D12.3 Interim report on JRA activities

Version 1.0 *(final)*

Date 21 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 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**

Author: **Carlo Meghini**

                                 CNR ISTI

Contributing partners: **Alessia Bardi, CNR ISTI**

                                 Pasquale Pagano, CNR ISTI

                                 Maria Theodoridou, FORTH

                                 Johan Finn, SDN

                                 Kai Salas Rossenbach, INRAP

Quality control: **Holly Wright, UoY-ADS**

Document History

- 30.11.21 - Document skeleton created by Carlo Meghini

- 1 to 13.12.2021 - Contributions by WP leaders of WP12 to WP16

- 17.12.20221 - Final version created by Carlo Meghini and submitted to project management for quality check

- 21.12.2021 – Revised Manuscript sent to Carlo Meghini for revision

Licensing Notice

Table of Contents

1 Executive Summary

This deliverable describes the progress on the Joint Research Activities (JRAs for short) of the ARIADNEplus project, and provides an integrated overview of all the scientific and technological developments that have taken place in JRA Work Packages. These Work Packages are:

- WP12 JRA1 - Data Integration and Interoperability (Lead beneficiary CNR)
- WP13 JRA2 - ARIADNEplus Infrastructure Operation and Management (Lead beneficiary CNR)
- WP14 JRA3 - The ARIADNEplus knowledge management system (Lead beneficiary FORTH)
- WP15 JRA4 - Innovative Services for Users (Lead beneficiary SND)
- WP16 JRA5 - Innovative Methods and Pilots (Lead beneficiary INRAP)

This deliverable covers Months 26-36 of the project and corresponds to the activities of task 12.6.

The document consists of one main Section (Section 3) which provides an overview of the progress in the JRA WPs, followed by a section for each of the WPs, from WP12 to WP16. Each WP Section has the same structure:

1. Goal of the WP
2. Planned schedule of activities (table)
3. Results obtained from Month 26 to Month 36 (bulleted list)
4. Problems encountered and open issues, if any
5. Future outlook

Progress in the JRA WPs was carried out on five main fronts:

- Infrastructure and Virtual Research Environments setup and operation
- Aggregation infrastructure
- The ARIADNE Content Cloud
- The ARIADNE Portal
- Services and pilots

On each of these fronts, the JRA activities of ARIADNEplus are proceeding according to the planned schedule. The only milestone in the third year of the project for the JRAs, MS19 *First four innovative pilots launched*, set for M35, has been achieved in spite of the difficulties created by the pandemic.

At the moment there is every reason to believe that the objectives set for the final year of the project will be fully achieved.

2 Introduction and Objectives

This deliverable describes the progress on the Joint Research Activities (JRAs for short) of the ARIADNEplus project, and provides an integrated overview of all the scientific and technological developments that have taken place in JRA Work Packages. These Work Packages (WPs for short) are:

- WP12 JRA1 - Data Integration and Interoperability (Lead beneficiary CNR)
- WP13 JRA2 - ARIADNEplus Infrastructure Operation and Management (Lead beneficiary CNR)
- WP14 JRA3 - The ARIADNEplus knowledge management system (Lead beneficiary FORTH)
- WP15 JRA4 - Innovative Services for Users (Lead beneficiary SND)
- WP16 JRA5 - Innovative Methods and Pilots (Lead beneficiary INRAP)

This deliverable covers Months 26-36 of the project. The activities carried out in the remaining period of the project lifetime, from Month 37 till the end of the project, will be detailed in the deliverables listed in the following Table:

Deliverable No.	Deliverable Title	Delivery Date
D12.4	Final report on data integration - JRA1	M46
D12.5	Final report on knowledge discovery and JRA activities - JRA1	M48
D13.3	Software Release Final Activity Report - JRA2	M46
D13.4	VREs Operation Final Activity Report	M48
D14.2	Final report on the ARIADNEplus knowledge management system - JRA3	M47
D15.2	Final report on the ARIADNEplus services - JRA4	M47
D16.2	Final report on the ARIADNEplus pilots - JRA5	M47

The present deliverable corresponds to the activities of task 12.6.

Given the objectives of the deliverable, this document has been designed to be a quick reference guide, that is concise and homogeneously structured. The detailed descriptions of the activities of the involved Work Packages will be given in the deliverables reported in the previous Table. The document consists of one main Section (Section 3) which provides an overview of the progress in the JRA WPs, followed by a section for each of the WPs, from WP12 to WP16. Each WP Section has the same structure:

1. Goal
2. Planned schedule of activities
3. Results obtained from Month 26 to Month 36
4. Problems encountered and open issues
5. Future outlook

3 Overview of the progress

This Section provides an overview of the progress achieved in the period covered by the deliverable. It is divided in two sub-sections: the first recapitulates the workplan for the JRA WPs, as presented in the project proposal, while the second identifies the main areas of work and elaborates on the progress achieved in each area.

3.1 Workplan

The following Figure (from the PERT Chart on page 32 of the ARIADNEplus Description of Work, Annex 1 (part B) to the Grant Agreement) schematises the Work Plan of the project and indicates the role of the different JRA WPs, also in relation to the other WPs of the project.
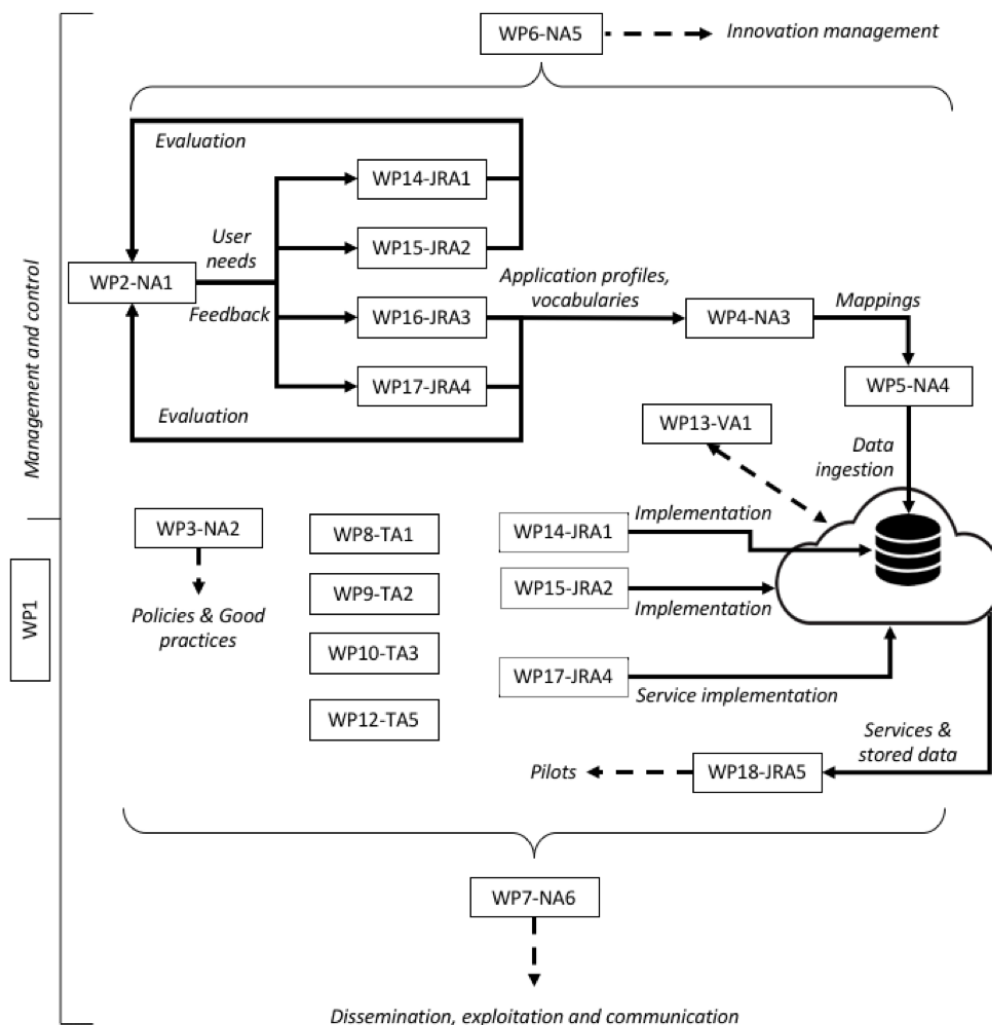

