



D15.1 – Mid-term interim report on ARIADNEplus services

Version 1.0 (final)

2021-01-22

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

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

Author:	Johan Fihn Marberg Swedish National Data Service
Contributing partners:	Alessia Bardi, Consiglio Nazionale delle Ricerche 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, Archaeological Museum in Zagreb Marco Potenziani, Consiglio Nazionale delle Ricerche Marion Lamé, Consiglio Nazionale delle Ricerche Massimiliano Assante, Consiglio Nazionale delle Ricerche Pasquale Pagano, Consiglio Nazionale delle Ricerche Sorin Hermon, The Cyprus Institute Valentina Vassallo, The Cyprus Institute
Quality control:	Stephanie Williams, PIN

Document History

- 2021-01-15 – Version 1.0
- 2021-01-22 – Quality control

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

Document History.....	3
Table of Contents	4
1. Executive Summary	5
2. Introduction and Objectives.....	6
3. Progress on service development	9
3.1 Task 15.1 – Design of ARIADNEplus services features.....	9
3.2 Task 15.2.1 – Visualization of archaeological imagery	14
3.3 Task 15.2.2 – Visual organisation of archaeological data.....	17
3.4 Task 15.2.3 – Visual documentation of an archaeological excavation	20
3.5 Task 15.3.1 – Archaeology text annotation.....	23
3.6 Task 15.3.2 – Image annotation	23
3.7 Task 15.4 – Implementing the ARIADNEplus text mining and NLP services.....	25
3.8 Task 15.5 – Implementing space-time services.....	27
3.9 Task 15.6 – Query services	33
3.10 Task 15.7 – Cloud Geoserver.....	34
4. References.....	38
Appendix A: Template for describing services	39

1. Executive Summary

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

The work package consists of several individual tasks and subtasks with the overall goal to develop and provide useful services to archaeologists. This means that the work package in itself 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 the European Open Science Cloud (EOSC) has been created. In connection to this, the ARIADNEplus AO-CAT ontology has been adapted to the requirements from the EOSC Resource Data Model.

The Visual Media Service (Task 15.2.1) has had a new format added, allowing for 2D visualization of LIDAR data in DEM format. In addition, three other standards have been added: glTF, ThreeJS and IIF, supporting various functionalities 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).

To integrate the EpHEMERA service (Task 15.2.3) into the ARIADNEplus infrastructure environment, a revision of the service is needed. As this revision needs to be based upon user needs and comments, tutorials have been created and a user survey has been prepared, distributed, and analysed. Revision of the service and implementation of new components have started.

A prototype of an image annotation service has been created based on the OpenAnnotation format, which is intended to be integrated with the Visual Media Service. The first prototype supports vector annotation of relightable images.

The GeoPortal service (Task 15.5) is a new REST service designed to manage complex space-temporal documents defined by metadata Profiles. It has been released as a component of the gCube framework. A prototype using the service has been 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 Aggregative Infrastructure has been established (Task 15.6): a full-text index service and a SPARQL endpoint. The full-text index service is based on Elasticsearch and supports the needed query functionality of the ARIADNEplus portal. The SPARQL endpoint allows to perform semantic queries on the RDF records in the ARIADNEplus data and knowledge cloud.

2. Introduction and Objectives

The overall goal of Work Package 15 is to provide to archaeologists useful services using ARIADNEplus data. This includes integrating existing services, developing new features to existing services, as well as development of brand-new services. All services should also be integrated in the ARIADNEplus infrastructure by making them available on the D4Science platform.

The work package consists of 10 individual tasks that cover services with a wide range of features that are useful for researchers, ranging from front-office services like visualization 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 Visualization of archaeological imagery (JRA4.2.1)
 - T15.2.2 Visual organization 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 is responsible for analysing the existing services available in the former ARIADNE Portal, as well as services already available at partners, and others available as open source. The task is also responsible for periodically surveying the market to find and possibly integrate other possible candidates into the ARIADNEplus infrastructure. A service design, in compliance with the cloud design developed in WP12, will also be produced to align the ARIADNEplus services with EOSC requirements.

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

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

In the context of the implementation of innovative visual services for users, the Task 15.2.2 aims at building new components that address the ARIADNEplus archaeological data management. More

specifically, the goal is to design and develop a web service for improving the visual organization of 3D archaeological datasets. The service, built in the framework of the above-described Visual Media Service, wants to link the 3D model of an artefact or monument to its archaeological documentation, visualizing it accordingly. The task will revise existing tools, and then implement the new service and port it in the ARIADNEplus infrastructure.

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

Task 15.3.1 provides support for archaeologists annotating 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, interactive web application text annotation tool that is easy to access (for example by pasting text content into a window). The tool will draw on vocabularies and multilingual resources resulting from Task 14.3. Thus, vocabulary concepts with URI identifiers should be output where possible. The tool should be easy to use and provide auto-suggestions that encourage but not force users to annotate with standard vocabularies.

Task 15.3.2 concerns the extension of the annotation tool DAP¹. The tool allows to annotate 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, where currently there is very little online except for some open-source support for relightable images (RTI). There is, however, nothing available for the annotation of such media. Standards for annotations of images 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 a semantic linking between archaeological reports (e.g., grey literature) and the linking of text 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 text material in the infrastructure with little original subject indexing. The task will build 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.

¹ <http://tss.isti.cnr.it/dap>

The management of complex collections of spatial documents requires a technical framework that can be tailored to the specific needs of the community in the European ARIADNEplus infrastructure, while remaining affordable. The design and implementation of a service capable of being maintained and operated through extendible Profiles and software handlers that can be customized per VRE is developed in Task 15.5, while the definition of a complete framework, which 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 will be established in Task 15.6. The multilingual query capabilities will be based on the multilingual vocabularies developed in Task 14.3. The services will be based on an RDF store containing the entirety of the ingested resources and an Elasticsearch service specifically designed to support the functionality of the ARIADNEplus portal.

3. Progress on service development

3.1 Task 15.1 – Design of ARIADNEplus services features

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 the follow-up ARIADNEplus project. 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". The role of domain infrastructures like ARIADNE in the realisation of the EOSC vision is fundamental: they 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 and interoperable along with the services of the other communities federated by EOSC.

Discoverability and interoperability with EOSC have therefore become fundamental requirements for the part of the AO-Cat Ontology that deals with the description of services. In order to implement these requirements, 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:



Figure 1: AO-CAT class taxonomy surrounding the class AO_Service

The properties currently defined on 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 necessary. 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 on data resources. In order 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 domain, pushing them down in the taxonomy of resources. This modification has no impact on the already ingested data, since 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 the values that a certain property can assume in a controlled and systematic way. 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 have introduced a general class in `AO-Cat`:

- `EOSC_Vocabulary`, the most general class encompassing every term from anyone of the above EOSC vocabularies. `EOSC_Vocabulary` is a sub-class of `AO_Concept`,

and one specific class for each vocabulary. Thus, we have:

- `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`

Similarly, 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 have introduced 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 as 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.

3.2 Task 15.2.1 – Visualization of archaeological imagery

The Visual Media Service² provides easy publication and presentation on the web of complex visual media assets. It is an automatic service that allows to upload visual media files on a server and to transform them into an efficient web format, making them ready for web-based visualization.

Visual media supported are images, relightable images (RTI) and 3D meshes (in .ply or .obj format). Once uploaded, they are automatically processed and the user can configure the presentation: visual style, background, which feature to enable, initial orientation of the model and type of navigation.

