Archeologia e Calcolatori
36.1, 2025, 493-504
doi 10.19282/ac.36.1.2025.28

A PRELIMINARY ANALYSIS OF ARIADNE'S ADMINISTRATIVE METADATA

1. Introduction

This paper presents a preliminary analysis of the data available in the Advanced Research Infrastructure for Archaeological Dataset Networking in Europe (ARIADNE), not focusing on the contents – i.e. archaeological data – but on administrative metadata, trying to outline a broader profile of the infrastructure and its principal agents, i.e. the institutions providing the data. The primary aim of ARIADNE as a European archaeological infrastructure is to support research by providing access to digital resources and fostering innovative services. These goals are achieved in many ways: maintaining a large catalogue of digital archaeological datasets, promoting best practices, offering training and guidance, and supporting the creation of services for archaeology. The study was performed within the larger framework of the H2IOSC project that fosters the creation of a federation of the Italian nodes of four distinct Research Infrastructures (RIs) in the domains of Humanities and Cultural Heritage – i.e. CLARIN, DARIAH, E-RIHS, and OPERAS. Understanding how agents contribute to the main existing RI for Archaeology can help with the future improved integration of Heritage data at the Italian and European level.

2. The ARIADNE portal

One of the most prominent merits of ARIADNE – developed and funded by the European Community during the period 2013-2017 and later broadened as ARIADNE*plus* (Niccolucci 2015, 2017; Richards 2023) – is to enable data providers to connect their resources, acting as a broker; it provides a wider framework where fragmented data can be uniformly described and accessed (Aloia *et al.* 2017; Meghini *et al.* 2017, 18:2; Vlachidis 2017; Richards 2023). The main access to the ARIADNE infrastructure is the ARIADNE Portal (https://portal.ariadne-infrastructure.eu/) (Meghini *et al.* 2017, 18:14-18:17). The Portal allows users to perform queries through a user-friendly interface that permits filtering by keyword, by place, by time period and other parameters[1].

[1] As of April 2025, the available filters are: a free text search for all the fields; 'place'; 'time period' (by year or by time periods, in the last case also with a list of 'temporal regions' and 'cultural periods'); 'title'; 'Getty AAT subject'; 'publisher'; 'contributor'; 'original subject'.

ARIADNE created its own data model, the ARIADNE Catalogue Data Model (ACDM), specifically to represent archaeological information (Debole *et al.* 2015); it also uses the Art & Architecture Thesaurus (AAT) of the Getty Research Institute, Pelagios/Pleiades and PeriodO to define subjects, places and chronological periods respectively, improving interoperability (Meghini *et al.* 2017, 18:8-18:9). ARIADNE also created its own ontology which was directly based on CIDOC CRM (Meghini *et al.* 2017, 18:21-18:22). ARIADNE*plus* introduced the AO-Cat ontology, derived from the ACDM and from the PARTHENOS Entities Model (PEM) (Felicetti *et al.* 2023, 6). The core of the AO-Cat is the *AO_Resource* class (subclass of *AO_Entity*) that includes several resource types (Services, Data Resources, Collections, Documents, Digital Images) organized in different sub-classes (Felicetti *et al.* 2023, 18).

3. ARIADNE Knowledge Base and SPARQL End Point

ARIADNE has a SPARQL End Point accessible from the ARIADNEplus Lab VRE | Virtual Research Environment (https://ariadne.d4science.org/web/ariadneplus_lab/). The access is available on-demand for scientific purposes. The Virtual Research Environment provides a JupyterHub environment, a DataMiner environment, an RStudio environment and a GraphDB environment. GraphDB allows to access the ARIADNEPlus Knowledge Base, an archaeological Linked Open Data set based on the ARIADNE ontology and offered by a global network of organizations (Bardi, Ottonello 2022; Bardi *et al.* 2024). Using SPARQL queries the ARIADNEPlus Knowledge Base can be thoroughly explored and assessed.