Figure 1: PERT of the ARIADNEplus project.

- In WP12 the ARIADNEplus Aggregative Infrastructure is set up and the main knowledge base of the project is established, the ARIADNE Content Cloud.

- In WP13 the ARIADNEplus Data Infrastructure is instantiated from D4Science, configured and further developed to serve the needs of the archaeological community. Moreover, the Virtual Research Environments (VREs) offering the services to support the R&D activities of the project are created on top of the Data Infrastructure, duly implemented and put into operations.

- WP14 focuses on the creation of the ARIADNEplus Knowledge Organization System (KOS), defining applications profiles (used in WP4 (NA3) for mappings and in general for managing the KOS), vocabularies and gazetteers.

- WP15 implements services, including visual, annotation, text mining and geo-temporal services. It also implements the back-office services that manage queries.

- WP16 implements a number of innovative pilots based on the ARIADNEplus technology and dataset integration.

All JRA WPs began in Month 1 with preparatory activities (planning, internal WP organisation, design, etc.) and had a WP kick-off meeting as part of the project Kick-Off. They will end at Month 48 contributing to the final reporting, including the project Progress Report.
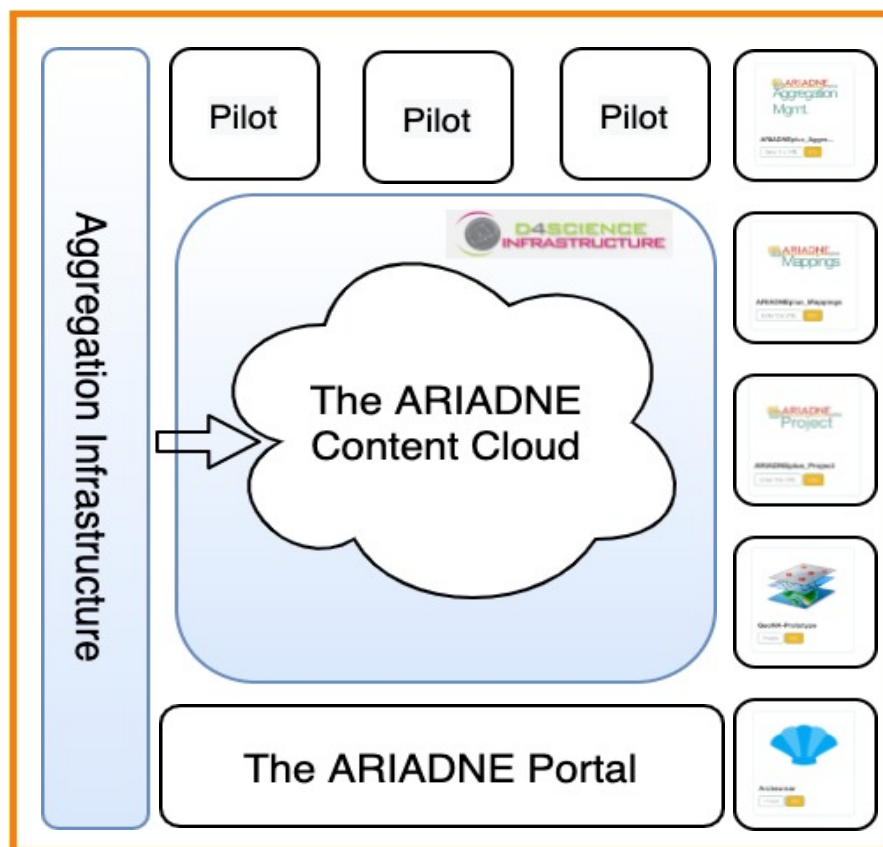
Figure 2: The different elements that make up ARIADNEplus.

3.2 Progress achieved

Progress in the JRA WPs was carried out on five main fronts (see Figure 2 above):

- Infrastructure and Virtual Research Environments setup and operation
- Aggregation infrastructure
- The ARIADNE Content Cloud
- The ARIADNE Portal
- Services and pilots

A separate sub-section is devoted to each.

3.2.1 Infrastructure and VREs setup and operation

This part of the work has been carried out exclusively within WP13.

During the third year of the project, the infrastructure for ARIADNEplus was maintained and enriched to cater to the needs of the project. A detailed account of the services added during this period is given in Section 5.3.

In addition, several VREs were operated: AggregationMgmt, Mappings, GeoNA-Prototype, Archeomar, and Project.

Finally, a new VRE, the ARIADNEplus_Lab, has been designed, configured and provisioned. ARIADNEplus_Lab allows any type of computations on data and metadata to be performed, and is open and accessible to all ARIADNEplus users. ARIADNEplus_Lab VRE also provides access to the ARIADNEplus Content Cloud stored in GraphDB via the standard SPARQL API.

A more detailed account of Infrastructure, including statistics on the accesses of the VREs, is given in Section 5.

3.2.2 Aggregation infrastructure

Data aggregation is essentially an inter-WP activity, carried out by WP12 in cooperation with WP14, and two WPs in the NA set of activities: WP4 *Integrating the datasets of the Archaeological Research Communities* and WP5 *Extending the ARIADNEplus Data Infrastructure* (ADI). The objective of this activity is to set up, implement and operate the ARIADNE Aggregative Infrastructure in order to create the ARIADNEplus Content cloud (AC).

At the beginning of the project, it was decided to re-perform the aggregation of the data already aggregated by the ARIADNE project, due to two main reasons:

1. The enhanced scope and functionality of ARIADNEplus required the development of an improved ontology than the ARIADNE Catalog Data Model, better aligned with the CIDOC CRM and more tightly related to the objectives of the project. As a consequence, the original data had to be transformed using the new ontology, which required a new aggregation process.

