

DISSEMINATION AND COMMUNITY ENGAGEMENT ACROSS H2IOSC PROJECT ACTIVITIES. INTRODUCTION TO THE SPECIAL SECTION

1. INTRODUCTION

The H2IOSC Project (Humanities and Cultural Heritage Italian Open Science Cloud), funded by the European Union through the Next-GenerationEU programme and by the Italian Ministry of University and Research within the framework of the National Recovery and Resilience Plan (NRRP), began in November 2022 and concluded in April 2026. Its primary objective was to establish a federated cluster integrating the Italian nodes of the CLARIN, DARIAH, E-RIHS, and OPERAS research infrastructures, all included in the ESFRI (European Strategy Forum on Research Infrastructures) roadmap and dedicated to fostering social and cultural innovation. The project (<https://www.h2iosc.cnr.it/>) was organised into eight Work Packages, with archaeologists and epigraphists playing a key role within the sub-packages related to the E-RIHS infrastructure. Beyond its official completion, the momentum generated by H2IOSC has ensured the continuity of the activities initiated during the project. This continuity enabled the further development of research topics that had been started within the project, allowing for a deeper exploration of their scientific potential. At the same time, it provided an opportunity to reflect on the future directions of the digital resources and services developed within H2IOSC, assessing their long-term sustainability and their possible integration into broader national and international research infrastructures.

This section brings together several articles presenting the latest outcomes of these ongoing activities across the Rome and Milan units of the H2IOSC project, which operate within the frame of the CNR-ISPC Open Science Group. The contributions illustrate recent methodological advances, new case studies, and innovative digital approaches and they demonstrate how the project's legacy extends beyond its formal conclusion, continuing to promote open science practices and interdisciplinary research in the fields of archaeology, epigraphy, and cultural heritage. These results stem from the continuous dissemination and engagement activities that the two units have carried out throughout the project, of which we wish to provide a brief overview below, demonstrating the interdependence between dialogue with the reference communities and the progress of research.

2. PUBLIC PRESENTATIONS OF DIGITAL PLATFORMS FOR ARCHAEOLOGICAL RESEARCH

The digital platforms developed by the Rome ISPC Research Unit, coordinated by A. Caravale and P. Moscati – DHeLO (Digital Heritage Landscaping Platform; <https://dhelo.cnr.it>), BiDiAr (Bibliography of Digital Archaeology; <https://bidiar.cnr.it>), and the Open Digital Archaeology Hub (ArchaeoHub; <https://open-archaeohub.cnr.it>; see also MANCUSO, PARACIANI, D'EREDITÀ in this section) – previously described in this Journal (CARAVALE, MOSCATI, ROSSI 2024a, 2024b, 2025a, 2026), have been the focus of a series of public presentations held between 2024 and 2026. These presentations took place in a variety of scientific contexts, highlighting complementary features and potential applications of the platforms. The events included both international conferences and institutional initiatives organised by the CNR, some of which were specifically devoted to disseminating the results of the H2IOSC project¹. These presentations, most of which are currently unpublished, are outlined below as they provide an overview of the most significant subsequent developments of the platforms and highlight their principal and distinct evolutionary trajectories (Fig. 1).

The first one took place during the *18th ArcheoFOSS Conference* (Chieti, December 2024), where, consistently with the mission of the initiative, particular emphasis was placed on the free and open-source software (FOSS) tools employed in the development of DHeLO and BiDiAr, the first databases on which the Rome team began working within the project framework (G. MANCUSO, A. D'EREDITÀ, *Two FOSS resources for Cultural Heritage Digital Archaeology in the H2IOSC project*). The presentation also provided an overview of the FOSS resources catalogued in DHeLO, thereby contributing to the dissemination of knowledge on open and collaborative practices in archaeological research: the adoption of such tools appears to reflect a growing awareness of the importance of sustainable research workflows and the reproducibility of data production processes, both of which are essential for ensuring transparency, interoperability, and the reusability of digital research outputs.

At *Digital Heritage 2025* (Siena, September 2025), the work was presented in a more mature form, with particular attention devoted to the infrastructural dimension of the project. The three platforms developed by the Rome Research Unit were illustrated as components of a coherent and interoperable ecosystem, with emphasis on the methodological choices that guided their evolution – from DHeLO's original relational model towards

¹ Among the latter were the H2IOSC General Meeting (Rome, February 2024); *H2IOSC incontra il settore pubblico e privato: presentazione di servizi e strumenti digitali avanzati per il settore culturale* (Naples, July 2025); and *H2IOSC: Strengthening Research Infrastructures, Consolidating the Federation* (Rome, April 2026).