Several SPARQL queries were tested before retrieving the final datasets on which this assessment is based: the SPARQL End Point appeared not to easily support large and complex queries with multiple variables able to provide ready-to-use datasets. The consequence was that raw rather than aggregated data were retrieved; the query selected to retrieve a dataset as similar as possible to the resources available on the ARIADNE Portal returned 4,075,233 overall rows and was performed on 14/03/2025. The query retrieved the URI of each resource that has rdfs:type aocat:Resource, the issue date, the name of the publisher, the ARIADNE subjects (if more than one, separated by a comma), and the original id (dataset1). It is important to stress that resources with one of these parameters missing were excluded from results. A second query returned 5,224,604 overall rows and was performed on 20/03/2025; it retrieved the URI and original ID of each resource that has rdfs:type aocat:Resource, the name of the publisher, the creator, the contributor, the owner and the responsible (dataset2). This second query, which generated a second distinct dataset, was necessary to

obtain more manageable datasets to be processed for different purposes. The further analyses described below were carried out on a local machine using RStudio.

For the reader's information, ARIADNE defines the previously mentioned roles as follows (Debole *et al.* 2015, 9):

1) publisher: '(…) an agent responsible for making the resource publicly accessible (via download, or API, or other)';
2) creator: '(…) an agent primarily responsible for creating the resource';
3) contributor: '(…) an agent primarily responsible for describing the resource in the Catalogue';
4) owner: '(…) an agent that is the legal owner of the resource';
5) (scientific) responsible: '(…) a person holding the scientific responsibility of the resource. It is the person who had enough competence and creativity to conceive the service or it is the person who gathered the data on a field'.

4. Assessment

A preliminary assessment of dataset1 revealed the following aspects. The overall number of resources per publisher is different from the one retrieved by a search through the online portal during the period in which the assessment was carried out. This misalignment has a couple of possible reasons: the query used to retrieve the dataset could not exactly match the data displayed on the Portal, or the ARIADNE Knowledge Base is not perfectly aligned with the online resources. The latter is possibly supported by the fact that the number of resources from specific publishers is indeed the same in both datasets.

A spatial analysis of the distribution of resources based on publishers' countries shows a major focus on Europe, with only the United States of America, Israel and Japan outside the continent (Fig. 1). Nevertheless, the distribution within the European countries is highly uneven, with countries like the United Kingdom, Ireland, Norway and Czech Republic being top publishers, countries like the Netherlands, Denmark, Sweden, France, Spain and Hungary being mid-tier publishers, while countries like Austria, Romania, Finland, Portugal, Germany, Italy, Slovenia, Serbia, Bulgaria, Greece and Cyprus having contributed less than 25,000 resources. A deeper focus on Italy shows that only three institutions are data providers for ARIADNE, and that they contribute a relatively small amount of data: the Istituto Nazionale di Fisica Nucleare (INFN, 43 resources), the Istituto Centrale per il Catalogo e la Documentazione (ICCD, 359 resources) and the Associazione Internazionale di Archeologia Classica (AIAC, 14,279 resources).

The number of publishers per each country is relatively small, with a mean value of less than 2 institutions that perform this role. The application