2. The infrastructure for ARIADNEplus aggregation is different from the one used in ARIADNE, based on a different technology and fully integrated with the D4Science platform. As a consequence, the inevitable update of the previously aggregated data required a re-definition of the aggregation process for these data, which made it effortless to re-aggregate the old data as well.

During the third year of the project, server-side data aggregation was carried out on three main fronts:

- aggregation of collection-level metadata from the ARIADNEplus partners into the ARIADNEplus Catalogue, a major component of the ARIADNEplus Content Cloud (described more fully in next Section); this is a continuation of the activity started at the beginning of the project;
- aggregation design of the item-level data from the ARIADNEplus partners into the domain knowledge bases, which altogether constitute the other major component of the ARIADNEplus Content Cloud; this is a new activity started in current reporting period, which has mainly involved the development of application profiles (see below for details) and of the mappings required to harmonise the item-level data provided by the partners according to these profiles;
- aggregation design of the item-level data from the THANADOS project into the burial knowledge base of the ARIADNEplus Content Cloud; this is also a new activity started in this reporting period, aimed at acquiring the data produced by the THANADOS project for the enrichment of the ARIADNEplus Content Cloud. The activity so far has tested the burial Application Profile for these data, and the design of the mappings for harmonising them with those provided by the ARIADNEplus partners.

Client-side data aggregation has resulted in the development, implementation, testing and release of the Activity Dash, a tool for monitoring the data provision and aggregation process, fully integrated in the D4Science environment. A tutorial on Activity Dash was given to 26 partners in February 2021.

Regular bi-weekly conference calls were held amongst representatives of the four relevant WPs to monitor the progress of data aggregation since the beginning of the pandemic.

3.2.3   The ARIADNE Content Cloud

The ARIADNE Content Cloud is the information repository collecting all the knowledge managed by the ARIADNE infrastructure.

From a technical point of view, the AC is an RDF dataset consisting of several RDF graphs. From an application point of view, the AC is a composite knowledge base that can be understood as consisting of several related parts, namely:

- the ARIADNE Catalog, giving collection-level knowledge about the resources of the AC, including data resources, services and the entities needed for contextualisation;
- several knowledge bases, each containing item-level knowledge about a specific archaeological domain.

The ARIADNE Ontology (AO for short) is the ontology axiomatising the terms for making the statements in the AC. The AO is a formal ontology of the resources managed by the research infrastructure, with a special focus on the archaeological domain and the ARIADNE infrastructure, developed by the ARIADNEplus project.

Mirroring the structure of the AC, the AO was structured into sub-ontologies, namely:

- a Catalogue ontology, named AO-Cat, which axiomatizes the terms for the statements in the ARIADNE Catalogue;

- several Application Profiles, axiomatizing the terms for the domain-specific parts of the AC.

During the third year of the project, the work on AO-Cat has mainly been focused on extending the ontology to cater to the new requirements coming from the data providers or from the Portal user group. Three main extensions are:

- adding images to archaeological records;
- allowing description of level of precision for spatial information;
- allowing description of provenance, authoring and versioning (PAV) information.

The first two extensions are described in detail in deliverable D5.3 because they are of interest to the content providers. In contrast, the integration of the PAV ontology has not been exposed to the content providers, as it was automatically derived by the aggregative infrastructure and kept for technical purposes, hence not offered as an access criterion. In fact, there is no application requirement for accessing the AC based on provenance, authoring and versioning information and consequently there is no functionality within the Portal for expressing queries on these aspects of the data. During the same period, work on Application Profiles was more intense, leading to the development and release of four Application Profiles:

- the Inscriptions Application Profile
- the Heritage Science Application Profile
- the aDNA (ancient DNA) Application Profile
- the Mortuary Application Profile

More details are given in Section 6 below.

3.2.4 The ARIADNE Portal

The development of the ARIADNEplus Portal is the objective of Task 12.3 in WP12. This Task was expected to extend the functionality of the Portal developed by the ARIADNE project, offering search and browse on a larger set of data and adding some (minor) functionalities based on the requirements collected from the users in WP2. But the activities regarding the Portal extend well beyond T12.3, as the data in the Portal must first be collected, then harmonised and inserted into the Content Cloud, from where they are queried through SPARQL to be published to Elasticsearch, where they are indexed and ready to be searched by users. During the third period of the project, work on the Portal addressed the following aspects:

- Improving the quality of the data displayed by the Portal v1.0 by extending the data provided by several partners through SPARQL queries. These queries compute values on the basis of other values, the AO-Cat ontology and the vocabularies employed. Another improvement has been obtained by passing more data from the Content Cloud, for example, spatial coverage and native period terms.
- Release of the new Portal in March 2021. The new Portal gives access to a larger Content Cloud than the previous one. It also provides access to higher quality data and richer visualisation, thanks to the improvements mentioned in the previous point.
- A new Portal was released in December 2021, with enhancements on the Content Cloud.

During the final year, two major releases of new features will be put into production, one before the summer and one before the end of year. The features of these releases will focus on temporal, topical, and multilingual search features, as well as preparing the portal software for a longer-term post-project administration.

3.2.5   The ARIADNE Services and Pilots

This is the area of the Project where the pandemic has caused more delays, as the development of services requires greater interaction between service providers, user communities and technology providers supporting the integration of services on the D4Science platform. For this reason, a workshop for relevant people was planned to be held at CNR in Pisa in the beginning of March 2020, to be followed by additional events. Due to the pandemic outbreak the workshop was cancelled and no other events were organised. Separate online tutorials between partners were held instead to explain the possibilities offered of the platform. These tutorials only addressed one aspect of the interaction however, and they could not replace the in-person interaction necessary for the sharing of data, tools and scenarios to plan the service integration.

Nevertheless, during the third period of the project the only Milestone about services, MS17 *Design of innovative ARIADNEplus user services* due at Month 28 was met, and several services were integrated on the platform, namely the TEXTCROWD service, the space-time service, the multilingual query service and the geoserver service. The integration of other services will follow in the final year. This includes the VisualMedia service and the tool for the visual organisation of archaeological data. More details can be found in Section 6.

The development of pilots also suffered during the pandemic, as it was difficult to understand requirements without in-person interaction.

Nevertheless, during the third year of the project a seminar within INRAP, the Leader of WP16, and the partners involved in the pilots, was held via Zoom. During the seminar, a common framework for the pilots, based on Virtual Research Environments and services already integrated into the platform, or to be integrated in the next few months, was shared, discussed and validated. As a result, the pilots could all be launched and milestone MS19 *First four innovative pilots launched* could be achieved (see Section 8).

