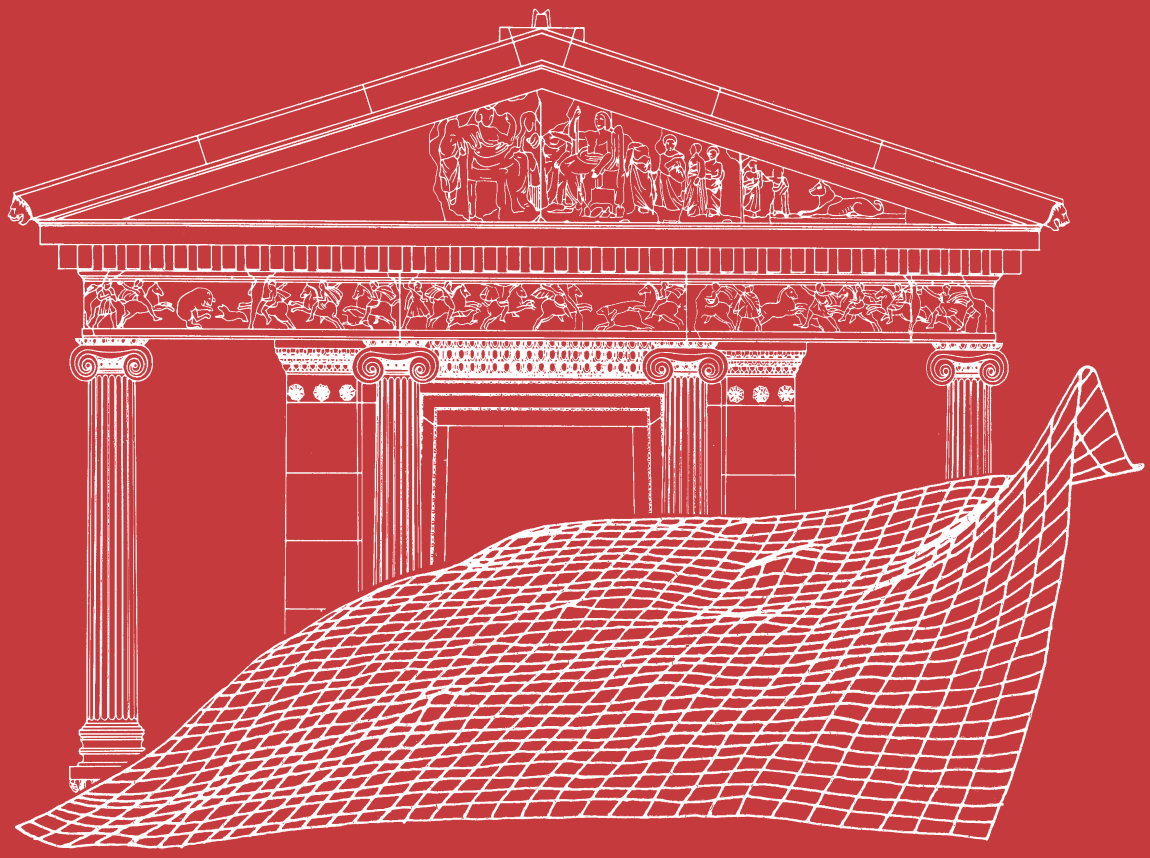


ARCHEOLOGIA E CALCOLATORI

35.1

2024

All'Insegna del Giglio

ARCHEOLOGIA E CALCOLATORI

DSU Dipartimento Scienze umane e sociali Patrimonio culturale

CNR – DIPARTIMENTO SCIENZE UMANE E SOCIALI, PATRIMONIO CULTURALE

ISTITUTO DI SCIENZE DEL PATRIMONIO CULTURALE

Rivista annuale open access e peer reviewed
fondata da Mauro Cristofani e Riccardo Francovich
già diretta da Paola Moscati (1990-2022)

Autorizzazione del presidente del Tribunale di Firenze n. 3894 del 6/11/1989

Indirizzo Redazione: Rivista «Archeologia e Calcolatori», CNR – ISPC, Area della Ricerca di Roma 1, Via Salaria Km 29,300, 00015 Monterotondo Stazione (RM)
Tel. +39.06.90672670 – Fax +39.06.90672818
E-mail: redazioneac.ispc@ispc.cnr.it
https://www.archcalc.cnr.it/

Edizione e distribuzione: Edizioni ALL'INSEGNA DEL GIGLIO s.a.s.,
Via Arrigo Boito 50-52, 50019 Sesto Fiorentino (FI)
Tel. +39.055.6142675
E-mail: redazione@insegnadelgiglio.it – ordini@insegnadelgiglio.it
https://www.insegnadelgiglio.it/

ARCHEOLOGIA E CALCOLATORI

35.1

2024

All'Insegna del Giglio

The publication of this journal's issue was funded 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.

Realizzazione grafica della sovracoperta di Marcello Bellisario
Rivista «Archeologia e Calcolatori» (ISSN 1120-6861, e-ISSN 2385-1953)
ISBN 978-88-9285-278-5, e-ISBN 978-88-9285-279-2

Sesto Fiorentino (FI), luglio 2024
Stampa, BDprint

Abbonamento 2024: 2 volumi, 35.1 e 35.2, € 96,00.
Spedizione: Italia, gratuita; estero, a carico del destinatario.
https://www.insegnadelgiglio.it/categoria-prodotto/abbonamenti/

INDICE

Necropoli etrusco-italiche: archeologia digitale e paesaggio funerario. Atti del workshop (Venezia 25 settembre 2023), a cura di Fiorenza Bortolami, Giovanna Gambacurta

Giovanna Gambacurta, *Premesse e prospettive di un workshop* 11

Paola Moscati, *Funerary archaeology and digital technologies: history and development of a successful cross-disciplinary approach* 15

Cristiano Putzolu, *The necropolis as a landscape of power: some reflections* 31

Antonella Massanova, Carmine Pellegrino, *Le necropoli di Pontecagnano: archivi e supporti digitali per la tutela e la ricerca a 25 anni dall'elaborazione del GIS* 41

Fernando Gilotta, Luca Lucchetti, Teresa Patriziano, *Esperienze da Monte Abatone a Cerveteri* 51

Alessandro Conti, Christian Mazet, Laura Maria Michetti, *'Ritorno a Vulci'. New tools for the study of the south-eastern necropolises* 67

Emanuele Taccola, Lisa Rosselli, Massimiliano Grava, *Per una ricostruzione dei paesaggi funerari in Etruria settentrionale: i casi di Pisa e Volterra* 81

Elisabetta Govi, Andrea Gaucci, Chiara Pizzirani, Anna Serra, Carlotta Trevisanello, Riccardo Vanzini, Enrico Zampieri, *Archaeological data management and spatial analysis tools in the study of necropolises: case studies from Bologna and Spina (6th-3rd cent. BC)* 99

Giovanna Gambacurta, Federico Bernardini, Fiorenza Bortolami, Martina Vanessa Filannino, Cecilia Moscardo, Angela Ruta Serafini, *Data management and reconstruction of funerary landscape in pre-Roman necropolises between Este and Padova* 117

*
* *

Images antiques et humanités numériques: section spéciale éditée par le programme ArcheoNum, sous la direction de Virginie Fromageot-Laniepce, Anne-Violaine Szabados

Virginie Fromageot-Laniepce, Anne-Violaine Szabados, *Images antiques et humanités numériques: une introduction* 137

Cécile Colonna, Federico Nurra, *Le* Digital Muret, *une édition numérique enrichie d'un recueil de dessins d'objets archéologiques du XIXe siècle* 139

Stéphanie Derwael, *Disséquer l'ornement, interpréter l'hybridité humano-végétale.* Diphuès – *Une base de données polyvalente* 157

Natacha Lubtchansky, Alban-Brice Pimpaud, *3D/4D visualisation for documenting and editing images of pre-Roman Italy: the ICAR database* 173

Anne-Violaine Szabados, *Le TAL pour les appellations d'œuvres figurées de l'Antiquité classique: évolution des ressources numériques du projet MonumenTAL* 193

*
* *

Giancarlo Lago, Lorenzo Cardarelli, Nicola Ialongo, *CQArchaeo: a Python package for Cosine Quantogram Analysis and Monte Carlo simulations* 215

Brunella Berzellini, Maurizio Marinato, Marco Piragnolo, *Un database condiviso ad accesso libero per lo studio archeologico dei cimiteri medievali in Italia settentrionale* 233

Mariacarmela Montesanto, *From the river to the sea of the setting sun: route networks between the Euphrates and the Mediterranean during the Iron Age (1200-600 BCE)* 251

Francesca Tomei, *Least-Cost Path analysis for the reconstruction of the communication networks between Thasian amphorae workshops and other sites in the 4th-3rd century BC* 269

Antonio Merola, *Analisi di visibilità delle fortificazioni d'altura di epoca arcaica nella Valle Peligna (Abruzzo). Ipotesi, analisi informatiche e ricostruzione* 285

Teresa Tescione, *Analytical approaches and digital methods in alluvial archaeology: the 'Ancient Shipyard' of Pisa-San Rossore as a case study* 305

Elena Pomar, *La villa romana di Matrice (CB) tra geofisica e scavo* 329

Fabio Cavulli, Carla Mannu, Massimo Vanzi, Michele Mazzurana, *Tecniche di modellazione 3D per una documentazione accurata delle incisioni rupestri: confronto tra Structure from Motion e stereofotometria* 349

Giulio Alberto Da Villa, Arturo Zara, Emanuela Faresin, Jacopo Bonetto, Cristiano Miele, Antonio Persichetti, *Methods and tools for generating the DTM of an archaeological site: the case-study of the photogrammetric survey of Nora (Sardinia, Italy)* 369

Alessandro Naso, Rodolfo Brancato, Martina Zinni, Simone Amici, *Application of integrated 3D survey technologies in an Etruscan necropolis: the case of Sasso Pinzuto (Tuscania, VT)* 389

Sofia Cingolani, Paolo Clini, Ramona Quattrini, Renato Angeloni, Francesca Angelo, Livia Sforzini, Antonella di Giovanni, Roberta Bollati, *Dalla replica digitale alla modellazione informativa. Un approccio scan-to-BIM alla documentazione del microscavo e restauro della tomba 27 di Colle Vaccaro (AP)* 407

Luca Doria, Caterina Previato, *BIM e archeologia: i casi studio degli anfiteatri romani di Verona e Pola* 427

Álvaro Corrales Álvarez, Rafael Hidalgo Prieto, *El Proyecto SIAPVA: un sistema de información arqueológica para el área de Palazzo de Villa Adriana (Tívoli)* 447

Paolo Clini, Renato Angeloni, Mirco D'Alessio, Giacomo Bardelli, Stefano Finocchi, *Un Virtual Immersive movie per la fruizione del patrimonio archeologico: il viaggio nella "Tomba della Regina" di Sirolo-Numana* 473

Alexandra Chavarría Arnau, Marco Bertuccelli, *L'archivio di comunità del Monte Baldo come esempio delle potenzialità di uMap-OSM per lo sviluppo di progetti partecipati sui beni culturali* 491

*
* *

The H2IOSC Project and its impact on digital antiquity within the E-RIHS infrastructure – I. Special section edited by Alessandra Caravale, Paola Moscati, Irene Rossi

Alessandra Caravale, Paola Moscati, Irene Rossi, *Landscaping and integrating Digital Archaeology and Digital Epigraphy resources: new challenges and future opportunities. Introduction to the Special section* 515

Giacomo Mancuso, Antonio D'Eredità, *DHeLO and BiDiAr: new digital resources within the H2IOSC Project* 521

Erica Scarpa, Riccardo Valente, *A resource hub for interoperability and data integration in Heritage research: the H-SeTIS database* 543

THE H2IOSC PROJECT AND ITS IMPACT ON DIGITAL ANTIQUITY WITHIN THE E-RIHS INFRASTRUCTURE – I

Special section edited by

Alessandra Caravale, Paola Moscati, Irene Rossi

Research conducted within 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.

Archeologia e Calcolatori
35.1, 2024, 515-520
doi 10.19282/ac.35.1.2024.30

LANDSCAPING AND INTEGRATING DIGITAL ARCHAEOLOGY AND DIGITAL EPIGRAPHY RESOURCES: NEW CHALLENGES AND FUTURE OPPORTUNITIES. INTRODUCTION TO THE SPECIAL SECTION

1. Introduction

The H2IOSC Project ('Humanities and cultural Heritage Italian Open Science Cloud'), funded by the European Union Next Generation EU and the Italian Ministry of University and Research as part of the National Recovery and Resilience Plan (NRRP), started in November 2022 with the aim to establish a federated cluster of the Italian nodes of the CLARIN, DARIAH, E-RIHS and OPERAS[1] research infrastructures in the ESFRI (European Strategy Forum on Research Infrastructures) domain focusing on Social and Cultural Innovation. This endeavour seeks to foster collaboration among researchers from various disciplines in the Humanities, Language technologies and Cultural Heritage sectors, enabling them to engage in data – and compute-intensive research. The Project (https://www.h2iosc.cnr.it/) is structured into eight Work Packages, with archaeologists and epigraphists playing a significant role in the sub-packages pertinent to the E-RIHS infrastructure.

This special section of the journal «Archeologia e Calcolatori» is dedicated to showcasing the research activities conducted by the CNR-ISPC Open Data, Open Knowledge, Open Science research group[2] as part of the H2IOSC Project. The featured research spans across three Work Packages: WP2 (Landscaping & Building Communities), WP4 (RIs Nodes and Resources Interoperability), and WP7 (Community Pilots: Innovative Cross-domain Services and Environments). Our editorial programme envisages publishing three special thematic sections in the journal's 35.1, 35.2 and 36.1 issues. The main objective is to promote and share the research outcomes by providing free access to scientific data, following an approach that aligns with our journal's tradition. At the same time, these publications will enrich the federated cluster of Italian nodes of the leading research infrastructures in the Humanities and Cultural Heritage, by creating awareness that knowledge and interpretation of the past can inform future technological decisions.

[1] http://dariah.cnr.it/; https://www.clarin-it.it/; https://www.e-rihs.it/; https://operas-eu.org/.
[2] https://www.ispc.cnr.it/it_it/2020/05/14/gruppo-open-data/.

2. Improve knowledge on Digital Archaeology resources and classification

As part of the WP2, research activity 2.4 focuses on 'Landscaping the Cultural Heritage and Heritage Science resources and needs panorama'. This activity is dedicated to surveying and detailing the existing resources and requirements within the Cultural Heritage (CH) and Heritage Science (HS) research communities participating in the Italian E-RIHS network. It involves a series of interconnected tasks aimed at identifying and highlighting the available resources and repositories, their specific disciplinary domains, the main services and tools developed, and the best practices, protocols, and standards employed. The objective is to ensure that the production and sharing of datasets adhere to the FAIR principles, promoting their interoperability.

Since the start of the H2IOSC Project, specific attention has been placed on classifying the main cross-cutting themes that characterise the Digital Archaeology research domain and its current integration within Digital CH and HS. In the context of the articles published in the international journal «Archeologia e Calcolatori», this evolution necessitates a fresh examination of the methods employed to classify the wide-ranging areas of application that now span a vast array of technological innovations, posing challenges to conventional classification frameworks. Thanks to the application of Machine Learning (ML) techniques, the implementation of a knowledge map has been instrumental in identifying topic clusters and keywords that exemplify the ten dominant technological approaches[3] to analyse and interpret archaeological data and in highlighting the pivotal contribution of archaeology – particularly computational archaeology – to the multidisciplinary domain of HS[4].

Within the activities illustrated in this special section, we delve deeper into the advancements made in the frame of WP2 and, therefore, into the landscaping activities. The definition of a collaborative framework between the disciplinary communities associated with the four research infrastructures helped us to establish general guidelines for conducting the survey, collecting information and mapping their technological landscape. The main focus of the activity was centred around three primary goals: the development of a web-based questionnaire structure, the selection of candidates for in-person interviews, and the initiation of a semi-automated search for existing resources, tools, software and services.

In addition to the active participation of the E-RIHS community in both the design and the dissemination of the questionnaire, which is set to be distributed

[3] Artificial Intelligence; GIS and Spatial Analysis; Imagery Analysis; Material Culture; Modeling and Simulation; Digital Cultural Heritage; Photogrammetry and 3D Scanning; Remote Sensing; Semantic Technologies; Virtual Reality.

[4] Caravale 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).

online via the Lime Survey service thanks to the licensing and server hosting provided by CNR (https://www.limesurvey.org/), our research group strategically developed two operational databases: DHeLO and BiDiAr (see Mancuso, D'Eredità, in this special section). They are designed to closely monitor the digital resources of the CH/HS domains, along with their extensive bibliographic references. These tools are intended to be both flexible and versatile, enabling the storage of data collected via the questionnaire distribution and facilitating interaction with other landscaping initiatives promoted by WP2, such as the CLARIN Virtual Language Observatory (VLO). Furthermore, the data collected can serve as input for the E-RIHS DIGILAB platform and contribute to implementing the Open Digital Archaeology and Epigraphy Hubs, which are being developed as an E-RIHS pilot project within H2IOSC's WP7.