Visual Media Service is built upon 3DHOP³ for 3D models featuring streaming, compression and, most importantly for large models, multiresolution: adaptive, seamless, view-dependent resolution rendering of the mesh available from Nexus⁴. Relightable images instead make use of Relight⁵, an efficient online representation of RTI. All of the software is developed by ISTI-CNR. Finally, the user can download the optimized models and the configuration files and either deploy the results in their own website or they can directly embed the viewer.

The goal of this task is to expand its features, enhance robustness and performances, and integrate the Visual Media Service in the ARIADNEplus portal and D4Science platform.

Lidar

Support has been added for a new format: 2D visualization of LIDAR data in DEM format (.las). The same tiled approach as for RGB images is used for visualization, allowing very large datasets to be inspected online interactively and automatically converted from the .las datasets. The encoding of 16-bit high precision elevation employs the different channels of an RGB image in the lossless PNG format with a good compression ratio. Shaders are used for hill shading and normal visualizations, in a similar way to how RTI relightable images are rendered. Other visualization modes, such as curvature, height dependent palette, etc. will be added in the remainder of the project.

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

³ <http://3dhop.net>

⁴ <http://vcg.isti.cnr.it/nexus>

⁵ <http://vcg.isti.cnr.it/relight>

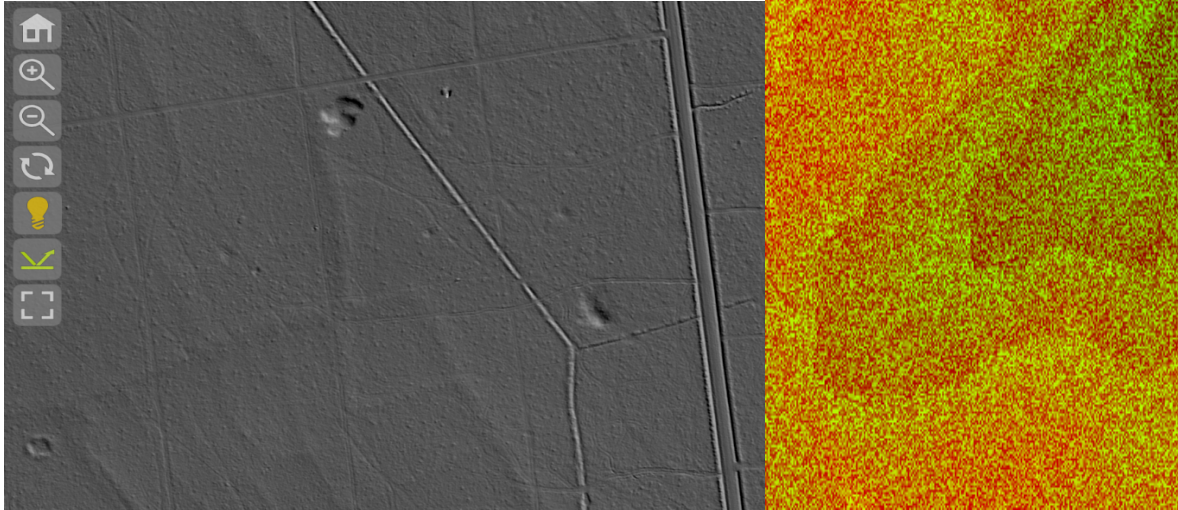


Figure 2: Interactive shading of the dem (left). Elevation 16 bits encoded in RGB channels (right)

Standardization

glTF⁶

Support has been added for an increasingly used 3D model format: glTF. It is a complex format optimized for network asset delivery and, most notably, it supports physically based rendering (PBR), a more general material representation. Draco compression is also supported in input and we plan to make available different compression options (Draco, MeshOpt) other than the library already in use (Corto⁷).

In order to support multiple textures, different materials and attributes, an extensive modification of both the Nexus format and the conversion and rendering software was required.

Work on an extension for the glTF format to support our multiresolution approach is ongoing. Currently, only a simple LOD extension has been proposed, however it is not suitable for large objects. Another approach targeting complex urban scenes (but recently extended also to large meshes such as photogrammetry results), has been developed by the Cesium consortium. Unfortunately, this approach does not support seamless multiresolution and is not natively supported by ThreeJS.

An optimal approach will fit nicely between the two described above, tailored for large single objects (or terrains), with an emphasis on performances.

⁶ <https://www.khronos.org/glTF/>

⁷ <http://vcg.isti.cnr.it/corto>

ThreeJS

The Nexus JavaScript renderer code has been rewritten so it integrates effortlessly into the ThreeJS application. The visualization of the datasets exported by the Visual Media Service becomes extremely easy to integrate into other viewers (e.g., Mirador).

International Image Interoperability Framework (IIIF)

IIIF is an increasingly adopted standard that allows communities to share images online. The Relight viewer adopted in Visual Media Service now supports the IIIF protocol; at the same time, the images submitted in the Visual Media Service can be served following the IIIF protocol. The task is being elaborated within the IIIF community, specifically in the “IIIF 3D Community Group” to extend the standard to support relightable images, 3D content and 3D annotations.

Visual Media Service as a service

The initial purpose of the Visual Media Service was to provide a simple way for users to convert their 3D content into a multiresolution format (.nxs) and easily configure the visualization interface. The service is now being made available to other frameworks. Support for D4Science was added through authentication and integration in the VRE (Virtual Research Environment). In future, this will allow to directly access 3D datasets hosted in this framework, and offer a visualization service that automatically performs the steps that are now carried out manually by the users. Following the same approach, the Visual Media Service will be integrated in the ARIADNEplus web portal.

This approach has been prototyped in cooperation with the Language Resource Switchboard⁸ where Visual Media Service is now one of the many tools available to the user.

Future development

Three main directions of development will be followed in the remainder of the project.

Annotations of 3D and 2D content capability will be added to the Visual Media Service. It will require enhancing the viewer software, development of user-friendly interfaces (see Task 15.2), management of the annotations in the server and finally integrating the annotations metadata with the ARIADNEplus ontology (see Task 15.3).

Support of standard formats and the definition of new standards will be especially important to ensure a healthy ecosystem allowing different tools to cooperate in creating, managing and presenting 2D and 3D content. The plan is to develop:

- a standard format for multiresolution 3D content in cooperation with glTF community
- a standard format for rich 2D content: LIDAR data, relightable images, multispectral data, in cooperation with CRS4 Visual Computing, Department of Computer Science, Verona, and KU Leuven
- an extension of IIIF protocols to 3D content and 3D annotations.

⁸ <https://switchboard.clarin.eu/>

Finally, the Visual Media Service is planned to be made accessible to other platforms, in terms of processing and embedding, through a set of API. This will allow processing and visualization of large datasets without extensive human intervention.

All of the developed software is and will remain open source.

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 aims at building new components addressing the ARIADNEplus archaeological data management. More specifically, the goal is to design and develop a web service to improve the visual organization of 3D archaeological datasets.

The service, built in the framework of the Visual Media Service [Ponchio et. al, 2016] (the platform already developed by CNR-ISTI's Visual Computing Lab), intends to link the 3D model of an artefact or monument to its archaeological documentation to visualise it accordingly.

The subtask roadmap starts with the revision of existing tools, then plans to implement the new service and finally port it in the ARIADNEplus cloud environment, adapting it to the service interface.

The activities started at M7 through an initial interaction with archaeological users to get a clear view of the functionalities needed and thus to derive specifications. The plan was to use the "hotspot" feature of 3DHOP [Potenziani et. al, 2015] (the framework developed by CNR-ISTI's Visual Computing Lab responsible for the 3D interactive visualization in the Visual Media Service) to allow archaeologists to add interactive links from the digital 3D model to the related documentation (in the form of text, images, or other media).