4 Overview on WP12 Data Integration and Interoperability

4.1 Goal

The goal of WP12 is to implement and deploy the components of the ARIADNEplus infrastructure that support the integration and interoperability of the data provided by the members of the consortium. The integrated data are made available to the ARIADNE Portal and the pilots developed in WP16 via two services: (1) the ARIADNEplus knowledge graph, which exposes data via a SPARQL API, and (2) an Elasticsearch service accessible via the Portal (see fig. 3).
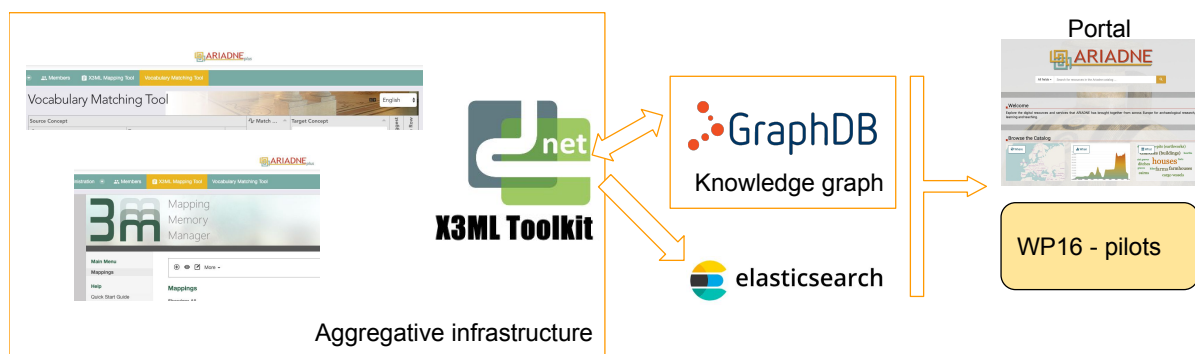

Figure 3: Services and tools for data integration and interoperability in the ARIADNEplus infrastructure

The aggregative infrastructure includes services and tools required to perform data collection, transformation, and harmonisation according to domain-specific vocabularies and ontologies:

- 3M Editor which facilitates the mappings from local metadata formats to AO-Cat;
- Vocabulary Matching Tool which facilitates the mappings from local subject terms to terms within the Getty AAT
- D-Net which is the ARIADNEplus aggregator based on the D-Net software toolkit. It collects provider XML records and integrates the X3ML toolkit to execute 3M mappings. It is configured to implement the aggregation workflows defined in collaboration with WP5.

The knowledge graph is a database using machine-readable and actionable Semantic Web-based tools and methodologies and the Resource Description Framework data model. It provides a Linked Data endpoint (using the SPARQL query language) and reasoning capabilities for advanced knowledge discovery services.

The Portal is the main entry point for humans wishing to search, browse and access the aggregated resources using standard textual or map-based searches.

4.2 Planned schedule

The following table covers M26-M36.

Goal	Planned delivery date	Status (textual description)
Portal release v1.0	M24	Released in M27 (March 2021)
Increase resources for GraphDB	M27	Completed
Define new Elasticsearch mapping to serve portal v2.0	M30	Completed
GraphDB upgrade to v9.8.0	M33	Completed
Portal release v2.0	M36	Released
Knowledge graph documentation	M36	Available on the Lab VRE and GitHub

4.3 Results obtained in M25-M36

4.3.1 Knowledge graph

- Tuned the configuration and added resources to the public instance of GraphDB (M27)

- Improved data quality within the portal v1.0 by: (i) passing more properties from GraphDB to Elasticsearch about spatial coverage and native periods; (ii) improving enrichment queries for several providers (M27)

- GraphDB staging and public instances upgraded from 9.0 to 9.8.0 (M33)

- Properties of the special provenance graph (which contains information about which endpoints and datasets were added to GraphDB, and when they were added) were mapped to the PAV ontology (M33)

- Documentation:

    o Sample queries and code for documentation purposes made available via the ARIADNE infrastructure GitHub repository (M33)

    o Documentation about the knowledge graph, including information about the organisation of data in named graph and sample SPARQL queries. The documentation is a Jupyter notebook that can be executed in the JupyterHub available in the

ARIADNEplus Lab VRE. Static versions (wiki/pdf) of the documentation were also made available in the VRE workspace. (M36)

4.3.2 Elasticsearch

- New Elasticsearch mapping defined to include additional properties from AO-Cat and support the new requirements of the portal (M30)

4.3.3 Portal

- Regular meetings of the Portal Requirements group were organised to test the Portal, identify enhancements to be implemented and issues to be solved

- M27 (March 2021): Release of the new portal v1.0

- M36 (December 2021): Release of the updated portal v2.0

4.4 Problems encountered and open issues

Nothing to report.

4.5 Outlook

4.5.1 Portal

During next year, two major releases of new features will be put in production, one before summer and one before end of year. The features of these releases will focus on temporal, topical, and multilingual search features, as well as preparing the portal software for a longer-term post-project administration.

4.5.2 Elasticsearch

The Elasticsearch mapping (aka schema) will be updated according to the indications of the portal user requirements group.

4.5.3 Aggregative infrastructure

No further main development activities are planned. If needed, additional collection protocols for the integration of data coming from sources that are not compliant with the standard protocols already supported by the aggregator will be developed.

4.5.4   Knowledge graph

The performances of GraphDB will be monitored, resources and configuration adapted and, if needed, a plan for the migration to the commercial version of GraphDB will be proposed to the JRA supervisor.

The queries for building the records for the Elasticsearch index will be updated if required by the planned versions of the portal.

Documentation will be updated based on the feedback of the consortium and external users.

5 Overview on WP13: ARIADNEplus Infrastructure Operation and Management

5.1 Goal

The goal of WP13 is to deliver and operate the ARIADNEplus Infrastructure.

This infrastructure is built by exploiting the computing and stored resources operated and provisioned by D4Science.org together with services for their management and administration. Upon those resources, data, tools, and services are deployed and made available to the research communities of the project for access and use, via an authentication and authorisation mechanism (also provisioned by this task) compliant with the EOSC identity federation.

The activities performed include:

1. the set-up of monitoring, alerting, and accounting services for all federated resources to guarantee the required Quality of Service (QoS);
2. the operation of a number of Virtual Research Environments (VREs) providing support for the exploitation of the provisioned storage and computing facilities to the other JRA Work Packages of the project via the production of training material and how-tos aiming at simplifying all phases of the preparation of new tools and services candidates for integration.

The above-mentioned storage and computing facilities are meant to be accessible via a VRE - exploiting the procedures and tools already validated and used in D4Science - by the tools and data required by the services developed in the context of WP15 (ensuring that the proper allocation of resources is guaranteed for their operation) and by WP16 activities.

As a complement, WP13 manages the software release process covering all stages from integration, through documentation and validation, up to provisioning in JRA work packages. Therefore, it defines release and provisioning procedures, establishes the release plan, coordinates the release process, and operates the tools required to support the release and provisioning activities by also taking care of the distribution of the software.

