects to manage the publication lifecycle, and for querying purposes.

## Case Studies

### Concessioni

In the first half of the project, a prototype was deployed and operated to manage archaeological excavation projects. Community users were able to use the prototype both in a confidential VRE and from a public user perspective in order to validate the service functionalities and disseminate the activity at the same time.

This has been useful to gather feedback from a direct experience of the user perspectives and needs, thus correcting the initial approach and adding features identified by the community.

The following activities have been performed to serve the case study:

- Creation and operation of 2 VREs: “GeoNa-Proto” and “GNA” (<https://ariadne.d4science.org/web/gna/>), for the delivery of the different evolution of the prototypes
- One public site, accessible at <https://gna.d4science.org/>, with public access to the portal excavation projects
- Development of custom project handler for specific model requirements
- Support activities:
  - Modelling and Configuration
  - BugFixing and Feature development
  - Custom handler for ad hoc logic

The community has supported the activity, resulting in:

- 458 Submitted Projects
- 35 Bug / Support issues
- Various emails and meetings for the discussion of feedback

The following screenshots show some views from the served case study

**GNA data entry facility**

New Project List of Projects

Informazioni di Progetto

Abstract Relazione di Scavo

Relazione di Scavo

Immagini Rappresentative

Posizionamento dell'Area di Indagine

Planta di Fine Scavo

Insert Information for: Informazioni di Progetto \* is required

\* Nome del progetto:

\* Introduzione:

\* Nome Autore, Email, Ruolo:

Repetable field

\* Contributore:

\* Titolare dei dati:

Repetable field

\* Responsabile:

Figure 39: Data Collection Form auto-generated by "Concessioni" Use Case Descriptor.

Show on Map Publication Report Edit Clone Project Publish/UnPublish Project Delete Project

List of Projects Sort by Name - ASC Reload Projects Search for Name type a text...

Name	Introduction	Author/s	Project Start/End Date	Created	Created by	Published with	Pub-UnPub
Elaiussa Sebaste Excavations	Le indagini condotte dalla Missione archeologica della Sapienza presso il sito di Elaiussa Sebaste a partire dal 1999 hanno lo scopo di indagare l'articolazione e le fasi di occupazione della città antica. Le campagne di scavo, della durata di circa due mesi l'anno, vengono condotte in collaborazione con un'équipe di archeologi e restauratori turchi.	Annalisa Polosa (annalisa.polosa@uniroma1.it) Asena Kulstalgirlio (asenakulstalgirlio@hotmail.com); Massimo Braini (massimo.braini@alice.it);	05-08-2017 / 28-10-2017	27-07-2021	annalisa.falcone	Success	Publish report
Deposito geo-paleontologico ex Cava a Filo (Croara, BO)	Le indagini intraprese nel giacimento di ex Cava a Filo a partire dal 2006 sono finalizzate all'acquisizione di nuovi dati crono-stratigrafici, paleoambientali e sulla componente faunistica inerenti al deposito fluvio-carsico, ricco di mammalofauna, che ha costipato ciclicamente le tasche erosive della cavità. Le indagini si sono localizzate nella serie basale del deposito (U.S. 99-100) riferibili all'ultimo come glaciale e delimita Cava Filo-1 (25000-24000 cal BP), Cava Filo-2 (22000-20000 cal BP).	Gabriele Nenzioni, gabriele.nenzioni@comune.sanlazzaro.bo.it, Direttore scientifico, Museo della Preistoria "L. Donini"; Paolo Reggiani, laboratorionem@provincia.it, Direttore di scavo, Museo di Storia Naturale di Venezia;	16-09-2019 / 30-09-2019	25-10-2021	chiara.travaglini	Success	Publish report
Via Appia Antica, V miglio. Mapping the via Appia	Il progetto, in corso dal 2011, ha come oggetto di indagine il tratto della via Appia Antica tra V e il VI miglio. Nel corso degli anni sono state effettuate ricognizioni, rilievi tridimensionali lungo l'intero tratto e saggi di scavo presso i Tumuli c.d. degli Orazi.	Prof. Eric Moorman, direttore di scavo; Dr. Stephan Mols, responsabile di scavo; Dott.ssa Christel Veen, responsabile di scavo;	13-07-2019 / 13-08-2019	26-10-2021	deleone.1542293b2b1d	Success	Publish report
Il sito paleolitico del Cimitero di Atella (Italia, Basilicata, Pz)	Le ricerche sul giacimento paleolitico del Cimitero di Atella (Basilicata, PZ) fanno parte di un progetto più ampio sulle prime dinamiche di popolamento in Italia (Progetto Paléo, scuola francese di Roma). Nella continuità delle ricerche portate avanti da più di 20 anni dal Prof. Borzatti dell'Università di Firenze, il nuovo gruppo di lavoro italo-francese, si è dato l'obiettivo di ottenere nuovi risultati sui modi di vita, il paesaggio e l'ambiente degli uomini che popolavano il territorio attuale di Atella 600.000 anni fa.	Dott.ssa Roxane Rocca, roxane.rocca@univ-paris1.fr, direttrice di scavo (Université Paris 1 Panthéon-Sorbonne, Umr 7041 ArScAn équipe ethnologie préhistorique); Dott. Daniele Aureli, danielereusn1@gmail.com, Vice direttore di scavo (Umr 7041 ArScAn équipe AntET);	01-08-2019 / 31-08-2019	28-10-2021	deleone.1542293b2b1d	Success	Publish report
Il progetto di ricerca e scavo archeologico si concentra su una delle aree più significative e storicamente qualificate del centro di Roma antica la cui adeguata conoscenza è stata, tuttavia,							