As will be extensively discussed, DHeLO's primary objective is to establish an observatory for digital products to provide a deeper insight into the dynamic scenario of CH/HS and better meet its needs in terms of digital tools. The database conceptual scheme enables efficient mapping of digital products, providing, in particular, an in-depth exploration of the progress of Digital Archaeology, including an examination of the ten thematic clusters mentioned above (see note 3). This ongoing data analysis progressively reveals common practices, shared standards, and similar workflows across various research projects and institutions. This holistic approach enhances our understanding of Digital Archaeology's trajectory, fostering improved cooperation and informed decision-making.

P.M.

3. Digital Archaeology: collection and sharing of bibliographic resources

In an effort to create new resources for the Italian node of the E-RIHS infrastructure, a new bibliographic reference tool primarily focusing on Digital Archaeology has been developed to be implemented in the H2IOSC Marketplace. Named BiDiAr, this tool is a product of the CNR Institute's 30 years of activity in the field of archaeological computing. It not only documents the most recent developments but also facilitates the preservation of past events and contexts. Indeed, in recent years, archaeological computing, a relatively young discipline, has seen significant growth in studies and research, due to the widespread integration of digital technology across all the different investigation areas. This has not only broadened the perspectives of analysis, but has also revolutionised the methods of collecting, preserving, and using archaeological data, opening up new horizons of comprehension, interpretation, and dissemination of the past through technology. Analysing the evolution of this discipline through its bibliography provides a historical and conceptual overview, starting from the earliest works focused on a small number of key topics up to the current research

panorama characterised by a diversified and multifunctional use of machines and technologies. The availability of a bibliographical corpus makes it possible not only to benefit from an organised and systematic framework of a constantly expanding discipline, but also to have a grasp of the sectors that today have achieved major relevance while identifying those that are less visited.

The CNR-ISPC Open Data, Open Knowledge, Open Science research group acknowledges the pivotal role of bibliography in understanding a specific disciplinary sector. Over time, the group has conducted different research in this domain, sharing their outcomes online. An important initiative is the Virtual Museum of Archaeological Computing (https://archaeologicalcomputing.lincei.it/), representing the multimedia apex of an international project dedicated to the historical development of computers in archaeology. This project originated from a fruitful collaboration between Accademia Nazionale dei Lincei and Consiglio Nazionale delle Ricerche. The website provides access to several bibliographies, beginning with those associated with prominent scholars in archaeological computing who have been active since the 1950s (J.-C. Gardin, A.C. Spaulding, D.L. Clarke, G.L. Cowgill, J.E. Doran, R. Ginouvès, F.R. Hodson). The portal also contains the Bibliography of Archaeological Computing of the 1990s, a rich repository comprising over 2700 titles spanning the period from 1989 to 2000, which has been the subject of a specific study[5]. The study aimed to describe the international cultural panorama of the 1990s, connecting it with the achievements of the previous decades and with the challenges of the subsequent years, which have deeply influenced the development of Digital Archaeology in the new millennium.

BiDiAr is a new research tool in the field of Digital Archaeology and Cultural Heritage (see Mancuso, D'Eredità, in this special section). It is developed in H2IOSC-WP2 using bibliographic data collected from recent volumes of «Archeologia e Calcolatori», which has been a reference point in this sector since 1990. As described in this special section, the repository archives over 6500 titles using the open-source software Zotero, providing access to a rich body of knowledge and offering an overview of studies that, although focused on recent years, provides a diachronic perspective on this research area, allowing its evolution and trends to be tracked over time. The tool's potential lies in its continuous implementation and updating, enabling ongoing monitoring of the discipline's status. Additionally, it can be linked with other tools developed within the same research group, such as the above-mentioned DHeLO and IADI (A&C Interactive Atlas of Digital Images; https://iadi.archcalc.cnr.it/), aimed at creating a digital ecosystem that offers an integrated approach to data consultation, fostering their sharing and dissemination.

A.C.

[5] Caravale A., Moscati P. 2021, *La bibliografia di informatica archeologica nella cultura digitale degli anni Novanta*, Firenze, All'Insegna del Giglio.

4. Towards a semantic framework for the interoperability of Cultural Heritage resources: the role of Digital Archaeology and Digital Epigraphy

Central to H2IOSC's objective of facilitating data integration and supporting research in the Humanities, Language technologies and Cultural Heritage sectors are the activities that CNR-ISPC is conducting in the frame of the Project's WP4. The CNR-ISPC Milan branch of the Open Data, Open Knowledge, Open Science group leads Activity 4.10, which aims at ensuring interoperability in the domain of Cultural Heritage and at contributing to the elaboration of a shared semantic framework among the four RIs involved in the Project to allow their resources to be onboarded into an integrated, cross-domain environment – the H2IOSC Marketplace. Activity 4.10 is related to the resources which will populate DIGILAB, E-RIHS's digital platform offering virtual access to tools and data in the field of Heritage Science. These activities encompass identification and mapping of metadata schemas, alignment to community standards, ontology design and application.

The contribution by Scarpa and Valente in this special section tackles the need for a survey of Knowledge Organisation Systems in the Heritage sector as a preliminary step for resources' harmonisation and integration, with a focus on semantic artefacts (i.e. knowledge models which are intelligible – and therefore automatedly processable – by machines, as the authors explain). The first result of this survey activity has been the setting up of H-SeTIS – Heritage - Semantic Tools and Interoperability Survey, an open access observatory of semantic artefacts in the Heritage domain, providing descriptions of resources and tools such as ontologies, metadata schemas, thesauri, application profiles and software. H-SeTIS, in turn, has been designed as a tool whose information is mapped to standard schemas and is exposed in interoperable formats through a web-service. The bibliographic catalogue of the semantic survey, whose entries are associated to the records of the H-SeTIS database, is maintained as an open Zotero group.

The use of keywords for describing the disciplinary domain(s) of each resource catalogued in H-SeTIS allows to draw some preliminary qualitative and quantitative reflections on the semantic landscape of the Heritage sector. This shows a longstanding interest and concerted effort towards a shared description of Cultural Heritage-related digital resources in the fields of Archaeology, Museum and Text studies. On the other hand, disciplines concerned with diagnostic analyses of cultural objects suffer from a lack of contents' formalisation and integration with existing semantic tools in the field of Cultural Heritage. To address such issue, CNR-ISPC will design an ontology for modeling the Heritage Science domain that integrates into the H2IOSC semantic platform. Crucial to this final aim is the acknowledgement, among the information catalogued in H-SeTIS, of the development/maintenance status of the described resources, which is an essential aspect to sift through them and provide an up-to-date

panorama of the reliable and usable semantic tools, while acknowledging the closed or abandoned initiatives to document the history of studies in the field.

The integration of cross-domain resources enabled by WP4's activities is going to be applied and tested in H2IOSC through a series of pilot projects, to be developed within WP7. As part of E-RIHS, CNR-ISPC is currently involved in the creation of several digital hubs which span and intersect the multiple disciplinary fields of Heritage Science. Activity 7.4 is working towards the implementation of two case studies related to the fields of Digital Archaeology and Digital Epigraphy that will benefit from the services provided by the H2IOSC RIs in terms of tools and resources for material heritage studies and text studies. Starting from the CNR decennial experience of Diamond Open Access publishing in the Digital Archaeology sector, represented by the journal «Archeologia e Calcolatori», the Open Digital Archaeology Hub will provide archaeologists with accessible resources to experiment with innovative technologies for the extraction of structured information from traditional textual and iconographic corpora, thus stimulating awareness of Open Science in the archaeological field[6]. The Open Digital Epigraphy Hub, on the other hand, aims at building an open platform for the discovering and sharing of resources, tools and methods in the field of Digital Epigraphy, leveraging the experience of the open access 'Digital Archive for the Study of pre-Islamic Arabian Inscriptions' (DASI: https://dasi.cnr.it/)[7].

Indeed, the peculiar nature of the inscriptions – which are both textual and material sources – has put the research on epigraphic corpora at the crossroads of Digital Humanities, Computational Linguistics and Heritage Sciences. 'Epigraph' appears thus as a complex object of study, one that is especially apt to test technological solutions developed within the Project for its interdisciplinary goals. The joint effort that WP7 and WP4 are undertaking to identify a common semantic framework of H2IOSC starts precisely from the case study of 'text-bearing objects', where epigraphs as resources represent the various facets of the textual, art-historical and archaeological inquiries onto which the diagnostic processes of the analytical sciences of Cultural Heritage can be applied. These activities will be described in forthcoming contributions in future issues of this journal.

I.R.

Alessandra Caravale, Paola Moscati, Irene Rossi
Istituto di Scienze del Patrimonio Culturale - CNR
alessandra.caravale@cnr.it, paola.moscati@cnr.it, irene.rossi@cnr.it

[6] Rossi I., Paraciani N. 2021, *IT applications to archaeology and the OA diamond journals' challenge. Enhancing access and reuse of textual and visual resources*, «Archeologia e Calcolatori», 32.1, 325-347 (https://doi.org/10.19282/ac.32.1.2021.18).

[7] De Santis A., Rossi I. (eds.) 2018, *Crossing Experiences in Digital Epigraphy: From Practice to Discipline*, Berlin-Boston, De Gruyter (https://doi.org/10.1515/9783110607208).

Archeologia e Calcolatori
35.1, 2024, 521-542
doi 10.19282/ac.35.1.2024.31

DHELO AND BIDIAR: NEW DIGITAL RESOURCES FOR THE H2IOSC PROJECT

1. Introduction

Within the extensive scope of the H2IOSC Project (https://www.h2iosc.cnr.it/), a primary goal of Work Package 2 (WP2 – Landscaping & Building Communities) is to conduct a comprehensive survey of the Italian digital landscape in Language Technologies, Humanities, and Heritage Science. This includes an analysis of existing research projects, resources, tools, communities, best practices, and current standards (see Caravale, Moscati, Rossi, in this special section). To achieve this, a dedicated multidisciplinary landscape research group was formed, including members from the four Italian nodes of ESFRI involved in the H2IOSC Project: CLARIN, DARIAH, E-RIHS, and OPERAS research infrastructures. The aim of this landscaping effort is to enhance the reliability of these infrastructures among scholars to further align their offer with the changing needs of the digital humanities, linguistics, and heritage science research communities.

The integration of various landscaping activities and different tools is leading to the establishment of a permanent observatory to monitor the status of these infrastructures in terms of new resources, best practices, technological and user needs (Luzietti *et al.* in press). This paper will discuss the initial results of the activities conducted by the Rome branch of the CNR-ISPC, as leader of Task 2.4 and in representation of the E-RIHS node, on the landscaping of Cultural Heritage (CH) and Heritage Science (HS) resources and needs panorama, that led to the development of two products: a web app titled DHeLO and a thematic bibliographic collection named BiDiAr.

G.M., A.D.

2. DHeLO. Digital Heritage Landscaping PlatfOrm

DHeLO is a web app developed as a part of the landscaping activities of the H2IOSC Project to collect, store and query metadata of research projects, products and digital tools used within the CH and HS domains. The main goal beyond its development is to take a first and consistent step towards the creation of a disciplinary observatory, able to collect and store metadata from multiple data sources in a structured and organized system, allowing complex and multivariate query, data indexing, and retrieval (Fig. 1).

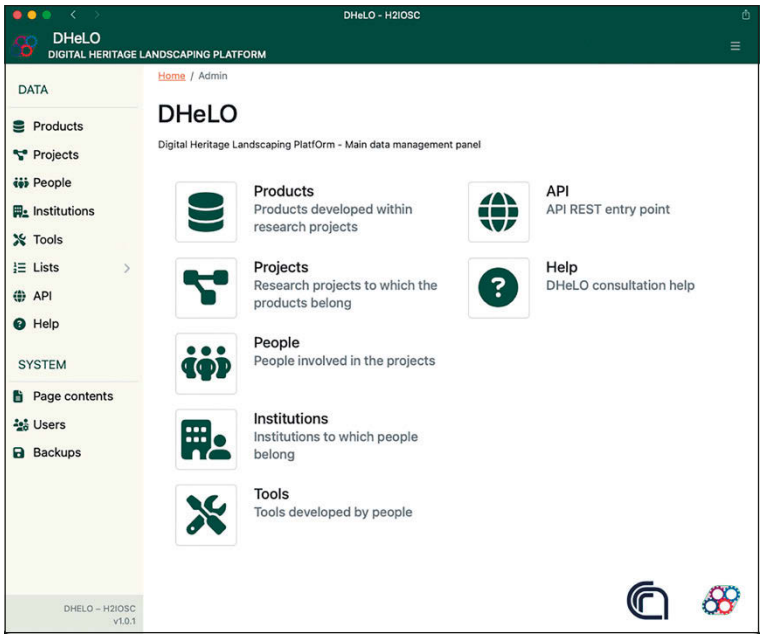


Fig. 1 – Screenshot of the DHeLO web app home page

2.1 *Premises*

It is generally acknowledged that over the past two decades, disciplines related to HS and CH have experienced significant growth in the creation of digital outputs, either as direct research products or through extensive digitization initiatives (Poulopoulos, Wallace 2022), that have significantly advanced the widespread implementation of the FAIR principles (Wilkinson *et al.* 2016). Among these noteworthy initiatives some are worth mentioning, such as CulturaItalia (Ben Hamida *et al.* 2009), launched in 2007, and Europeana (Silva, Terra 2024), started in 2008; two projects that populated virtual portals aggregating digital resources from national cultural foundations (the former), and European institutions (the latter). Among the projects more focused on archaeological data, Ariadne and Ariadne+ (Richards, Niccolucci 2019) certainly stand out. Launched in 2013 and 2019 respectively, they aimed at integrating archaeological data infrastructures within an online platform, open to users and researchers. With a more strictly geographical approach, the recent Geoportale Nazionale per l'Archeologia (GNA) must also be mentioned as an Italian national initiative designed to integrate and provide access to archaeological digital information and resources across the

country (Calandra *et al.* 2021). On an extra European Union (EU) front, it is worth mentioning the Archaeology Data Service (AdS), as a long-withstanding repository for UK archaeology and historic environment data, active for over 25 years (Richards 2017). Similar US-founded initiatives are The Digital Archaeological Record platform (tDAR), a repository designed to host the records of archaeological investigations (McManamon *et al.* 2017), and the Open Context service, dedicated to the storage, publication and consultation of CH and archaeological data (Kansa 2022).

Furthermore, the creation of wide digital infrastructures that aggregate, manage and disseminate data has provided researchers with new tools for the reuse, combination and analysis of datasets (Taylor 2023). A meaningful Italian example is D4Science, a data infrastructure service that provides comprehensive virtual research environments, enabling collaboration, data management, and advanced analytics for researchers and organizations across various scientific disciplines, HS included (Assante 2019).

In addition to these projects, it is also important to acknowledge how the growing use of DOI-providing services for sharing and storing data, such as Zenodo, has facilitated the dissemination of digitally native research data over the last ten years. This open access repository, supported by the EU through the OpenAIRE initiative, currently (May 2024) houses over 313,000 datasets, including 2890 tagged under archaeology, 5427 under CH, and 25,357 under HS. Furthermore, for HS researchers, a dedicated portal is accessible at Heritage Science on OpenAIRE (https://heritage-science.openaire.eu/), offering a wealth of information, publications and tools. Similarly, the Iperion HS gateway on OpenAIRE (https://iperionhs.openaire.eu/) houses products developed within the IPERION HS project (Pallot-Frossard 2016; Caliri *et al.* 2020), including publications, research software, and data. Lastly, it is worth mentioning that the focus on FAIR data management is a central goal of the Institute of Heritage Science (CNR-ISPC). The DataSpace (DS), launched in 2022 (https://dataspace.ispc.cnr.it/), represents the Institute's first approach to data lifecycle management in HS through a Linked Open Data (LOD) strategy. As part of the H2IOSC Project, a second digital platform, DIGILAB, will be introduced for the E-RHIS network. In accordance with FAIR principles, DIGILAB will enable access, interoperability, and reuse of data, tools, and digital services to foster the creation of new shared knowledge among various stakeholders involved in the multidisciplinary field of HS (Bucciero *et al.* 2022).

In this dynamic and complex data management environment, the fairification process is becoming increasingly vital and the creation of a disciplinary observatory for HS and CH could be a key step in this process. It would serve as a foundational platform to systematically catalog data from various sources, thus ensuring broad accessibility on the web, and to track evolving

trends and topics within these disciplines. The development of a specialized web application, DHeLO, with a data model specifically tailored for metadata enrichment, arose from the identification of some needs emerged during the initial phases of WP2 work, and particularly from the insights gained through the questionnaire launched at the beginning of the project (Luzietti *et al.* in press). This survey manifested a huge lack of common data management practices and confirmed the strong heterogeneity of the CH/HS data, often consisting of multiple file types (images, texts, tables, surveys, point clouds, etc.) developed in multiple formats and aggregated through a consistent chronological or topographical framework.