Fig. 1 – Some meetings presenting the digital platforms developed by the Rome Research Unit.

an architecture based on the principles of Linked Open Data, the adoption of Omeka S as a shared environment for both DHeLO and BiDiAr, and the aggregative role of ArchaeoHub as a unified access point to heterogeneous resources (textual, bibliographic, visual, datasets, projects, etc.) (MANCUSO *et al.* 2025). Particular emphasis was placed on the grounding of these design choices in the needs identified during the community landscaping phase, carried out through questionnaires and interviews, as well as on the overall coherence of the system with the FAIR principles and with the broader objectives of the H2IOSC project, and in particular of WP2, in which the Rome Research Unit was involved, within the framework of European research infrastructures (see also MANCUSO *et al.* in this section).

The Conference *Linking Pasts and Sharing Knowledge. Digital Archaeology, Data Integration and Semantic Technologies for 3D Modelling Historical Landscapes*, organised and hosted by the Department of Humanities of the University of Naples Federico II (November 2025), brought together international scholars to discuss methods, tools, and perspectives in digital archaeological research, with particular attention to data integration, 3D modelling, and semantic technologies for the study of historical landscapes. On this occasion, DHeLO, BiDiAr, and ArchaeoHub were presented from a more analytical perspective, shifting the focus from their descriptive features to their potential as instruments for observing and interpreting digital archaeological research.

Specifically, two concrete examples of data-driven analysis were illustrated. The first concerned the datasets catalogued in DHeLO, examined in terms of their typologies, access modalities (with an almost equal distribution between open and closed resources), significantly influenced by European funding policies, and the software tools used for their production, revealing a progressive consolidation of shared standards across proprietary and open-source environments. The second example concerned the citation networks within BiDiAr, analysed through bibliographic coupling and co-citation analysis applied to the corpus of the last four years of «Archeologia e Calcolatori». The analysis made it possible to identify recognisable thematic clusters, such as those related to Virtual Archaeology, BIM, and open-source GIS, and opened a reflection on the relationship between open access availability and centrality within the citation network.

The ISPC Open Science Group promotes the principles of Open Science and Open Access in scientific publications, disseminating their fundamental features and the best practices to be adopted in the production, management, sharing, and enhancement of research outputs. Within this framework, a seminar organised in Naples at the University Federico II and dedicated to *Open Science and Archaeology* (May 2026) provided an opportunity for P. Moscati, A. Caravale, and G. Mancuso to retrace the theme of Open Science in its historical development and its most recent evolutions across different domains related to classical studies and archaeological publishing, with particular attention to the ArchaeoHub platform as an innovative tool for knowledge aggregation and integration.

In this context, the development of ArchaeoHub (G. MANCUSO, *La piattaforma ArchaeoHub nell'ambito del progetto PNRR H2IOSC*) was presented, with particular emphasis on the methodological choices behind its conceptual and functional framework. In particular, the dual nature of the demand for open data was explored, emerging from the analysis of the scientific community's needs and constituting the starting point of the project, articulated between a bibliocentric and a datacentric perspective. The platform's navigation model was also illustrated, based on free-text search, thematic collections, and geographical access through a gazetteer. A key feature of ArchaeoHub is its visual apparatus. In this respect, a significant contribution derived from the collaboration with the activities of Spoke 8 of the PNRR CHANGES project, which enabled the metadata enrichment of the images hosted on the platform. These images, partly drawn from the IADI visual atlas (Interactive Atlas of Digital Images; <https://iadi.archcalc.cnr.it/>), were predominantly described using controlled vocabularies from the Getty Art & Architecture Thesaurus (AAT) (see BUSCEMI, FAZIO 2024; PARACIANI 2026) and are now also accessible through the European digital library Europeana.

The activities of the ISPC Open Science Group have been further explored and disseminated through two works currently in press (MOSCATI 2026, in press), which examine the process that led to the gradual establishment of open science principles and open data practices within archaeological research. In both works, the results achieved within the framework of the PNRR, and, in particular, the data interoperability that characterises the ArchaeoHub platform, are presented as a key milestone in a methodological and scientific approach oriented toward openness, sharing, and the enhancement of the discipline. ArchaeoHub is thus interpreted not merely as a technological platform, but as the tangible expression of a research approach centred on collaboration, digital integration, and the open sharing of knowledge.

The most recent presentation is linked to the ISPC conference *Voci dalla rete* (Naples, May 2026). Here, the platforms were framed within the broader activities of the Open Science Group, alongside the Open Digital Epigraphy Hub (EpiHub; <https://open-epihub.cnr.it/>; ROSSI, SALVADOR 2024), developed in parallel within the same project by the Milan Research Unit coordinated by I. Rossi. Within this context, particular emphasis was placed on the systemic dimension of the work: ArchaeoHub was presented not as an isolated product, but as a node within an expanding information ecosystem that aggregates and enhances the data from DHeLO and BiDiAr, and is designed to progressively integrate with new journals, archives, and external platforms. Special attention was devoted to the aggregation model without data duplication, based on harvesting and API-based querying, as well as to its scalability as a solution for a still highly fragmented landscape of digital resources. Overall, the presentation outlined a coherent image of an evolving project, further shaped by continuous engagement with the scholarly community. DHeLO, BiDiAr, and ArchaeoHub thus emerge not as closed tools, but as components of a digital infrastructure intended to grow and fully enhance its potential.