5.2 Planned schedule

Goal	Planned delivery date	Status (textual description)
ARIADNEplus e-infrastructure Gateway	M1	Completed
MS12 First Virtual Research Environments deployed	M20	Completed
ARIANDEplus Laboratory	M30	Completed

5.3 Results obtained

- Operation of the ARIADNEplus e-infrastructure Gateway available at
   https://ariadne.d4science.org;
- Operation of the AggregationMgmt, Mappings, GeoNA-Prototype, Archeomar, and Project VREs;
- Integration of new computing resources to support the configuration and deployment of a Kubernetes cluster and a computational cluster;
- Adoption of the new Identity and Access Management (IAM) supporting the standard OpenID Connect (OIDC). This activity required the porting of existing users, roles, VREs settings, services privileges, etc. to the new technological framework;
- Release of the new geographical Map Viewer allowing increased and customisable zoom levels, tailored Web Feature exploitation and role-driven access to geographical features;
- Release of the new Map Explorer allowing access to textual descriptions, images, and any other related information associated with a geo-package;
- Exploitation of the new version of the Spatial Data Infrastructure (SDI) provided by D4Science. The SDI is composed of a set of technologies for the storage, indexing, cataloguing, discovery, and access to geographical datasets according to OGC (Open Geographic Consortium) standards. This activity required the porting of existing datasets, styles, VREs settings, access privileges, etc. to the new technological framework;
- Configuration and deployment of JupyterHub technology in the Kubernetes cluster. This technology enabled the activation and exploitation of a JupyterLab service and the exploitation of computational environments and resources without burdening users with installation and maintenance tasks. This JupyterHub environment is preconfigured with libraries and packages to ease the execution of common data analytics tasks, and provides access to the Workspace enabling sharing of resources with other members much easier;
- Configuration and deployment of RStudio services in the newly added computational cluster. RStudio provides an integrated development environment for R. It includes a console and a syntax-highlighting editor, and enables code execution. Tools for plotting are also included. This RStudio environment is *(i)* preconfigured with libraries and packages to ease the

execution of common data analytics tasks; and *(ii)* provides seamless access to the Workspace enabling sharing of resources with other members much easier;

- Configuration and activation of the Analytics engine framework (DataMiner) to permit the execution of an array of analytics methods by transparently relying on distributed computing infrastructure. Executions can run either on multi-core machines or on distributed working nodes. New software can be integrated by using the dedicated Software Importer (SAI);
- Integration of GraphDB. GraphDB provides users with access to the ARIADNEplus Knowledge Base, an archaeological Linked Open Dataset modelled according to the ARIADNE ontology and provided by an international network of organisations. With GraphDB any user can explore the knowledgebase with the available web GUI or programmatically with SPARQL queries;
- Design, configuration, and provisioning of a new VRE named ARIADNEplus_Lab which hosts all users needing to perform any type of computations on data and metadata. This new environment is open and accessible to all ARIADNEplus users that can easily create a new notebook, for example, using RStudio, or integrating their software and running it as-a-service by using the Web Processing Service (WPS) standard API. ARIADNEplus_Lab VRE also provides access to the GraphDB.

5.4 Encountered problems and open issues

Nothing to report.

5.5 Outlook

| Goal	Planned delivery date	Status (textual description)
Design, configuration and delivery of the GNA VRE	M38	This new environment will be open access and provide access to Italian archeological excavations. It will collect spatial stratigraphic excavation data from the entire national territory. Data from stratigraphic excavations can be visualised either via a map-based application or a more traditional catalogue view.
Enhancement of the ARIADNEplus Lab VRE	M42	The ARIADNEplus Lab VRE will be enhanced to include most services, potentially all of them, developed by the ARIADNEplus community.
Adoption of the new Gateway service	M46	A new version of the Gateway service will be designed and rolled out. The new Gateway service will bring enhancement such as the adoption of modern OSGi standards and it will be broken down into many modules to benefit from the Modular Development Paradigm.
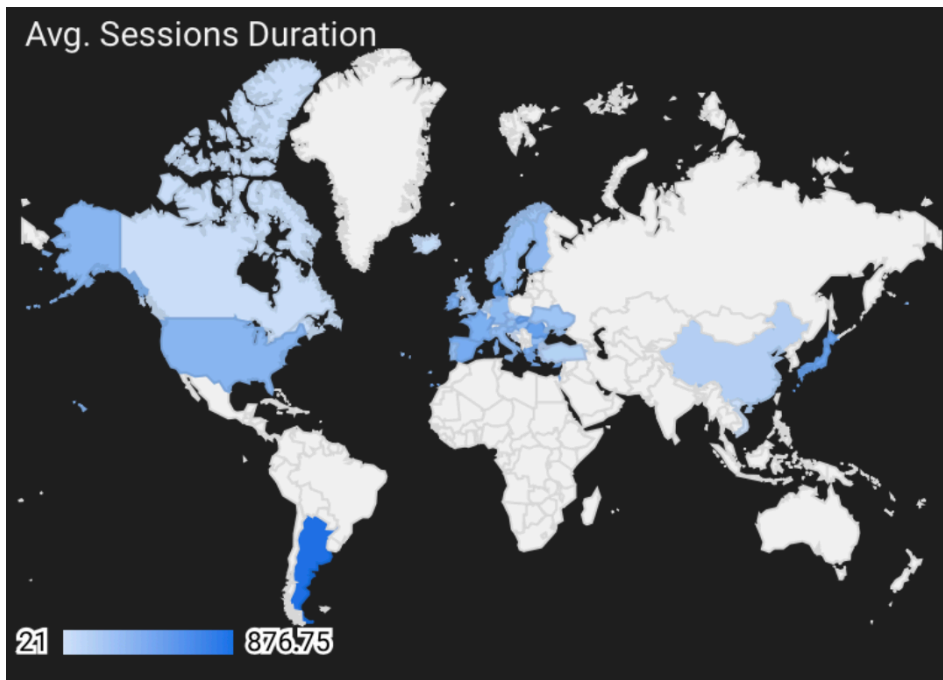
Figure 4: Detailed VRE accesses.
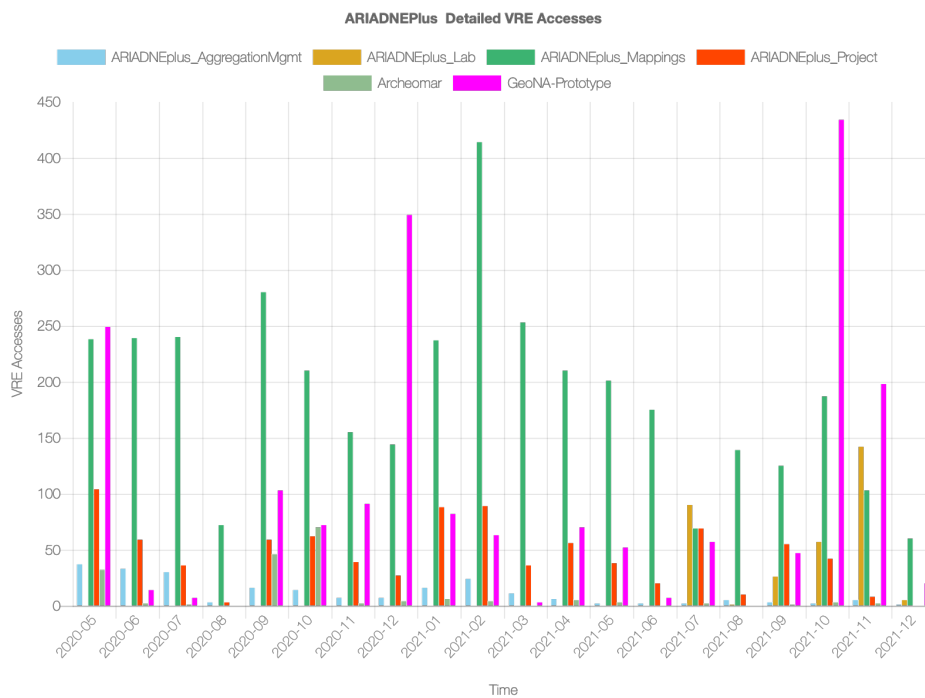
Figure 5: Average Session Duration within the ARIADNEplus Data Infrastructure.