This awareness led us to the decision to direct the landscaping activities toward datasets rather than individual resources, for which the development of specific ontologies would have been necessary. During this preliminary exploratory phase, the work group also noted the absence of an existing platform capable of fulfilling the role of a disciplinary observatory, since most of the existing ones (previously mentioned), exhibited certain limitations with respect to the final objective. For instance, the distinct need to map datasets rather than raw data led to exclude some of the more inclusive disciplinary platforms like Europeana and CulturaItalia. On the other hand, other products with a higher level of data granularity, such as Ariadne+ or Zenodo, were either too narrowly focused thematically or burdened by overly generic metadata.

All these preliminary considerations led to the idea of creating a system dedicated to enrich existing open datasets with specific metadata, while simultaneously allowing map data not yet deposited but described in literature. The goal is to provide a clear picture of what exists and what is being done in terms of digital products within the fields of CH/HS. Additionally, there was a need to implement thematic classification of research products and their geographical and chronological indexing, which is crucial for these disciplinary sectors and missing from many reference platforms. The general metadata scheme was based upon the guidelines suggested by the Dublin Core Metadata Initiative (DCMI), to maximize the overall interoperability toward future integrations. Ultimately, the necessity became apparent for analytical tools that could assist in measuring and analyzing these data, to assess, for example, the impact of certain types of tools on specific products or workflows. Lastly, the need to make all these metadata widely available and machine-readable led to the creation of JSON rest APIs for the DHeLO web app, to ensure future data harvesting and interoperability, enabling seamless integration and communication with other software systems and tools such as the previously mentioned DIGILAB, and the Open Digital Archaeology and Epigraphy Hubs, both to be developed during the H2IOSC Project.

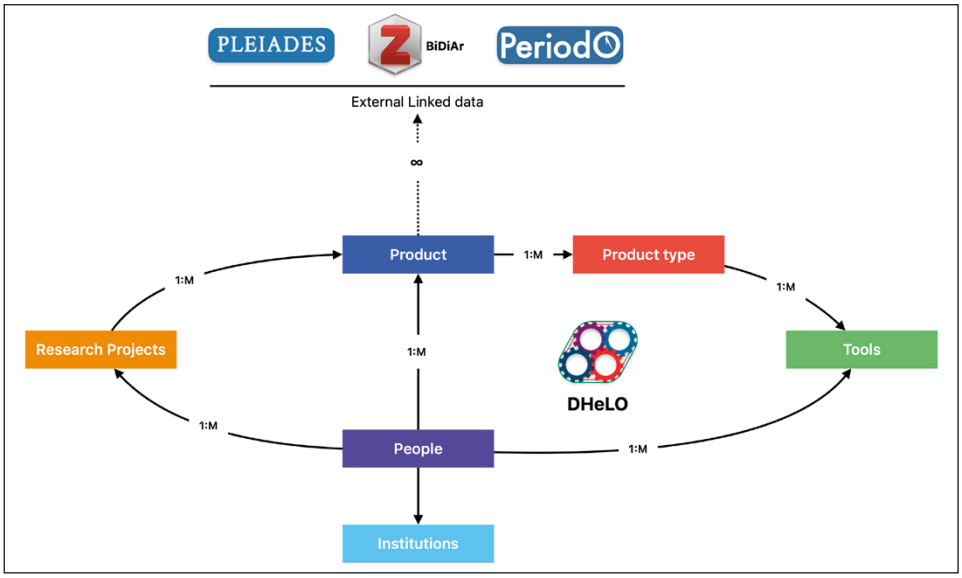


Fig. 2 – Overview of the DHeLO table schema structure.

2.2 *DHeLO web app. A technical description*

From a technical standpoint, the web app consists of a MySQL database with an HTML interface that can be used for data entry and retrieval. The database is organized into six main tables, with the 'Products' table serving as the core component (Fig. 2). In this context, 'Products' (as in research products) refer to datasets or collections of datasets that are grouped together based on a shared framework, such as chronological period or geographic region. This solution recalls the approach used by Zenodo for managing user datasets. In this specific case it proves particularly efficient because it does not limit the ability to input data in the system based on specific and individual data types, allowing the possibility to index complex products made of multiple kinds of resources (e.g. database with images or 3D models with semantic annotations).

Within the 'Products' table, additional metadata are entered to facilitate the indexing and the retrieval of items. These include classification and subject matter, data ownership, licensing details, the location of data storage (if data are shared), research projects, and bibliographic references describing them. For the classification tags, it was decided to test the criteria developed using Natural Language Processing (NLP) and Machine Learning techniques on the abstracts from the journal «Archeologia e Calcolatori» (A&C - Caravale

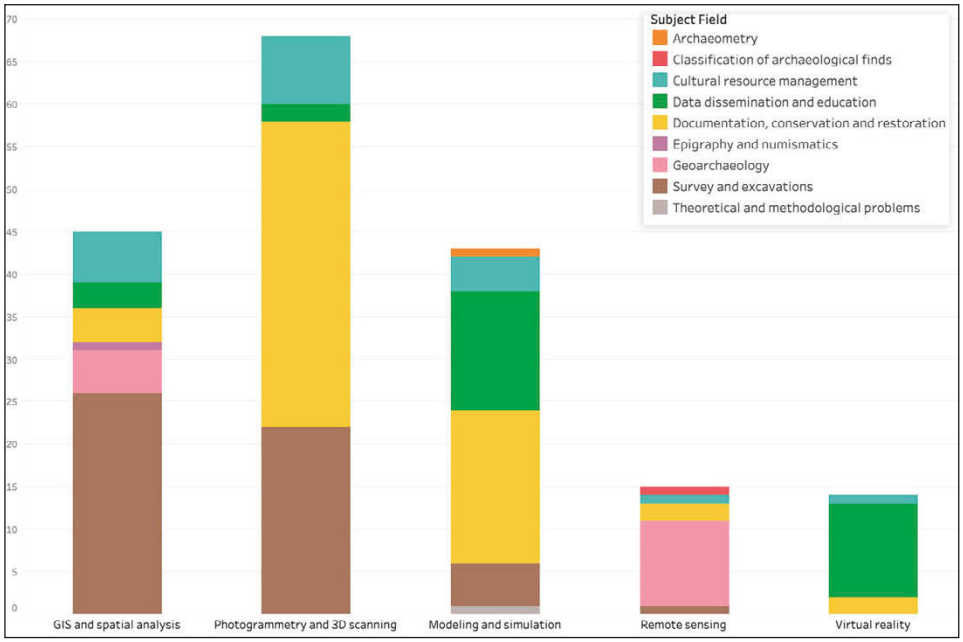


Fig. 3 – Bar chart of a selection of Products related to the most popular categories of technological applications in the digital CH/HS sectors, colored according to the archaeological research fields primarily concerned.

et al. 2023), since CH/HS disciplinary topics are broadly represented both in the journal and in DHeLO (Fig. 3). Ultimately, this choice proved effective, even though some of the classes seem poorly represented within DHeLO; additional subject details are also provided by the 'Subject' field. This dual-layered approach, a long-term strategy within the A&C journal (Moscati 1999), ensures a comprehensive description that captures the technological aspect, provided by the 'Category' field, while contextualizing its disciplinary applications within the 'Subject' field.

For bibliographic references, only a link to the Zotero library of BiDiAr is included (see *infra* § 3). The approach of linking to an external Zotero library streamlines bibliographic management, allowing for dynamic updates and centralized reference handling. This method also simplifies the maintenance of the reference data, making it easier for users to access up-to-date bibliographic information. For geographical and chronological classification, the system relies on two gazetteers: Pleiades and PeriodO. By referencing gazetteers, accuracy and standardization in geographic and temporal data are ensured, a critical aspect for chronological and spatial query (Fig 4). Similar to the

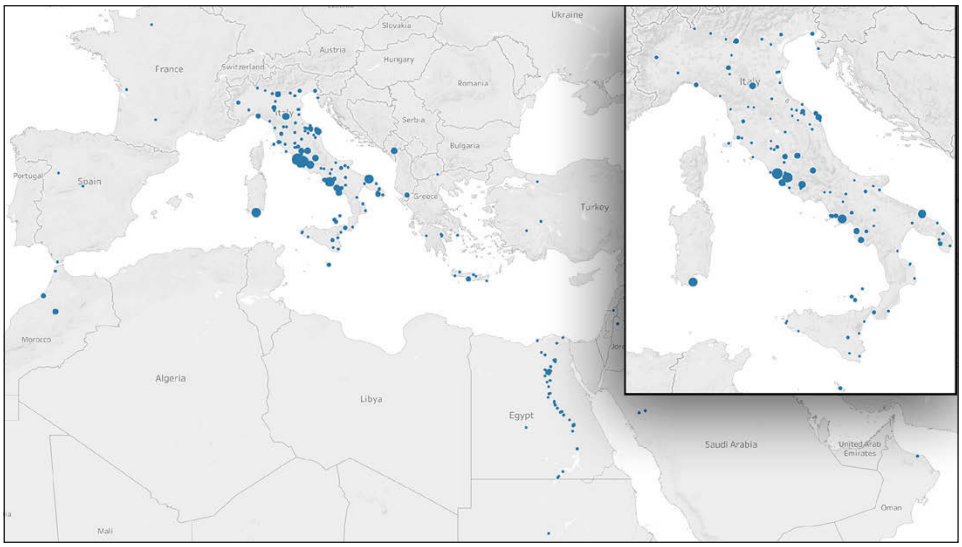


Fig. 4 – Map of the Products in the DHeLO database, geotagged through the Pleiades references.

solution adopted for the bibliography, these data are contained in relation tables, which allow the construction of many-to-many relationships, enabling the precise indexing of datasets that contain references to multiple places, periods, or are described by multiple bibliographic resources. The authorship of the dataset is described through two distinct fields that map the curation and production aspects; both fields draw data from the 'People' table.

The technical contents of the records in the 'Product' table are further detailed through a second table called 'Product Type,' in which the different types of products that make up the dataset are specified. This additional table allows for a more nuanced classification and management of the information, so, for instance, a photogrammetric survey product record can be further detailed in all the data types that were created during the process (e.g. pictures, point cloud, mesh, 3D model) and shared. Due to the relationship with the 'Product' table, the individual types of products that make up the dataset inherit their metadata and are thus individually indexable and searchable enhancing the database's overall functionality and user experience in conducting searches based on data types over specific territorial or chronological frameworks.

All the product types that make up the dataset can be further detailed through a many-to-many relationship with the 'Tools' table, that records the digital tools used for the creation, display, or interaction with the data. The 'Tools' table then records several important pieces of information about the type of tool (e.g., software, plugin, web-app), its availability, its license

and its developers. This setup enhances the database's utility by providing comprehensive details on the technological aspects of product data handling, allowing, for instance, to assess the prevalence of open-source software in relation to different product types.

The 'Project' table, which is closely linked to the 'Products' table, is designed to gather information on research projects. This table records various details such as the project's title and acronym, a brief description, its duration, the reference website, and the people involved. The relationship between this table and the products allows the records in this table to act as an umbrella under which all products developed during a single research project are collected. This structure enhances the organization and tracking of outputs directly associated with specific research initiatives. Two additional tables, 'People' and 'Institutions', are used to track the individuals and institutions involved in the creation of products, projects, and tools. The 'People' table contains basic personal information (name, surname, email) easily accessible online. It also allows for linking a contact record with the corresponding institution and the author's ORCID to facilitate future automatic updates. The 'Institutions' table focuses on the department or institute, and maps out details including the institution, department, and the associated website, providing a comprehensive reference for each entity involved in the database.

2.3 *First results and statistical analysis*

As of now, although the gathering process is still ongoing, sufficient data has been collected to attempt an initial analytical approach based on the collected information. The data collection process began with scraping the products and research projects carried out by the ISPC and the E-RHIS infrastructure, along with the incorporation of results from the landscaping questionnaire launched in the early phases of WP2. Data was subsequently gathered from major data-sharing platforms relevant to the disciplines of CH/HS, such as Zenodo, Ariadne+, HS-Openaire, Iperion HS, AdS, tDAR, and OpenContext. Following this, the parsing process was continued based on sector-specific literature, focusing primarily on works from the last five years, to reflect an updated state of the art. For the Digital Archaeology (DA) field were considered products described in peer-reviewed journals such as «Archeologia e Calcolatori», «Virtual Archaeology Review», «Open Archaeology», and «Internet Archaeology». An in depth focus on Digital CH was achieved by indexing products from the «Journal of Cultural Heritage» and «Digital Applications in Archaeology and Cultural Heritage», while the HS domain was explored through «Heritage Science» and «Heritage» journals. Additionally sector-specific proceeding series (e.g. Computer Application in Archaeology, MetroArchaeo) within the last five years were taken into account.

During this process, due to H2IOSC Project requirements, priority was given to Italian research products, with consideration also given to those developed abroad, but when conducted by Italian researchers or research groups (Fig. 4). Nowadays (May 2024), on the DHeLO web app 265 records are registered on the project table, hosting 1121 individual product types, in addition to metadata from 121 research projects and 114 tools. This data allows us to make some preliminary observations.

One key insight that emerged through reviewing sector-specific literature is the lack of a direct correlation between datasets and research projects. In fact, out of the 265 cataloged products, 38.8% are not linked to a specific research project. Further analysis of products categorized by classes and organized by whether they are part of a research project reveals that the ratio among various classes does not undergo radical changes. Some minor shifts are observed when comparing the types of products developed within and outside structured research projects. The data is interesting because it suggests that the production of these kinds of datasets is becoming a regular part of disciplinary workflow, without necessarily requiring research projects that involve various interdisciplinary skills.

Based on the data concerning the prevalence of different data types in various products (in or out research projects), there is a clear dominance of five kinds of resources, which together make up approximately 62% of the mapped resource types: 3D models (18.7%), databases (9.59%), GIS data (10.33%), pictures (11.7%), and point clouds (11.62%) (Fig. 5). This information is particularly noteworthy as it highlights a significant component of products oriented towards three-dimensionality (3D models and point clouds), which, when combined with relevant data from BIM models (1.48%), DEM and DTM (2.4%), GPR data (2.4%), and tomography (0.74%), generate a group that account for nearly 40% of all mapped product types. Further analysis of the distribution of these resources according to their classification shows that most of these products (ca. 50%) are related to survey operations, especially photogrammetric, where 3D meshes and point clouds are closely linked in the operational workflow (D'Eredità 2020).

It is also observed that the values for point cloud types relevant to survey operations are slightly higher, including data from surveys conducted solely by laser scanning and without the production of polygonal meshes. The remaining 3D data primarily fall into three categories: modeling & simulation (ca. 31%), digital cultural heritage (ca. 6%), which includes surveys of objects and 3D diagnostic products, and virtual reality (12%), where 3D models used in VR applications are cataloged. The first category is quite extensive and encompasses 3D models used for virtual reconstruction or simulation of buildings and ancient contexts. Collectively, the products in this category

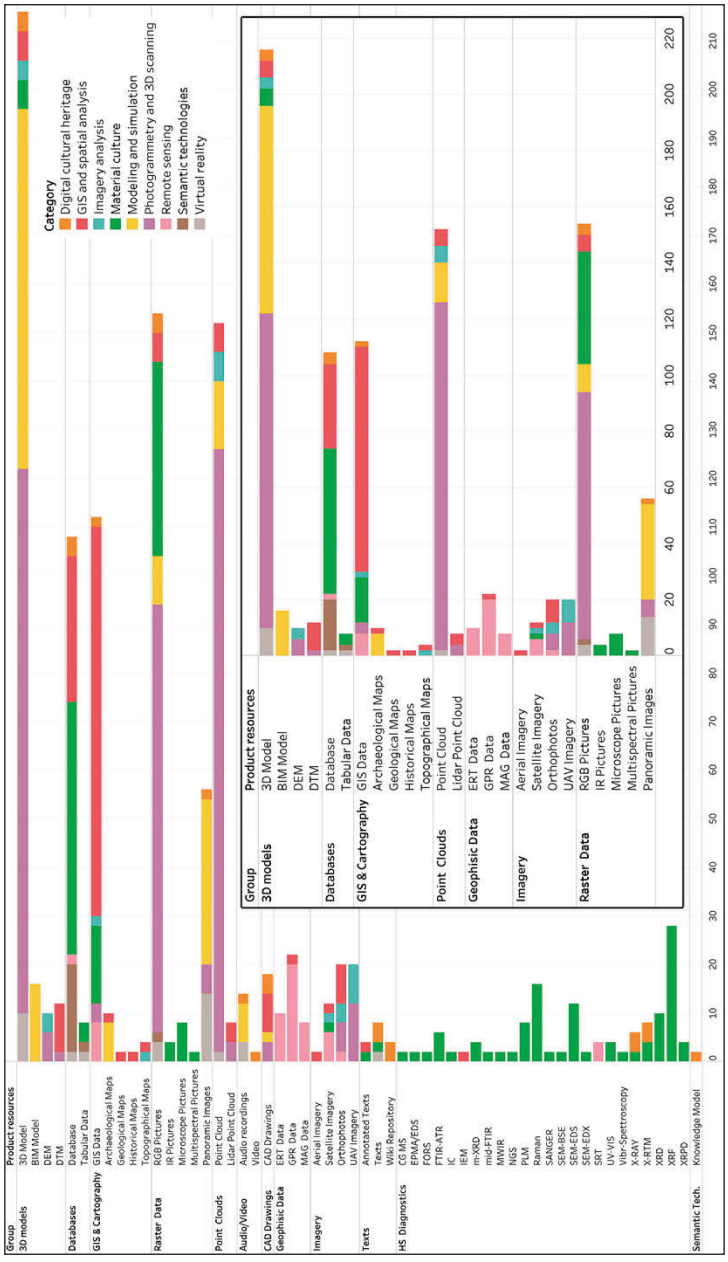


Fig. 5 – Data types recorded in the DHeLO database thematically grouped and color-coded by category. On the right, within the black box, a selection of the most popular for the CH sector.

account for nearly 8% of the overall total. This high percentage is particularly interesting because it quantifies a trend that is clearly observable in the literature sector and is rapidly increasing in terms of data volume (see *infra* § 3.2). It should also be noted that within this trend, marked by the presence of polygonal meshes suitable for rendering, there is also growth in informative models created using HBIM (2%; Mammoli *et al.* 2022), Extended Matrix (3%; Demetrescu, Ferdani 2021), and ArchaeoBIM (0,5%; Garagnani *et al.* 2021) processes.