A.C.

3. ENGAGING COMMUNITIES AROUND INTEROPERABILITY IN THE HERITAGE AND EPIGRAPHY DOMAINS

Within the H2IOSC project, the Milan operating unit of the CNR-ISPC, coordinated by I. Rossi, was entrusted with two tasks for the E-RIHS infrastructure: building the foundations for semantic interoperability in Heritage Science (HS), aimed at the development of a domain ontology, and creating a pilot hub for the discovery and sharing of digital epigraphy resources. Over the course of the project, two convictions matured within the team: that both activities should advance in constant dialogue and that the boundaries of a research domain – whether ‘epigraphy’ or the broader ‘Heritage Science’ – are an object of continuous negotiation. Defining what counts as an inscription,

or what Heritage Science covers, proved to be neither a preliminary formality nor an abstract exercise: it was the substance of the modelling work, with direct and lasting consequences for how data could be described, integrated, and reused. Dissemination and community engagement were thus conceived not as activities downstream of the research, but as the method through which its results were produced.

The first thread of actions in this sense concerned the development of the Heritage Science ontology, which we decided to build around a shared pivot across the disciplines and infrastructures participating in the H2IOSC project, namely the ‘inscribed object’. Within the Work Package dedicated to interoperability, the ‘inscribed object’ was deliberately chosen as the case study around which the federation pilot of the four Italian nodes of the European research infrastructures participating in H2IOSC was built, as CLARIN, DARIAH, OPERAS and E-RIHS gravitate towards the linguistic, textual, editorial, and physical dimensions of Cultural Heritage assets respectively. The inscribed object thus became a productive site of friction, exposing the divergent descriptive logics that different traditions bring to what is, ostensibly, the same object of study.

It was precisely to surface and govern these divergences that the *Heritage Science Ontology RoundTable* (H/SORT; <https://doi.org/10.5281/zenodo.15166624>; Fig. 2) was established, conceived as a permanent interdisciplinary working group bringing subject-matter experts into face-to-face debate with knowledge engineers. The first roundtable of the cycle *Heritage Science and Semantic Web*, which took place at Ca’ Foscari University of Venice (January 2025), gathered experts from archaeology, history, and epigraphy around the application of Semantic Web technologies to large datasets; epigraphy, and the deceptively simple question of what an inscription is (SCARPA, VALENTE 2025), stood at its centre. Two further roundtables in Milan widened the circle to specialists from the applied sciences for Heritage – chemistry, natural sciences, geology (March 2025; SANSONETTI *et al.* 2025) – and to archaeologists and experts in the sustainable use of cultural heritage (April 2025).

The complementary, and at times conflicting, perspectives that emerged were treated not as obstacles but as data to be preserved: the H/SORT hub (<https://hsort-ec1af2.gitlab.io/>) became a living space where these discussions, recast as concise ‘topics’ and ‘case studies’, remain openly accessible as an invitation to contribution. The negotiation of seemingly elementary notions – provenance *versus* provenience (SCARPA 2025), or whether a carved rock surface is natural, cultural, or both (VALENTE 2025) – showed that even foundational concepts are products of community agreement rather than self-evident facts. This approach informed the methodology and early modelling choices of the HS ontology – presented at the XIV AIUCD conference (Verona,



Fig. 2 – Flyer of the *Esperienze di epigrafia digitale a confronto* workshop (19, 26-28 May 2025).

June 2025; SCARPA, VALENTE, ROSSI 2025a) – most notably the decision to characterise Heritage through the tangible/intangible pair rather than the traditional nature/culture opposition.

As regards the current landscape of semantic artefacts for Heritage, the systematic survey conducted through the Heritage – Semantic Tools and Interoperability Survey (H-SeTIS; <https://h-setis.cnr.it/>; SCARPA, VALENTE 2024) showed that many of them fail to comply with basic FAIR and Linked Data tenets, and that a substantial share remains effectively orphaned (see SCARPA in this section). This evidence motivated the H/RADIOsa initiative (Heritage: Recommended Approaches for the Development of Interoperable Open Semantic Artefacts; <https://h-radiosa-c71003.gitlab.io/>), presented and published at *Digital Heritage 2025* (Siena, September 2025; SCARPA, VALENTE, ROSSI 2025b), which offers practitioners a toolkit of recommended practices (semantic versioning, changelogs, standardised metadata, clear licensing).