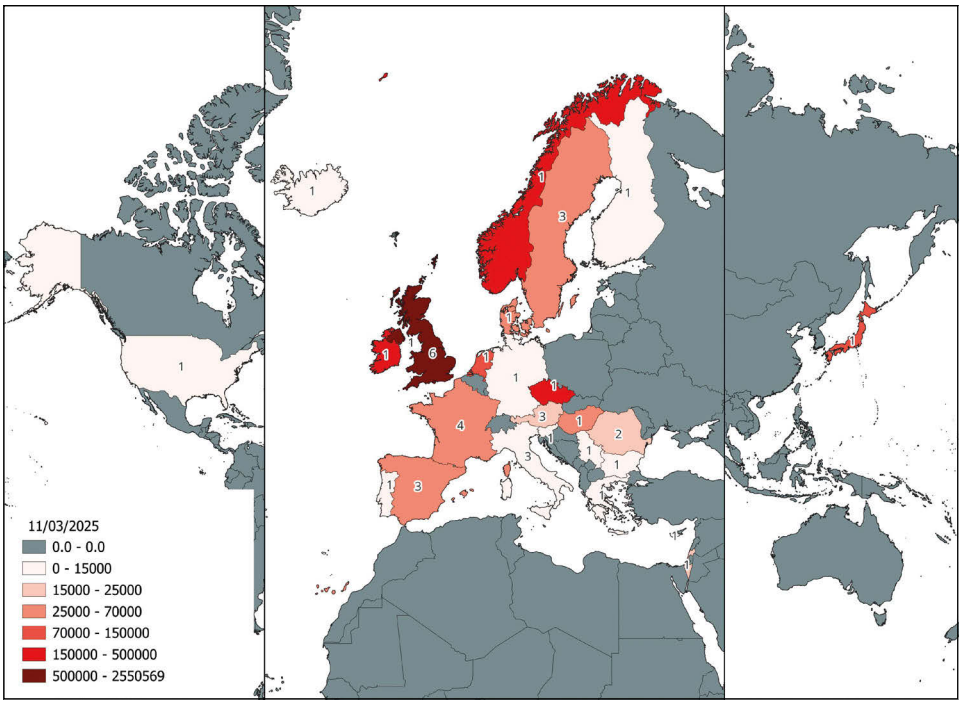

Fig. 1 – Density map of ARIADNE resources per publisher's country; digits in the map show the number of institutions involved.

of the Herfindahl-Hirschman Index (HHI, normally used to assess competition among firms) to the publishers from each country gave the results displayed in Tab. 1.

Further analyses were carried out by narrowing the initial dataset to only publishers that had more than 25,000 resources: this filtering returned fifteen institutions[2].

[2] The fifteen publishers are: Archaeology Data Service (ADS); Archaeological Information System of the Czech Republic (AIS CR); Aarhus University (AU); British Museum (BM); Centro para el Estudio de la Interdependencia Provincial en la Antigüedad Clásica - Universitat de Barcelona (CEIPAC); Data Archiving and Networked Services (DANS); Historic England (HE); Historic Environment Scotland (HES); Hungarian National Museum (HNM); Institut National de Recherches Archéologiques Préventives (INRAP); Nara National Research Institute for Cultural Properties (NARA); National Monuments Service, Department of Housing, Local Government and Heritage, Republic of Ireland (NMS); Royal Commission on the Ancient and Historical Monuments of Wales (RCAHMW); Swedish Rock Art Research Archives (SHFA); Museum of Cultural History, University of Oslo (UIO KHM).

Country	HHI
United Kingdom	0.344
Germany	0.593
Spain	0.697
Sweden	0.877
Italy	0.947
Romania	0.989
Austria	0.974
France	0.999
Bulgaria, Cyprus, Czechia, Denmark, Finland, Greece, Hungary, Iceland, Ireland, Isle of Man, Israel, Japan, Netherlands, Norway, Portugal, Serbia, Slovenia, United States of America	1

Tab. 1 – Application of the Herfindahl-Hirschman Index (HHI): values tending towards 1 indicate a high concentration of ARIADNE resources contributed by one single publisher.

A comparative analysis of the distribution of the ARIADNE resources and ARIADNE subjects revealed some relevant aspects of the composition of the ARIADNE Knowledge Base. From a numeric point of view, the Archaeology Data Service (ADS) with 1,097,289 resources and the British Museum (BM) with 945,229 resources are the two main providers/publishers, while the third one, Historic Environment Scotland (HES) with 334,636, and the following institutions are long distanced. The ADS is also by far the publisher that made available the greatest number of resources labelled with different ARIADNE subjects, highlighting their diversity; conversely, there are publishers (AU, CEIPAC, HE, INRAP, NARA, NMS, SHFA, UIO KHM) whose resources are all assigned to the same subject, revealing their high level of specialization (Fig. 2).The analysis of the distribution of resources per issue date is another aspect investigated: the overview appears twofold, with almost half of the publishers whose resources are distributed over a range not lower than twenty-five years, while the resources of the remaining half were issued in a single year (Fig. 3).

Another assessment was carried out looking for a possible correlation among the overall number of resources per publisher, the number of distinct issue years and the number of distinct combinations of ARIADNE subjects (Fig. 4). Results clearly show that most of the publishers are connected to relatively small datasets, focused on few subjects and issued during a relatively small span of time. The few exceptions are represented among the others by the ADS, which can count on a large number of resources belonging to many different subjects, and by the Swedish Rock Art Archives with resources that were issued during a very long period from 1800 onwards (with some previous years recorded as well).