This may be the most interesting data, indicating a process of increasing autonomy of the 3D model as a container of information and as an independent product. In general, it is difficult to explain the significant growth of 3D data production, especially based on data still being collected. However, unlike the trend observable in GIS datasets, the presence of open-source software does not seem to be a discriminating factor. In fact, only a small percentage (15%) of 3D products are created/processed within a free open-source software environment, while the remaining portion uses proprietary paid tools. In this context, Metashape is particularly notable for its widespread adoption within the scientific community, a dominance that can perhaps be explained by its ease of use and adoption within the archaeological documentation since the early 2010s (e.g. Doneus *et al.* 2011).

At the outset of the landscaping activity, GIS data and databases emerged as significant components, accounting for notable percentages of the data collected. This prominence is further amplified when these categories are linked to related types of products. For GIS specifically, when data related to historical cartography, DEM and DTM, geological maps, satellite imagery, orthophotos, and drone photography are aggregated, the percentage increases substantially (ca. 18%). This trend is not surprising, considering the longstanding application of GIS in archaeological contexts and its efficacy in integrating diverse data sources (Caravale, Moscati 2021, 60-67). In terms of software utilization, two primary trends were identified. The first is the extensive adoption of QGIS as desktop software for creating GIS datasets (92%), frequently supplemented with specialized plugins. Regarding shared data, there is a noticeable preference for tailor-made solutions within web platforms and web apps (80%); standards like WMS and WFS seem to be less common among scholars, even though that could reflect the partial nature of the data collection to date. Similarly, the use of databases, a staple in archaeological computing (Caravale 2022) shows a preference for diverse strategies, geared towards online data dissemination. The lack of a broadly accepted consumer software standard in this area underscores the variety of approaches used in archaeological data management.

G.M.

3. BiDiAr: Bibliography of Digital Archaeology

BiDiAr is a relational database that houses a comprehensive collection of bibliographic items relevant to the field of DA and its relationships with the wider domains of Digital CH and HS. The primary objective of this database is to gather, reference and tag bibliographic entries, ensuring their convenient accessibility through the open-source tool Zotero (Puckett 2011).

3.1 *BiDiAr: a technical description*

As part of the landscaping activities conducted in WP2, it was decided to create a bibliographic corpus focused on digital applications in archaeology and CH. The chosen source was the scholarly journal A&C, which has been a reference point in this disciplinary fields for over 30 years. A&C is data provider of the Open Access Infrastructure for Research in Europe (OpenAIRE) that indexes all the journal's articles, provided with metadata harvested from the OAI repository and with links to the full-text (Rossi, Paraciani 2021). Moreover, CulturaItalia and Europeana expose over 900 resources from the journal's archive. The tool selected for the realization was Zotero, an open-source software designed for accessing, organizing, and referencing bibliographic sources. Zotero facilitates collaboration and synchronization functionalities across devices, promoting teamwork among multiple actors. This method mirrors the approach adopted within the DARIAH research infrastructure, which has incorporated a Digital Humanities bibliography into its portal using the same tool (https://www.zotero.org/groups/744474/dariah/library) along with TaDiRAH taxonomy to categorize bibliographic data (Borek *et. al.* 2021). A similar experience, although primarily focused on English-language literature related to Digital Archaeology, is The Digital Archaeology Collection on ScienceOpen (https://www.scienceopen.com/).

In BiDiAr (https://www.zotero.org/groups/5293298/bidiar/library), all gathered bibliographic citations pertain to two main categories. Firstly, articles published in A&C from 1990 to 2023, including both regular issues and Supplements, totaling over 1258 records, structured according to the DCMI (https://www.archcalc.cnr.it/). Secondly, all bibliographic citations included at the end of each article published in the nine volumes released between 2019 and 2023, amounting to nearly 6500 units of data. While the focus is primarily on recent years, this approach offers a diachronic perspective of the discipline, enabling the observation of its evolution and trends over time (Fig. 8).

In order to effectively manage, query, and utilize this tool, it is essential to organize information according to specific criteria. Therefore, it was decided to initially incorporate thematic classification entries relevant to computer science techniques and archaeological application fields used for the A&C journal's articles since the 1990s (Moscati 1999) into the software's dedicated 'Tag'

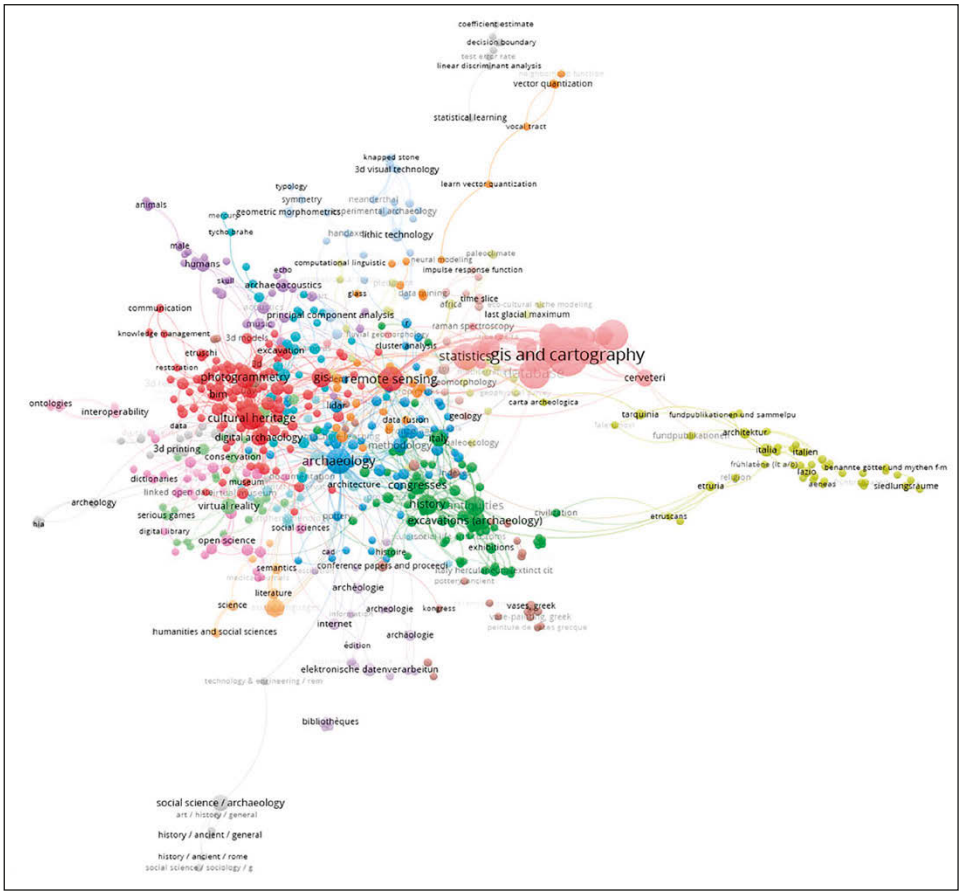


Fig. 6 – Network visualizations of the keywords associated to bibliographic references in BiDiAr. Worth noticing 4 main clusters to explore with further analysis: GIS and cartography (pink), archaeology (blue), photogrammetry (red), excavations (green).

feature[1]. This classification not only serves as a guide for searching topics of interest and linking them with HS needs and purposes but also aids in understanding their trends over time, thereby monitoring tradition, innovation,

[1] Regarding computer science, these categories include: Computer Graphics IP CAD, Data encoding and metadata, Database, GIS and cartography, History of applications and research projects, Multimedia and web tools, Remote Sensing, Simulation AI, Statistics, while for archaeological classification: Archaeometry, Classification of archaeological finds, Cultural Resource Management, Data dissemination and education, Documentation, conservation and restoration, Epigraphy and Numismatics, Geoarchaeology, Survey and excavations, Theoretical and methodological problems.

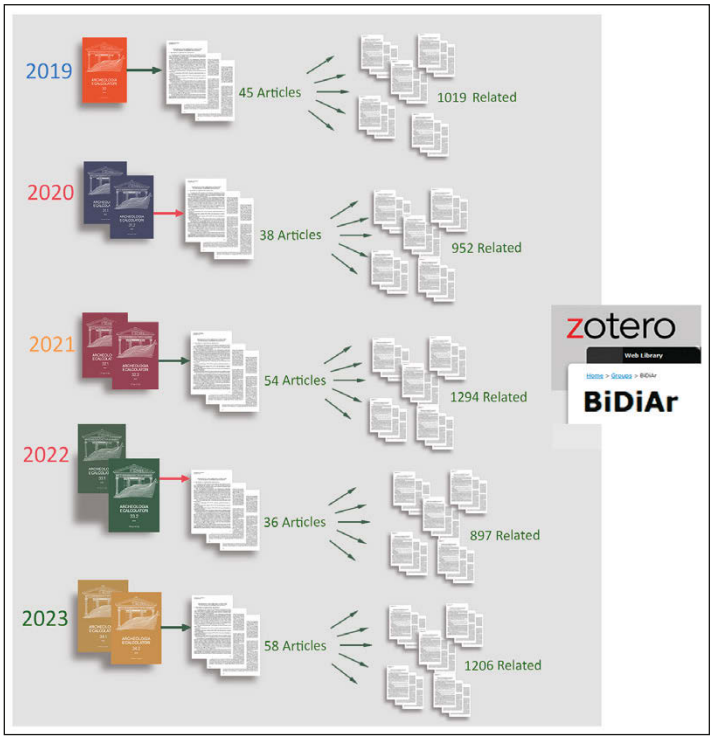


Fig. 7 – Graph illustrating the relationships between A&C papers and their correlated records within BiDiAr.

and technology development for the study of antiquity and cultural heritage more broadly. Furthermore, we leveraged another feature of the Zotero software, which is its ability to establish correlations between texts in the virtual library, to structure and associate the inserted information. This enables an interconnected view of the topics discussed and facilitates the evaluation of which texts are of greater reference and impact within the E-RIHS research community context through network analysis tools. Each A&C article was thus correlated with the bibliographic references cited at the end of the text, creating a network of knowledge that fosters critical analysis and understanding of research dynamics (Fig. 6).

3.2 *First results and statistical analysis*

The period from 2019 to 2023 was selected to identify recent trends in the use of computer tools and to understand progress through increasingly specific statistics. The focus on DA also seeks to gain new insights to update

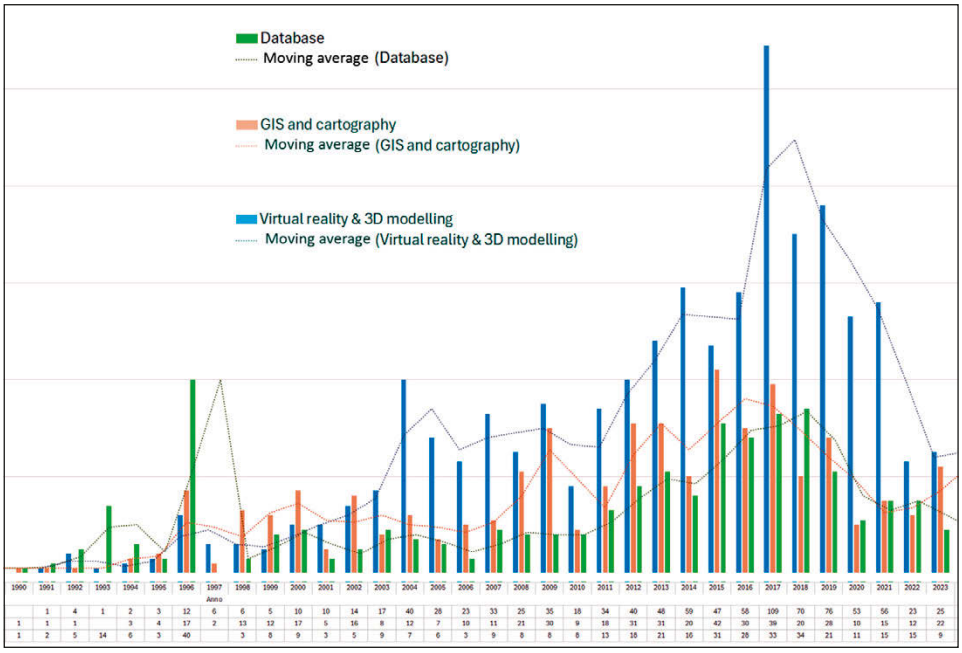


Fig. 8 – Graph illustrating the distribution over time of the number of publications concerning: Virtual reality and 3D modelling, GIS and cartography and Database.

the scholarly classification established over twenty years ago, which has evolved with the discipline (Cantone, Caravale 2019, and, lastly, Caravale *et al.* 2023). In the last five years, nine issues were published (n. 30, 31.1-31.2, 32.1-32.2, 33.1-33.2, 34.1-34.2) for a total amount of 231 articles and 5638 correlated bibliographic references (Fig. 7). In addition to the yearly submissions by authors providing an up-to-date overview of the most significant application sectors in DA, these nine issues include four special thematic issues and six international conference proceedings (https://www.archcalc.cnr.it/pages/special_issues.php) that contribute to elucidating how the topics discussed well align with the HS domain. Issue 30, commemorating the journal's 30th anniversary, presents a comprehensive view of archaeological computing development and highlights the research activity defining its editorial journey. Volume 32.1 features a special insert dedicated to the pioneering field of archaeomusicology, a multidisciplinary research area that applies archaeological methods to study music and musical life in the ancient world. Volume 32.2's special issue, titled 'From Pottery to Context: Archaeology and Virtual Modeling', showcases the outcomes of a research project that focuses on merging funerary archaeology and DA.

In the same five-year period, the Proceedings of the 12th and 16th editions of the ArcheoFOSS Workshop (Free, Libre and Open Source Software and Open Format in Archaeological Research Processes) were published in volumes 30 and 34.1 respectively, while the Proceedings of the conference 'Milan International: the Territorial Fragility of Archaeological Contexts' were published in volume 31.2. Furthermore, two special sessions of the Florence and Trento editions of the IMEKO TC-4 International Conference on Metrology for Archaeology and Cultural Heritage (MetroArchaeo) were published, one dedicated to 'Logic and Computing: The Underlying Basis of Digital Archaeology' (volume 31.2) and the second to 'Archaeological Computing: Selected Papers from the 2020 IMEKO TC-4 MetroArchaeo International Conference' (volume 32.2). Finally, volume 34.1 contains the Proceedings of a special session of the 7th Landscape Archaeology Conference (Iasi, Bulgaria, 10-15 September 2022), dedicated to 'Modeling the Landscape: From Prediction to Postdiction'.

Starting from this extensive database, still under implementation, the analysis directed toward the investigation of thematic trends over time. This first line of inquiry aligns well with already established reflections in the archaeological literature, often carried out on the A&C journal's data (Caravale, Moscati 2021). The innovative aspect here was not to confine the analysis on a single journal, considering the entire network of its citations. In this case, the most challenging aspect was to uniform the tags and the classifications of all the records. Meaningful keywords per topic were uniformly searched in the titles, among the author's tags and in the available abstracts, with the intent to isolate significant notions, lately counted over year. This process helped to widen the scope of the A&C journal classification, including secondary themes in addition to the primary categorization; in this initial step this process was carried out for three themes: Virtual reality and 3D modeling, GIS and cartography, and Database (Fig. 8).

