Sub-task 4.4.10. – Public Final Report

Version 1

10th December 2021

Grant Agreement number:	**823914**
Project acronym:	**ARIADNEplus**
Project title:	**Advanced Research Infrastructure for Archaeological Dataset Networking in Europe - plus**
Funding Scheme:	**H2020-INFRAIA-2018-1**
Project 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.

Authors: **David Novák, ARUP-CAS**

Jan Hasil, ARUP-CAS

(and individual data providers)

Versions: 1 – initial version

Table 1. List of institutions referenced in the report:

Partner	***Country***	***Full title***
AIAC	Italy	International Association for Classical Archaeology
ARUP-CAS	Czechia	Institute of Archaeology of the Czech Academy of Sciences, Prague
CARARE	Ireland	Connecting Archaeology and Architecture in Europe
CNRS (UT)	France	National Centre for Scientific Research
CONICET	Argentina	National Scientific and Technical Research Council
DGPC	Portugal	Directorate-General for Cultural Heritage
FI	Iceland	Institute of Archaeology, Iceland
HES	Scotland	Historic Environment Scotland
HNM	Hungary	Hungarian National Museum
INP	Romania	Institute of Archaeology of Vasile Parvan
INRAP	France	National Institute for Preventive Archaeological Research
KHM-UO	Norway	Museum of Cultural History
PP	Greece	Athena Research Center
SND	Sweden	Swedish National Data Service
UH	Finland	University of Helsinki
UoY-ADS	UK	Archaeology Data Service
ZRC-SAZU	Slovenia	Research Centre of the Slovenian Academy of Sciences and Arts

1 Initial assessment

The initial goal of T4.4.10 was to analyse in which form the spatial data is stored by the individual partners and how it can be integrated into the ARIADNE Catalogue (AO-Cat) at the item level. It was clear from the outset that T4.4.10 overlaps in its focus with other subtasks within T4.4, as spatial information is an integral part of many different dataset types. These datasets are being analysed and integrated into AO-Cat based on needs defined by specific Application Profiles (AP). Regardless of this apparent limitation, T4.4.10 has attempted to obtain information from the partners involved about the nature of the spatial data under their management and the possibilities for its use within the project.

To this end, a questionnaire was designed and distributed (see Annexe 1). It became apparent that there are several national standards for data description in place (HNM, FI, SND). PP planned to make their datasets compliant with CIDOC-CRM, while other data stay non-standardized (UoY-ADS, DGPC, ARUP-CAS). Geospatial features representations in GIS datasets are usually variable, without any methodological specification (UoY-ADS, FI, SND, ARUP-CAS, DGPC). The only exceptions are HNM (on the national level) and PP (on the institutional level) with a transparent strategy of geodata classification and representation.

Temporal (chronological) attributes often appear in the data, however on different levels. Item level temporal data were usually present in all datasets stored by partners, but only datasets of HNM and PP are described according to a shared standard. In other datasets, temporal data were included according to the decision of the data provider and their automatic collection or harmonization is currently impossible. DGPC, UoY-ADS, SND and (partly) ARUP-CAS were able to provide standardized temporal data on the collection level.

There are plenty of formats used by partners for geospatial data storage (SHP, DWG, DWF, GeoTIFF, GeoJSON, TIN, GRID, KML). ESRI shapefiles, GML and PostGIS databases seems to be prominent options chosen for data handling as well as for long-term preservation. Three partners were ready to publish (or publishing already) their geodata as web services (WMS/WFS) using GIS servers (DGPC, FI, ARUP-CAS).

With only one exception (FI), all the datasets were described by metadata, which usually follows institutional (UoY-ADS), national (HNM) or international standards (DGPC, PP, SND). Further systemization of the geospatial data during the ARIADNE+ project is planned only by DGPC and PP and considered by FI and ARUP-CAS.

Two partners expressed their will to include geospatial data on the item level (PP, ARUP-CAS), with FI possibly joining them later. Due to various reasons, all the other partners were able to provide their data on the collection level only.

*

As a result, we identified a high level of heterogeneity limiting the possibilities of integration of a wide range of existing GIS data. As many datasets use descriptions developed at a project-to-project basis with no shared ontology or vocabularies in place, without a selective approach, it would be hardly

possible to find a common ground for harmonisation of specific items in geospatial databases provided by different partners. Harmonisation of GIS datasets would need direct and time-consuming intervention into the datasets to add shared metadata for each item. It might be even a complicated step in the case of repositories where license rights and authorship of the datasets is kept by the external providers.

It was decided to change the strategy for spatial data integration. We decided to pursue further a possibility of "passive" integration, using geospatial services (WMS, WFS, WCS) provided by ARIADNE+ partners, using existing or newly established GIS servers. This way of data sharing corresponds with collection level integration, is flexible and may be continuously enhanced as partners will make their data available in future.

