https://doi.org/10.26069/greynet-2025-000.763-gn

MOD (Mappa Open Data): A Digital Repository Enhancing Open Access to Archaeological Grey Literature and Data

Lorena Bravi[2], Salvatore Basile[1], Chiara Giovannetti[2], Filippo Sala[1]

[1]University of Pisa, [2]Sapienza University of Rome

Keywords: archaeological research, open data, digital repository, interoperability, DOI

Abstract

MOD (Mappa Open Data) is an open data repository designed to enhance the dissemination, accessibility, and reuse of archaeological grey literature and data. Its core strength lies in the heterogeneity of the data it contains not only textual files, such as excavation reports, but also images, spatial data, archaeometric analyses, 3D models, and other data produced throughout the archaeological research process. Additionally, MOD hosts datasets from the Open Access scientific journal ArcheoLogica Data (ALD), which publishes original archaeological datasets and research.

Every dataset on MOD is associated to a DOI (Digital Object Identifier). The assignment of a DOI to each dataset enhances the searchability and accessibility of its contents, while also allowing grey literature to be cited in other projects.

Hosted within the University of Pisa Digital Library, and developed on a robust Open Source technological framework following FAIR principles, MOD offers structured metadata descriptions adhering to the Dublin Core metadata standard. The metadata schema has now been semantically enriched, improving semantic interoperability, data discoverability and compliance with international standards. The ontology used was provided by the Ariadne Research Infrastructure, a major data aggregator in the field, into which MOD data and grey literature will soon be integrated.

MOD addresses legal and ethical issues of Open Data sharing practices, providing explicit guidelines on intellectual property, authorship attribution, privacy, and usage policies. By adhering to clearly defined open licences (CC BY, CC BY-SA), it significantly enhances scholarly communication by serving as a comprehensive repository for archaeological datasets and grey literature.

https://doi.org/10.26069/greynet-2025-000.763-gn

1. Introduction

Over the past two decades, the archaeological sector has undergone a profound transformation, driven by the progressive digitalisation of research processes and the large-scale adoption of the open data philosophy (Anichini, Gattiglia, & Gualandi, 2021, pp. 9–10). The emergence of practices such as open access, open data, and data sharing has not only generated new and significant opportunities for scientific innovation, but has also raised a number of cultural, technical, and institutional challenges, many of which remain only partially resolved (Anichini & Gattiglia, 2015, pp. 299–302). Within this framework, the so-called *grey literature* in archaeology, namely, the heterogeneous body of technical reports, excavation records, unpublished documentation, and raw datasets not disseminated through conventional publishing channels, plays a strategic role in building genuinely open and shared scientific knowledge (Anichini & Gattiglia, 2012, p. 53). However, the enhancement and management of grey literature raise unavoidable issues concerning accessibility, interoperability, standardisation, and long-term sustainability (Anichini & Gattiglia, 2015, pp. 309–316).

Against this backdrop, the MOD (MAPPA Open Data Archive), conceived and developed by the MAPPA Laboratory (*Metodologie Digitali Applicate all'Archeologia*) of the University of Pisa, and now integrated into the Digital Library of the Department of Civilisations and Forms of Knowledge (CFS), stands as the first Italian repository of archaeological data structured in accordance with open data principles (Anichini, Gattiglia, Gualandi, & Noti, 2013, p. 47). MOD is not only an advanced tool for the preservation and dissemination of archaeological data, but also an emblematic case study of the potential and limitations inherent in the practical implementation of digital infrastructures for the systematisation and management of grey literature.

This paper aims to present and analyse the role, potential, and critical aspects of the MOD portal, with particular reference to the strategies adopted for the management, publication, and semantic enrichment of archaeological grey literature. Furthermore, it will examine issues relating to the system's technical architecture, the organisation of metadata, and prospects for future development.

2. Towards a sustainable digital ecosystem for archaeological data

Archaeology, like many other disciplines in both the humanities and sciences, is undergoing a profound transformation triggered by the so-called "digital revolution" (Anichini & Gattiglia, 2015). The exponential expansion of the volume of available data, coupled with the progressive adoption of digital methodologies, is substantially reshaping research practices, making the

https://doi.org/10.26069/greynet-2025-000.763-gn

collection, management, and analysis of information increasingly streamlined, collaborative, and accessible (Anichini et al., 2013; Anichini et al., 2021). Within this context, online repositories, specialist databases, and collaborative platforms constitute central components, as they enable the interconnection of heterogeneous datasets, both in content and in informational structure, thereby fostering new forms of collaboration and knowledge dissemination.