Siena's Conference also seeded a collaboration with the French node of E-RIHS that connected H-SeTIS to the Heritage Science Portal (<https://hsportal.espadon.net/>), which was presented at the panel dedicated to the Italian node of E-RIHS closing the CNR-ISPC conference *Voci dalla rete* (Naples, May 2026; E. SCARPA, R. VALENTE, I. ROSSI, *Heritage – Semantic Tools and Interoperability Survey*). Sustainability for these initiatives beyond the project was pursued in parallel through the ongoing integration of H-SeTIS into the E-RIHS Catalogue of Services at the European level.

Finally, in order to engage the community of Heritage professionals and digital practitioners and help them explore foundational and emerging topics and initiatives in the field, the *abzû* observatory of digital Heritage tools and practices was set up (<https://abzu-36ba19.gitlab.io/>). Curated by E. SCARPA, it offers articles and news that highlight emerging trends, projects, and best practices across the research and Heritage landscape; a glossary of key concepts, technologies, and methodologies that underpin our current digital ecosystems; and curated collections offering a guided entry point into key areas such as web development, semantic technologies, digital stewardship, or legal and ethical approaches to Heritage.

The second thread of actions carried the approaches described above into the practice of digital epigraphy. The workshop *Esperienze di epigrafia digitale a confronto* (19 May 2025, online; 26-28 May 2025, Brienza), co-organised by CNR-ISPC and CNR-ISMed in the persons of I. ROSSI and F. DI FILIPPO, under the H2IOSC and PRIN2022 UrbE projects (Fig. 3), was conceived as a meeting-ground for epigraphic traditions that rarely speak to one another, from third-millennium BCE cuneiform to Aegean and Cypriot scripts, N-W Semitic and pre-Islamic Arabian alphabetic *corpora*, and the epigraphy of ancient Italy. The encounter made vivid a problem that classical epigraphy – where most digital tools were born – tends to obscure because of the homogeneity of its cultural background: the epigraphic concepts on which databases are modelled differ considerably from tradition to tradition, hindering interoperability. The very concept of 'epigraph' proves to be problematic, or – at least – fluid (see VALENTE in this section).

To offer a tool for navigating the complexity of epigraphic concepts, the initiative REMEDIA – Registry for Epigraphic Metadata: Encoding, Description, Integration and Annotation (<https://remedia-36e4c4.gitlab.io/>; doi.org/10.5281/zenodo.15525097), was set up and was first publicly presented at the Brienza workshop (E. SCARPA, R. VALENTE, *Epigrafia digitale e principi FAIR*). The dialogue with colleagues developing other databases at a European level confirmed that the registry answered shared community needs. The need for a cross-disciplinary approach to interoperability in digital epigraphy had already been brought to the round table *Ancient Texts and New Technologies* organised in the frame of the bicentennial symposium of the Museo



Fig. 3 – Landing page of the H/SORT – *Heritage Science Ontology RoundTable* hub.

Egizio (Turin, December 2024), where the paper *Different epigraphies, same challenges: Crossing borders of disciplinary domains in digital epigraphy* (I. ROSSI, C.E. SALVADOR) jointly discussed two reference archives for the Ancient Near Eastern and Egyptological epigraphic communities – i.e. DASI - The Digital Archive for the Study of pre-Islamic Arabian Inscriptions (<https://dasi.cnr.it>), and SITH - Système d'Indexation des Textes Hiéroglyphiques of the Karnak project (<http://sith.huma-num.fr>). At the same venue, we announced the development of the Open Digital Epigraphy Hub (EpiHub; <https://open-epihub.cnr.it/>; ROSSI, SALVADOR 2024), intended to support the community in navigating the vast array of digital resources available for epigraphy. The contents of EpiHub have since multiplied, as shown in the presentation by R. VALENTE at the CNR-ISPC conference *Voci dalla rete* (Naples, May 2026). At the same congress in Naples, the paper *Epigrafia digitale e modellazione semantica: verso una descrizione dinamica degli oggetti iscritti* (E. SCARPA) allowed to share the team's experience with the community of researchers working on written objects, in the frame of the roundtable *Supporti scrittori in trasformazione: diagnostica, scrittura, miniature e tecnologie digitali*.

The need to not only compare datasets of different epigraphic traditions, but also widen the scope of reference epigraphic databases to written sources previously excluded or only sporadically recorded within each tradition, because regarded by major epigraphic traditions as ‘utilitarian’ or ‘manuscript’, led us to assess the data model enrichment needed for the DASI archive, taken as a case-study, to accommodate sources like written wooden sticks, ceramics and coins, and graffiti – a feasibility study that was presented at the *Egitto e Vicino Oriente antichi* Conference (Turin, June 2024), and was later implemented (ROSSI 2025). This effort made it possible to include around 400 Ancient South Arabian graffiti in DASI, which were digitised in the frame of training activities for the next generation of digital epigraphers. These involved both high-school students (Liceo ‘E. Fermi’ in Desio) within the Italian Ministry of Education’s PCTO programme (‘Percorsi per le competenze trasversali e per l’orientamento’) for the years 2024 and 2025, and university students of Ca’ Foscari, Venice, through the Laboratory of Epigraphy of the Ancient Near East and the Semitic Philology course of the year 2024 (ref. Prof. P. Corò).