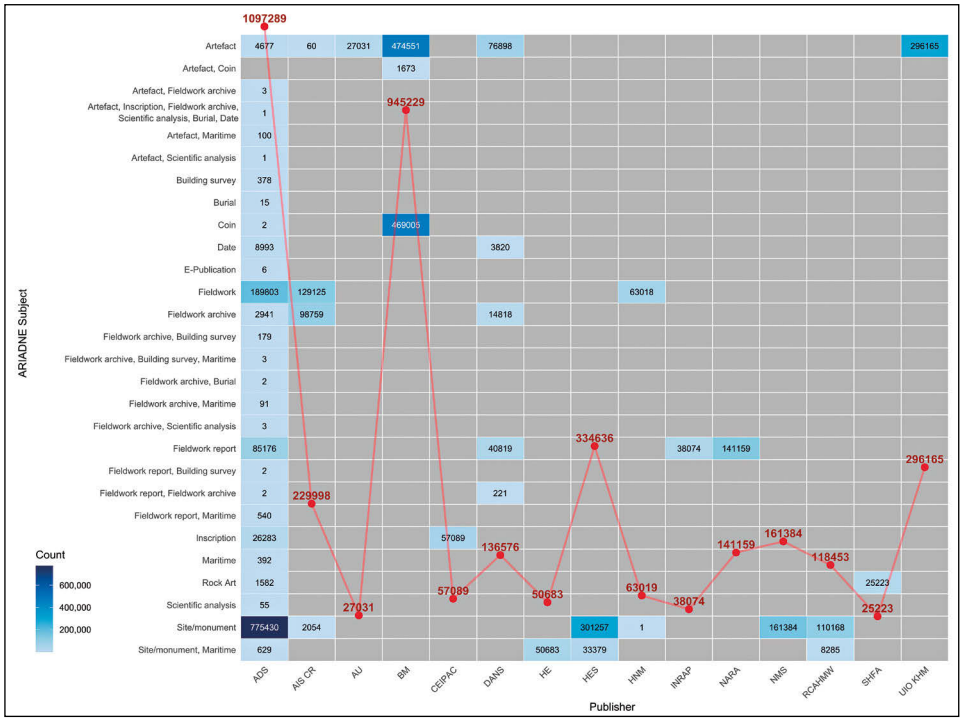

Fig. 2 – Heatmap of the number of ARIADNE subjects per publisher; the red line shows the overall number of resources per publisher.

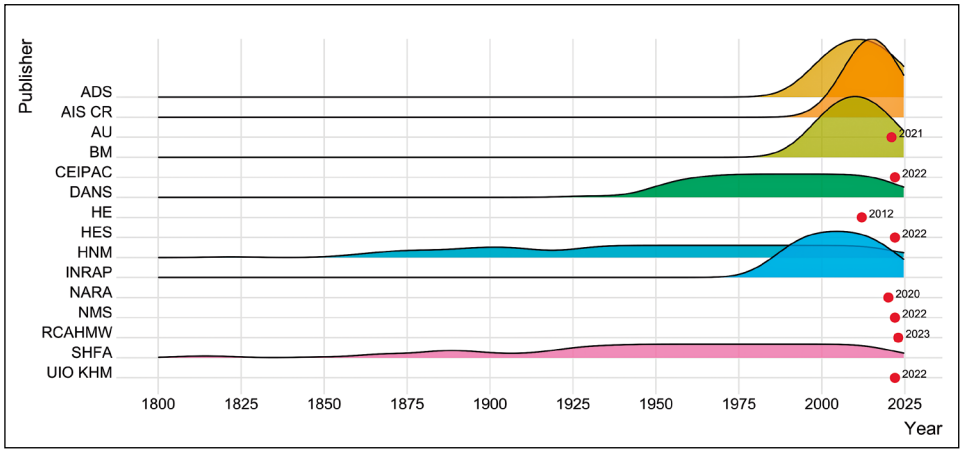

Fig. 3 – Distribution of ARIADNE resources per issue date.