Despite these advances, the creation of a mature and fully operational open data ecosystem in the cultural heritage sector, and in archaeology in particular, remains complex and, in some case, incomplete (Kansa, 2012; Huggett, 2015). Anichini and Gattiglia (2015) highlight persistent and significant challenges concerning the quality and usability of information: archaeological data are often fragmented and inconsistent, stored in formats that are not machine-readable, lacking complete and structured metadata, and seldom compliant with shared standards. Such shortcomings significantly limit the potential for dataset reuse and integration, thereby hindering the development of robust interpretative models and the adoption of advanced computational approaches.

A further obstacle is posed by the phenomenon known as *data hoarding*, namely the reluctance to share complete and verifiable datasets. The reasons for this are various: fear of losing a competitive advantage or future recognition, lack of time for data cleaning and structuring, and insufficient training on the requirements and potential of open data (Anichini & Gattiglia, 2015; Kansa, 2012, pp. 507–508). Even in the presence of a generally positive attitude towards data openness, as demonstrated by the 2012 MAPPA survey, in which 57% of respondents expressed support and 91% viewed the creation of an open archaeological data archive as a benefit to the community, actual sharing remains limited. This is due, in part, to misconceptions regarding data "ownership", which is frequently attributed to the *Soprintendenze* (Anichini & Gattiglia, 2015, pp. 313–314).

This issue is further compounded by the lack of infrastructures, genuinely oriented towards reuse: repositories organised in non-functional ways, the use of formats such as PDF or Excel that hinder interoperability, insufficient metadata, inconsistent standards, and only partial application of the FAIR principles (Findable, Accessible, Interoperable, Reusable) (Anichini & Gattiglia, 2015; Nicholson et al., 2023).

From an institutional perspective, the adoption of open access practices can help to reduce inequalities between well-resourced institutions and those that are financially more vulnerable,

particularly due to the high Article Processing Charges (APCs) imposed by publishers, which limit publication opportunities for authors and organisations without adequate funding (Borrego, 2023). The difficulties in developing shared vocabularies and platforms, as demonstrated by the experience of the CAMIS project, further illustrate the sector's fragmentation and the absence of a unified strategy (Berzellini, Marinato, & Piragnolo, 2024).

In recent years, however, concrete strategies have emerged to address these challenges: the adoption of open licences such as Creative Commons Attribution 4.0, the dissemination of repositories aligned with the FAIR principles, and the use of ontologies such as CIDOC-CRM have contributed to significant progress. Nicholson (2023) stresses the importance of complementing the FAIR principles with the CARE principles (Collective Benefit, Authority to Control, Responsibility, Ethics), thereby paying due attention to the ethical implications of data management in archaeology. The *Odyssey* project (Sá et al., 2024) represents an example of an interoperable spatial infrastructure, built on a GeoNode foundation and enhanced with machine learning models, capable of supporting the annotation and advanced visualisation of archaeological data. In Italy, Iacopini (2024) describes an initiative for the digitisation of archives and their publication on interoperable open platforms, with metadata structured according to the Dublin Core standard. FAIR-compliant publication via Zenodo has been adopted by the Swedish Heritage Laboratory within the framework of the IPERION HS project, as reported by Larsson, Gelfgren, Kilbrink, Bergström, Lindgren, and Lerm (2025), to ensure sustainable models for managing cultural heritage data. The PERAIA project (Laguna-Palma, 2024), in turn, proposes a web application dedicated to archaeological sites in the Eastern Mediterranean, integrating legacy data and satellite imagery in accordance with the LOUD and FAIR principles. Finally, Kansa and Kansa (2022) promote high-quality practices for data reuse, encouraging the abandonment of a narrow conception of sharing as a mere temporary obligation.

In parallel, the practice of *Open Methods* is becoming increasingly widespread; this entails the publication of scripts, workflows, and complete datasets, often developed in languages such as R or Python, promoting transparency and reproducibility (Marwick et al., 2017). Data papers also play a significant role: these are short articles that clearly and concisely describe a research dataset, illustrate its potential for reuse, and formalise its scholarly recognition, thereby enhancing both the value of primary documentation and the acknowledgement of its authors (Anichini & Gattiglia, 2015, p. 309).

Despite the challenges that remain in terms of accessibility, standardisation, and sustainability, the growing adoption of shared tools, infrastructures, and practices is laying the foundations for

a more effective and enduring open data ecosystem. Such an ecosystem has the potential to make visible data and documents that have so far remained hidden or difficult to access, with particular attention to grey literature, which constitutes a valuable resource even when lacking a uniform structure. Open access to research-generated data enables the review, reorganisation, and reinterpretation of past studies, thereby stimulating significant advances in knowledge production. The aggregation of sources of different natures and formats, including those from other disciplines, allows for a richer and more articulated analysis of research subjects.