Hotspots are one of the most useful features of 3DHOP, since it gives the content creator the possibility to place specific clickable elements of interest somewhere in the 3D scene, thereby linking them with particular content present on the HTML page. In order to meet this need, 3DHOP provides a way to create geometrical hotspots and a series of events to detect user clicks on 3D models and hotspots.

Unfortunately, the implementation of this feature is not very straightforward for unskilled users, since they need to add a new kind of element in the scene definition, manually placing geometric hotspots in the 3D scene (defining translation, rotation, and scale through parameters in the source code), and finally registering handler functions to detect user clicks on these geometries.

This complexity is the reason why hotspot is currently one of the few 3DHOP features not provided by the Visual Media Service. To overcome these issues, a visual wizard able to guide Visual Media Service users in adding hotspots to the 3D scene easily and quickly has been defined in this subtask.

Initially implemented locally in 3DHOP, the beta version of the defined extension will allow archaeologists to create interactive links from the digital 3D model to the related documentation without writing any source code.

In the current version of this visual wizard, the authoring tool allows the content creator to interactively select a point on one of the models composing the 3D scene where a spherical geometric hotspot is placed. The wizard also allows to control the size of the sphere (by editing its radius), to define a group ID for the hotspot (to create visibility groups), and to associate a predefined camera position (i.e., a bookmarked view) to it. All these fine-tuning actions can be performed after the creation of the hotspot, without any limitations in terms of time/number of changes.

The process is recursive, in the sense that it enables the sequential creation of several spherical hotspots, giving content creators the possibility to remove them whenever they want.



Figure 3: The beta version of the sphere-based annotation service. The panel on the left contains the button for adding the sphere hotspots (in light pink) in the 3D scene. Through the panel on the right is possible to modify the hotspot setup (radius, groups ID).

All the hotspots setup parameters can also be stored in a JSON structure to facilitate importing and exporting them on the 3D scene.

After a testing period, the current implementation of the annotation service is now ready to be integrated in the Visual Media Service, where the logic for the hotspot creation will be linked with the web form for the setup of additional information (already defined in the Visual Media Service), and finally adapted to the web service interface.

In addition to the annotation feature, other possibilities have also been investigated with the project partners. More specifically, CNR-ISTI's Visual Computing Lab discussed with the Cyprus Institute (CYI) possible extensions for some of their datasets to be integrated in ARIADNEplus, as well as for some of the research in their related field (Sub-task 4.4.13 – Inscriptions; epigraphic research).

Specifically, CYI's activity focused on the identification of user needs for the implementation of the service planned in subtask 15.2.2. A list of desiderata has been created and proposed to CNR-ISTI's Visual Computing Lab:

1. Enhancing filter for visual exploration. A rendering mode applied to the 3D model which makes features such as signs appear more clearly. Its intensity can be controlled with a bar that appears when it is activated. An important example would be “Radiance Scaling”, which is already employed in several 3D online viewers for inscriptions.
2. Paint and drawing tool to highlight or mask areas of interest with semi-transparent colour. A user can choose between a polygon drawing tool or a paintbrush tool, as well as different sets of colours and levels of transparency. This will allow users to highlight both larger areas, such as lines or groups of signs (polygon tool), and individual strokes in an accurate and controlled way (brush tool). There should also be a tool to delete. Highlighted areas do not need to be saved permanently and are not visible to other users.
3. Possibility to take snapshots. Users can use this tool to save a snapshot of the current position and of the potential highlighted areas or measurements in place, and download the picture. The saved picture can be used within other software and further annotated.

Another extension has been implemented locally in 3DHOP, while there are still ongoing discussions aimed at defining the feasibility and evaluating costs and benefits of the implementation over the existing infrastructure of the features in 1., following the indications in 2., and inspired by the platform described in Apollonio et al. [Apollonio et al., 2017].

Still aiming at enriching 3D datasets, this experimental feature will provide users with a simple drawing tool based on polylines for highlighting or marking areas of interest with semi-transparent colours. Developed following the scheme of the sphere-based annotation service, the current design of this extension provides the same features of the former: interactive/recursive creation, customization, deletion, setup import/export. Since it is in the validation stage, this feature still needs to be refined before proceeding with its integration in the Visual Media Service platform.



Figure 4: The experimental version of the polyline-based annotation service. The panel on the left contains the button for adding the polyline hotspots (in light green) in the 3D scene. Through the panel on the right is possible to modify the hotspot setup (groups ID, associated view), as well as remove them.

Finally, the feature identified in point 3. has been developed and officially released with the 3DHOP v4.3 release; this extension is also ready to be integrated in the Visual Media Service platform.

Over the next months, the subtask roadmap plans to finalize the work carried out so far, integrating the developed extensions in the Visual Media Service platform as an online service, thus incorporating these tools in the ARIADNEplus services' provision.

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 MEditerRAnea area)⁹ developed at CYI to visualize archaeological excavations, ancient buildings, and their related documentation [Abate et al., 2017] in 3D. In EpHEMERA, it is possible to: visualize 3D architectural and archaeological models (classified according to a specific type of risk) online and through standard web browsers; query the database system and retrieve metadata attached to each digital object; extract geometric and morphological information. The service is conceived as a platform open to multidisciplinary research.

⁹ <http://ephemera.cyi.ac.cy>

Technical characteristics

The EpHEMERA online portal is being developed to address the Open Access (OA) guidelines for providing online access to scientific information, and it is built on top of the open-source Drupal Content Management Framework (CMF). A key feature of the system is managing large, high-resolution 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 huge amount of 3D data and other media files. The storage system is scalable and expandable, depending on the availability of new 3D assets. A Content Model has been developed mainly with the aim to support 3D models and their relevant data (text file, pictures, panoramic images, external links, and resources), and to further map their attributes to existing cultural heritage metadata schema. More than 45 fields and a dedicated taxonomy connected to the customized Content Model have thus been created. Through the database architecture, three-dimensional models stored in a separate NAS server are available online via the open-source WebGL based PoTree viewer. PoTree¹⁰ is a point-based rendering solution specifically developed for visualizing large point clouds using standard web-based technologies [Schutz, 2014; Schutz & Wimmer, 2015]. It can easily provide a responsible interactive viewer that only requires a traditional web browser (WebGL enable), freeing the user from any configuration issue or specific software. A list of command utilities is provided to navigate, interact, and perform geometric analysis of the visualized 3D assets. Two different types of users have been *a priori* identified, and privileges have been respectively assigned: Non-authenticated (or Anonymous) user (by default, any user who accesses the EpHEMERA service can browse the content, interact with the 3D models, and perform geometric analysis); and Registered user (this authenticated user can log in to the system to create or edit content through a dedicated user interface).

Planning of the activities and tasks management

The system needed a revision before it could be ported in the ARIADNEplus cloud environment and adapted to the service interface. For this reason, activity planning and task management has been carried out. First, it was decided that the implementation of the service would be based on users' comments. Therefore, a feedback framework has been planned and built. The framework's main steps consisted of creating and distributing a survey, gathering and analysing the responses, and identifying possible new components to be implemented. This was finally followed by the implementation and integration in the ARIADNEplus cloud environment.