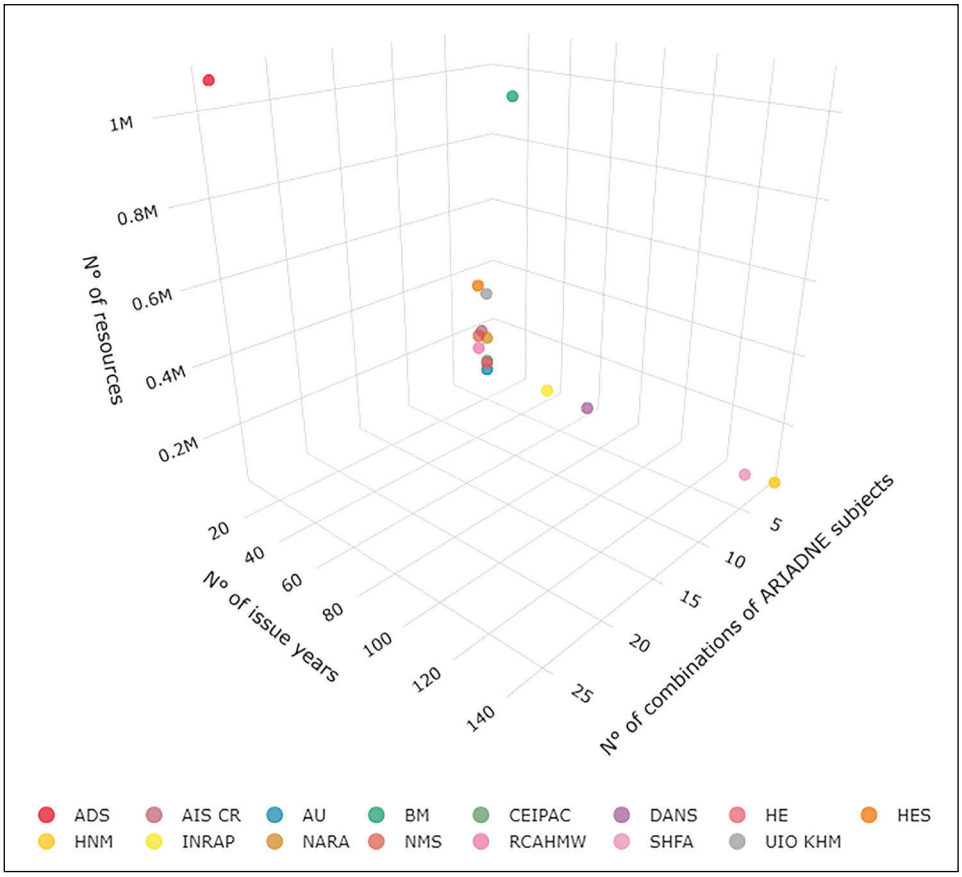

Fig. 4 – 3D scatterplot correlating the number of ARIADNE resources, distinct issue years and combinations of ARIADNE subjects.

A different kind of analysis was carried out on the second dataset (dataset2). Each ARIADNE resource has five main roles for what concerns the curatorship and authorship (see §2.1). These five different roles can all be potentially fulfilled by the same agent; alternatively, each role can be fulfilled by a distinct agent. Among these two ends, fifty other possible combinations can exist: for instance, an institution could fulfill the roles of publisher and creator, while another institution the roles of contributor, owner and responsible; or an institution could fulfill the role of publisher, a second institution the roles of creator and contributor, and a third institution the roles of owner and responsible. The analysis retrieved the agents' names and compared them considering the fulfilled roles and the possible combinations; if a role was