Furthermore, the enhancement of grey literature not only makes it possible to reconsider the results of previous research, but also to understand and reuse its methodologies in different contexts and for different purposes. Among the most promising areas is the training of Artificial Intelligence models: the availability of such documentation facilitates the development of machine learning systems capable of automatically acquiring archaeological data and efficiently recognising specific objects and patterns, thus accelerating research timelines and improving overall quality (Anichini & Gattiglia, 2015).

3. MOD and the Open Data revolution in Italian archaeology

The emergence of the MOD (MAPPA Open Data Archive) marks a decisive turning point in the history of Italian archaeological documentation, as the first national repository entirely dedicated to the publication and sharing of archaeological data in open format and in full compliance with the FAIR principles (Anichini, Gattiglia, Gualandi, & Noti, 2013, p. 45). Established in 2011 within the framework of the MAPPA Project (*Metodologie Applicate alla Predittività del Potenziale Archeologico*) at the University of Pisa and modelled on the international best practice of the Archaeology Data Service (ADS) at the University of York, MOD responds to the need for an open, interoperable digital repository capable of hosting heterogeneous datasets originating from a wide range of research activities. This initiative is rooted in the awareness that archaeological practice, by its very nature intrinsically destructive, renders every piece of data collected in the field unique, irreplaceable, and therefore of inestimable value (Anichini, Ciurcina, & Noti, 2013, p. 136; Anichini & Gattiglia, 2015, pp. 298–300).

In this light, MOD's primary mission is the long-term digital preservation, full accessibility, and maximum reuse not only of final interpretative documentation, but also of the wide array of materials falling under the category of so-called *grey literature*, including preliminary reports, excavation diaries, site plans, photographs, and raw datasets. Safeguarding scientific memory

and ensuring the free circulation of information are at the core of a strategy that combines digital sustainability with methodological innovation (Anichini, Gattiglia, & Gualandi, 2021, p. 10).

A survey conducted by the MAPPA Laboratory in 2012 revealed that, despite a general favourability towards the adoption of open data, the Italian archaeological community was deeply divided over the modalities of its implementation. Many archaeologists expressed concerns regarding the protection of intellectual authorship and feared a lack of scholarly recognition resulting from the open publication of raw data (Anichini & Gattiglia, 2012, p. 54; Anichini & Gattiglia, 2015, pp. 312–313). Such caution reflects a broader cultural resistance to the adoption of the open science philosophy within the humanities, often associated with a reluctance to disseminate materials perceived as incomplete or not yet ready for formal publication (Anichini et al., 2021, p. 12).

To address these issues, the MOD repository has adopted a model of virtuous social oversight, designed to foster widespread transparency and the active participation of the scholarly community. The broad availability of data enables scholars and researchers to carry out in-depth verification, update existing datasets, and propose new interpretations, thereby continuously improving the quality of archaeological documentation (Anichini, 2013, p. 131). This approach is supported by concrete tools ensuring traceability and author recognition: the assignment of distinct DOIs for each digital object, the use of Creative Commons licences to guarantee proper attribution, formally structured citations, and versioning systems that record every modification (Anichini & Gattiglia, 2012).

In this way, MOD addresses a dual need: on the one hand, to overcome Italy's structural delay in the publication and sharing of archaeological data; and on the other, to provide an operational platform capable of hosting not only final interpretative results, but also the valuable technical documentation produced in the field, material too often confined to private archives and inaccessible to the scholarly community (Anichini & Gattiglia, 2012, pp. 51–53; Anichini & Gattiglia, 2015, pp. 300–302). Through this integrated approach, MOD transforms the ephemeral and fragile value of archaeological documentation into a shared, safeguarded, and fully accessible asset for the entire academic and professional community.

3.1 MOD: conceptual foundations

The conceptual structure of MOD is based on a hierarchical model organised into three levels: archives, collections, and digital objects. Archives constitute thematic or chronological macro-collections, within which collections bring together materials from specific excavation

campaigns, defined geographical areas, or research projects (Anichini, Ciurcina, & Noti, 2013). At the heart of this system are the digital objects, which may include texts, images, videos, audio files, vector data, and three-dimensional models, resources that ensure not only archiving but also the possibility of interpretation and critical reuse of the documentation (Anichini, Gattiglia, & Gualandi, 2021, pp. 10–11).

MOD recognises archaeological data, the precious remains of an intrinsically destructive process such as excavation, as a common good and an essential infrastructure for any historical-archaeological reconstruction. From this perspective, it seeks to move beyond the traditional proprietary mindset, distinguishing between intellectual property, which should not restrict the circulation of data, and intellectual authorship, which is safeguarded through correct citation practices and the assignment of distinct DOIs not only to scholarly publications but also to interpretative reports and primary datasets (Anichini, Ciurcina, & Noti, 2013, p. 48).