Tutorial preparation

Since the survey was developed so it would be based on both the users' interaction with the platform (online survey dedicated to large groups) and on task performance (in-presence survey dedicated to small groups of individuals), it was decided to create a series of tutorials, documents and visual demonstrations. More specifically, the reason for this rested on the fact that the surveyed users were not selected among specific professional categories or with specific backgrounds (for instance, only archaeologists), nor according to their expertise or level of acquaintance with digital tools. The idea was to maintain an unrestricted approach to the application and leave space for multidisciplinary

¹⁰ <http://potree.org>

comments, since the tool fits multidisciplinary use (e.g., archaeology, conservation and restoration, climate, art history, chemistry).

The first step was the preparation of a tutorial to support users with a list of all the available command utilities and relative descriptions to navigate, interact and perform geometric analysis of the visualized 3D archaeological excavations and ancient buildings within the EpHEMERA service (EpHEMERA Navigation tools and Geometric Analysis tools). Then a document was prepared concerning the back-end of the tool and its main technical specifications. Finally, both a short video demonstration providing a general overview of the EpHEMERA application, and a series of video tutorials with the explanation of the command utilities and functions were prepared.

The material thus created and attached to the survey (tutorials, video tutorials and technical specifications) has been published online¹¹.

It was decided to create a YouTube channel to share the video tutorials¹² and keep them available also after the survey, so they could eventually be updated with new content and provide material support for the users of the application after its publication in the services section of the ARIADNEplus portal.

Moreover, this documentation supports the WP16 pilot on sharing archaeological excavation data in 3D via a web-based platform.

Preparation and distribution of the Survey

The aim of the survey was to test the main functionalities of the EpHEMERA platform, identify strengths and weaknesses, and possible new components to be implemented. The first step was, therefore, to develop the methodology behind the construction of the survey. Two approaches have been developed according to the types of users the survey aimed to reach: an online survey dedicated to large groups, and task-based trials for small groups to be carried out in presence. Unfortunately, because of the pandemic situation, it was not possible to organise meetings for task-based trials dedicated to testing the EpHEMERA platform among small groups of people. For this reason, the analysis has been solely based on the online survey.

The online survey aimed at testing the EpHEMERA platform and getting feedback from different multidisciplinary specialized users has been submitted and shared¹³. Accessing the current link of the platform¹⁴, and with the support of the tutorials attached to the survey, participants were invited to test it and reply to a few simple questions based on system's usability and functionality. A second section, composed of open questions, gave the participants the possibility to provide their own perspective on the platform, suggestions and desiderata based on their specific background.

The successive step consisted of analysing the results of the survey, following which a list of feedback and possible new components suggested by the users has been prepared.

¹¹ A copy of the text documents can be found at the following link:

https://drive.google.com/drive/folders/1gpZ7jZm0QCxHC7JJ9tjlv9ME_rLT-VF0?usp=sharing

¹² <https://www.youtube.com/channel/UCgkKhBZfMOKFSku2KtQY9tA/featured>

¹³ The survey is online at the following link: <https://forms.gle/ZSmthfTt3S1jg6ja8>

¹⁴ <http://ephemera.cyi.ac.cy/>

Review and update of the EpHEMERA platform

This part of the work consists of reviewing the platform by taking into consideration the survey's feedback, analysing what is already available and what is feasible in terms of tools and instrument development for the interaction with the content. Currently, the implementation of some components is under finalisation. The next step will be the integration within the ARIADNEplus service interface.

As a final point, there are plans to publish an article regarding the whole set up methodology, the survey, its outcomes and effects, the implementation and the final results of the integration into ARIADNEplus portal.

3.5 Task 15.3.1 – Archaeology text annotation

The main part of this task, developing an interactive text annotation tool for archaeological reports, is scheduled for the second part of the project. However, some initial consideration of the development's scope and software platform has taken place.

The task is intended to be completed within the next year. The annotation tool will be driven by archaeological vocabularies expressed as SKOS and will follow a web service architecture. Suggestions will be generated with an expression of confidence, allowing the ranking of suggestions for easy intellectual judgment and confirmation. A client annotation tool will be developed based on the SKOS web services.

3.6 Task 15.3.2 – Image annotation

The goal of this task is to develop an online tool for annotating archaeological 2D media: images and relightable images.

Currently, there is very little online except for some open-source support for relightable images (RTI), but there is nothing available for the annotation of such media. Standards for annotations of images are not widely supported, so the need to connect the metadata annotations with ARIADNEplus ontology and vocabularies poses another challenge.

It is our intention to eventually make the tool available in the Visual Media Service.

As a first step, we determined how the tool should represent, store and export image annotations. The final decision was to support vector annotations encoded in SVG and support OpenAnnotation format¹⁵, as well as IIIF metadata API¹⁶ that is in turn based on the first standard.

¹⁵ <http://www.openannotation.org/spec/core/specific.html#SvgSelector>

¹⁶ <https://iiif.io/api/metadata/1.0/#api-advanced-non-rectangular>

The second step has been to develop a prototype of the web interface and plan its integration into the Visual Media Service with ISTI-CNR. The prototype, called *Dapper*¹⁷, supports vector annotation of relightable images, albeit specialized to annotate symbols and letters. It supports different alphabets and exports the annotations in SVG format.

Catalogue: 23.15 - Inventory: 12012 front (back)

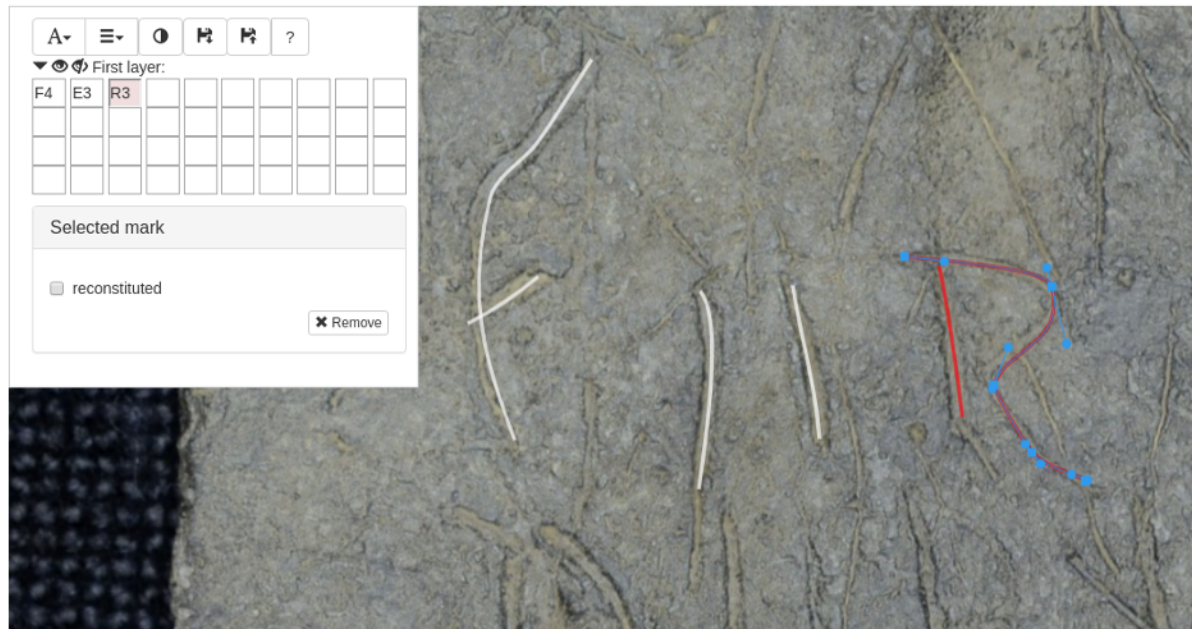


Figure 5: Vector annotation of relightable images

The Tesserarum Sisciae Sylloge collection (TSS – the initial database was created by Federico Ponchio and Marion Lamé, with Ivan Radman’s feedback and funding by the Croatian Ministry of Culture). Featuring over one thousand Roman lead tags and kept in the collections of the Archaeological Museum in Zagreb, these will be used as a proof of concept and test bed for the functionality of the annotation tool. Forty tags of the collection were previously digitized (an undertaking funded by a grant from the European Association of Digital Humanities) and are now available as RTI images.