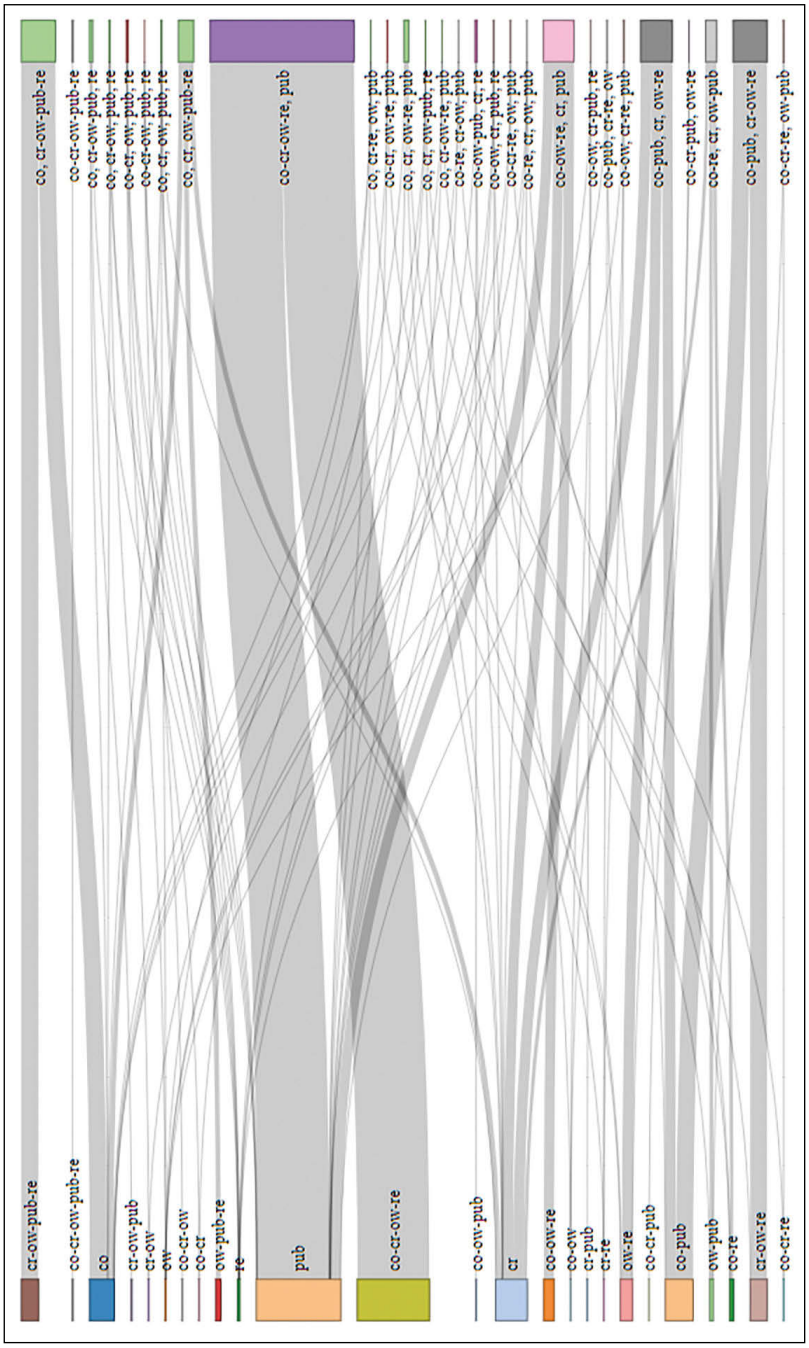

Fig. 5 – Sankey diagram showing the existing combinations of roles of ARIADNE resources. On the left side, the roles or role combinations fulfilled by the same institution; on the right side, roles and combinations as they appear in ARIADNE resources ('co' = contributor; 'cr' = creator; 'ow' = owner; 'pub' = publisher; 're' = responsible;)

covered by more than one agent, their names were alphabetically sorted and combined in a single value, to keep consistency over the entire dataset (Fig. 5).

5. Discussion

The analysis of the resources available on the online ARIADNE Portal highlights some trends. One of the most visible is that most data are provided by a few publishers in each country. Despite the reduced size of the dataset, the results of the application of the HHI suggest that there are usually one or two main institutions that provide most, if not all, of the resources. This implies that the competences or the responsibility to provide digital archaeological data are normally concentrated in one leader institution. This is partially due also to the nature of data: Heritage data, unlike most data from physical and chemical sciences, are normally managed at regional and/or national level, as they are more related to the territories where they are located in (Richards 2023).