At present, the portal hosts tens of thousands of digital images and hundreds of reports and datasets relating to over 170 archaeological interventions, most of which concern the territory of Pisa and Tuscany, although its scope is increasingly extending to extra-regional contexts (Anichini, Gattiglia, & Gualandi, 2021, p. 13). Operational management is based on a single national archive of archaeological open data, in which each intervention, regardless of its scale or commissioning body, is accompanied by a descriptive page, fully downloadable datasets, grey literature reports, and essential metadata compliant with the ICCU MAG 2.0.1 profile (see § 3.2).

All content is released under Creative Commons licences (CC BY or CC BY-SA), in line with the FAIR principles, and is accompanied by transparent guidelines for the protection of privacy, cultural heritage, and personal data (Anichini & Gattiglia, 2015, pp. 316–318). This approach encourages the multidisciplinary reinterpretation and innovative reuse of materials, allowing them to be given “new and infinite lives”, optimising the use of already-produced resources, and ensuring widespread social oversight of the quality and preservation of the documentary heritage.

3.2 MOD: Technical and Operational Structure

Since its establishment, MOD has been designed as a modular and adaptable technological platform, built on well-established and extensively tested open-source components, particularly the LAMP stack (Linux, Apache, MySQL, PHP), and hosted on the servers of CISIAU (*Centro Interdipartimentale di Servizi Informatici per l’Area Umanistica*) at the University of Pisa (Anichini & Gattiglia, 2015, p. 362; Anichini et al., 2013, p. 46; Anichini & Gattiglia, 2012, p. 53). Each

archaeological intervention uploaded to the platform is described using a consistent set of essential metadata, structured according to the Dublin Core™ Metadata Initiative (DCMI) (About DCMI, n.d.), which includes fields such as title, authors, location, year of production, and DOI (Anichini & Gattiglia, 2015; Anichini, Ciurcina, & Noti, 2013).

Each record is accompanied by a digital archive containing files in MAG format, comprising texts (excavation reports, lists of stratigraphic units, tables, diaries), images (plans, photographs, sections), multimedia materials (audio, video, 3D models), and tabular or geographic data such as databases, shapefiles, CSV, DBF, or XML files (Anichini et al., 2013, pp. 47–49).

Metadata management is one of MOD's key strengths. From its inception, a flexible schema has been adopted and periodically updated to ensure compatibility with national and international standards and to guarantee both technical and semantic interoperability. At present, the portal employs the MAG 2.0.1 profile, developed by the ICCU (*Istituto Centrale per il Catalogo Unico*), in XML format, fully compatible with OAI-PMH (Open Archives Initiative Protocol for Metadata Harvesting). This choice enables integration with national aggregators such as *Internet Culturale* and international platforms such as Europeana (Anichini et al., 2021, p. 12; *Chi siamo*, n.d.; *Home - Internet Culturale*, n.d.).

From a functional perspective, the platform provides two areas of access: an administrative section, reserved for managers and authors, for uploading, managing, and validating content; and a public section, intended for open consultation, viewing, and downloading of materials. Search tools, structured into free search, thematic or chronological browsing, and advanced metadata search, are designed to facilitate access for non-specialist users as well, thus promoting transparency and the dissemination of archaeological knowledge (Anichini et al., 2013, pp. 50–51).

To support scientific quality, MOD adopts a model of virtuous social oversight (§3), encouraging collective verification and the continuous improvement of content. To this end, systems for activity tracking (accounting), open peer review procedures, and participatory validation workflows have been progressively integrated, in line with the FAIR principles, to acknowledge the plurality of actors involved and promote the sustainability of the archive (Anichini et al., 2021, p. 14).

Since 2021, MOD's inclusion in the Digital Library of the Department of Civilisations and Forms of Knowledge has marked a strategic step, situating the archive within a long-term preservation infrastructure based on Fedora Commons, compliant with the OAIS model (Open Archival

https://doi.org/10.26069/greynet-2025-000.763-gn

Information System, ISO 14721:2003), and built on open-source architectures. The platform also makes use of the Mirador viewer, compliant with the IIIF standard, for high-resolution images, and of the OpenCMS content management system, which enables advanced interactions with digital objects.

The internal organisation follows a defined workflow, including preliminary checks, specialised metadata creation, and responsible publication, always accompanied by open licences (CC BY or CC BY-SA) and transparent policies for the protection of privacy and sensitive data. Integration with Fedora and OAI-PMH flows ensures direct connection with national and international data harvesting infrastructures, such as Europeana (*Chi siamo*, n.d.) and OpenAIRE (Administrator, n.d.), thereby guaranteeing maximum exposure and reusability of content. In this context, MOD stands as one of the core digital collections of the Digital Library, contributing to an interdisciplinary ecosystem aligned with the principles of Open Science (Gattiglia, 2021).