A dataset of the corpus was created with the transcriptions of the tags and associated drawings of the inscriptions and letters, as well as various indices of personal names, textile products, prices and abbreviations mentioned on the tags.

This data corpus was converted into a KOS (knowledge organization system) backed by the SKOS format, and a tool called Heterotoki [Lamé et al., 2019] was developed to support the editors in the formalization of the vocabulary to ensure consistency, fill in missing concepts and labels, and define hierarchical relationships. Heterotoki supports mapping with external terminologies such as the Getty AAT vocabulary, which is being used in support of the ARIADNEplus ontology.

¹⁷ <http://tss.isti.cnr.it/dap>

The next planned steps involve mapping the SKOS terminology created for the TSS dataset to the AAT terminology, then connecting the annotation interface with the SKOS vocabularies, and finally to integrate the prototype functionalities into the Visual Media Service viewer.

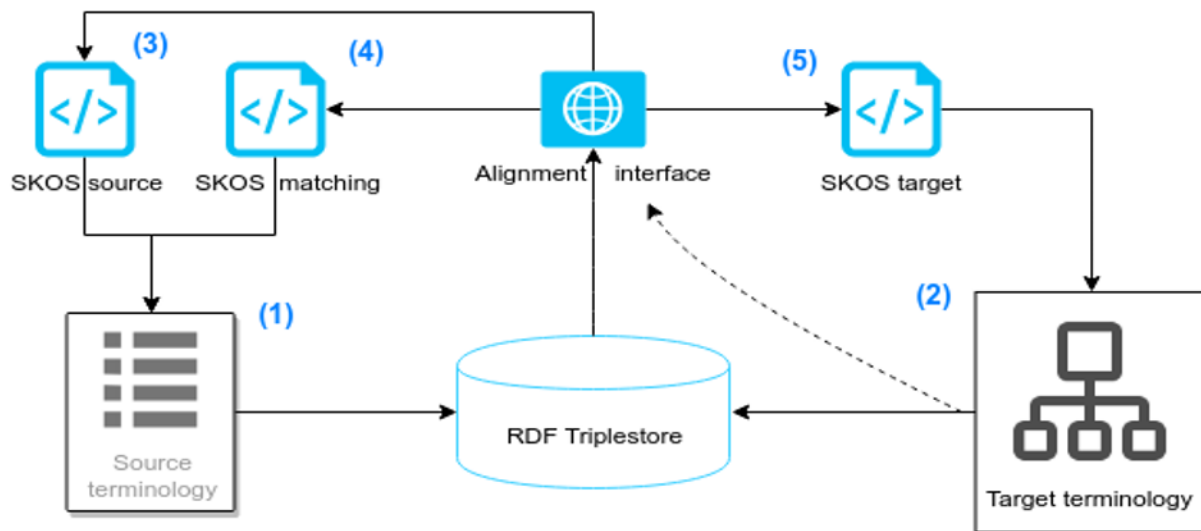


Figure 6: Heterotoki structure

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

Initial work for this task is outlined, together with plans for the second half of the project.

One strand of work in this task aims to build on the NLP outcomes from the preceding ARIADNE project. USW collaboratively produced a Cooperation Work Plan (CWP) with Andreas Vlachidis (who developed the ARIADNE NLP pipelines) as part of the Cooperation Agreement between Associate Partner UCL and ARIADNEplus¹⁸. Following the CWP, we reconstituted and deployed the archaeological Named Entity Recognition NLP pipelines developed within the GATE¹⁹ framework, as ARIADNE WP16 outcomes for ARIADNEplus purposes on the GATE cloud²⁰. Services are provided in Dutch, English and Swedish for general archaeological Named Entity Recognition and also for dendrochronological entities in particular.

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. A tool is being developed that automatically performs normalization based on the wide set of date patterns and variant expressions for time spans commonly used in archaeology. A normalized set of date spans is produced for expressions of dates, both in terms of years and as named periods, in different languages. Archaeological dataset records

¹⁸ <https://ariadne-infrastructure.eu/associate-partners>

¹⁹ <https://gate.ac.uk>

²⁰ <https://cloud.gate.ac.uk/shopfront#tagged=Archaeology>

often give 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, sometimes expressed in different languages. There can be prefixes present such as 'Circa', 'Early', 'Mid', 'Late' - and suffixes such as 'A.D.', 'B.C.', 'B.P.' that may influence the dates intended. Named periods may also be encountered and different countries employ different standard sets of time periods (see discussion in D14.1). 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 the general point:

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

A case study was conducted for Norwegian language temporal expressions in particular. A meeting was held with ARIADNEplus partners KHM-UO at the Museum of Cultural History, Oslo in Sept 2019 to discuss a prototype multilingual date span matching web application. In collaboration with KHM-UO, USW elaborated Norwegian textual patterns for time periods and date spans for use in the multilingual date span matching tool. Again, some examples illustrate the general point:

eldre steinalder	9500 BC	4000 BC
tidlig metalltid	2000 BC	300 BC
mellommessolitikum	8000 BC	6000 BC
310-340 e.Kr	310 AD	340 AD
500 f.Kr-500 e.Kr	500 BC	500 AD

Planned activities for the second half of the project

- Continue to refine the prototype multilingual timespan utility with a view to providing it as a temporal text mining web service that would recommend temporal metadata for archaeological reports.
- Building on the CWP with UCL, make the general archaeological NER pipelines services accessible via the D4Science platform.
- Investigate case studies applying the multilingual timespan and the general NER services to archaeological report abstracts and descriptions, in collaboration with relevant ARIADNEplus data provider partners using different languages. The services will recommend additional

subject metadata to enhance resource discovery and reuse. They are seen as particularly applicable to abstracts or summaries, where instances of the information extraction are likely to be relevant and important subject metadata for the overall resource.

3.8 Task 15.5 – Implementing space-time services

The GeoPortal Service is a new REST service designed to manage complex space-temporal Documents defined by metadata Profiles. It has been released as a component of the gCube [Assante M. et al., 2018] framework, and is powered by SmartGears²¹.

It has been designed to manage the publication lifecycle of complex spatio-temporal Documents, supporting their materialization and indexing in different platforms (Databases, Catalogues, OGC Services), thus maximising reusability.

Its main features are:

- Documents Publication Management:
 - CRUD (Create, Read, Update, Delete) operations, publication lifecycle, Profiles-based validation;
 - Role-based access to resources and metadata sections;
 - Cooperative workflow;
- Automatic Indexing & aggregation based on Profiles annotations:
 - GIS indexes (both centroids & multi-polygon), with dynamic aggregation based on zoom level;
 - Text indexes;
 - Time indexes and aggregation;
- Extensible Document model:
 - Generic meta-model Profile;
 - Custom extensions of the model;
- Extensible behaviour
 - Core engine based on a meta-model;
 - Publication management implemented by an extendable set of Event Handlers;
 - Behaviour configured in Profiles by declaring Event Handlers call in cascade.

Users can define their model as a Profile in the VRE and then manage their Documents publication lifecycle both via the 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 Profile (defaults apply) in order to be invoked in the different phases of the publication lifecycle.