6 Overview of WP14: the ARIADNEplus knowledge management system

6.1 Goal

The goal of WP14 is to define the ARIADNEplus semantic framework which:

- monitors the data provision and aggregation process, through a convenient and easily accessed system. This system offers a flexible and friendly collaborative environment, to be used by many users simultaneously for managing workflow activities and tracking their progress. The system acts as a reference point for any ARIADNEplus stakeholder to check the status of any process and provide the appropriate feedback;

- defines appropriate disciplinary Application Profiles required for the diverse types of data handled by ARIADNEplus. Application Profiles are specialisations expressed as CIDOC CRM extensions for the different thematic groupings broadly corresponding to those identified in Task 4.4, and aggregated according to similarity;

- provides domain specific vocabularies, a space-time gazetteer and a periodisation system to address the interrelation between space and time.

The work done in WP14 will be assessed in Task 14.4 in close collaboration with T4.4 and will provide feedback and changes where necessary.
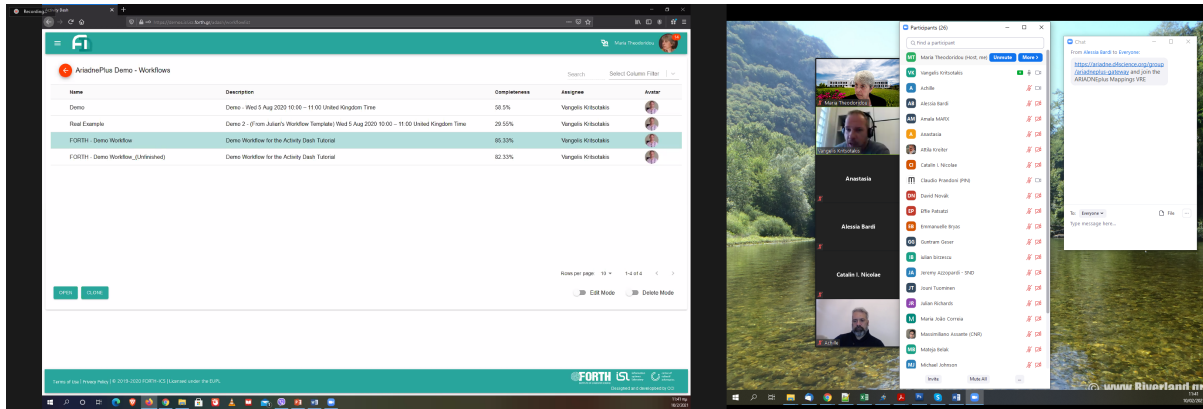
6.2 Planned schedule

Goal	Planned delivery date	Status
Finalisation of the implementation of the Activity Dash v1.0	M24	Completed
Integration of the Activity Dash on the Aggregative Infrastructure	M30	Completed
Provision of the tool to users of WP4 in order to monitor their aggregation activities.	M32	Completed
Testing and Validation of the tool	M48	Started M32 - Ongoing
Definition of the Inscriptions and Heritage Science Application Profiles	M24	Completed

Goal	Planned delivery date	Status
Definition of the aDNA data Application Profile	M34	Completed
Definition of the Mortuary data Application Profile	M35	Completed
Usage of the Application Profiles (Inscriptions, Heritage Science, Mortuary data, aDNA) to provide data to the ARIADNE Cloud	By M48	Started M33 - Ongoing
Analysis and development of other application profiles	By M48	Started M25 - Ongoing
Substantial space-time vocabulary via PeriodO (based on T4.4 data integration)	By M48	Started M25 - Ongoing
Supporting data partners as they enter their specific time periods into PeriodO	By M48	Started M25 - Ongoing
Liaise with the Getty Research Institute on requirements for proposal of specialist ARIADNEplus archaeological concepts and multilingual synonyms for AAT	M36	Completed
Assessing Application Profiles, CRM extensions	By M48	Started M25 - Ongoing

6.3 Results obtained in months M25-M36

6.3.1 Activity Dash

Activity Dash v1.0, the tool to monitor the data provision and aggregation process, was released for use by all partners in M24. Activity Dash was integrated into the D4Science environment and users could access the tool directly using their D4Science credentials. In February 2021 (M26) a workshop was organised, during which a detailed tutorial of Activity Dash was given to all partners. The workshop was attended 26 partners.

Figure 6: A template workflow was implemented according to which the workflow of each data source is defined. Currently there are 33 workflows.

There is continuous support for partners when they encounter problems in using the tool. Moreover, there is continuous work on optimising, debugging and improving the functionality of the tool, taking into consideration their feedback.

6.3.2 Application Profiles

Several application profiles are under development and testing is ongoing. During months M25-M36 the work on the four Application Profiles has matured significantly:

Application Profile for inscriptions, marks and graffiti:

- CRMtex v1.0 is stable and currently under revision from the CRM SIG (https://cidoc-crm.org/crmtex/)
- The Application Profile has been enriched and is ready to be tested and used for the encoding of the epigraphic information available in ARIADNEplus

24

- The new enriched version of the epigraphic model (described in D4.2 https://zenodo.org/record/4916299#.Yanxe9BByid) is currently being evaluated on epigraphic data provided by the Cyprus Institute and MiBAC
- The first extensive tests have shown that the model is suitable for describing the complex phenomena and entities of the epigraphic domain

Application Profile for heritage science data:

- CRMhs v0.85 (https://data.d4science.net/Mnmn)
- A preliminary test of the CRMhs model on C^{14} data has been carried out
  Evaluation of CRMhs for item level integration of heritage science information has continued
- The CRMhs model was updated based on feedback resulting from testing using different research scenarios
- The alignment between CRMhs and the aDNA model was verified and validated
- PIN continued its collaboration with FORTH on the alignment of the aDNA and CRMhs to make them suitable for item level integration

Application Profile for ancient DNA data

- Presentation of the aDNA application profile in the 2nd VRE workshop
- The aDNA application profile is described in D4.2
   https://zenodo.org/record/4916299#.Yanxe9BByid
- Preliminary tests are ongoing for item level integration

Application Profile for Mortuary Data

- Presentation of the Mortuary application profile to the 1st VRE workshop

- The Mortuary data application profile reached a first stable version and was formalised

- Preliminary tests are ongoing for item level integration

The development of new Application Profiles for Metal Detector Surveys (subtask 4.4.7), Remote Sensing (4.4.8), Geospatial Data (4.4.10) and Maritime and Underwater Archaeology (4.4.11) has proved to be unnecessary, both because AO-Cat was found to be sufficient to describe their entities, and because different strategies for data integration were implemented. These include the extensive use of vocabularies for terminological integration and the use of dedicated services (e.g., geographic systems) to establish data integration layers.