3.3. ArcheoLogica Data (ALD)

In 2020, a new and powerful tool was introduced alongside the MOD portal: *ArcheoLogica Data* (ALD), conceived to address a gap identified as early as 2011, namely, the absence in Italy of a platform dedicated to the systematic publication and preservation of archaeological data (Anichini, Gattiglia, & Gualandi, 2021, p. 9). Archaeology is inherently data-driven, generating large volumes of information that are unrepeatable due to the destructive nature of excavation. With the advent of digital technologies, the management of such data, from acquisition to preservation and sharing, has become crucial, enabling the continuous reworking and reinterpretation of collected information.

In 2012, to respond to this need, the MOD portal was launched with the aim of promoting archaeological data as common goods, freely accessible and reusable. However, the transition to truly open publication was hindered by cultural resistance linked to a persistent proprietary perception of data (Anichini, Gattiglia, & Gualandi, 2021, p. 10). To overcome this obstacle and combine dataset dissemination with more traditional publishing formats, essential for academic recognition and career advancement, the *MAPPA Data Books* were introduced in 2015. These volumes offered an interpretative synthesis accompanied by raw data in open format, establishing an initial model for integrating scientific narrative with primary documentation (Anichini, Gattiglia, & Gualandi, 2021, p. 11).

From the evolution of this experience, *ArcheoLogica Data* was launched in 2021 as an open-access journal that inseparably integrates data and interpretation. Each published article

requires the permanent availability of its corresponding dataset in the MOD repository, thereby ensuring scientific verification, direct comparison, and the possibility of generating new interdisciplinary readings (Anichini, Gattiglia, & Gualandi, 2021, p. 12).

ALD also stands out for its promotion of small-scale investigations and emerging research, offering space and visibility to professionals and early-career scholars who are often disadvantaged by the unpublished status of their theses or dissertations. The journal, published annually and bilingually (Italian and English), adopts an inclusive approach to archaeological data, without chronological or thematic constraints, encouraging interdisciplinary reflection and promoting data reuse (Anichini, Gattiglia, & Gualandi, 2021, p. 13).

From its second volume (2022) onwards, the journal has also introduced monographic thematic sections, alongside miscellaneous contributions and theoretical articles specifically oriented towards digital archaeology and archaeological data management (Anichini, Gattiglia, & Gualandi, 2021, p. 14).

4. The role of grey literature in MOD

One of the most distinctive and innovative features of MOD is its systematic focus on the collection, preservation, and enhancement of archaeological grey literature. From its earliest development stages, the portal set as a strategic goal the creation of a digital archive capable of gathering, organising, and making accessible hundreds of excavation reports, preliminary assessments, technical records, diaries, site plans, and specialist databases. These materials are often produced by universities, professionals, and *Soprintendenze*, yet are rarely disseminated through traditional publishing channels (Anichini & Gattiglia, 2012, p. 51; Anichini et al., 2013, pp. 46–48). This approach has made it possible not only to safeguard a fundamental component of Italy's archaeological memory, but also to foster a culture of sharing, transparency, and innovation in digital research.

The added value of grey literature lies not merely in the informational richness of individual documents, but above all in the ability to aggregate, compare, and reinterpret large volumes of data, applying methodologies such as predictive modelling, diachronic analysis, and multidisciplinary approaches. The availability of such a varied and often unique documentary heritage offers opportunities for comparative analysis and scholarly reinterpretation which,

without such an infrastructure, would be irretrievably lost (Anichini et al., 2021, pp. 10–11; Anichini & Gattiglia, 2015, pp. 300–302).

MOD assigns grey literature a formative and fully acknowledged role in scientific production. Each document is subject to classification and metadata creation, assigned a DOI, and distinguished in its constituent parts, such as interpretative report and dataset, in order to ensure formal traceability and proper author attribution (Anichini et al., 2013, pp. 47–49; Anichini & Gattiglia, 2012, p. 55). In this way, these documents are regarded as an integral part of scholarly output, thereby contributing to their recognition as a common good and a reusable resource, fully aligned with the FAIR principles.

This potential can be fully realised only through the adoption of a semantic framework and advanced metadata systems, both central elements in MOD's technical and scientific strategy (Anichini et al., 2013, pp. 50–51). From its inception, the platform has employed dynamic, shared metadata schemas based on Dublin Core (§3.2) and developed in dialogue with the scholarly community, which are continually updated to ensure traceability, reusability, and the proper contextualisation of data and sources (Anichini & Gattiglia, 2015, p.362).

From this perspective, metadata creation is not limited to a superficial or administrative description, but integrates technical, descriptive, and semantic levels, enabling multidimensional searches and analyses of the content (Anichini et al., 2021, p. 12). A particularly innovative feature lies in the fact that grey literature is also subjected to detailed metadata creation, thus acquiring the same degree of traceability, citability, and interoperability as the most structured datasets (Anichini & Gattiglia, 2015, pp. 308–310).