The analysis on the issue date of ARIADNE resources brings to light a possible incoherent data entry, that may be due to a couple of reasons. The 'dct:issued' attribute appears to be mandatory in the ACDM and should show the «Date of formal issuance (e.g., publication) of the resource»(Debole *et al.* 2015); the AO-Cat ontology confirms that the 'was_issued' property is mandatory and shows «the date of formal issuance (e.g., publication) of the resource by the publisher» (Felicetti *et al.* 2023, 37). This information appears to have been variously interpreted by the ARIADNE publishers. Those institutions whose resources are assigned a single specific year as the issue date likely interpreted the 'date of formal issuance' as the moment in which resources were provided to ARIADNE. Conversely, those institutions whose resources span over a range of years (such as HNM or SHFA, whose archival resources display issue dates spanning two centuries) likely interpreted it as the initial date of issuance of the resource. The second reason seemingly lies in the structure of data itself. In the AO-Cat Ontology the 'was_issued' property has as Range a xsd:dateTime value; the data present different formats instead, including xsd:dateTime, xsd:datetime (probably a misspelling of the previous format), xsd:date, xsd:string and rdf:langString. The result is a high variability, which is evident even within a single data format: for instance, xsd:dateTime data includes values like '2021-02-26', '28/01/2019', '07 Dec 2011', 'Fri May 15 12:00:00 BST 2020', '2017-11-30T09:14:59.000+01:00', '1992', '1990/92', '1949 Augusti', leading to inconsistency and hindering a possible efficient interoperability of data. Some other minor issues can be also detected, such as wrong dates or a mismatch between the issue date provided by the ARIADNE Portal and the one provided on the publisher's database.

The analysis of the combinations of fulfilled roles of the agents of each resource revealed that the most frequent combination is that of an agent as publisher and a second agent as creator, contributor, owner and responsible[3]. In second position there is the combination of an agent as contributor and a second agent that fills the other roles, while in third position there is the combination of an agent as publisher and contributor and a second agent as creator, owner and responsible. The analysis also reveals a relevant degree of role scattering: for instance, 3,552 resources have a distinct agent for each of the five roles (14th position), while 8,917 resources have an agent as contributor-creator and three other distinct agents for the other three roles (12th position). Nevertheless, most resources appear to be associated with a maximum of three agents: the most frequent combination is that of an agent that creates, describes, owns and holds the scientific responsibility of a resource, while another one makes it publicly available. Although the performed analysis does not reach this level of detail, it is quite likely that the ADS is the publisher of a relevant part of the resources that match the combination of 'publisher, contributor-creator-owner-responsible'.

6. Conclusions

The analyses described in the article target ARIADNE Portal as the major digital infrastructure for archaeological data and are focussed on the administrative metadata to gain a view of the contribution approaches and challenges of archaeological data exchange. Most of its resources have a publisher based in Europe and are contributed by a few countries; in turn, each country usually has one or a few institutions that act as publishers, showing a high concentration of data curation. Most institutions are also focused on specific fields, as highlighted by the analysis of the combinations of the ARIADNE subjects. The majority of the resources were issued over the last forty-fifty years, showing that older archival materials are less frequently available in ARIADNE, although some notable exceptions exist. The assessment of the institutions' policies, their approach to digital data, their use of metadata schemas or semantic resources prior to participating in the ARIADNE initiative, will be the further steps to be addressed, to better frame the European overview of archaeological data management and distribution.

The analysis revealed also that some inconsistencies exist and some misalignments have been detected between data referred to a resource as published in ARIADNE and data referred to the same resource in the provider's portal. However, this work especially focussed on administrative metadata

[3] In this case, as previously stated, the 'agent' could include more than one institution/person as it is considered as a combination of terms.

in the ARIADNE Knowledge Base accessed via SPARQL Endpoint, that are presumably much less involved in the researches carried out by common users on the ARIADNE Portal interface. Despite these inconsistencies, ARIADNE currently allows to retrieve information about more than four million archaeological resources, an unmatched result for what concerns archaeological data.