Together, the dissemination and engagement efforts described above make explicit the wager of this section: that interoperability in Heritage research is, before anything else, a matter of communities learning to describe their objects in ways that others can recognise, reuse, and sustain.

I.R.

4. AI-DRIVEN TERMINOLOGICAL ANALYSIS OF HERITAGE SCIENCE AND DIGITAL ARCHAEOLOGY LITERATURE IN THE H2IOSC CONTEXT

To complete the overview of the activities conducted by the Rome operating unit of the CNR-ISPC Open Science Group within the H2IOSC project, it is important to consider another research trajectory that concerns the application of Artificial Intelligence (AI) techniques to the analysis of scientific literature in the fields of Heritage Science and Digital Archaeology. Although this work began before the launch of H2IOSC, it has informed several methodological choices and has provided the intellectual background for the terminological analysis discussed below.

The first step in this direction was undertaken in 2019, shortly after the establishment of CNR-ISPC. In collaboration with SIRIS Academic, NLP and Machine Learning techniques were applied to build a Heritage Science taxonomy (CARAVALE *et al.* 2022). The project drew on varied sources, including Scopus, CORDIS, Creative Europe, the CNR institutional repository, and the open-access repository of the journal «Archeologia e Calcolatori» (A&C). Text mining and expert validation identified key research areas, showing AI’s value in mapping the intellectual landscape of Heritage Science and guiding future research directions.

Building on this work, a second study examined the thematic structure of Digital Archaeology (CARVALE *et al.* 2023). The analysis considered publications from A&C and the CAA conference and journal series for the period 2011-2020. Three AI approaches were applied to analyse titles and abstracts: supervised classification using the ACM taxonomy, topic modelling based on BERT language models, and Named Entity Recognition. Together, these methods produced a data-driven overview of the technologies and archaeological domains represented in the corpus (MOSCATI 2026).

The results highlighted the growing importance of spatial analysis, remote sensing, photogrammetry, 3D modelling, and open data practices, alongside significant methodological shifts within the discipline. The study has since been extended to the full textual corpus of A&C, with the aim of tracing the long-term evolution of themes, methods, and technologies. Initial results will be presented by Paola Moscati and Giacomo Mancuso (*Digital Archaeology through the lens of Artificial Intelligence: developments and science policy perspectives*) at the XXI UISPP World Congress (Poznań, 31 August-4 September 2026).

This cumulative experience with AI-driven methods led to the exploration of tools for detailed examination of terminological patterns in domain-specific textual collections. Among these, Voyant Tools was selected. Developed by Stéfan Sinclair and Geoffrey Rockwell and maintained as an open-source community resource, it offers analytical functions (frequency analysis, KWIC concordances, and distributional trends across document segments) that support distant reading while complementing qualitative close reading. Within the Digital Humanities landscape (CIOTTI, BALDI 2022), Voyant occupies a distinctive position: rather than prioritising statistical robustness or large-scale scalability, it foregrounds interpretive transparency and interactivity, enabling researchers to explore terminological patterns that conventional reading might miss. This makes Voyant particularly suitable for corpora that are large enough to benefit from computational analysis while remaining small enough for qualitative interpretation.

To provide a concrete example, the following case study applies Voyant Tools to analyse the texts produced by the Open Science Group within the H2IOSC project, focusing on the open-access volume collecting all contributions published in A&C (<https://doi.org/10.19282/h2iosc.2026>). This approach builds on the previous AI-assisted analyses but shifts in granularity: from macro-level mapping of an entire scientific domain to micro-level analysis of the vocabulary used by a research community illustrating its own work.

The corpus was normalised to ensure linguistic consistency. All texts were written in English, except for one originally published in Italian, which was translated using Anthropic Claude with instructions to preserve the original terminology and argumentative structure. The resulting corpus was uploaded

to Voyant Tools and analysed with a customised stopword list designed to filter high-frequency function words and allow domain-specific terms to emerge more clearly. To facilitate the interpretation of distributional data, individual texts were uploaded using their article IDs.

Word cloud analysis highlights a set of terms that are representative of the H2IOSC research framework. Among the most frequent words, data (462 occurrences), digital (374), resource* (350), research (294), and open (212) stand out. These terms are semantically central, reflecting the conceptual axes around which H2IOSC activities have been organised: the management, sharing and reuse of research data; the role of digital infrastructures and tools; the identification and description of resources; and the broader commitment to Open Science.