Figure 40: Data Management Form for the use case Concessioni.

Search for Any field type a text...

Sort Results

CAN19 (Indagini archeologiche e territoriali nella Lucania Occidentale 2019)

Il progetto si pone come obiettivo quello di indagare il popolamento protostorico della Lucania occidentale, ed in questo particolare caso quello di comprendere al meglio la cronologia di utilizzo e lo scopo, nonché della relazione con l'ambiente circostante e con il comprensorio Lucano, dell'abitato di Cannalichio, sito nel territorio di Castelgrande(PZ), a sud dell'odierno paese di Castelgrande; lo studio stratigrafico del contesto, relazionale al territorio e ed all'ambiente circostante, ci permette di effettuare una ricostruzione del territorio durante la protostoria.

Share Expand

04-08-2019

Open Details

Data Inizio Progetto 15-07-2019

Data Fine Progetto 04-08-2019

Parole chiave Libere

abitato  
insediamento  
Lucania  
emirodria

Figure 41: GIS Collection Viewer and Project Viewer with an auto-generated Project display.



Figure 42: Project Viewer: management of projects files, images, and detailed maps respectively.

## Pilot

The cloud GeoServer is also used to support the GIS pilot described in T16.6. A UCD supporting the pilot will be identified and defined in the new tailored VRE (named Esquiline, <https://ariadne.d4science.org/web/esquiline/>) hosting the cloud GeoServer. The VRE will then be populated with the data and metadata identified by the pilot case.