The set of Event Handlers is dynamically loaded by the service allowing support for custom implementation of Event Handlers.

²¹<https://wiki.gcube-system.org/gcube/SmartGears>

The provided Java client library allows for dynamic model Serialization/Deserialization of custom provided Java model classes, allowing for easy integration of different systems.

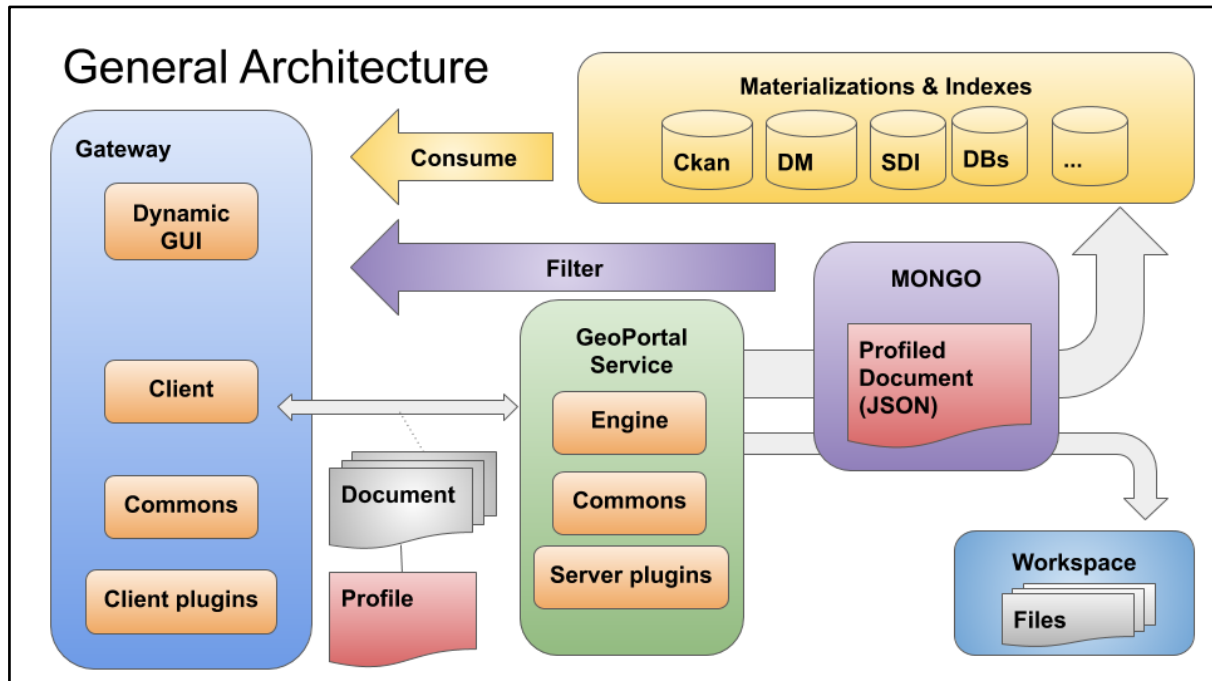


Figure 7: GeoPortal integration in Cloud Geoserver

Extensible model

Dataset schemata are defined as Profile, and every managed collection of documents refer to a specific Profile. A complex geo-temporal dataset linked to a Profile is called Profile Document.

This approach allows for the definition of tailored Profiles that allow the management of a heterogeneous collection of documents. Moreover, it supports evolution by intrinsically enabling the extensions of any Profile.

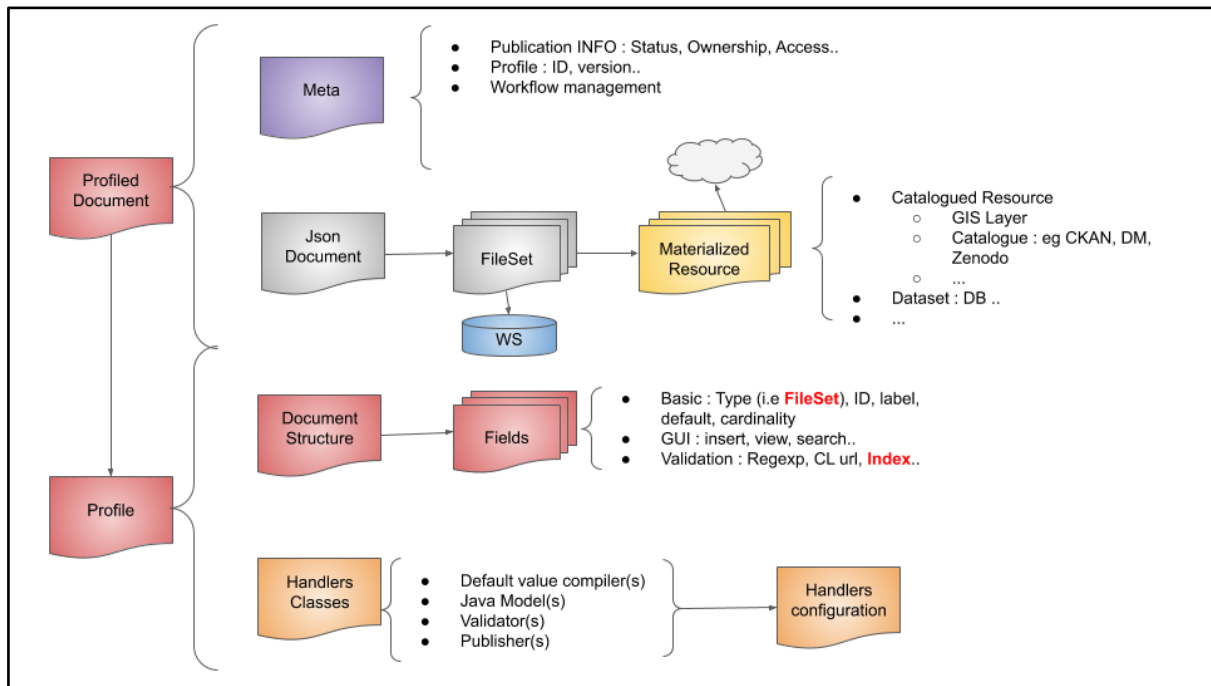


Figure 8: Profile and Profiled Document

Profiled Documents are stored as:

- Meta: core section for publication management, containing:
 - Ownership
 - Identification
 - Versioning
 - Status
 - Phase
 - User-oriented messages
- JSON Document metadata;
- FileSets uploaded along with metadata;
- Materialised Resources obtained by publishing FileSets.