This process makes it possible to overcome the historical invisibility of grey documentation, granting it a new status as a primary scientific source, endowed with its own DOI and accompanied by clear, formally recognised citation policies (Anichini et al., 2013, p. 48).

In a context of progressive openness and alignment with best international practices, MOD is working towards full compliance with FAIR standards and the principles of Linked Open Data (LOD), with the aim of enhancing semantic interoperability. The publication and exposure of metadata through harvesting protocols such as OAI-PMH (§3.2) also ensure full integration into major national and international aggregators, including Europeana, thereby increasing the visibility, findability, and reusability of content (Anichini et al., 2013, pp. 50–51; *Chi siamo*, n.d.).

https://doi.org/10.26069/greynet-2025-000.763-gn

The advanced semantic structure adopted by MOD not only facilitates the consultation and discovery of sources, but also enables interoperable and multidimensional analyses, making full use of the potential of the collected documentary heritage. In this respect, MOD stands at the forefront compared with traditional documentary archives, anticipating the needs of contemporary research and contributing to the creation of a truly open, transparent, and reusable digital ecosystem (Gattiglia, 2021; Anichini & Gattiglia, 2015).

A particularly significant development is the current phase of evolution: MOD is in fact migrating towards an even more advanced metadata system, based on the AO-Cat standard (*ARIADNE Ontology – Catalogue*) (Felicetti, Meghini, Richards, & Theodoridou, 2025). This sub-ontology, developed within the framework of the ARIADNE Research Infrastructure and the European ARIADNEplus (A+) project, formally describes the resources in the ARIADNE catalogue, including datasets, providers, and collection metadata, and was designed to ensure richer and more structured semantic integration (Felicetti, Meghini, Richards, & Theodoridou, 2025).

This transition marks a qualitative leap in the semantic modelling of archaeological resources, enabling integrated, multi-level, and interoperable management of both structured and unstructured data, including all the grey literature hosted on the portal. This is made possible by the adoption of formally defined ontologies linked to the CIDOC Conceptual Reference Model (CRM), the international ISO 21127:2014 standard, which provides a high-level ontology for the semantic integration of cultural heritage data (*Home | CIDOC CRM*, n.d.). CIDOC CRM offers a common language for modelling entities such as objects, people, events, places, and concepts, as well as the relationships between them, thereby facilitating interoperability between heterogeneous sources and ensuring semantic consistency. The archaeological extensions of CIDOC CRM also ensure terminological and structural compatibility between different digital archives, making it possible to connect and reuse data from diverse systems and contexts (*Home | CIDOC CRM*, n.d.).

AO-Cat has been specifically developed to meet the particular needs of archaeological documentation, allowing for the modelling of complex relationships between different entities, such as sites, contexts, artefacts, sources, and analytical methods, and for connecting heterogeneous sources through cross-referencing mechanisms and persistent URIs (Felicetti, Meghini, Richards, & Theodoridou, 2025). This semantic framework enables MOD to fully align with the Linked Open Data (LOD) approach, contributing to the construction of an archaeological web of data based on shared, open, and reusable standards.

Through this integration, MOD is preparing to effectively address the challenges posed by Open Science and the Big Data ecosystem, ensuring traceability, interoperability, and extensive reuse of documentation, with particular attention to grey literature and primary datasets, in line with the most advanced international paradigms (Richards, 2017).

MOD's strategic approach, grounded in data openness and advanced semantic enrichment, generates significant effects in scientific, cultural, and institutional domains. It strengthens research quality, fosters interdisciplinary cooperation, and promotes sustainable and responsible access to the digital archaeological heritage.

From a scientific perspective, the availability of digital archives that include structured grey literature, raw data, and interoperable metadata constitutes a strategic resource for ensuring the verifiability and reproducibility of research, two requirements that are increasingly central to the debate on the quality of scholarly output in archaeology (Anichini et al., 2021; Richards, 2017). Access to interoperable primary sources not only enables the reconsideration and redefinition of previous interpretations but also paves the way for the development of new data-driven lines of inquiry, such as data mining, predictive modelling, diachronic analysis, and geospatial studies (Huggett, 2015).

From a cultural and social standpoint, MOD's adoption of an open access and interoperable model encourages active public engagement, enhances transparency in the management of cultural heritage, and stimulates collaborative innovation processes involving institutions, heritage protection bodies, professionals, and local communities (Anichini & Gattiglia, 2015).