## 4. References

- Abate, D., Avgousti, A., Faka, M., Hermon, S., Bakirtzis, N., & Christofi, P. (2017). An Online 3d Database System for Endangered Architectural and Archaeological Heritage in the South-Eastern Mediterranean. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Volume XLII-2/W3.
- Ashan, M., Altea, G., Bettio, F., Callieri, M., Camarda, A., Cignoni, P., Gobetti, E., Ledda, P., Iutzu, A., Marton, F., Mignemi, G., Ponchio, F. (2022). Ebb & Flow: Uncovering Costantino Nivola's Olivetti Sandcast through 3D Fabrication and Virtual Exploration. *Eurographics Workshop on Graphics and Cultural Heritage*. doi:10.2312/gch.20221230
- Assante, M., Candela, L., Castelli, D., Cirillo, R., Coro, G., Frosini, L., Lelii, L., Mangiacrapa, F., Marioli, V., Pagano, P., Panichi, G., Perciante, C., Sinibaldi, F. (2018). The gCube system: delivering virtual research environments as-a-service. *Future generation computer systems* doi:10.1016/j.future.2018.10.035.
- Binding, C., Tudhope, D., Vlachidis, A. (2019). A study of semantic integration across archaeological data and reports in different languages. *Journal of Information Science*, 45(3), 364-386. <https://doi.org/10.1177/0165551518789874>
- Cunningham, H., Tablan, V., Roberts, A., Bontcheva, K. (2013). 'Getting more out of biomedical documents with GATE's full lifecycle open source text analytics', *PLoS Computational Biology*, Vol. 9, No. 2.
- Lamé, M., Ponchio, F., Robertson, B., Radman-Livaja, I. (2017). Teaching (Digital) Epigraphy, in Silvia Orlandi, Raffaella Santucci, Francesco Mambrini, Pietro Maria Liuzzo (eds.), *Digital and Traditional Epigraphy in Context*, Proceedings of the EAGLE 2016 International Conference. Rome, Italy, 267-282.
- Lamé, M., Pittet, P., Ponchio, F., Markhoff, B., & Sanfilippo, E. (June, 2019). Heterotoki: Non-Structured and Heterogeneous Terminology Alignment for Digital Humanities Data Producers. *Open Data and Ontologies for Cultural Heritage*. Rome, Italy.
- Ponchio, F., Potenziani, M., Dellepiane, M., Callieri, M., & Scopigno, R. (2016). ARIADNE Visual Media Service: easy web publishing of advanced visual media. *CAA 2015 - 43rd Annual Conference on Computer Applications and Quantitative Methods*, Siena, Italy.
- Potenziani, M., Callieri, M., Dellepiane, M., Corsini, M., Ponchio, F., & Scopigno, R. (2015). 3DHOP: 3D Heritage Online Presenter. *Computers & Graphics*, Volume 52, 129-141.
- Radman-Livaja, I. (2014). *Plombs inscrits de Siscia, Musei Archaeologici Zagradiensis Catalogi et Monographiae*, Vol. IX/1-IX/2. Zagreb.
- Ronzino, P., Niccolucci, F., & Hermon, S. (2012). 'A metadata schema for Cultural Heritage documentation'. *Electronic Imaging & the Visual Arts: EVA 2012 Florence*, 36-41. Firenze University Press.
- Schütz, M. (2014). PotreeConverter - Uniform partitioning of point cloud data into an octree. Retrieved from [http://potree.org/downloads/converter\\_documentation.pdf](http://potree.org/downloads/converter_documentation.pdf)
- Schütz, M., & Wimmer, M. (2015). Rendering large point clouds in web browsers. *Proc. 19th CESC*, 83-90.
- Schütz, M. (2016). Potree: Rendering Large Point Clouds in Web Browsers.

Vassallo, V., Ronzino, P., Damjanovic, U., & Hermon, S. (2016). 'Un processo aperto di comunicazione della conoscenza'. Stanco F. & Gallo G. (eds) *Proceedings of ArcheoFOSS: Free, libre and open source software e open format nei processi di ricerca archeologica: VIII Edizione*, Catania 2013, 91–100. Archaeopress.

Vlachidis, A., Tudhope, D. (2022). A Method for Archaeological and Dendrochronological Concept Annotation using Domain Knowledge in Information Extraction. *International Journal of Metadata, Semantics and Ontologies*, 15(3), 192-203. doi:10.1504/IJMSO.2021.123042

## Appendix A: Template for describing services

EOSC is promoting a profile for describing EOSC Resources (mainly services). This profile is described in detail here:

<https://eosc-portal.eu/providers-documentation/eosc-provider-portal-resource-profile>

This profile describes the ARIADNEplus services between our initiatives and EOSC.