Their distribution across individual contributions indicates that these terms recur throughout the corpus and that openness, digital resources, and data management constitute a shared conceptual framework underpinning the different research activities. Some localised peaks and lower concentrations are nonetheless observable: the former reflect papers that explicitly foreground the infrastructure's core terminology, whereas the latter correspond to those that approach the project through more specialised or tangential themes, methods, or case studies. Overall, Voyant's distributional evidence confirms the internal coherence of the corpus while also accounting for the thematic specificity of individual works.

A closer look at the two most frequent terms, data and digital, further refines this initial observation. The term data (Fig. 4) shows a uniform distribution, with its strongest lexical correlations involving research, digital, and the lexical family of resource, underscoring its centrality both as a research output and as an entity to be described, structured, shared, and integrated within digital infrastructures. Although relatively limited in frequency, the FAIR lexical family (Fig. 5) remains connected within this semantic field through terms such as data, resource, and principle, thereby reflecting a key conceptual horizon in H2IOSC. The term digital is lexically correlated with archaeology and epigraphy, suggesting that it functions not merely as a technical qualifier but as a conceptual lens through which these disciplinary practices are framed within the corpus. Similarly, resource is embedded within a wider lexical network that includes information and metadata, referring to the processes through which research materials are organised, described, and made accessible for reuse.

The Mandala visualisation corroborates this picture. In Voyant Tools, Mandala is a conceptual visualisation in which each search term, or 'magnet', attracts documents according to their relative frequency in the corpus. Here, the position of the articles suggests thematic convergence, with individual contributions retaining their own terminological profile while clustering around

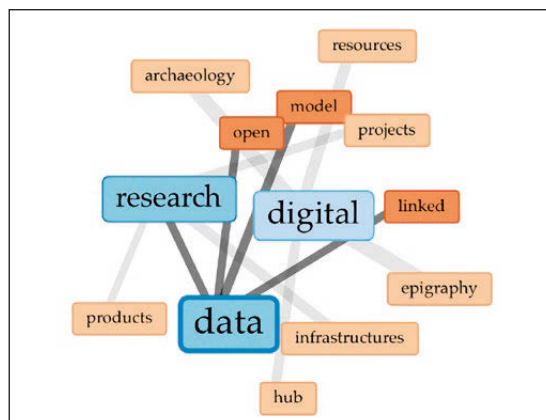


Fig. 4 – Co-occurrence network of the most frequent terms in the corpus.

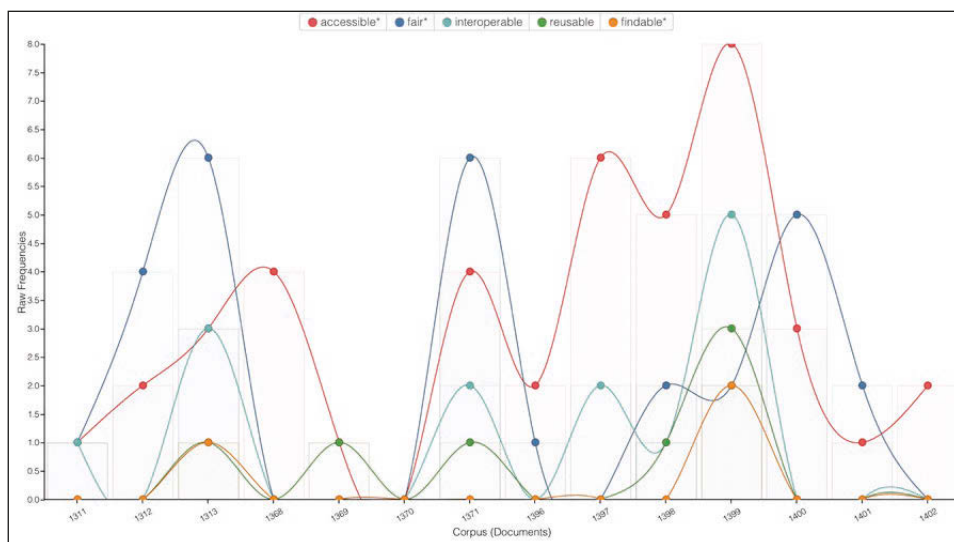


Fig. 5 – Distribution of the main FAIR-related terms across the articles in the corpus.

a shared set of core concepts: digital research practices, data, resources, and Open Science as applied to archaeological and epigraphic domains.

The Topics tool provides an additional exploratory perspective on the corpus by generating clusters of terms and showing their distribution across documents, based on Latent Dirichlet Allocation (LDA). Words are initially

assigned to a specified number of topics and then reorganised through a series of iterations according to co-occurrence patterns. Therefore, the results should not be understood as fixed or definitive thematic categories but rather as interpretive indicators: each run may produce slightly different clusters, although the resulting topics generally retain internal coherence. In the present analysis, six topics were generated, each represented by the ten most relevant words.