Riccardo Valente
Istituto di Scienze del Patrimonio Culturale - CNR
riccardo.valente@cnr.it

Acknowledgements

This research was supported by H2IOSC Project - Humanities and cultural Heritage Italian Open Science Cloud funded by the European Union NextGenerationEU – National Recovery and Resilience Plan (NRRP) - Mission 4 'Education and Research' Component 2 'From research to business' Investment 3.1 'Fund for the realization of an integrated system of research and innovation infrastructures' Action 3.1.1 'Creation of new research infrastructures strengthening of existing ones and their networking for Scientific Excellence under Horizon Europe' - Project code IR0000029 - CUP B63C22000730005. Implementing Entity CNR. The author wishes to express gratitude to Erica Scarpa and Irene Rossi for their insightful comments and editorial assistance.

REFERENCES

Aloia N., Debole F., Felicetti A., Galluccio I., Theodoridou M. 2017, *Mapping the ARIADNE catalog data model to CIDOC CRM: Bridging resource discovery and item-level access*, «SCIRES-IT - SCIentific RESearch and Information Technology», 7, 1, 1-8 (https://doi.org/10.2423/i22394303v7n1p1).

Bardi A., Baglioni M., Artini M., Mannocci A., Pavone G. 2024, *The ARIADNEplus Knowledge Base: A Linked Open Data set for archaeological research*, in *SEBD 2024. Symposium on Advanced Database Systems 2024*, Villasimius, CEUR-WS.org (https://ceur-ws.org/Vol-3741/paper16.pdf).

Bardi A., Ottonello E. 2022, *User Guide for the ARIADNEplus Knowledge Base (GraphDB)* (https://data.d4science.org/workspace-explorer-app/workspaceexplorerapp/DownloadServlet?ids=bb9a9337-2090-4248-89c5-810ef3892edf;&viewContent=true&redirectonerror=true).

Debole F., Meghini C., Aloia N. 2015, *Specification of the ARIADNE Catalogue Data Model v. 2.6*. (http://ariadne-support.dcu.gr/files/ACDM_Version_2.6.pdf).

Felicetti A., Meghini C., Richards J., Theodoridou M. 2023, *The AO-Cat Ontology v 1.2*. (https://doi.org/10.5281/zenodo.7818375).

Meghini C., Scopigno R., Richards J., Wright H., Geser G., Cuy S., Fihn J., Fanini B., Hollander H., Niccolucci F., Felicetti A., Ronzino P., Nurra F., Papatheodorou C., Gavrilis D., Theodoridou M., Doerr M., Tudhope D., Binding C., Vlachidis A. 2017, *ARIADNE: A Research infrastructure for archaeology*, «Journal on Computing and Cultural Heritage», 10, 3, 1-27 (https://doi.org/10.1145/3064527).

Niccolucci F. 2015, *Un'infrastruttura di ricerca per l'archeologia: il progetto ARIADNE*, «Archeologia e Calcolatori», Suppl. 7, 41-44 (https://www.archcalc.cnr.it/supplements/articles/808).

Niccolucci F. 2017, *ARIADNE e gli Open Data: come trasformare i dati archeologici da open a "FAIR"*, in M. Serlorenzi, I. Jovine (eds.), *Pensare in rete, pensare la rete per la ricerca, la tutela e la valorizzazione del patrimonio archeologico. Atti del IV Convegno di Studi SITAR (Roma 2015)*, «Archeologia e Calcolatori», Suppl. 9, 141-150 (https://doi.org/10.19282/ACS.9.2017.13).

Richards J.D. 2023, *Joined up thinking: Aggregating archaeological datasets at an international scale*, «Internet Archaeology», 64 (https://doi.org/10.11141/ia.64.3).

Vlachidis A. 2017, *Enabling European archaeological research: The ARIADNE E-Infrastructure*, «Internet Archaeology», 43 (http://doi.org/10.11141/ia.43.11).

ABSTRACT

This article provides a preliminary analysis of the administrative metadata of the ARIADNE resources. ARIADNE, an acronym for Advanced Research Infrastructure for Archaeological Dataset Networking in Europe, is a European archaeological infrastructure that supports research. The presented analyses were performed accessing the ARIADNE Knowledge Base via its SPARQL Endpoint. The distribution of available resources per publishers' country, of archival resources, of combinations of ARIADNE subjects and of combinations of agents (publishers, creators, contributors, owners, scientific responsibles) were estimated and commented, tracing an up-to-date profile of the infrastructure.