Thus Services are expected to be described by the following attributes:

- **Name\***: Brief and descriptive name of the Service as assigned by the Provider.
- **Resource organisation\***: The name (or abbreviation) of the organisation that manages or delivers the Service, or that coordinates Service delivery in a federated scenario.
- **Resource providers\***: *The list of name(s) (or abbreviation(s)) of Provider(s) that manage or deliver the Service in federated scenarios.*
- **Description\***: A high-level description in fairly non-technical terms of a) what the Service does, functionality it provides and Resources it enables users to access, b) the benefit to a user/customer delivered by a Service; benefits are usually related to alleviating problems (e.g., eliminating undesired outcomes, obstacles or risks) or producing gains (e.g. increased performance, social gains, positive emotions or cost saving), c) list of customers, communities, users, etc. using the Service.
- **Tagline\***: Short catchphrase for marketing and advertising purposes. It is usually displayed close to the Resource name and should refer to the main value or purpose of the Resource.
- **Multimedia**: Link to video, screenshots or slides showing details of the Service.
- **Use Cases**: list of brief descriptions of use cases supported by this Service.
- **Service Classification attributes**:
  - **Scientific Domain**: The branch of science, scientific discipline that is related to the Service. For the list of controlled values to be used see Resource Scientific Domain/Subdomain.
  - **Scientific Subdomain**: The sub branch of science, scientific subdiscipline that is related to the Service. For the list of controlled values to be used see Resource Scientific Domain/Subdomain.
  - **Category**: A named group of Services that offer access to the same type of Service. For the list of controlled values to be used see Resource Category, Subcategory (and Supercategory).
  - **Subcategory**: A named group of Services that offer access to the same type of Service, within the defined Service category. For the list of controlled values to be used see Resource Category, Subcategory (and Supercategory).
  - **Target Users**: Type of users that commissions a Provider to deliver a Service. For the list of controlled values to be used see Resource Target Users.
  - **Access Type**: The way a user can access the Service (Remote, Physical, Virtual, etc.). For the list of controlled values to be used see Resource Access Type.
  - **Access Mode**: Eligibility/criteria for granting access to the Service to users (excellence-based, free-conditionally, free etc.). For the list of controlled values to be used see Resource Access Mode.



- **Contact Information:**
  - **Main contact/Service Owner**
  - **Public contact**
  - **Helpdesk email**
  - **Security contact email**
- **Maturity information:**
  - **Technology readiness Level:** The Technology Readiness Level of the Service. For the list of controlled values to be used see [Resource TRL](#).
  - **Life Cycle Status:** Status of the Service life-cycle. For the list of controlled values to be used see [Resource Life Cycle Status](#).
  - **Certifications:** List of certifications obtained for the Service (including the certification body or URL if available).
  - **Standards:** List of standards supported by the Service.
  - **Open Source Technologies:** List of open source technologies supported by the Resource.
- **Dependencies information:**
  - **Required Resources:** List of other Resources required to use this Service.
  - **Related Resources:** List of other Resources that are commonly used with this Service.
  - **Related Platforms:** List of suites or thematic platforms in which the Service is engaged or Providers (Provider groups) contributing to this Service.
- **Attribution information:**
  - **Funding Body:** Name of the funding body that supported the development and/or operation of the Service. For the list of controlled values to be used see [Resource Funding Body](#).
  - **Funding Program:** Name of the funding program that supported the development and/or operation of the Resource. For the list of controlled values to be used see [Resource Funding Program](#).
  - **Grant/Project Name:** Name of the project that supported the development and/or operation of the Resource.
- **Management information:**
  - Helpdesk page
  - User manual
  - Terms of use
  - Privacy policy
  - Access policy