The first cluster highlights the organisational and infrastructural dimension, focusing on data management and resource availability. The second summarises the broader conceptual framework of the project, bringing together digital research practices, Open Science, and archaeological applications. The third foregrounds issues related to accessibility, metadata description, repositories, and the circulation of visual and textual resources within wider European infrastructures. The fourth topic is disciplinary, as indicated by terms reflecting the specific role of digital epigraphy within the corpus and its connection with linked data. The fifth shifts the focus towards the semantic and infrastructural modelling of cultural heritage data. Finally, the sixth cluster addresses the conceptual and technical frameworks used for structuring and integrating heritage information.

Taken together, these topics do not define clearly separated thematic areas, since several terms and concepts recur across different clusters. Instead, they reveal the internal articulation of a shared vocabulary in which disciplinary, infrastructural, and semantic dimensions are closely intertwined. Overall, this first level of analysis based on a small and deliberately bounded corpus shows that the most frequent terms correspond closely to the main objectives of H2IOSC and the broader Open Science agenda. However, the uneven distribution of some terms across the volume reveals the thematic diversity of the contributions and the different conceptual registers through which a shared vocabulary is adapted to methodological and domain-specific contexts.

P.M.

ALESSANDRA CARAVALÉ, PAOLA MOSCATI, IRENE ROSSI

Istituto di Scienze del Patrimonio Culturale - CNR
alessandra.caravale@cnr.it, paola.moscati@cnr.it, irene.rossi@cnr.it

REFERENCES

- BUSCEMI F., FAZIO L. 2024, *Verso la costituzione di linee guida per l'esposizione di risorse visuali negli aggregatori europei*, in CARAVALÉ, MOSCATI, ROSSI 2024b, 483-502 (<https://doi.org/10.19282/ac.35.2.2024.50>).
- CARAVALÉ A., DURAN-SILVA N., GRIMAU B., MOSCATI P., RONDELLI B. 2023, *Developing a digital archaeology classification system using Natural Language Processing and Machine Learning techniques*, «Archeologia e Calcolatori», 34.2, 9-32 (<https://doi.org/10.19282/ac.34.2.2023.01>).

- CARAVALE A., MILIANI C., MOSCATI P., SFAMENI C. 2022, *La sfida delle competenze per il patrimonio culturale: complementarità, integrazione, interazione*, in A. FILIPPETTI (ed.), *Le Scienze Umane, Sociali e del Patrimonio Culturale nell'era delle grandi transizioni*, Roma, CNR Edizioni, 65-86.
- CARAVALE A., MOSCATI P., ROSSI I. (eds.) 2024a, *The H2IOSC project and its impact on digital antiquity within the E-RIHS infrastructure – I*, «Archeologia e Calcolatori», 35.1, 515-562.
- CARAVALE A., MOSCATI P., ROSSI I. (eds.) 2024b, *The H2IOSC project and its impact on digital antiquity within the E-RIHS infrastructure – II*, «Archeologia e Calcolatori», 35.2, 463-523.
- CARAVALE A., MOSCATI P., ROSSI I. (eds.) 2025, *The H2IOSC project and its impact on digital antiquity within the E-RIHS infrastructure – III*, «Archeologia e Calcolatori», 36.1, 431-504.
- CARAVALE A., MOSCATI P., ROSSI I. (eds.) 2026, *The H2IOSC Project and its Impact on Digital Antiquity within the E-RIHS Infrastructure*, Firenze, Edizioni All'Insegna del Giglio (<https://doi.org/10.19282/h2iosc.2026>).
- CIOTTI F., BALDI A. 2022, *Macchine per leggere: promuovere la lettura con il distant reading*, in S. BALLERIO, M. TOGNINI (eds.), *La letteratura e la rete. Alleanze, antagonismi, strategie*, «Enthymema», 30, 173-192 (<https://doi.org/10.54103/2037-2426/19557>).
- MANCUSO G., CARAVALE A., D'EREDITÀ A., MOSCATI P. 2025, *Merging knowledge and tools in Heritage Science and Digital Archaeology. Practices from H2IOSC WP2*, in S. CAMPANA, D. FIRDANI, H. GRAF, G. GUIDI, Z. HEGARTY, S. PESCARIN, F. REMONDINO (eds.), *Digital Heritage 2025 (Siena 2025)*, Goslar, The Eurographic Association (<https://doi.org/10.2312/dh.20253320>).
- MOSCATI P. 2026, *Tracciare il passato, aprire il futuro: archeologia, open data e open science*, in L. CAPPUCCINI, A. GAUCCI, C. TREVISANELLO (eds.), *L'Appennino tra passato e presente: risorse, ambiente, Open Data. Atti del Workshop (Bologna 2025)*, Insediamenti d'Altura 3, Firenze, All'Insegna del Giglio.
- MOSCATI P. in press, *Archeologia digitale, letteratura scientifica e scienza aperta: il percorso di un'idea*, in *Dati archeologici nell'era digitale: gestione, comunicazione, sostenibilità. Atti del Convegno internazionale (Lecce 2026)*, «Studi di Antichità».
- PARACIANI N. 2026, *Archivi e repository visuali di «Archeologia e Calcolatori»*, in A. CARAVALE (ed.), *Una storia per immagini. L'archeologia digitale attraverso IADI, l'Interactive Atlas of Digital Images*, Firenze, All'Insegna del Giglio, 25-38 (<https://doi.org/10.19282/FA.8.2026>).
- ROSSI I. 2025, *Building an ecosystem of digital resources on the written heritage of ancient Arabia*, in CARAVALE, MOSCATI, ROSSI 2025, 469-480 (<https://doi.org/10.19282/ac.36.1.2025.26>).
- ROSSI I., SALVADOR C. 2024, *An observatory of epigraphic resources on the web: The Open Digital Epigraphy Hub*, in CARAVALE, MOSCATI, ROSSI 2024b, 503-523 (<https://doi.org/10.19282/ac.35.2.2024.51>).
- SANSONETTI A., CONTI C., BOTTEON A., COLOMBO C. 2025, *Characterization in Heritage Science*, in *Heritage Science Ontology RoundTable (<H/SORT>)* (<https://hsort-ec1af2.gitlab.io/posts/heritage-characterization/>; <https://doi.org/10.5281/zenodo.15877897>).
- SCARPA E. 2025, *Provenance and provenience*, in *Heritage Science Ontology RoundTable (<H/SORT>)* (<https://hsort-ec1af2.gitlab.io/posts/provenance/>; <https://doi.org/10.5281/zenodo.15166678>).
- SCARPA E., VALENTE R. 2024, *A resource hub for interoperability and data integration in Heritage Research: The H-SeTIS database*, in CARAVALE, MOSCATI, ROSSI 2024a, 543-562 (<https://doi.org/10.19282/ac.35.1.2024.32>).
- SCARPA E., VALENTE R. 2025, *Case study #2: What is an epigraph?*, in *Heritage Science Ontology RoundTable (<H/SORT>)* (<https://doi.org/10.5281/zenodo.15167879>).