Implementation deserved partners to identify existing GIS servers in place and to provide a specification of services available, together with a description compliant to AO-Cat. Such information was to be later included in the ARIADNE Catalogue and made public by ARIADNE Portal. It implicated further technical development of the ARIADNE Portal, implementing tools for working with geospatial layers, cartographic symbology, and geospatial queries (which are, however, tasks out of the T4.4.10 scope).

2 Designing application profile for geospatial data

Most of the archaeological material has its spatial properties and therefore collecting spatial data is a natural component of archaeological data processing. The specific spatial definition is dependent on various factors as a type of recorded entity, tools used for the spatial properties acquisition, chosen spatial representation or selected data format. All these factors make it difficult to find a common ground to integrate spatial data from different sources, if not linked with the specific topic (i.e., site locations, remote sensing data, etc.). Nevertheless, spatial data is being produced and shared in many fields and are being exchanged and re-used online on a daily basis. To this end, a system of standardized protocols has been already developed by the Open Geospatial Consortium (OGC) that allow geospatial data to be shared between individual producers and users. These protocols are mostly known as web GIS services of specific types, like WMS (Web Map Services) for maps/images, WFS (Web Feature Services) for vector data and WCS (Web Coverage Services) for raster data sharing.

Many of the existing geospatial services are relevant for the field of archaeology; some are provided by the ARIADNE+ partners with existing or newly established GIS servers, others are available for specific countries and managed by non-archaeological stakeholders. Geospatial services can be considered equivalent to standard APIs developed for geospatial data, which are made available throughout the Internet and easily readable by most of the GIS software or webGIS libraries.

Developing an alternate model for sharing generic spatial data would be a pointless effort for ARIADNE+ partners if there is an existing model which overcomes all the practical issues brought by the needs of data harmonization. This approach does not prevent direct integration of spatial data into ARIADNE Portal for specific fields but gives a common ground to share datasets of the unknown structure easily for view and re-use. Therefore, if exploitation of spatial data should be facilitated throughout archaeology on the international level, we need to identify such services for specific countries or regions and make them findable for the users in the ARIADNE Portal.

To be accessible and findable, geospatial services need to be described according to the AO-Catalog (AO_Cat) and mapped to the ARIADNE+ data model. After evaluation of available services, it is possible to directly integrate them into the ARIADNE Portal for data visualisation (WMS) and (possibly) for querying (WFS, WCS) in appropriate technical WPs.

*

We define five types of entities relevant for description and integration – geospatial service, geospatial layer, feature class, coverage, and service provider.

Geospatial service represents an instance *AO_Data_Resource*, which is structured according to relevant international standards and provided as an online service. It is described by metadata available upon request and may be queried for results. Depending on the service type and the specific request, a response of the query can be a geocoded image, set of geospatial records or structured metadata describing archaeological situations and/or their context. To query a service, it is needed to know its location defined as URL (which serves as the base address for HTTP requests used to get the data or metadata), version and type. Type of service is information important both for further implementation into the ARIADNE Portal GIS environment and for description in the AO_Cat. Different

types of services offer different solutions for viewing and querying, therefore, they should be linked to a fixed vocabulary of service types. Required information on the service can be easily described using standard properties in the AO_Cat (*has_type*, *has_description* and *has_landing_page*).

Geospatial layers and **feature classes** are specific outputs provided by geospatial services (WMS, respectively WFS) with uniform content, style and format. A layer may be composed of sublayers, a feature class is composed of individual features. Both entities correspond to the *AO_Collection*. A layer or feature class usually provides an individual dataset on a specific topic. It can be described by *AO_Spatial_Region_BBox*, *AO_Temporal_Region* and also by the attributes of its content (types of described objects etc.). For a better user experience and if appropriate technical prerequisites are met, it should be possible for a layer or feature class to be visualised in the ARIADNE Portal GIS environment upon user request.

Coverages are raster datasets provided by WCS services corresponding to the *AO_Individual_Data_Resource* entity. A raster dataset usually provides continuous and uniform geospatial data (grid) on selected topics in machine-readable and processable format. Same way as layers or feature classes, it can be described by *AO_Spatial_Region_BBox*, *AO_Temporal_Region* or content attributes and can be visualised in the ARIADNE Portal webGIS.

A **service provider** is a standard *AO_Cat* entity classified as *AO_Agent* and interconnected with geospatial service by *has_publisher* property. It represents the institutions and persons responsible for the service provisioning and for the dataset it offers. The detail of the provider description should be like the description of any other data providers represented in the ARIADNE Portal.

*

The suggested solution for geospatial data builds on existing standards of data exchange proven by international and interdisciplinary practice. Analysis shows that the geospatial data application profile needs no specific amendments to the standard AO_Cat, however, the data integration and possible re-use are dependent on the upcoming development of ARIADNE+ services, esp. the ARIADNE Portal.