Profiles define collections in a VRE, specifying:

- Document structure
 - Fields (cardinality, type, constraint, defaults, indexing, GUI declaration)
- Publication lifecycle handlers
 - Default values
 - Java model for specific business logic
 - Validators
 - Publishers
- Handlers configuration

Document Structure

The Document Structure is implemented in the current version as a 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 of a descriptive, spatial, temporal kind, 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>
```

Figure 9: gCube Metadata Profile

- *Spatial* data - can be specified by using the *dataType* values:
 - (i) *GeoJSON*²²;
 - (ii) *File* - by uploading a spatial File (e.g., a shapefile).
- *Temporal* data - can be specified by using the *dataType* values (based on ISO-8601²³):
 - *Time*;
 - *Time_Interval*;
 - *Times_ListOf*;

Document Lifecycle

Document lifecycle is managed by invoking handlers. Default handler implementations can be reused as fitting, while they can be extended. Handlers are invoked in cascade to enhance reusability by modularisation.

²² GeoJSON: <http://geojson.org/geojson-spec.html>

²³ ISO-8601: <http://www.iso.org/iso-8601-date-and-time-format.html>

Spatio-temporal complex documents are materialised, published and indexed according to the collection's needs, exploiting the underlying third parties' engines.

Default Document Lifecycle handler

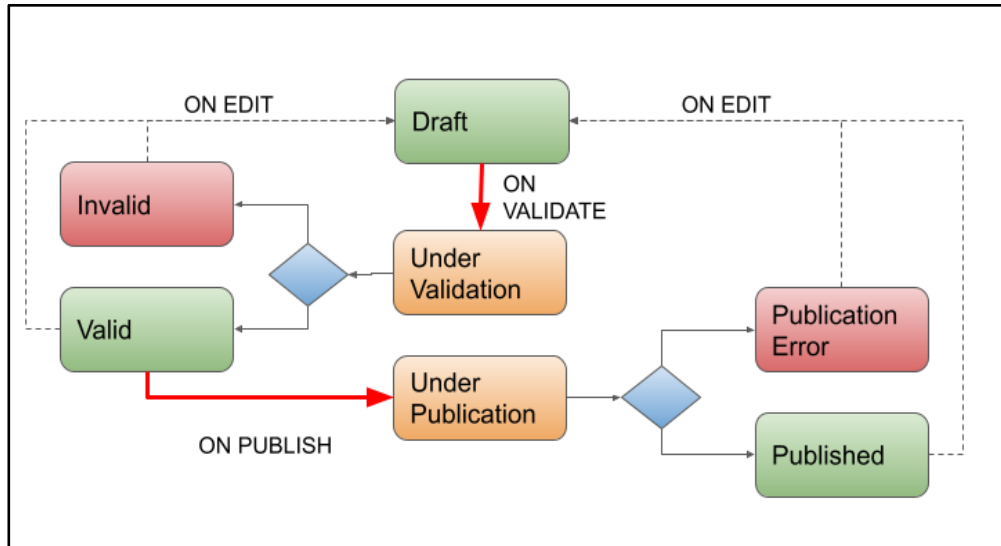


Figure 10: Default Document Lifecycle

The section Status of a Profiled Document reports information regarding the document publication lifecycle, both as:

- Phase: The name of the current lifecycle phase in which the document is;
- Messages: Collection of user-oriented messages pertinent to the Phase value;

Provided default publication handler is aimed at managing a simple 3-phase publication lifecycle comprised of:

- Draft: Document has been registered, and default values are applied if any;
- Valid: Document has been validated against schema and custom validation rules, and it's ready to be published;
- Published: Document has been materialised, published and indexed.

Additional steps are introduced to support asynchronous interrogation of the Document status:

- Declare ongoing processing of the document (i.e., Under Validation, Under Publication)
- Notify errors depending by the Document (e.g., Invalid) or by unexpected internal causes.

Default Geospatial Publication Handler

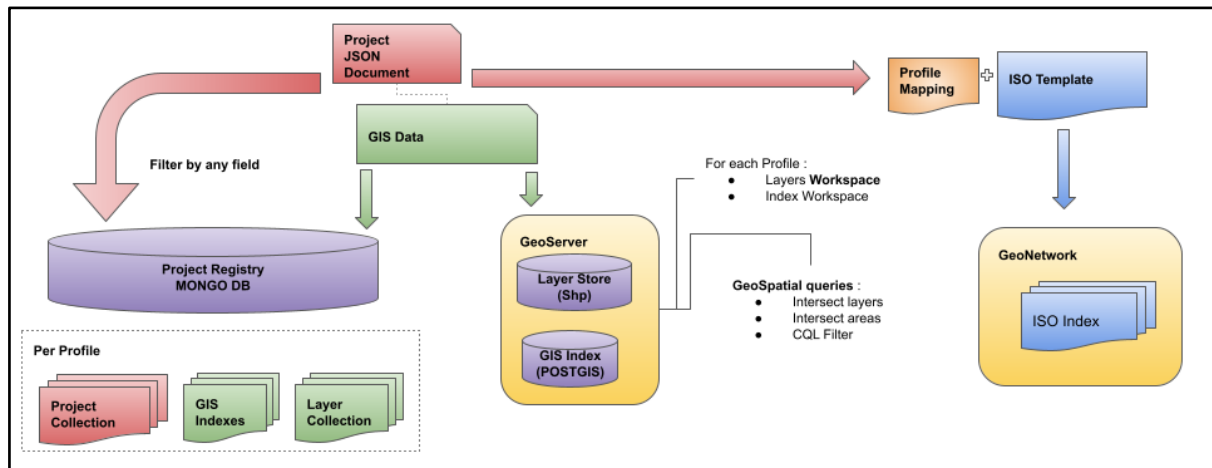


Figure 11: Default Publication Handler

Default publication Handler provides support for the publication of standard geospatial dataset formats as layers in the gCube SDI:

- ESRI Shape FileSet (.shp, .shx)
- GPKG
- NetCDF

The gCube SDI exposes OGC interfaces (e.g., WMS, WFS, WCS, CSW) relying on:

- GeoNetwork for ISO19139 XML Metadata formats;
- Geoserver, Thredds for rendering
- Postgis, MongoDB for storage

It also allows for managing geospatial indexes as layers with both centroids and multi-polygon geometries to navigate the published Documents collections.

Indexes are configured in the Profile by:

- Annotation in the model's fields to be indexed
- Declaring aggregation type (i.e., Aggregated geometries, Centroids)
- Declaring aggregation level (i.e., layer per keyword, layer per project)

ISO19139 XML Metadata is generated by the template mapping, allowing for configurable behaviour defined in the document's Profile.

Future plans

For the second half of the project, the plan is to fully implement GeoPortal service as described above, by:

- Extending Profile metadata in order to manage the entire management lifecycle;
- Incrementally modifying the developed prototype in order to reach its reusability in different application scenarios operated via one or more ARIADNEplus VREs.

Additionally, the plan is to expand the default handler set to:

- Implement multi polygon GIS aggregation and Indexing;
- Implement GIS metadata indexing
- Enable support for Thredds servers;
- Implement Text default indexing;
- Implement Time default indexing;
- Support for collaborative validation lifecycle.

3.9 Task 15.6 – Query services

Query services

In the first two years of the project, the task, in collaboration with WP12, delivered two services for querying the RDF AO-Cat metadata records aggregated by the ARIADNEplus Aggregative Infrastructure: a full-text index service and a SPARQL endpoint.

The full-text index service is based on Elasticsearch²⁴ version 7.4.0. Elasticsearch is used for persistence of the metadata of the resources. It is a fit-for-purpose database with a built-in search engine where all metadata is stored as JSON²⁵ documents. Elasticsearch has been developed as a scalable database especially focused on searching large amounts of data in near real-time. An Elasticsearch mapping, which contains rules of the structure of the documents, has been developed to ensure all documents have the same structure and are interpreted in the same way. The mapping document also provides various structures, which facilitate the various search options that the ARIADNE portal offers. Elasticsearch also offers an API for search and browse.

The SPARQL endpoint, based on a GraphDB²⁶ server (version 9, free edition), allows to perform semantic queries on the RDF records in the ARIADNEplus data and knowledge cloud (AC).

²⁴ <https://www.elastic.co/products/elasticsearch>

²⁵ <https://www.json.org/json-en.html>

²⁶ <https://graphdb.ontotext.com/>

Multilinguality

Some initial investigation and experimentation of multilingual development possibilities has taken place and multilingual requirements for the portal have been drafted. This includes the complementary use of Getty AAT vocabulary concepts and Wikidata multilingual expressions of archaeological concepts. This is discussed further in the report D14.1, as WP14 activity feeds into this task.