Other application profiles are under finalisation. There is continuous assistance and support to all the partners involved in the development and definition of application profiles for their research domains.

6.3.3   Vocabularies and gazetteers

Preliminary work, with some initial investigation of PeriodO and also multilingual possibilities using vocabulary mappings and other multilingual resources as Getty's Art & Architecture Thesaurus (AAT), Wikidata, and so on, has taken place. Data partners were supported when entering their period data into PeriodO. Moreover, data partners were supported in creating the mapping of their native subjects

to the AAT. The process for communicating with the Getty has been established. The work supporting partners in implementing PeriodO collections for ARIADNEplus and in mapping local subject concepts to AAT continues as part of the routine integration activity. The resulting set of partner period definitions will yield the definitive PeriodO / ARIADNE data collection.

6.3.4 Assessing the CRM extensions

- Completed mapping the burial database "In Touch with the Dead", testing the ontology through mortuary data item-level integration case study in WP 4.4.14. The mortuary data AP was written up as a reference document for partners. Further testing was undertaken by using the Mortuary Data AP for integration of partner datasets, starting with THANADOS
- Continued testing CRM extensions concerning excavation data item-level integration

6.4 Encountered problems and open issues

No specific issue has to be reported.

6.5 Outlook

Goal	Planned delivery date	Status
Testing and Validation of AD tool	By M48	Ongoing
Usage of the Application Profiles (Inscriptions, Heritage Science, Mortuary Data, aDNA) to provide data to the ARIADNE Cloud	By M48	Ongoing
Analysis and development of other application profiles	By M48	Ongoing
Substantial space-time vocabulary via PeriodO (based on T4.4 data integration)	By M48	Ongoing
Supporting data partners as they enter their specific time periods into PeriodO	By M48	Ongoing
Supporting data partners mapping their native subjects to AAT	By M48	Ongoing
Assessing Application Profiles, CRM extensions	By M48	Ongoing

7 Overview of WP15: Innovative Services for Users

7.1 Goal

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

The services cover a wide range of features useful for researchers, ranging from front-office services like visualisation services, various types of annotations and documentation, to more back-office oriented services like the multilingual query service used by the ARIADNEplus Portal for querying the underlying aggregated partner resources.

Task 15.1 is responsible for analyzing 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 finding and integrating other possible candidates as services in the ARIADNE infrastructure, and to produce a service design for the other services to use.

Tasks 15.2 through 15.5 cover the front-office services and Tasks 15.6 and 15.7 are the back-office services, which are developed under the ARIADNEplus umbrella.

Task 15.2.1 will build upon the VisualMedia EOSCPilot Science Demonstrator. The VisualMedia service, which enables displaying of archaeological information in the form of images and 3D models, will be adapted to the ARIADNEplus infrastructure and made available as an ARIADNEplus service.

Task 15.2.2 will make the pre-existing 3DHOP service at CNR, which links archaeological documentation to the 3D model of an artefact or monument and visualize it accordingly, adapted to the ARIADNEplus infrastructure.

Task 15.2.3 will rework the Ephemera service provided by CYI. The service visualizes the layers of archaeological excavations in 3D, together with related documentation. The service will also be adapted to the ARIADNEplus infrastructure.

Task 15.3.1 concerns the development of an annotation tool for archaeological reports and other related texts, building upon existing open-source tools.

Task 15.3.2 will extend the image annotation tool DAP, developed by CNR and AMZ.

Task 15.4 concerns the implementation of the text mining and NLP service, based on the previous ARIADNE NLP tool later developed into TEXTCROWD, a cloud-based Science Demonstrator within the EOSCpilot EU project. Work will consist in porting the previous ARIADNE NLP tool in the cloud environment, following what has been done for TEXTCROWD, and extending the NLP functionality to other languages beyond Italian (as done in TEXTCROWD), English and Dutch (as done in the ARIADNE tool). The task will make extensive use of the vocabularies, gazetteers and time period vocabularies.

Task 15.5 will implement the usual space-time services present in GIS systems and make it available through the ARIADNEplus infrastructure.

Task 15.6 concerns the development of a multilingual query service used by the ARIADNEplus Portal for querying the underlying aggregated partner resources.

Task 15.7 will implement a geoserver for geographical information provided by partners.

7.2 Planned Schedule

Goal	Planned delivery date	Status
MS17 Design of innovative ARIADNEplus user services defined	M28	Completed
Deployment of the VisualMedia service within the ARIADNEplus infrastructure	M48	Ongoing
Deployment of the tool for visual organisation of archaeological data within the ARIADNEplus infrastructure	M48	Ongoing
Deployment of the Ephemera service within the ARIADNEplus infrastructure	M48	Initial activities
Deployment of the archaeological text annotation tool within the ARIADNEplus infrastructure	M48	Initial activities
Deployment of the DAP tool within the ARIADNEplus infrastructure	M48	Initial activities
Deployment of the TEXTCROWD service within the ARIADNEplus infrastructure	M48	Completed
Deployment of the ARIADNE NLP tool within the ARIADNEplus infrastructure	M48	Initial activities
Deployment of the space-time service within the ARIADNEplus infrastructure	M48	Completed
Deployment of the multilingual query service within the ARIADNEplus infrastructure	M48	Completed
Deployment of the geoserver service within the ARIADNEplus infrastructure	M48	Completed

7.3 Results obtained

- ARIADNEplus user services design document finished

- The VisualMedia service authentication module and visualisation interface was updated to be more easily integrated as an ARIADNEplus service

- LiDAR visualisation component designed and released for the VisualMedia service. A definition of a common framework for 3D, LiDAR and image annotation expansion planned in the VisualMedia service to meet the needs of the ARIADNEplus partners involved

- Improvement and integration of Nexus tools for the visualisation of large 3D models in the VisualMedia Service

- A new web-based advanced image viewer, OpenLime, has been developed and published in connection to the VisualMedia service

- A number of new extensions were implemented locally in 3DHOP. The new features allow archaeologists to, *e.g.*, create interactive links ("hotspots") from the 3D model to related documentation. The new features will be integrated into the VisualMedia service.