The thematic organization and annual breakdown of the data infer some general observations, certainly refinable with the prosecution of the data entry process. Observing the data from the 1990s to the 2000s, it is evident that these three technologies have remained fundamentally represented in almost equal measure, with minimal fluctuations observed year by year. It is also unsurprising that there are no mentions of virtual reality technologies before 1991, the year of publication of the article *Toward a virtual archaeology* (Reilly 1991), which first opened the debate on the use of these technologies in archaeology[2]. The growing attention to virtual archaeology recorded in the scientific debate from 2000 to 2010 (e.g. Barcelò *et al.* 2000) emerges in a steady and considerable increase in articles addressing this topic, clearly

[2] Regarding the concentration of articles on the use of GIS in the biennium 1996-1997, this grouping has already been highlighted and explained in Caravale, Moscati 2021, 67-75.

visible within our data. Concurrently, contributions related to the use of GIS and databases, although steadily increasing (a sign of their ever greater and widespread adoption), seem to run almost parallel. This outcome is not surprising, considering the close relationship between these technologies. It is equally unsurprising that it is possible to count more contributions related to databases, a predominance explained by the presence of digital archives linked to the cataloging and classification of artifacts without a specific spatial analysis. From 2010 to today, this trend seems to continue. The exponential increase in products related to virtual reality and 3D modeling, already seen in the last five years on the data collected in DHeLO (see *supra* § 2.3), can perhaps be explained by two factors. Firstly, the epistemological definition of virtual archaeology following the London and Seville charters (2009, 2012), with the consequent increase in contributions related to products and research projects developed in these areas. Secondly, the increasingly disruptive use of 3D photogrammetric modeling, whose description and application have occupied a significant space in the literature dedicated to archaeological computing.

Further analysis will be available over time with the prosecution of the data entry process as well as the keyword normalization procedures, to map minor trends over time. Additionally, this process will be integrated with network analysis tools, to evaluate reference publications through their citation weight, an important tool for new scholars to identify key publications in a fast-growing literature. An additional future development could be to link bibliographic data to geospatial information, by geotagging sites and ancient regions described within the literature with the objective to highlight, enhance and quantify thematic relations among specific places through network analysis. In this context, the choice to use Zotero as the data container for BiDiAr proves to be successful, as it allows the integration of tools and plugins developed by the community for this specific purpose such as Zotero2Map (Iacopini in press).

G.M., A.D.

4. New frontiers of archaeological information systems: GIS, databases and 3D models integration to document the multidimensional nature of archaeological data

From the early analysis of DHeLO and BiDiAr data, it becomes evident that the intrinsic complexity of archaeological data demands a variety of digital solutions for its visualization and interpolation. This requirement underscores the need for advanced digital tools that can effectively manage and represent the complex layers of data, ensuring that users can explore the physical spatial and temporal contexts of the findings.

Particularly notable is the increasing choice to represent the multidimensionality of the data using 3D elements combined with various degrees

of metadata that facilitate navigation through its dimensions (spatial and chronological) and properties (authors, actions, etc.). Data emerging from the landscaping process confirm that three product categories have become pivotal in archaeological documentation over the last five years: GIS, databases, and 3D models, both for documentary and management purposes (survey, HBIM) and reconstructive aims (polymeshes, EM, ArchaeoBIM). 3D models represent a significant innovation over the past decade compared to the use of GIS and databases, slightly older more established tools with a longer history of use (Caravale, Moscati 2021; Moscati 2021). It is interesting to note that the informative component is also growing in relation to 3D models. Playing a particularly key role here are the capabilities to construct BIM of archaeological (ArchaeoBIM) or heritage context (HBIM) or to enrich polygonal models semantically. Sector literature also indicates a cautious move towards combining these three components, creating GIS enriched with 3D and tabular data (Dell'Unto, Landeschi 2022). This trend towards integration reflects a broader effort to leverage diverse data types and modeling techniques to enhance the depth and usability of archaeological research from data collection to information visualization and virtual recreation (Zoni 2017; Pansini 2022; Poggi *et al.* 2022; Mancuso 2024). This holistic approach not only optimizes data utility but also paves the way for more nuanced analyses and interpretations within the field, combining multiple data types within a single complex environment. The integration of 3D GIS with informational models could effectively lead toward the creation of holistic data aggregator, offering a unique tool to represent and analyze complex data across multiple dimensions, thus making it an essential instrument for archaeological studies, from field data collection to virtual recreation.

G.M., A.D.

5. Conclusion

Based on the data presented above, it seems palpable that tools like DHeLO and BiDiAr, when used as part of a cohesive information synergy, can facilitate navigation and enhance the understanding of developments in digital technologies applied to CH/HS research areas. The future implementation of these tools, with data covering longer time spans and increasingly broad thematic interdisciplinary coverage, will lead to a more comprehensive and coherent overview of topics, themes, and tools interconnected with these disciplines. With the aim of rendering data coalition feasible, these databases will ensure the continued monitoring of resources and the increase in visibility, usability, and longevity of the data and their source material.

Tito Orlandi masterfully expressed the importance of this analytical course thirty years ago when he stated that classification processes and

bibliographic tools are the blood lymph of each discipline. Significantly, in the introduction to the book *Bibliografia di Informatica Umanistica* he stated that a discipline could be considered mature – meaning easily recognizable and active – when it is possible to conceive and implement a bibliographic tool that demonstrates the interconnection between its various sectors and levels, and a certain number of common perspectives to which its scholars are committed, beyond the differing views they propose in their contributions (Orlandi 1994). In this regard, the effort behind the construction of these two related systems in the H2IOSC environment is to connect research products to this vast bibliographic knowledge network, considering them both as primary sources for the arguments and information produced and presented in the literature, and as independent and autonomous knowledge outputs.

G.M., A.D.

Giacomo Mancuso, Antonio D'Eredità
Istituto di Scienze del Patrimonio Culturale - CNR
giacomo.mancuso@cnr.it, antonio.deredita@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 authors wish to express their gratitude to Paola Moscati and Alessandra Caravale for their insightful comments and and to the senior developer Matteo Gallo for the technical support in the development of the DHeLO web app.

REFERENCES

Assante M., Candela L., Castelli D., Cirillo R., Coro G., Frosini L., Lelii L., Mangiacrapa F., Pagano P., Panichi G., Sinibaldi F. 2019, *Enacting open science by D4Science*, «Future Generation Computer Systems», 101, 555-563 (https://doi.org/10.1016/j.future.2019.05.063).

Barceló J.A., Forte M., Sanders D.H. (eds.) 2000, *Virtual Reality in Archaeology*, BAR International Series 843, Oxford, Archaeopress.

Ben Hamida K., Di Giorgio S., Buonazia I., Masci E., Marlitti D. 2009, *CulturaItalia: aspetti tecnico-scientifici,* «DigItalia. Rivista del digitale nei beni culturali», 4, 1, 83-102.

Borek L., Hastik C., Khramova V., Illmayer K., Geiger J.D. 2021, *Information organization and access in Digital Humanities: TaDiRAH revised, formalized and FAIR*, in C. Wolff, T. Schmidt (eds.), *Information between Data and Knowledge. Information Science and its Neighbors from Data Science to Digital Humanities. Proceedings of the 16th International Symposium of Information Science (ISI 2021)*, Glückstadt, Verlag Werner Hülsbusch, 321-332.

Bucciero A., Chirivì A., Fanini B., Massida M., Pescarin S., Taurino F. 2022, *Soluzioni integrate web-based per il Patrimonio Culturale: una prospettiva*, in C. Volpe, M. Mieli (eds.), *Conferenza GARR 2022 - Condivisioni - La rete come strumento per costruire il futuro (Palermo 2022)*, Viterbo, Associazione Consortium GARR, 94-102.

Calandra E., Acconcia V., Boi V., Falcone A. 2021, *Il Geoportale Nazionale per l'Archeologia. Uno strumento per la tutela, la divulgazione e la fruizione del Patrimonio archeologico italiano*, in F. Cambi, D. Mastroianni, V. Nizzo, F. Pignataro, S. Sanchirico (eds.), *Landscapes. Paesaggi culturali, Atti della Giornata di Studi (Roma 2019)*, Roma, Fondazione Dià Cultura, 353-361.

Caliri C., Romano F.P., Doherty B., Rosi F., Miliani C. 2020, *MOLAB: A European mobile laboratory enabling advanced studies in archaeometry*, «Archeometriai Műhely», 17, 3, 46-49.

Cantone F., Caravale A. 2019, *Archeologia e Calcolatori. Classificazione geografica e tematica per la condivisione della conoscenza,* «Archeologia e Calcolatori», 30, 93-107 (https://doi.org/10.19282/ac.30.2019.07).

Caravale A. 2022, *Le banche dati archeologiche. Articolazione e formalizzazione delle conoscenze*, Firenze, All'Insegna del Giglio.

Caravale 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., Moscati P. 2021, *La bibliografia di informatica archeologica nella cultura digitale degli anni Novanta*, Firenze, All'Insegna del Giglio.

Dell'Unto N., Landeschi G. 2022, *Archaeological 3D GIS*, New York, Routledge.

Demetrescu E., Ferdani D. 2021, *From field archaeology to virtual reconstruction: A five steps method using the Extended Matrix,* «Applied Sciences», 11, 5206 (https://doi.org/10.3390/app11115206).

D'Eredità A. 2020, *Il rilievo fotogrammetrico di Doclea*, «Archeologia e Calcolatori», 31.1, 213-230 (https://doi.org/10.19282/ac.31.1.2020.10).

Doneus M., Verhoeven G.J., Fera M., Briese C., Kucera M., Neubauer W. 2011, *From deposit to point cloud. A study of low-cost computer vision approaches for the straightforward documentation of archaeological excavations*, «Geoinformatics FCE CTU», 6, 81-88 (https://doi.org/10.14311/gi.6.11).

Garagnani S., Gaucci A., Moscati P., Gaiani M. 2021, *ArchaeoBIM. Theory, Processes and Digital Methodologies for the Lost Heritage*, Bologna, Bononia University Press.

Iacopini E. in press, *Zotero2map: sviluppo di un nuovo strumento ITC per la condivisione e la pubblicazione dei dati bibliografici per la ricerca storica e archeologica*, «Archeologia e Calcolatori», 35.2.

Kansa E. 2022, *On infrastructure, accountability, and governance in digital archaeology*, in K. Garstky (ed.), *Critical Archaeology in the Digital Age. Proceedings of the 12th IEMA Visiting Scholar's Conference (Buffalo, NY 2019)*, Los Angeles, UCLA Cotsen Institute of Archaeology Press, 141-152.

Luzietti R.B., Spadi A., Giampietro N., Mancuso G., Caravale A., D'Eredità A., Caradonna M., Moscati P., Quochi V., Monachini M., Degl'Innocenti E. in press, *Digital Humanities and Heritage Science: Moving from landscaping to a dynamic research observatory in an Open Science Cloud*, Poster presented at *MeTe digitali. Mediterraneo in rete tra testi e contesti (Catania 2024)*.

Mancuso G. 2024, *ArchaeoBIM ed Extended Matrix. Analisi e potenzialità di due processi per l'elaborazione di modelli informativi,* «Archeologia e Calcolatori», 34.2, 123-142 (https://doi.org/10.19282/ac.34.2.2023.07).

McManamon F.P., Kintigh K.W., Ellison L.A., Brin A. 2017, *tDAR: A Cultural Heritage archive for twenty-first-century public outreach, research, and resource management*, «Advances in Archaeological Practice», 5, 3, 238-249 (https://doi.org/10.1017/aap.2017.18).

Moscati P. 1999, *"Archeologia e Calcolatori": dieci anni di contributi all'informatica archeologica,* «Archeologia e Calcolatori», 10, 343-352 (https://www.archcalc.cnr.it/journal/id.php?id=284).

Moscati P. 2021, *Digital Archaeology: From interdisciplinarity to the 'fusion' of core competences,* «magazén. International Journal for Digital and Public Humanities», 2, 2, 253-274 (http://doi.org/10.30687/mag/2724-3923/2021/04/004).

Orlandi T. 1994, *Prefazione*, in G. Adamo (ed.), *Bibliografia di Informatica Umanistica*, Roma, Bulzoni, vi-vii.

Pallot-Frossard I. 2016, *From CHARISMA to IPERION CH and E-RIHS: Towards a European research infrastructure for Heritage Science*, «Technè. La science au service de l'histoire de l'art et de la préservation des biens culturels», 43, 46-49 (https://doi.org/10.4000/techne.641).

Pansini R. 2022, *Rilievo, documentazione e ricostruzione di un centro monumentale con fotogrammetria e modellazione tridimensionali: il caso della città romana di Sala*, in A. Arrighetti, R. Pansini (eds.), *Sistemi e tecniche di documentazione, gestione e valorizzazione dell'architettura storica. Alcune recenti esperienze*, «Archeologia e Calcolatori», 33.1, 239-256 (https://doi.org/10.19282/ac.33.1.2022.13).

Poggi G., Falchi F., Russo L., Buono M. 2022, *From surfaces to volume: Towards a volumetric reconstruction of the archaeological deposit*, «Archeologia e Calcolatori», 33.2, 197-214 (https://doi.org/10.19282/ac.33.2.2022.11).

Poulopoulos V., Wallace M. 2022, *Digital technologies and the role of data in Cultural Heritage: The past, the present, and the future*, «Big Data and Cognitive Computing», 6, 3, 73-92 (https://doi.org/10.3390/bdcc6030073).

Puckett J. 2011, *Zotero: A Guide for Librarians, Researchers, and Educators*, Chicago, Association of College and Research Libraries.

Reilly P. 1991, *Towards a virtual archaeology*, in S. Rahts, K. Lockyear (eds.), *CAA90. Computer Applications and Quantitative Methods in Archaeology 1990*, Oxford, Archaeopress, 133-139.

Richards J., Niccolucci F. (eds.) 2019, *The Ariadne Impact*, Budapest, Prime Rate Kft.

Rossi I., Paraciani N. 2021, *IT applications to archaeology and the OA diamond journals' challenge. Enhancing access and reuse of textual and visual resources*, «Archeologia e Calcolatori», 32.1, 325-347 (https://doi/org/10.19282/ac.32.1.2021.18).

Silva A.L., Terra A.L. 2024, *Cultural heritage on the Semantic Web: The Europeana data model,* «IFLA Journal», 50, 1, 93-107 (https://doi.org/10.1177/03400352231202506).

Taylor J. 2023, *Digital infrastructures and their impact on data acquisition*, «Current Swedish Archaeology», 31, 53-58 (https://doi.org/10.37718/CSA.2023.05).

Wilkinson M.D., Dumontier M., Aalbersberg I.J. *et. al.* 2016, *The FAIR guiding principles for scientific data management and stewardship*, «Scientific Data», 3, 1 (https://doi.org/10.1038/sdata.2016.18).

Zoni F. 2017, *L'uso della tecnologia DEM nella documentazione archeologica. Alcune applicazioni in casi di scavo stratigrafico e nello studio dell'edilizia storica,* «Archeologia e Calcolatori», 28.1, 219-238 (https://doi.org/10.19282/ac.28.1.2017.13).

ABSTRACT

This paper explores the initial outcomes of the H2IOSC Project, specifically within Work Package 2 (WP2 - Landscaping & Building Communities), which aims to survey the Italian digital landscape in Language Technologies, Humanities, and Heritage Science (HS). A significant outcome of the efforts of the Rome branch of CNR-ISPC is the development of two key resources: the DHeLO web app and the BiDiAr bibliographic collection. DHeLO (Digital Heritage Landscaping Platform) is designed to collect, store, and query metadata of

research projects, products, and digital tools in Cultural Heritage (CH) and Heritage Science (HS). It aims to create a comprehensive disciplinary observatory by integrating data from multiple sources into a structured system that allows for complex queries and data indexing. This platform supports the FAIR principles (Findability, Accessibility, Interoperability, and Reusability) and includes metadata standards based on the Dublin Core Metadata Initiative (DCMI). BiDiAr (Bibliography of Digital Archaeology) functions as a relational database within Zotero, an open-source bibliographic tool. It compiles bibliographic entries relevant to Digital Archaeology, integrating themes and research outputs from the «Archeologia e Calcolatori» journal. This database aids in thematic trend analysis and network analysis by linking bibliographic citations, enhancing the understanding of research dynamics and impacts within the E-RIHS community. Analyzing these resources reveals an exponential increase in virtual reality and 3D modeling products, driven by epistemological developments and the disruptive use of photogrammetric modeling These tools not only enhance data accessibility and usability but also support interdisciplinary collaboration and innovation in digital heritage and archaeology.

Archeologia e Calcolatori
35.1, 2024, 543-562
doi 10.19282/ac.35.1.2024.32

A RESOURCE HUB FOR INTEROPERABILITY AND DATA INTEGRATION IN HERITAGE RESEARCH: THE H-SETIS DATABASE*

1. The Humanities and Cultural Heritage Italian Open Science Cloud (H2IOSC) Project

H2IOSC is a project led by the Consiglio Nazionale delle Ricerche (CNR), actively involving several of its Institutes (https://www.h2iosc.cnr.it/). Its main objective is to create a federated and inclusive cluster of the Italian nodes of the four European Research Infrastructures (RIs) in the field of Humanities and Cultural Heritage: DARIAH for Humanities, CLARIN for language sciences, OPERAS for scientific communication in the field of Humanities and Social Sciences, and E-RIHS for Heritage Science (HS). Their nature is very heterogeneous, including both physical instrumentation and repositories such as archives and databases, computing and communication systems that are essential for research purposes. The entire H2IOSC Project activity aims to support data-driven research and the digital transformation of the cultural and creative industries sectors. H2IOSC indeed promotes a data-centric approach, with data made accessible through an integrated digital environment designed according to FAIR principles (Findable, Accessible, Interoperable, Reusable; Wilkinson *et al.* 2016).