The bulk of the work with multilinguality in this task falls in the second half of the project. The plan is to provide multimedia support for the development in the portal of a semantic concept-based search capability via the AAT. This would entail support for an extension of the ARIADNE portal auto-complete functionality (with the option of basic string search, if the offered concepts are not taken) on the search box in the portal. ARIADNE allowed an English language search to return records with metadata in other languages (swords → spada). In ARIADNEplus, we hope to allow an initial search entry in other languages leading to AAT concept search (spada → swords).

This augmented auto-complete functionality would make use of (i) the AAT mappings, (ii) AAT's own multilingual resources (iii) wikidata multilingual resources and (iv) any partner language synonym sets for their subject metadata. USW plans to provide this multimedia resource to assist the portal developers produce the augmented auto-complete functionality.

This will require an expanded auto-complete functionality and a choice of language by some means in the portal. It will also require an index of AAT concepts (URIs), ideally the subset used in the current build of the ARIADNEplus data, in order to avoid zero matches.

3.10 Task 15.7 – Cloud Geoserver

GeoPortal is a feature-complete framework enabling the publication, access and management of GIS projects consisting of multiple payloads (as documents, images, and datasets), each described with tailored metadata.

Its main features are:

- 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.

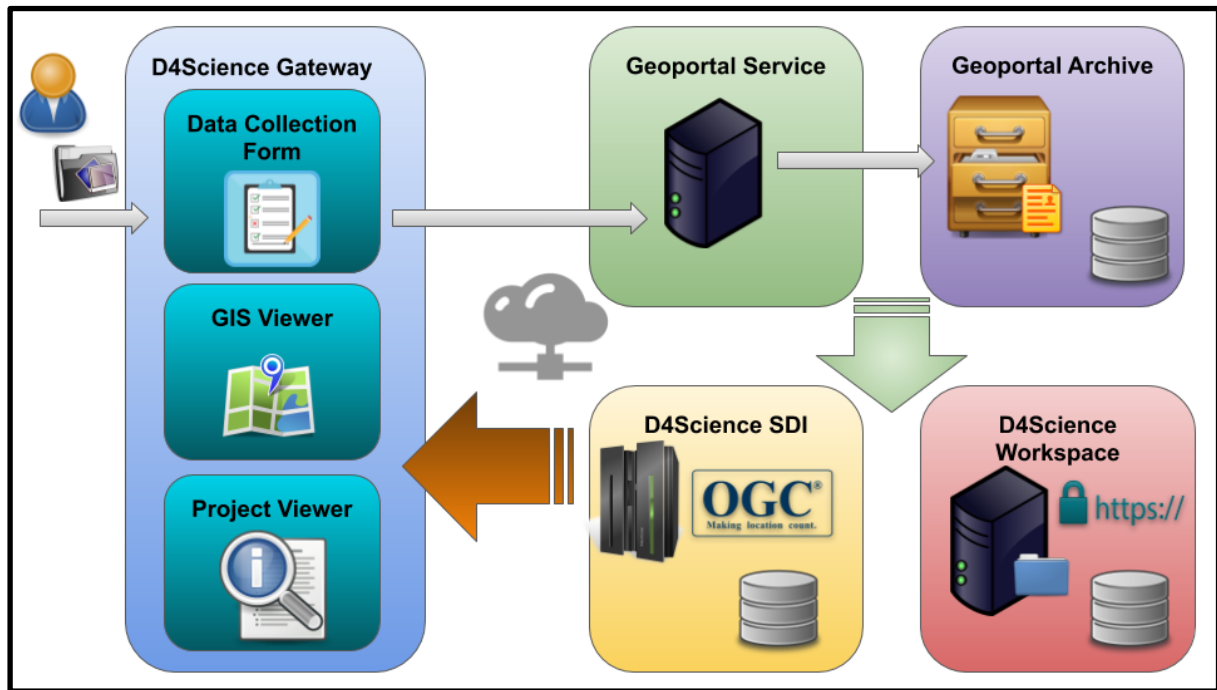


Figure 12: Cloud Geoserver Framework

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

- *Data Collection Form* assisting users to publish 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, 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 Profiled Document to manage publication lifecycle and for querying purposes.

Case Study

A prototype has been deployed and operated to manage archaeological excavation projects. The following screenshots are the current result of the prototype.

Insert Information for: Pianta di Fine Scavo * is required

* Valutazione della qualità:	<input type="text"/>	i
* Metodo di raccolta dei dati:	<input type="text"/>	i
* Scala di acquisizione dei dati:	<input type="text"/>	i
* Autore:	<input type="text"/>	i
	Repeatable field	
* Politica di accesso:	OPEN	i
* ID Licenza:	CC-BY-4.0	
* File:	Browse... No file selected.	i
	Repeatable field	

Create

Figure 13: Data Collection Form auto-generated by "Pianta di Fine Scavo" Profile

Heraion Foce Sele

Share View

Dal 1987 il Dipartimento di Studi Umanistici dell'Università degli Studi di Napoli Federico II ha ripreso le indagini presso il santuario di Hera sul fiume Sele, nel territorio extramurale dell'antica Paestum/Partheneion. Le ricerche, negli ultimi anni, si sono concentrate nella Zona C, situata 500 metri a Sud-Est dell'area sacra, con l'obiettivo di indagare le evidenze strutturali.

Data Inizio Progetto	02-07-2019
Data Fine Progetto	02-08-2019
Parole chiave Libere	Paestum Sele Heraion santuario
Parole chiave Cronologia	età arcaica età lucana età romana
Contributore	prof. Bianca Femara
Editore	Università degli Studi di Napoli Federico II
Responsabile	prof. Bianca Femara
Ricerche Correlate	
Titolari dei dati	prof. Bianca Femara Università degli Studi di Napoli Federico II
Soggetti	Research Excavation Università degli Studi di Napoli Federico II
Editore	Università degli Studi di Napoli Federico II

Figure 14: GIS Viewer and Project Viewer auto-generated displaying a Profiled Document

Future plans

The described cloud Geoserver will be used to support the Temporal GIS pilot described in T16.6. A profile supporting the pilot will be identified and defined in a new tailored VRE hosting the cloud Geoserver. Such VRE will then be populated with the data and metadata identified by the pilot case.

We plan to expand cloud Geoserver framework's offered features by providing additional GUI components that will:

- Extend the GIS Viewer capabilities to include:
 - add a layer to a map;

- overlay of layers;
 - export to images of a complex multi-layer visualization;
- Enable the temporal dimension by supporting historical contexts, periods, and actual timeframe;
- Support users in:
 - Managing registered documents
 - Navigating the registered collections of documents

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*.
- Assante, M., Candela, L., Castelli, D., Cirillo, R., Coro, G., Frosini, L., . . . 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
- 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*,. ISSN: 0097-8493, Volume 52, 129-141.
- Schutz, M. (2014). *PotreeConverter - Uniform partitioning of point cloud data into an octree*. Retrieved from http://potree.org/downloads/converter_documentation.pdf
- Schutz, M., & Wimmer, M. (2015). Rendering large point clouds in web browsers. *Proc. 19th CIESCG*, 83-90.

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>

Relying on this profile to describe ARIADNEplus services aims at promoting the exchange of information about services between our initiative 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 to access, b) the benefit to a user/customer delivered by a Service; benefits are usually related to alleviating pains (e.g., eliminate 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 will be 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