- A tutorial, with the list of all the command utilities and descriptions to navigate, interact and perform geometric analysis of the visualised 3D archaeological excavations within the Ephemera tool has been released

- Short movies explaining of each of the command utilities of the Ephemera tool was released

- Survey for the preparation of user-trials aimed at testing the Ephemera tool and getting feedback from several groups of specialised users was launched

- New components of the Ephemera service are being developed, based on feedback from a user survey

- Expansion of Norwegian textual patterns for time periods and references to timespans for use within a multilingual timespan matching tool

- Development and GitHub deployment of the ReMatch multilingual timespan matching application with interactive web demonstration

- A pilot study on temporal information extraction from French grey literature was conducted.

- The Tokenizer tool was improved to deal with uncertain readings of inscriptions, as well as surface damage to incised letters

- The Elasticsearch service was updated to the latest version and a new, updated index mapping was deployed

- ARIADNEplus Portal was updated to facilitate the new version of Elasticsearch and new index mapping, including new multilingual features and spatial search features

- The Geoportal Service was enhanced and a reimplemented as a REST-based service. It has been designed to manage the publication lifecycle of complex spatio-temporal documents, supporting their materialisation and indexing within different platforms (Databases, Catalogues, OGC Services, etc. In particular, MongoDB was integrated, tested and validated.

7.4 Problems encountered and open issues

Several of the tasks in Work Package 15 were planned for integration into the ARIADNEplus infrastructure, *i.e.,* deployment on the D4Science platform. A workshop to increase knowledge about D4Science was planned to be held at CNR in Pisa in the beginning of March 2020, but due to the pandemic, the workshop was cancelled. Individual online tutorials between partners were held instead. It is however anticipated that continuous contacts during the lockdown period will avoid damage or excessive delays in the completion of work.

7.5 Outlook

The development of services is being finalised, but several new components have already been developed and released.

Goal	Planned delivery date	Status
Finalisation of texture processing and Nexus standardisation in the VisualMedia service	M48	Nearly completed
Deployment of the Openlime editor on ARIADNEplus infrastructure	M48	To be started
Make all extensions of 3DHOP publicly available	M48	Nearly completed
Establishment of an automatic ingest system to upload external users' data for visualisation and management through the Ephemera platform.	M42	To be started
Deployment of the Ephemera service within the ARIADNEplus infrastructure	M48	To be started

Goal	Planned delivery date	Status
Deployment of the archaeological text annotation tool within the ARIADNEplus infrastructure	M48	To be started
Deployment of the DAP tool within the ARIADNEplus infrastructure	M48	To be started
Refinement of the prototype multilingual timespan utility with a view to developing it as an NLP service recommending temporal metadata for archaeological reports (particularly abstracts/summaries).	M48	Ongoing
Deployment of the ARIADNE NLP tool within the ARIADNEplus infrastructure	M48	To be started
Refinement of the Elasticsearch index mapping based on portal requirements	M42	Ongoing
Enhancement of the Dynamic GUI to expose the new version of the Geoportal Service	M48	To be started

8 Overview on WP16: Innovative Methods and Pilots

8.1 Goal

The general goals of the WP are to:

• Define the innovative methods enabled by ARIADNEplus services for archaeological research communities

• Test the services and innovative methods in pilots using real use cases

• Demonstrate the advantages of using ARIADNEplus to the archaeological user communities

• Showcase how ARIADNEplus data and services are building applications for professionals, heritage managers and the public at large

The general approach consists of:

• Adopting an End-User perspective

• Being in line with the first objective of the project, namely, that the infrastructure is used, useful and innovative

• Creating elements that reinforce the archaeological perspective of the infrastructure

• Developing pilots demonstrating the innovation potential of the project results, methodology and tools

• Following a Case study approach

The pilots include seven case studies involving fourteen partners.

8.2 Planned schedule

Goal	Planned delivery date	Status
MS19 First four innovative pilots launched	M35	Completed

8.3 Results obtained

A collective seminar with the leaders of the pilots was held. During the seminar, the common framework of the pilots was shared, discussed and validated. The main points that have emerged from the seminar are:

- Task 16.1 *Establishment of a common framework for the ARIADNEplus pilots*. The framework has been drafted and presented to involved partners

- Task 16.2 *ARIADNEplus for Airborne-LiDAR data use and reuse*. The data to be tested is the LIDAR data from the partner data in the framework of geoserver service
- Task 16.3 *ARIADNEplus for a Historic Environment Spatial Data Infrastructure*. The data to be tested is the Historical environment spatial data from the ARIADNEplus partnership data in the framework of geoserver service
- Task 16.4 *ARIADNEplus for sharing archaeology in 3D*. The data to be tested is 3D data from the partner the framework of the VMS and of the multilingual query service on the infrastructure (cross search)
- Task 16.5 *ARIADNEplus for public/community archaeology*. The data to be tested is nonprofessional community archaeology data in the framework of the multilingual query service on the infrastructure (cross search)
- Task 16.6 *ARIADNEplus for understanding ancient and present cities: scanning the heart of Rome*. The data to be tested is a GIS from Roman data in the framework of geoserver service
- Task 16.7 *Exploiting the ARIADNEplus Data Infrastructure to understand complex phenomena of the Past*. The data to be tested is Research Data on Amphora marks, in the framework of the multilingual query service on the infrastructure (cross search).
- Task 16.8 *Exploiting the ARIADNEplus Data Infrastructure for preventive archaeology*. The data to be tested is French Excavation Reports using the framework of the TEXTCROWD service and the multilingual query service on the infrastructure (cross search).

8.4 Problems encountered and open issues

Nothing to report.

8.5 Outlook

- Pilots are finalised

- Seven Virtual Research Environments (VRE) are created, one for each pilot. These VREs will be the formal environments for the related pilots.

- Define and implement the visibility strategy for these VREs on the Portal.

9 Conclusions

The deliverable presents the status of the JRA activities at M36 of the project in a homogeneous and synthetic way, in line with the deliverable goal of providing the fundamental elements necessary to monitor and evaluate the JRA project activities at M36.

After introducing the objectives of the JRA work packages in ARIADNEplus, and the structure of the workplan (Section 2), Section 3 has given an overview of the progress, focusing on the five main areas of work:

1. Infrastructure and Virtual Research Environments setup and operation
2. Aggregation infrastructure
3. The ARIADNE Content Cloud
4. The ARIADNE Portal
5. Services and pilots

The detailed work performed in each Work Package is given in Sections 4 to 8, each devoted to a single Work Package. These Sections are structured in a homogeneous way to facilitate reading the document.

Overall, the JRA activities of ARIADNEplus are proceeding according to the planned schedule. The only milestone in the third year of the project for the JRAs, MS19 *First four innovative pilots launched*, set for M35, has been achieved in spite of the difficulties created by the pandemic.

At the time being, the project progress in the activities in WP12, 13, 14, 15 and 16 is on schedule, despite the difficulties created by the COVID-19 pandemic as the impossibility of meeting in person, the closure of academic and research centres, offices and so on. This confirms that it can be reasonably that the objectives set for the final year of the project will be fully achieved.