The Istituto di Scienze del Patrimonio Culturale (CNR-ISPC) is directly involved in the project as part of the E-RIHS network; the Milan branch of ISPC, as leader of Task 4.10 'Resources interoperability: DIGILAB resources (E-RIHS)' within WP4 'RIs Nodes and Resources Interoperability', is in charge of a general survey of semantic tools for the Heritage sector, of designing strategies to assure data interoperability, and of the integration of digital resources within the Heritage domain and the wider H2IOSC common semantic framework.

2. Heritage Science and semantic tools

2.1 *Heritage Science: a long-standing and yet novel discipline*

Heritage Science (HS) is an interdisciplinary research field that combines social sciences and natural sciences applied to cultural and natural Heritage. It is a research area whose scope and objectives are well known, but whose formalization as an autonomous sector has not yet been completed. This

* Both authors have equally contributed to the content of this paper.

definition of HS is quite recent, as it was jointly developed in 2019 by E-RIHS and the International Centre for the Study of the Preservation and Restoration of Cultural Property (ICCROM)[1], while the term itself was defined by the Science and Technology Select Committee of the British House of Lords and dates back to 2006 (House of Lords Science and Technology Select Committee 2006; Kennedy 2015, 214-215; Strlič 2018, 7260; Kennedy *et al.* 2024). The new definition aims to overcome disciplinary barriers, especially between the fields of Cultural Heritage Conservation, which has always focused on the technical aspects of assessing and controlling degradation, restoration, and protection of tangible Cultural Heritage, and that of social sciences, which focuses on the study of material evidence and the relationship between humans and the environment from a historical perspective, such as archaeology, art history, and anthropology, among others (Carman, Sørensen 2009).

In recent decades, the increasing use of scientific techniques involving remote data acquisition has enhanced the analytical potential and application of HS diagnostic methodologies. This is because the fragility of the materials under study often discourages or prohibits more invasive interventions and direct sampling. The chance to acquire data without physically contacting the object of study has expanded the number of measurements and the fields of investigation, increasing also scientific results (Kennedy 2015, 220-221).

The identification of the HS domain aims to shift the focus from individual disciplines, each with its specificities, to the object of research, namely 'Heritage'. By changing the perspective, there is greater integration among researchers and contextual improvements compared to individual research projects, thanks to the contribution of multiple expertise as well as a coordinated approach. The need for significant shared data repositories and appropriate management tools has been emphasized by several authors (Kennedy 2015, 224-225; Bordalo, Bottaini, Candeias 2020; Castelli, Felicetti, Proietti 2021). Although this is a common requirement for the entire scientific community, it plays an even more significant role in HS due to the complex articulation of possible fields of investigation and its multidisciplinary nature.

2.2 *Assessing the use of semantic technologies for the Heritage field*

Building on the emerging focus on collaboration and overcoming disciplinary divides within the HS domain, the surge in specialized software and data management tools becomes even more crucial. By leveraging semantic technologies, these tools can bridge the gap between disciplines and facilitate the contribution of multiple areas of expertise to a common object of research.

[1] https://www.e-rihs.eu/e-rihs-in-a-nutshell/.

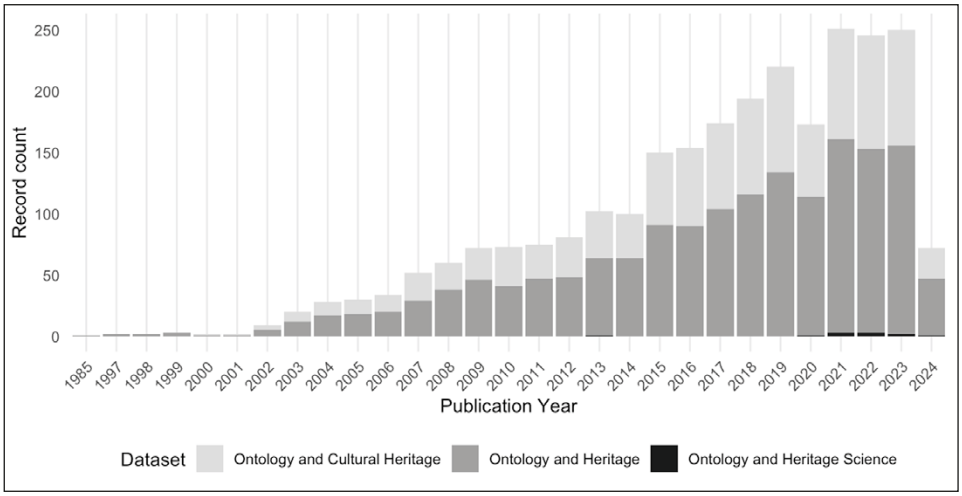


Fig. 1 – Annual distribution of research products (articles, conference proceedings, book chapters) related to the combination of keywords indicated in the legend (data source: Scopus).

However, as the field embraces these advancements, ensuring accessibility and reusability of the data and tools themselves becomes paramount.

With this regard, an initial analysis of relevant literature revealed a rapidly changing landscape (Fig. 1)[2]. Since 2010, controlled vocabularies, taxonomies, ontologies, and specialized software designed for this field became popular, and the use of semantic technologies improved the integration of interoperability principles. Within the domain of semantic tools, which extends beyond Heritage practitioners, a consensus is emerging regarding the necessity for such tools to comply with accessibility and reusability standards, as many resources, particularly ontologies and vocabularies, still fall short of these principles. The recently published report within the FAIR-IMPACT project (Le Franc *et al*. 2020, 11-12), which is part of the initiatives undertaken within the framework of the implementation plan of the European Open Science Cloud (EOSC), highlights the main issues that characterize, for example, the development of ontologies (Garijo, Poveda-Villalón 2020). Often, they lack proper documentation, version control, and are not published and maintained following Linked Data (https://www.w3.org/DesignIssues/LinkedData.html) and FAIR data principles.

[2] Data for all charts and graphs included in the article can be found on Zenodo (https://doi.org/10.5281/zenodo.11388725).

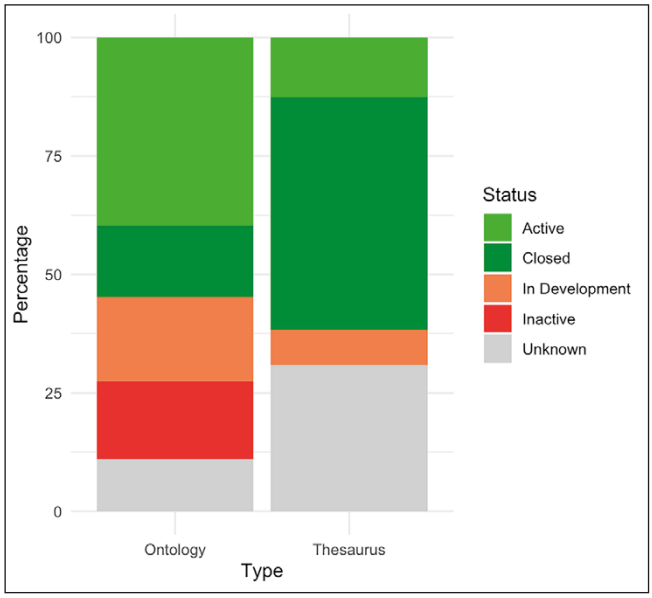


Fig. 2 – Percentage distribution of semantic artefacts cataloged in H-SeTIS according to their curation status (i.e., completeness and maintenance).

More concerningly, many resources lack essential metadata regarding their creation, purpose, usage, and maintenance. This absence of information significantly hinders both understanding and reusability. Garijo, Poveda-Villalón (2020, 7) recommend including 23 metadata elements within an ontology's documentation, with 12 of these being optional. Notably, a significant portion of these belong to the Dublin Core standard, including dc:title, dc:author, dc:contributor, and dc:description. Unfortunately, most resources fail to provide metadata related to creation and modification dates, namespaces, and bibliographic references. Versioning, a fundamental tool in digital tool development, allows tracking a resource's evolution over time and identifying stable versions. However, versioning information is often missing from most of the resources' URIs and, when present, it rarely adheres to semantic versioning principles (https://semver.org/).

The Heritage sector mirrors this broader trend: in-depth documentation exceeding essential metadata remains scarce for the ontologies and the tools cataloged thus far. Most only present traditional documentation: the scientific development process for a semantic resource and the final product are often published in academic journals or conference proceedings. However, the lack of proper documentation and the inability to locate them using

unique identifiers (URIs) significantly hinders or even impedes their reuse. It is worth noting that among the roughly 200 semantic tools cataloged so far, a significant portion have a 'negative' status, indicating incompleteness, lack of updates, or irretrievability (Fig. 2).

Recognizing the critical need for well-documented and accessible tools, the H-SeTIS database (Heritage - Semantic Tools and Interoperability Survey, see below §3) aims to create a collection of existing semantic and interoperable resources for the Heritage domain and will serve as an up-to-date toolkit for developing similar tools. Such overview plays a crucial role in creating a knowledge model specifically designed to integrate E-RIHS data into the H2IOSC semantic framework, thereby promoting interoperability and data integration within the Heritage domain.

This state-of-the-art review informed key decisions about the H-SeTIS database design and data entry process. The adoptions of platforms like the Linked Open Vocabularies (LOV) catalog and OntoPortal software are hampered by the limitations of the data itself. LOV (Vandenbussche *et al.* 2017) offers robust functionalities for visualizing semantic vocabularies. OntoPortal (Jonquet *et al.* 2023) focuses on creating comprehensive catalogs of diverse semantic resources. Both leverage semantic technologies to automatically generate metrics and statistics. However, the fragmented nature of information associated with many cataloged resources in H-SeTIS hinders their effectiveness in this context.

In addition, H-SeTIS goes beyond simply cataloging existing semantic tools in the Heritage sector. It also aims to capture information and data about the scientific process behind their creation, even if these do not meet the minimum quality requirements of accessibility and reusability previously mentioned. This additional objective provides valuable insights into the development process of these tools, even if they may not be fully functional or well-maintained.

2.3 *Geospatial analysis of institutions involved in the research of semantic tools for Heritage*

As a preliminary overview of the international scientific community involved in the development of semantic tools within the field of Heritage, a geospatial analysis focused on the involved institutions was carried out using available repositories.

The initial bibliographic dataset was obtained from Web of Science (WoS), a repository selected for the completeness of the raw data made available to the user. The query compiled in the 'Topic' field, which includes titles, abstracts, and author keywords, is as follows: TS=(heritage) AND (TS=(ontolog*) OR TS=("semantic web")); the term 'heritage' was selected to maximize results related to Heritage (Huang 2024): the English term stresses the importance

of inheritance passed down through generations, encompassing both Cultural and Natural Heritage. The query therefore returned bibliographic records in which the term 'heritage' was associated with 'ontolog*' (which includes both 'ontology' and 'ontologies') or alternatively with 'semantic web' in titles, abstracts, or keywords. Associating these terms was necessary to disambiguate queries that could have been misleading. For example, the term 'ontology' alone returns hundreds of results dealing with ontologies from a strictly philosophical point of view or, conversely, with computer ontologies applied to any possible theme. The association with the terms 'heritage' and 'semantic web' allowed a more relevant dataset to be obtained, increasing the focus on the Cultural Heritage domain – the main focus of the H2IOSC Project – and retrieving those contributions that also deal with semantic technologies for Heritage, resulting in a final dataset of 1248 bibliographic records[3].

The obtained dataset was used to carry out a quantitative network analysis focused on the institutions involved in the research on ontologies and semantic tools for the Heritage. This dataset was initially preprocessed using VOSviewer software, specifically designed for network analysis and clustering of bibliometric data (van Eck, Waltman 2010). Institutions were selected based on authors' affiliations, with the requirement that they had relations with each other and appeared at least twice to refine the results further. The institutions thus filtered count 128. VOSviewer automatically processed also the related network of relations, connecting institutions based on their appearance in scientific contributions with authors from different affiliations.

To conduct more in-depth analyses, the output data from VOSviewer were subsequently post-processed using Gephi, R, and QGIS software. In Gephi, the network of relations was regenerated, refining it by combining institutions that were considered separate by VOSviewer due to discrepancies in the input data; names were then normalized to overcome any residual ambiguities. With the tidygeocoder package in R (Cambon *et al.* 2021), it was possible to automatically retrieve the geographical coordinates of individual institutions, except for a few cases where manual intervention was necessary. The obtained data were then loaded into a GIS environment, where the initially generated network from VOSviewer was georeferenced, allowing for further advanced analyses (Fig. 3).

[3] The access to large bibliographic repositories is of great assistance to research, with the awareness, however, that the provided data are not always complete or entirely accurate. A comparison of results from various queries has highlighted how difficult it is to extract completely reliable datasets, except perhaps in extremely delimited disciplinary sectors, and how the search system itself likely applies optimizations on results that are hardly controllable. The query used here is the one that has proved to be more flexible compared to others that are more restrictive and accurate, albeit at the cost of omitting a large amount of data. However, it is possible that not all resulting institutions are actually engaged in the field of ontologies and semantic web for Cultural Heritage.

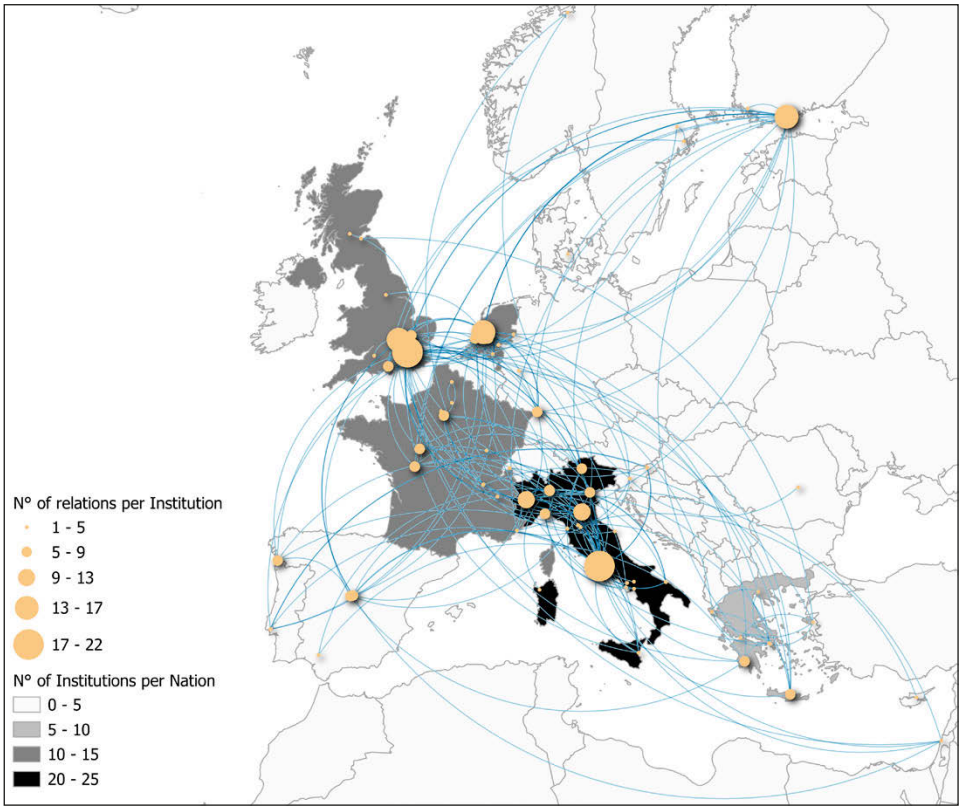


Fig. 3 – Spatial distribution and relations of the institutions involved in the research of ontologies and semantic web for Heritage; the size of nodes is proportional to the number of relations.

From the observation of the obtained result, several preliminary considerations can be drawn. The majority of research institutions involved in the development of ontologies for Cultural Heritage are located in Europe (about 84%)[4], while among the remaining countries, institutions from the United States, Australia, Israel, Qatar, Pakistan, China, Hong Kong, Taiwan, South Korea, and Vietnam are represented. Among European countries, the most involved one in research activities, as appreciable by the number of its involved institutions, is certainly Italy, with almost double the number of entities compared to the Netherlands, which is second, with France, the

[4] In this case, when referring to Europe, the continent is meant rather than the political Union, thus including countries such as the United Kingdom, Switzerland, Norway, and the Mediterranean basin with Israel.