In terms of sustainability and scientific memory, MOD plays a crucial role in the preservation, visibility, and valorisation of archaeological documentation. In particular, it breathes new scientific life into grey literature, often destined to remain invisible, by ensuring its traceability, citability, and potential for future reuse, in full alignment with the FAIR principles and European Open Science strategies (European Commission, 2018; Anichini & Gattiglia, 2015).

Furthermore, MOD does not merely safeguard archaeological evidence; it actively facilitates its use by researchers, professionals, small and medium-sized cultural institutions, and enthusiasts. Through an open and modular archive, it supports specialist studies, conservation initiatives, and educational projects, thereby laying the foundation for the sustainable, long-term management of documentary heritage.

https://doi.org/10.26069/greynet-2025-000.763-gn

5. Conclusion

The MOD experience demonstrates how the full-scale opening of archaeological documentation, with systematic attention to grey literature, can mark an epistemological turning point in the discipline. The availability and traceability of materials unpublished in traditional editorial channels represent not only an antidote to the loss of knowledge but also living resources for contemporary research (Anichini et al., 2013, pp. 46–48; Anichini & Gattiglia, 2012, pp. 51–53).

When grey literature is properly documented, metadata-enriched, and made accessible, it becomes a catalyst for new discoveries: it enables the validation and updating of established interpretations, the reopening of seemingly closed cases, the addition of new interpretative layers, and, in some instances, a fresh start on firmer foundations (Anichini, Gattiglia, & Gualandi, 2021, pp. 10–11). From this perspective, open data in archaeology is not merely a technical issue, but a sphere of cultural and social experimentation, in which the participation and accountability of the scholarly community are essential to ensuring quality, equity, and innovation (Anichini & Gattiglia, 2015, p. 315).

Thanks to its advanced architecture and participatory governance, MOD emerges not only as a technological infrastructure but also as a true laboratory for sustainability, transparency, and collective knowledge production. The opening of data and the valorisation of grey literature are transforming the ways in which archaeology is conducted, communicated, and interpreted in the 21st century, paving the way for a data-driven, inclusive, and rigorous ecosystem.

The MOD experience shows that it is now possible to counteract the fragmentation, inaccessibility, and dispersion of Italian archaeological documentation through a digital infrastructure built on systemic vision, advanced technical architecture, and participatory governance (Anichini et al., 2013, pp. 45–46; Anichini & Gattiglia, 2015, pp. 300–302). The challenge of open data in archaeology transcends the technological dimension, directly involving the culture and politics of the discipline and urging the scholarly community to reconsider the paradigms of knowledge production, sharing, and validation (Anichini, Gattiglia, & Gualandi, 2021, p. 10).

Only through the systematic opening, transparency, and sharing of data, particularly by enhancing and enriching the metadata of grey literature, can truly collaborative, equitable, and innovative research be achieved. From this perspective, grey literature ceases to be a marginal repository and becomes a living archive and a laboratory of knowledge, capable of validating,

reinterpreting, and continually expanding the archaeological information heritage, paving the way for new discoveries and unexplored research avenues (Anichini & Gattiglia, 2015, p. 315).

The model tested by MOD suggests that the digital archaeology of the future must be based on integrated sustainability, not only technical but also social, systematically involving the academic community, institutions, and civil society in the collective production of knowledge (Anichini & Gattiglia, 2015, p. 315). In this context, the enhancement of grey literature, made accessible through rigorous metadata and interoperability practices, becomes the essential prerequisite for a more transparent, open, and continuously renewable discipline, reaffirming its role in contemporary society and in safeguarding collective memory (Anichini et al., 2021, p. 11).

Bibliography

About DCMI. (n.d.). DCMI. Retrieved August 10, 2025, from https://www.dublincore.org/about/

Administrator. (n.d.). *About*. OpenAIRE. Retrieved August 10, 2025, from https://www.openaire.eu/about

Anichini, F. (2013). *MAPPA survey: Gli Open Data nell'archeologia italiana*. In F. Anichini, N. Dubbini, F. Fabiani, G. Gattiglia, & M. L. Gualandi (Eds.), *MAPPA. Metodologie applicate alla predittività del potenziale archeologico* (Vol. 2, pp. 121–132). Edizioni Nuova Cultura. https://doi.org/10.4458/0917-08

Anichini, F., & Gattiglia, G. (2012). *Il progetto MAPPA: verso un archivio open data di archeologia italiana*. In F. Anichini, N. Dubbini, F. Fabiani, G. Gattiglia, & M. L. Gualandi (Eds.), *MAPPA. Metodologie applicate alla predittività del potenziale archeologico* (Vol. 1, pp. 51–56). Pisa University Press. https://doi.org/10.4458/0917-07