3 Registering geospatial services

Creating an interactive catalogue of the geospatial services available online would greatly enhance compatibility and rise attractiveness of the ARIADNE Portal not only for archaeologists but also for other scientific domains (esp. for environmental research, geosciences, heritage management, etc.) in which online published geospatial layers are a common way of data exchange. The positive outcome is very much connected with the willingness of individual partners to share their local expertise and provide AO_Cat with a list of the most relevant services. For T4.4.10, it became crucial to rise discoverability of both:

1) **Archaeological data,** which would be available in a standardized form and easy to integrate with other spatial data for any use-case even outside archaeology.
2) **Proxy-data** (e.g.: elevation data, historical climate data, hydrology data, old maps, etc.), useful for archaeologists in specific fields and regions, which can contextualise their research data and are made available by non-archaeological service providers.

In Dec–Feb 2020, a questionnaire to collect information on available services was created and distributed to all members of the ARIADNE+ consortium. The goal of the survey was to specify the nature of available and relevant geospatial services from the perspective of content, licensing as well as technical and administrative aspects. Answers in the questionnaire were to represent individual services. Sixteen answers were delivered, and the subtask leaders achieved a general overview of the possible character of data services, which shall be integrated into the ARIADNE Portal.

Based on the experience gained, it was decided to include the topic of spatial data in the first VRE Use Cases Workshop. In addition to spatial data, it focused on burial data, took place at the end of January 2021 and was attended by 52 people. It presented the findings to date, the revised objectives of T4.4.10 (including specific examples of services requested) and proposed a further information gathering process, which included a new round of questionnaire survey among partners. This was combined with the offer of individual consultations to better understand the correct format of the data to be collected and the purpose of recording it. These consultations and the follow-up survey were implemented in June 2021.

The last round of the survey was supported by the provision of sample data for the Czech Republic (a total of 25 geospatial services described). It was also backed by the above-mentioned consultations. The subsequent data collection took more than a month, after which the questionnaire was closed and the data evaluated. Ten responses from six ARIADNE+ members were recorded in the questionnaire, and only 3 of the records described the available geospatial services in a standard way. The usual problem was that providers gave a description not for a standard geospatial service, but for webGIS applications used to view specific datasets, which are not machine-readable and cannot be re-used by other services based on standard protocols or included in the ARIADNE Portal. Thus, the majority of the valid results consist of the sample provided for the Czech Republic alone and 43 responses provided by an external partner (HES) describing 18 services (some in multiple versions - WMS/WFS/ArcGIS Server/ATOM Feed; all valid results are summarised in Annexe 2).

4 Final summary

The original goal of the T4.4.10 was to find out possibilities of the item level integration of geospatial databases and GIS datasets into the ARIADNE Portal. During March-May 2019, an initial information assessment on available datasets was conducted among the concerned partners. The survey results showed a vast diversity of data on the item level as well as the intersection of transversal categories with other subtasks. In June 2019, ARUP-CAS suggested an update of subtask T4.4.10 focus, aiming at integration of the archaeological and archaeologically relevant online geospatial data services in the ARIADNE Portal.

Based on this decision, two rounds of questionnaire surveys were organised (2020 and 2021), which resulted in the collection of valid information on 53 available geospatial services. However, their representation is heavily imbalanced, with a predominance of data for Scotland and the Czech Republic, due to the more intensive involvement of local partners. It appears that the European archaeological community is currently not fully prepared to integrate archaeological data using geospatial services, or does not always understand this need in a standard way. The exact reasons for this situation cannot be easily identified, but from the perspective of the authors of this report it is a combination of:

(1) Low awareness of existing standards for geodata exchange and presentation.
(2) Low level of standardisation of spatial data and its storage, preserved predominantly offline or as a part of larger datasets with unknown internal structure.
(3) Consequences of the popularity of building custom-made webGIS portals and applications to present archaeological data, which are often very user-friendly and of high quality when utilized by human users, but do not envisage machine-to-machine re-use.
(4) Licensing and other related barriers to sharing and publishing archaeological data.
(5) The limited competence of archaeologists to create and use OGC geospatial services.

With these points in mind, European archaeology should not limit itself to finding its own ways when it comes to spatial data and its management but can be well inspired by the procedures now standard in the earth sciences or land administration. Archaeological data are by their nature intertwined with these types of data in many ways and adopting proven concepts could provide archaeology not only an easier path to modernisation but also the opportunity to better communicate and share spatial data on the interdisciplinary basis.

From the point of view of the ARIADNE+ project, we consider that the tasks set out in T4.4.10 have been fulfilled, albeit with not a fully satisfactory outcome. However, this is clearly in line with the current state-of-the-art and only future systematic efforts to counteract the problems related to the points outlined above could bring improvements. The output of T4.4.10 is thus on the one hand a summary of the state-of-the-art presented in this report, but also an organized set of sample geospatial data layers that can be used for testing purposes while extending the functionalities of the ARIADNE Portal and to test model use-cases in corresponding work packages.

Annexes

Annexe 1 – Responses to initial assessment

Annexe 2 – List of services valid for further re-use