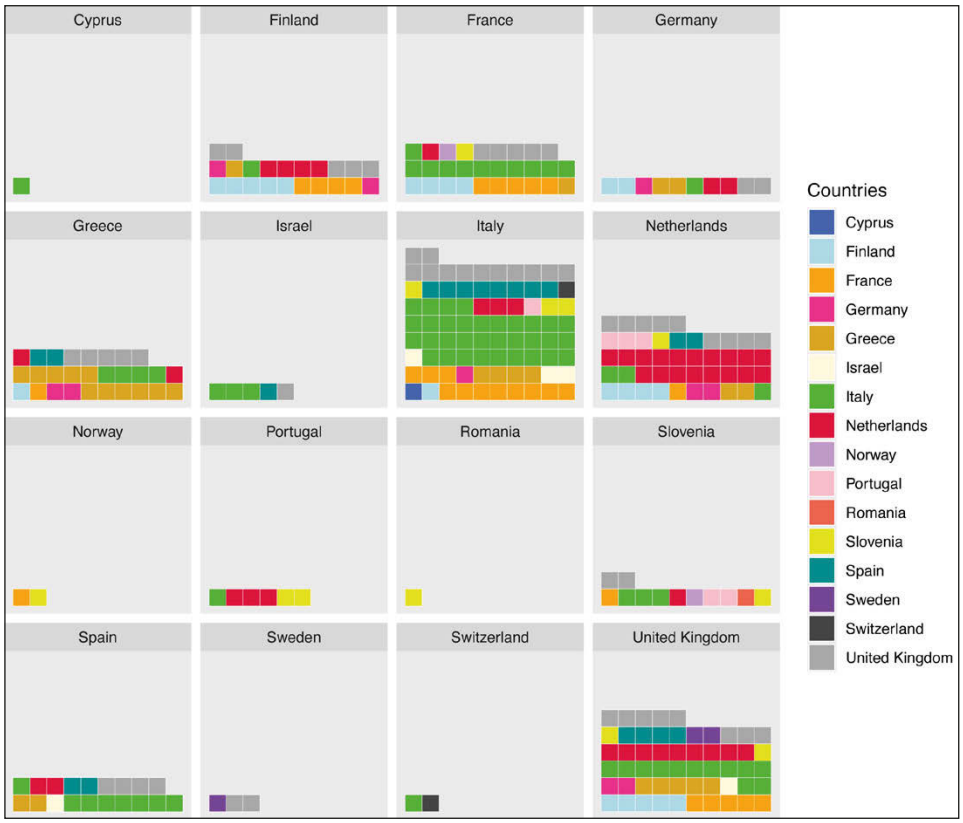


Fig. 4 – Waffle chart displaying national and international collaborations among the institutions mapped in Fig. 3.

United Kingdom, and Greece following. Interconnections between institutions from the same country, especially among those from Italy, the Netherlands, and the United Kingdom, are quite evident. In terms of relationships between individual institutions, the highest number of collaborations is achieved by the CNR, followed by the University College London, the Vrije Universiteit Amsterdam, the Open University (UK), and the Aalto University (Finland). The most significant collaborations at the level of individual countries are found between Italy and the United Kingdom, Italy and France, the United Kingdom and the Netherlands, and Italy and Spain.

From this analysis, it appears quite evident that Italy stands out as the country that contributes the most, at least in terms of number of institutions and collaborations, to research in the ontological field for the

Cultural Heritage domain. The debate itself appears, at present, substantially Eurocentric. It is also possible to assert that there are countries where the number of collaborations among national institutions is considerably high, as is the case with Italy, the Netherlands, and Greece, while for other nations this ratio is more skewed towards relationships with foreign institutions (Fig. 4).

2.4 *Future steps: towards an ontology for Heritage Science*

At present, HS is focused on tangible Cultural Heritage of any nature and scale, from individual objects to landscapes; however, future expansion to intangible Heritage is not to be excluded (Skublewska-Paszkowska *et*

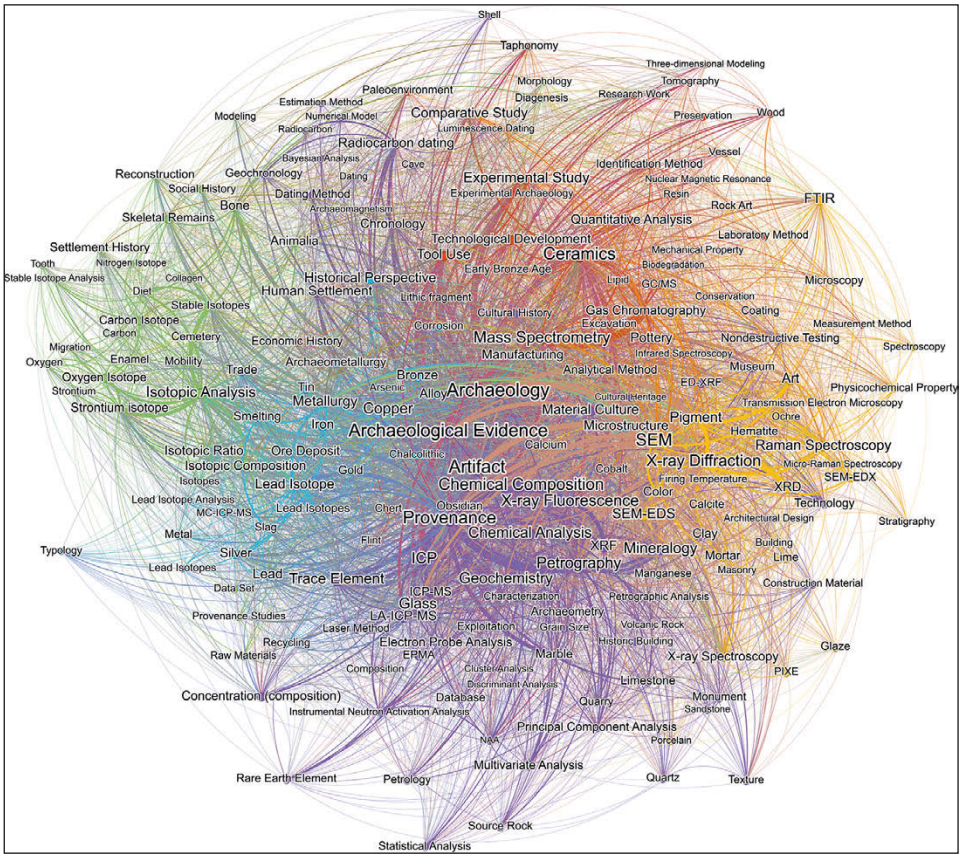


Fig. 5 – Keyword network for the journal «Archaeometry» (2242 articles, 1958-2024), generated with VOSviewer and reprocessed with Gephi, spatialized using Force Atlas 2. Color is based on VOSviewer clustering, while node size is based on Eigenvector centrality factor.

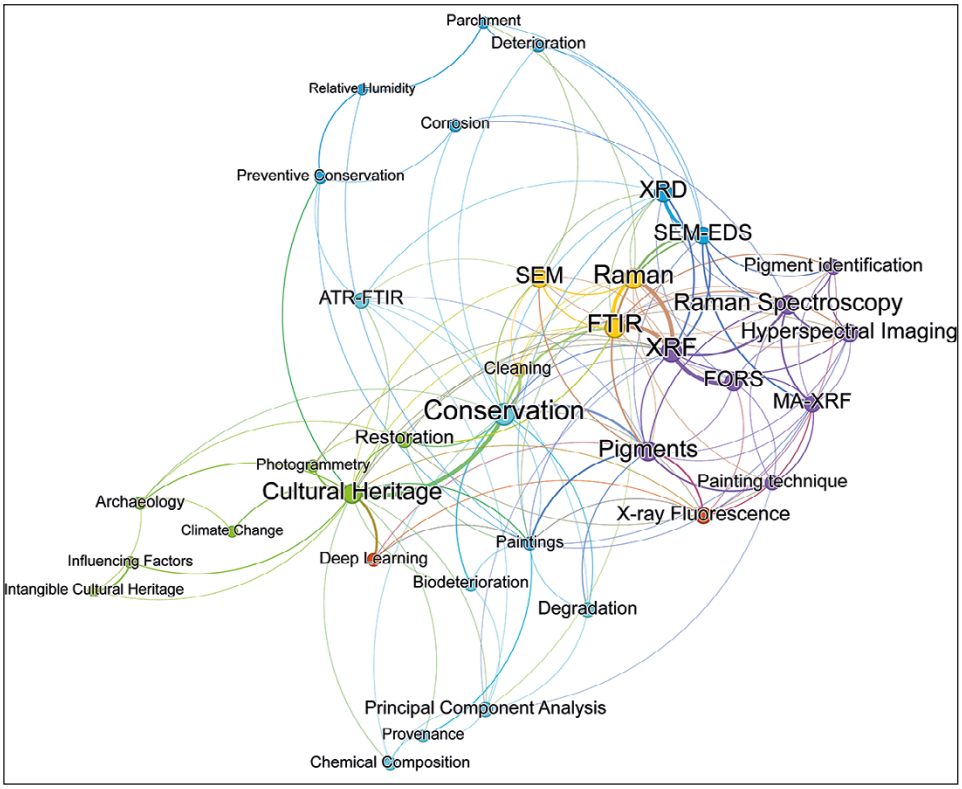


Fig. 6 – Keyword network for the journal «Heritage Science» (1264 articles, 2013-2024), generated with VOSviewer and reprocessed with Gephi, spatialized using Force Atlas 2. Color is based on VOSviewer clustering, while node size is based on Eigenvector centrality factor.

al. 2022). The more theoretical aspects of HS are still under development (Strlič 2018); therefore, the design of a dedicated ontology can significantly contribute to the formal definition of the discipline itself. As already observed, there is still no conceptual model that formalizes the domain related to diagnostics for cultural Heritage, and the data produced are often non-interoperable (Castelli, Felicetti, Proietti 2021, 280). There are resources from related fields and disciplines that participate in HS, although they are not specific to this particular domain, which can be reused, such as geographic and spatial resources, but none of these present specific features for this domain.

The lack of semantic web technology-related keywords in relevant literature underscores the current state of the field. Analyses of author-chosen keywords from two prominent journals, «Archaeometry» (Fig. 5) and «Cultural

Heritage» (Fig. 6), indicate a complete absence of semantic technologies. The dataset, gathered from Scopus using the ISSN of each journal, was visualized through network processing by VOSviewer with Gephi.

The commitment of the CNR to the development of a formal ontology for HS and its computer implementation, a challenging endeavor that is by its very nature ongoing, is based on its multi-decadal experience in the specific field, with coordination from the Italian node of E-RIHS. As previously stated in 2.3, CNR is already one of the mostly involved institutions in terms of national and international collaborations for what concerns this research field. In this regard, it is interesting to recall also the programmatic document of the Science and Technology Select Committee, which coined the term 'Heritage Science' in 2006 and pointed to CNR as a virtuous example of a research institute where basic and applied research are «(…) inextricably intertwined» (House of Lords Science and Technology Select Committee 2006, 24).

3. H-SeTIS database: the resource hub

3.1 *Preliminary conceptual structure*

The H-SeTIS database centers around five key objects: 'ontologies', 'metadata standards', 'thesauri', 'application profiles', and 'software'. While software itself does not constitute a 'semantic artefact' (see below), its inclusion within the cataloging framework is warranted due to its role in facilitating the implementation of the aforementioned four items. Collectively, these resources are referred to as 'semantic tools', reflecting their capacity to structure and enable the representation of knowledge in a machine-understandable format.

The term 'semantic artefact' emphasizes the ability of ontologies, metadata standards, thesauri, and application profiles to be processed by computers (Le Franc *et al.* 2020, 11-17). They represent the latest stage in the evolution of Knowledge Organization Systems (KOS). While many contemporary KOS applications are digital, the concept encompasses a broader range of tools, both physical and digital, designed to organize knowledge (Hodge 2000, 5). Classic examples of KOS include the Linnaean taxonomy for classifying animals and the Dewey Decimal Classification system, both predating modern computers. Despite their long history, the scientific community continues to debate the precise definitions and classifications of these tools, regardless of their machine readability (Souza, Tudhope, Almeida 2012). Hodge (2000, 4-5), for example, distinguishes between controlled vocabularies (authority files, glossaries, dictionaries, gazetteers), classifications and categorizations (subject headings and taxonomies), and lists of relationships (thesauri, semantic networks, ontologies). Hedden (2010,

1-15) prefers to use the single term 'taxonomies' distinguishing between controlled vocabularies, hierarchical taxonomies, thesauri, and ontologies.

For this reason, terms like taxonomy, thesaurus, ontology, and controlled vocabulary, although all falling under the umbrella of KOS, are frequently used interchangeably to describe various forms of knowledge representation. The classification and description of these models also exhibit significant variation in the literature due to the frequent entanglement of their characteristics, objectives, and specific use cases. Therefore, these entities can be conceptualized as distinct classification systems situated along a spectrum of rising complexity, each possessing unique characteristics and fulfilling diverse purposes (Souza, Tudhope, Almeida 2012, 183).

Drawing upon this overview, H-SeTIS employs a pragmatic and incremental approach to classifying semantic artefacts. This approach prioritizes simplicity and flexibility, enabling it to accommodate the diversity of these tools. This strategy facilitates the addition of new information and the integration of emerging semantic tools, ensuring the long-term scalability of the classification system. Aligned with FAIR principles and the Linked Data paradigm, H-SeTIS exposes its information through APIs and describes resource attributes by integrating standard metadata schemas such as Dublin Core and schema.org, thereby promoting data interoperability, reusability, and discoverability[5].

3.2 *The five semantic tools: an overview*

As mentioned above, H-SeTIS focuses on cataloging five key types of semantic tools: 'ontologies', 'metadata standards', 'thesauri', 'application profiles', and 'software' (Fig. 7). For each tool, information is provided to assess its compliance with the FAIR principles and its alignment with the Linked Data paradigm. References to research documenting each tool are listed within its record. The relevant bibliography is maintained through a public Zotero group (https://www.zotero.org/groups/5434475), which currently includes about 340 references: H-SeTIS utilizes Zotero's APIs to query the bibliographic data and leverages the unique identifiers provided by Zotero to manage and retrieve the individual references associated with each semantic tool. The project's bibliographic dataset will also be seamlessly integrated into the database's user interface.

Records for thesaurus-type artefacts encompass a diverse group of controlled lists with varying characteristics. An example of authority lists, a special type of controlled vocabulary that lists standardized forms of proper

[5] The numeric IDs referenced in the following notes correspond to the identifier that will appear in the URI of each resource within the H-SeTIS public website.

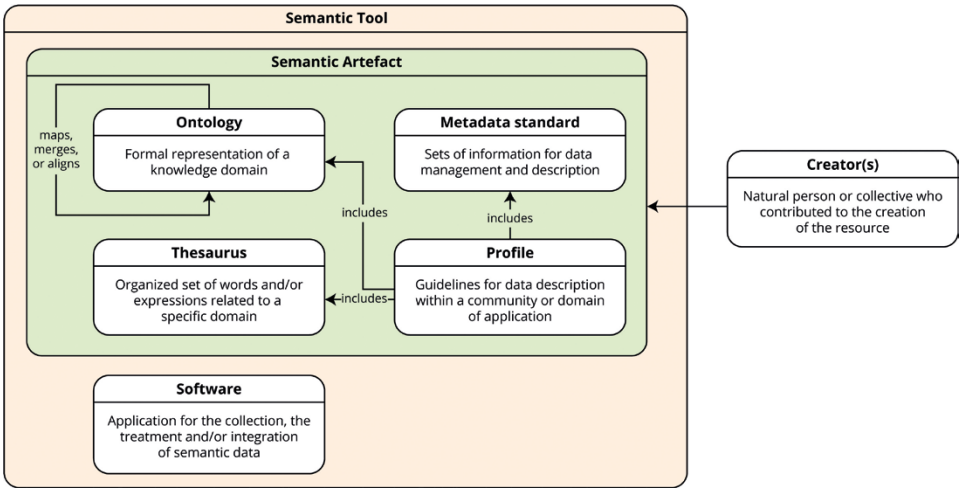


Fig. 7 – Preliminary Conceptual Structure of H-SeTIS.

nouns, is the *Liste d'autorités 'Auteurs'* (ID 109) used to populate the 'Auteur' field of the Joconde database (https://www.pop.culture.gouv.fr/search/list?base=%5B%22Collections%20des%20mus%C3%A9es%20de%20France%20%28Joconde%29%22%5D), which includes proper nouns referring to «une personne physique, un groupe de personnes physiques, une personne morale, une population, une civilisation, etc». Gazetteers, on the other hand, are generally limited to place names only. In this regard, the *Liste d'autorités 'Lieux'* (ID 131) used to populate the 'Lieux' field of the Joconde database falls into the category of gazetteers. As mentioned, it is possible to combine multiple characteristics of these artefacts, as in the case of the *Cairo Gazetteer* (ID 15): the latter presents a list of historical sites in Cairo hierarchically ordered by type and for which coordinates and descriptions are provided. The *Cairo Gazetteer* also presents semantic associations between the terms that compose it, a typical characteristic of thesauri.