Anichini, F., & Gattiglia, G. (2015). MAPPA open data metadata: The importance of archaeological background. In F. Giligny, L. Costa, & F. Djindjian (Eds.), *CAA2014: 21st century archaeology: Concepts, methods and tools. Proceedings of the 42nd Annual Conference on Computer Applications and Quantitative Methods in Archaeology* (pp. [inserire intervallo di pagine]). Oxford: Archaeopress. https://digital.casalini.it/9781784911010

Anichini, F., Ciurcina, M., & Noti, V. (2013). *Il MOD: l'archivio Open Data dell'archeologia italiana*. In F. Anichini, N. Dubbini, F. Fabiani, G. Gattiglia, & M. L. Gualandi (Eds.), *MAPPA. Metodologie applicate alla predittività del potenziale archeologico* (Vol. 2, pp. 133–160). Edizioni Nuova Cultura. https://doi.org/10.4458/0917-09

Anichini, F., Gattiglia, G., & Gualandi, M. L. (2021). *Idee, dati, condivisione. Gli Open Data in archeologia non sono mai abbastanza*. *ArcheoLogica Data, 1*, 9–15. https://doi.org/10.13131/ARCHEOLOGICADATA-KP5G-0J59

Anichini, F., & Gattiglia, G. (2015). *Verso la rivoluzione. Dall'Open Access all'Open Data: la pubblicazione aperta in archeologia*. *Post-Classical Archaeologies, 5*, 298–326.

Berzellini, B., Marinato, M., & Piragnolo, M. (2024). Un database condiviso ad accesso libero per lo studio archeologico dei cimiteri medievali in Italia settentrionale. *Archeologia e Calcolatori, 35*(1), 233–250. https://doi.org/10.19282/ac.35.1.2024.16

Chi siamo. (n.d.). Retrieved August 10, 2025, from https://www.europeana.eu/en/about-us

European Commission, Directorate-General for Research and Innovation. (2018). *Turning FAIR into reality: Final report and action plan from the European Commission expert group on FAIR data*. Publications Office. https://doi.org/10.2777/1524

Felicetti, A., Meghini, C., Richards, J., & Theodoridou, M. (2025). The AO-Cat Ontology. Zenodo. https://doi.org/10.5281/zenodo.15348438

Huggett, J. (2015). A Manifesto for an Introspective Digital Archaeology. *Open Archaeology*, *1*(1). https://doi.org/10.1515/opar-2015-0002

Home | CIDOC CRM (n.d.). Retrieved August 10, 2025, from https://cidoc-crm.org/

Home—Internet Culturale. (n.d.). Retrieved 10 August 2025, from https://www.internetculturale.it/

Iacopini, E. (2024). Digitization of archival data and metadata in archaeology: the case of Ancona. *Proceedings, 96*(1), 9. https://doi.org/10.3390/proceedings2024096009

Kansa, E. (2012). Openness and archaeology's information ecosystem. *World Archaeology*, *44*(4), 498–520. http://www.jstor.org/stable/42003546

Kansa, E. C., & Kansa, S. W. (2022). Promoting data quality and reuse in archaeology through collaborative identifier practices. *Proceedings of the National Academy of Sciences, 119*(43), e2109313118. https://doi.org/10.1073/pnas.2109313118

Laguna-Palma, D. (2024). Enhancing archaeological heritage management: The PERAIA Project's open data database and web application. *Journal of Open Humanities Data, 10*(1), 2. https://doi.org/10.5334/johd.158

Larsson, Å. M., Gelfgren, S., Kilbrink, L., Bergström, K., Lindgren, S., & Lerm, J. (2025). Developing practices for FAIR and linked data in heritage science. *npj Heritage Science, 13*, 53. https://doi.org/10.1038/s40494-025-01598-x

Marwick, B., Barton, C. M., Bates, L. A., Baxter, M., Bevan, A., Bollwerk, E. A., ... & Wren, C. D. (2017). Open science in archaeology. *SAA Archaeological Record*, *17*(4), 8-14. https://doi.org/10.17605/OSF.IO/3D6XX

Nicholson, C., Kansa, S. W., Gupta, N., & Fernandez, R. (2023). Will it ever be FAIR? Making archaeological data findable, accessible, interoperable, and reusable. *Advances in Archaeological Practice, 11*(1), 63–75. https://doi.org/10.1017/aap.2022.40

Sá, R., Gonçalves, L. J., Medina, J., Neves, A., Marsh, F., Al-Rawi, M., ... Moreira, J. (2024). Odyssey: A Spatial Data Infrastructure for Archaeology. *Journal of Computer Applications in Archaeology*, *7*(1), 225–236. https://doi.org/10.5334/jcaa.115

Richards, J. D. (2017). Twenty Years Preserving Data: A View from the United Kingdom. *Advances in Archaeological Practice*, *5*(3), 227–237. https://doi.org/10.1017/aap.2017.11