- SCARPA E., VALENTE R., ROSSI I. 2025a, *Modeling an ontology for Heritage Science: Challenges and key strategies*, in S. REBORA, M. ROSPOCHER, S. BAZZACO (eds.), *Diversità, equità e inclusione: sfide e opportunità per l'Informatica Umanistica nell'era dell'Intelligenza Artificiale, Proceedings del XIV Convegno Annuale AIUCD2025*, Quaderni di Umanistica Digitale, Verona, AIUCD, 499-504 (<https://doi.org/10.5281/zenodo.18505398>; <https://doi.org/10.6092/unibo/amsacta/8380>).
- SCARPA E., VALENTE R., ROSSI I. 2025b, *H/RADIOA: Recommended Approaches for the Development of Interoperable Open Semantic Artefacts in the Heritage Domain*, in S. CAMPANA, D. FERDANI, H. GRAF, G. GUIDI, Z. HEGARTY, S. PESCARIN, F. REMONDINO (eds.), *Digital Heritage 2025 (Siena 2025)*, Goslar, The Eurographics Association (<https://doi.org/10.2312/dh.20253041>).
- VALENTE R. 2025, *Nature and culture*, in *Heritage Science Ontology RoundTable (H/SORT)* (<https://hsort-ec1af2.gitlab.io/posts/nature-culture/>; <https://doi.org/10.5281/zenodo.15167149>).

ABSTRACT

This paper provides an overview of the main initiatives carried out by the Rome and Milan research units of CNR-ISPC within the H2IOSC project, highlighting how the work has continued beyond its official completion while contributing to the promotion of Open Science practices in the field of Cultural Heritage/Heritage Science. It outlines the evolution of the digital platforms developed by the Rome unit (DHeLO, BiDiAr, and ArchaeoHub), focusing on the methodological choices underpinning interoperability, Linked Open Data, and FAIR principles, as well as on the role of scientific dissemination in fostering dialogue with the research community. In parallel, it presents the work carried out by the Milan unit on semantic modelling for Heritage Science and digital epigraphy through initiatives aimed at developing shared ontologies, advancing interoperability, and implementing open resources such as H-SeTIS, H/RADIOA, REMEDIA, and EpiHub. The final section explores the application of Artificial Intelligence and terminological analysis to the project's scientific output, showing how text mining and distant reading approaches can be used to identify and structure the core concepts emerging from H2IOSC research.