Thesaurus artefacts, in the strictest sense of the term, feature a more structured organization and provide detailed information on the relationships between terms. These relationships include hierarchical links, associations (i.e., related concepts), and equivalences (i.e., synonyms). One of the best-known is the *Getty Art & Architecture Thesaurus* (ID 2), in which the terms are hierarchically organized and the most common spelling among various synonyms indicated together with a definition. Another type of thesaurus is the *Digitizing Early Farming Cultures* (ID 23), developed by the Austrian Centre for Digital Humanities for their project and relating to the cataloging

of Neolithic and Chalcolithic sites and finds in Greece and Anatolia (ca. 7000-3000 BCE; https://defc.acdh.oeaw.ac.at/). The DEFC presents a hierarchical organization of concepts and includes SKOS broader/narrower relationships, but also horizontal relationships between terms and definitions. It is therefore halfway between a taxonomy and a thesaurus. Compared to traditional taxonomies, thesauri place greater emphasis on the interconnections between terms, offering a more dynamic representation of semantic relationships.

Expanding on the previously mentioned types of thesauri, hierarchical taxonomies are also included in this category. Characterized by a tree-like structure, taxonomies feature broader terms encompassing more specific ones. Similarly to controlled vocabularies, taxonomies are domain-specific but generally simpler than thesauri: they omit equivalence and association relationships, focusing solely on presenting the preferred term chosen by the creators. Interestingly, despite their hierarchical structure, many of the resources described earlier often self-define as thesauri. This highlights the interchangeable use of these terms in practice.

Ontologies represent a further evolution in terms of complexity compared to other semantic artefacts. They incorporate logical relationships between terms to comprehensively represent a domain of knowledge. Consequently, they are classified as a separate artefact from thesauri. Gruber (2009, 1963) defines an ontology as «a set of representational primitives with which to model a domain of knowledge or discourse». Through inference, ontologies allow for reasoning about concepts, extracting new knowledge based on the encoded concepts, relationships, and rules.

Ontologies are also an annotation tool for which reusability is one of the distinguishing features. This promotes efficiency, consistency, and interoperability, and is primarily (but not only) realized through three modalities: aligning, merging, or mapping multiple ontologies. Alignment aims to identify correspondences between different ontologies in an automated or semi-automated manner. The merging method involves creating a new unified ontology by combining elements from multiple source ontologies, while mapping establishes relationships between concepts in different ontologies (Narula *et al*. 2018). In this regard, the H-SeTIS structure allows the ontology-type artefact to be related to itself in order to collect in a structured way the reuses that have occurred between various ontologies.

The CIDOC Conceptual Reference Model (CIDOC-CRM), a core ontology, is the most widely cited ontology in the Heritage sector. A 'core ontology' serves to express the basic concepts according to which a domain of knowledge is modeled. Due to this characteristic, it is scalable, meaning that it can be extended as needed. A 'foundational' ontology (also defined as 'upper' or 'top-level') instead models categories so general that they can be considered independent of any specific domain. This category also includes DOLCE

(Descriptive Ontology for Linguistic and Cognitive Engineering), created by the Istituto di Scienze e Tecnologie della Cognizione (CNR-ISTC) to reproduce the ontological categories of natural language and common sense (Gaio *et al.* 2010). The Architecture of Knowledge ontology (ArCO, ID 42) indirectly reuses two light versions of DOLCE, DOLCE-zero and DOLCE+DnS. EpiONT, a specialization of CIDOC-CRM for the epigraphic domain (ID 76), aims to achieve the same result (Cantone *et al.* 2019, 155). The ICON ontology also reuses some DOLCE classes (Sartini *et al.* 2023, 14-15).

As investigated by Moraitou *et al.* (2019, 623-624, Tab.2), CIDOC-CRM has been extended over the years through mapping, merging, or extension. A recent example of mapping is the adaptation of the Italian national standards for coding archaeological information, developed and maintained by the Istituto Centrale per il Catalogo e la Documentazione (ICCD), to CIDOC-CRM within the ARIADNE project (Felicetti *et al.* 2013; Moraitou *et al.* 2019, 617). CIDOC-CRM also has several official extensions, including CRM-archaeo, developed for the conceptual representation of the excavation process and related activities (Christaki *et al.* 2024).

In addition to Knowledge Organization Systems (KOSs), H-SeTIS also encompasses metadata standards, a distinct type of semantic artefact. While KOSs focus on vocabulary control and knowledge representation, metadata standards provide a structured and consistent way to describe various entities, from individual objects to entire databases, and offer a method for organizing, describing, tracking, and ultimately improving access to information (Gilliland 2008, 2-3). Similarly to thesauri and ontologies, the term 'metadata' can carry different meanings: within the Heritage domain, it typically refers to a set of supplementary information designed to organize, describe, track, and enhance access to information about a cultural asset and its associated physical collections. One of the pioneering examples is the schema developed by the Art Museum Image Consortium (AMICO), established by the Association of Art Museum Directors, aiming to standardize and regulate the reuse, distribution, and reproduction of digital images archived in the digital catalogs of various museums. Notably, AMICO, active from 1997 to 2005 was initially mapped by CIDOC-CRM (https://cidoc-crm.org/lrmoo/Resources/the-amico-data-model)[6].

Application profiles, on the other hand, are collections of practices, schemes, and guidelines adopted by a specific community or domain of application to describe a certain type of resource. Essentially, they provide instructions on how to effectively utilize metadata schemes within a particular domain (Baca 2008, 73): the Dublin Core Metadata Initiative Usage Board defines a profile as «a document (or package of documents) which describes

[6] See also the AMICO Data Specification (https://www.amico.org/AMICOlibrary/dataspec.html) and the AMICO Data Dictionary (https://www.amico.org/AMICOlibrary/dataDictionary.html).

a metadata application in order to facilitate broader reuse of its metadata» (https://www.dublincore.org/specifications/dublin-core/profile-review-criteria/).

An example of this type of semantic artefact is the Europeana Collection Profile, which integrates the Europeana Data Model (EDM) by reusing and extending existing classes and properties. As such, application profiles are positioned within H-SeTIS in relation to the other three semantic artefacts (ontologies, thesauri, metadata standards) to map the reuse of existing resources.

H-SeTIS also incorporates data standards alongside application profiles. Data standards function as comprehensive guidelines, defining the structure, formats, ontologies, and vocabularies employed for data management within a specific domain. Compared to application profiles, data standards are more general in scope: an example of a data schema is MIDAS Heritage, which aims to formalize data documentation for historical sites in the United Kingdom (https://historicengland.org.uk/images-books/publications/midas-heritage/midas-heritage-2012-v1_1/).

H-SeTIS further aims to gather information on software that enables the utilization of semantic artefacts, with a preferential but not exclusive focus on software developed in the Heritage sector. Examples of such software include Arches, an open-source platform for managing cultural heritage data (https://www.archesproject.org/), and Omeka S (https://omeka.org/s/), a software for creating virtual exhibitions developed by the Digital Scholar project of the Roy Rosenzweig Center for History and New Media.

Finally, H-SeTIS also includes record types pertaining to resource creators, encompassing both individuals and collectives who contributed to the development of a particular tool. This type of record will enable the construction of an up-to-date overview of the key players actively involved in the creation of semantic resources for the Heritage domain.

3.3 *H-SeTIS preliminary keyword analysis*

Each semantic artefact cataloged in H-SeTIS has been manually associated with one or more keywords. Both the artefacts and keywords were analyzed using Gephi (Fig. 8). The greatest aggregation revolves around the concept of ‘Material Culture’, which is expected given that a large part of Heritage is tangible. The only area currently covered by intangible or immaterial Heritage is essentially music, to which few ontologies and independent thesauri are associated. ‘Material Culture’ also serves as a macro-keyword for other closely associated concepts that specify it, such as ‘Archaeology’, ‘Architecture’, ‘Chronology’, ‘Geography’, ‘History’, ‘Museum Collections’. It is noticeable how archaeology is well-represented at a semantic level, with about a fifth of thesauri and ontologies directly related to the discipline itself, as well as Archaeometry, Epigraphy, Egyptology, Numismatics.

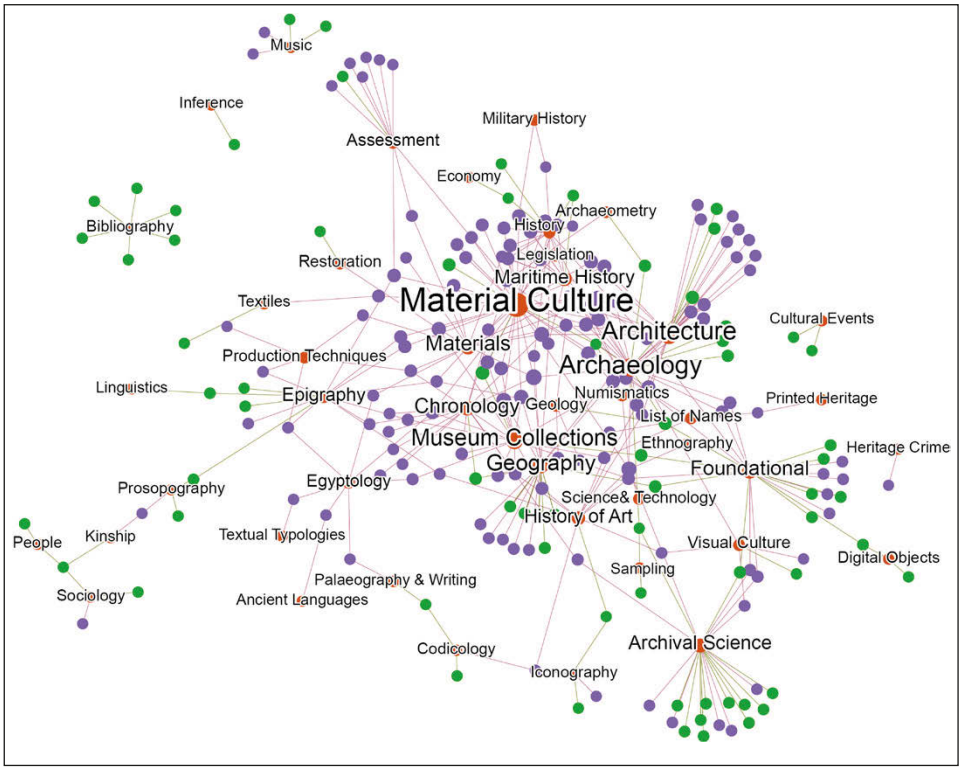


Fig. 8 – Keyword network for the first 200 records added to H-SeTIS. Green nodes refer to ontologies, purple nodes to thesauri, red nodes to keywords.

The clustered spatialization allows for the identification of at least three other subgroups: those of 'Foundational' artefacts, those of 'Archival Sciences', and those related to 'Library Science'. The first ones pertain to resources that, despite belonging to different fields, are fundamental for the description of other contents, such as Dublin Core (https://www.dublincore.org/specifications/dublin-core/dcmi-terms/) or the GS84 Geo Positioning RDF schema (http://www.w3.org/2003/01/geo/wgs84_pos#). Resources related to archival sciences are noteworthy both for the number of artefacts and for their relative independence from other keywords, although they are not entirely isolated like bibliographic ones: this, as well as their proximity to 'Foundational' resources, is partly due to the reason that a significant portion of ontological and taxonomic activity has traditionally been the responsibility of archivists and librarians, among the categories most involved, especially at a practical level, in organizing knowledge.

4. Final remarks

The H2IOSC Project is going to have a significant impact on HS both theoretically, improving to define its boundaries, and practically, increasing the interoperability and the machine-readability of data. At a preliminary overview, the core institutions engaged in researching semantic artefacts for Heritage are located in Europe and namely in Italy, that plays an important role in the Heritage research as a whole: within this landscape, CNR is actively involved with its long-standing and multidisciplinary expertise on it.

The H-SeTIS database, a central deliverable of the H2IOSC Project, serves as a foundational digital repository for semantic resources in Heritage studies. This includes ontologies, metadata standards, thesauri, application profiles, and software. The database will not only ensure a continued monitoring of these resources but will also facilitate the development of a comprehensive Heritage studies ontology. In this regard, it will provide the essential knowledge for integrating and mapping these resources with existing ones, offering a clear picture of the areas already covered by existing semantic artefacts. From these initial research steps, a more comprehensive HS ontology will be made available to the scientific community, promoting FAIRer data and potentially leading to a wider impact for the entire discipline.

Erica Scarpa, Riccardo Valente
Istituto di Scienze del Patrimonio Culturale - CNR
erica.scarpa@cnr.it, 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 H-SeTIS implementation benefited from the technical support of the IT specialist Matteo Gallo, who was entrusted with the software development service within this endeavor. The authors wish to express their gratitude to Irene Rossi and Chiara Salvador for their insightful comments and editorial assistance.

References

Baca M. (ed.) 2008, *Introduction to Metadata*, Los Angeles, Getty Research Institute.

Bordalo R., Bottaini C., Candeias A. 2020, *A framework design for information management in Heritage Science laboratories*, «Journal on Computing and Cultural Heritage», 14, 1, 1-14 (https://doi.org/10.1145/3417304).

Cambon J., Hernangómez D., Belanger C., Possenriede D. 2021, *Tidygeocoder: An R package for geocoding*, *R package version 1.0.5*, «Journal of Open Source Software», 6, 65, 3544 (https://doi.org/10.21105/joss.03544).

Cantone D., Nicolosi-Asmundo M., Santamaria D.F., Cristofaro S., Spampinato D., Prado F. 2019, *An EpiDoc ontological perspective: The epigraphs of the Castello Ursino Civic Museum of Catania via CIDOC CRM*, «Archeologia e Calcolatori», 30, 139-157 (https://doi.org/10.19282/ac.30.2019.10).

Carman J., Sørensen M.L.S. 2009, *Heritage Studies – An outline*, in J. Carman, M.L.S. Sørensen (eds.), *Heritage Studies. Methods and Approaches*, London, Routledge, 11-28.

Castelli L., Felicetti A., Proietti F. 2021, *Heritage Science and Cultural Heritage: Standards and tools for establishing cross-domain data interoperability*, «International Journal on Digital Libraries», 22, 3, 279-287 (https://doi.org/10.1007/s00799-019-00275-2).

Christaki E., Doerr M., Felicetti A., Hermon S., Hiebel G., Kritsotaki A., Masur A., May K., Ore C.-E., Ronzino P., Schmidle W., Theodoridou M., Tsiafaki D. 2024, *Definition of the CRMarchaeo. An Extension of CRMbase to support the archaeological excavation process* (https://www.cidoc-crm.org/crmarchaeo/sites/default/files/CRMarchaeo_v2.1%28site%29.pdf; last accessed 29/05/2024).

Felicetti A., Scarselli T., Mancinelli M.L., Niccolucci F. 2013, *Mapping ICCD archaeological data to CIDOC-CRM: The RA Schema*, in V. Alexiev, V. Ivanov, M. Grinberg (eds.), *CRMEX 2013 Practical Experiences with CIDOC CRM and its Extensions*, CEUR WS, 11-22.

Gaio S., Borgo S., Masolo C., Oltramari A., Guarino N. 2010, *Un'introduzione all'ontologia DOLCE*, «AIDA informazioni: Rivista di scienze dell'informazione», 28, 107-125 (https://doi.org/10.1400/212462).

Garijo D., Poveda-Villalón M. 2020, *Best Practices for Implementing FAIR Vocabularies and Ontologies on the Web* (https://doi.org/10.48550/arXiv.2003.13084).

Gilliland 2008, *Setting the stage*, in M. Baca (ed.), *Introduction to Metadata*, Los Angeles, Getty Research Institute, 1-19.

Gruber T. 2009, *Ontology*, in L. Liu, M.T. Özsu (eds.), *Encyclopedia of Database Systems*, Boston, Springer, 1963-1965 (https://doi.org/10.1007/978-0-387-39940-9_1318).

Hedden H. 2010, *The Accidental Taxonomist*, Medford, Information Today.

Hodge G.M. 2000, *Systems of Knowledge Organization for Digital Libraries: Beyond Traditional Authority Files*, Washington, Digital Library Federation.

House of Lords Science and Technology Select Committee 2006, *Science and Heritage. Report with Evidence*, London, The Stationery Office Limited.

Huang Y. 2024, *Bibliometric analysis of GIS applications in heritage studies based on Web of Science from 1994 to 2023*, «Heritage Science», 12, 1, 57 (https://doi.org/10.1186/s40494-024-01163-y).

Jonquet C., Graybeal J., Bouazzouni S., Dorf M., Fiore N., Kechagioglou X., Redmond T., Rosati I., Skrenchuk A., Vendetti J.L., Musen M. 2023, *Ontology repositories and semantic artefact catalogues with the OntoPortal technology*, in T.R. Payne, V. Presutti, G. Qi, M. Poveda-Villalón, G. Stoilos, L. Hollink, Z. Kaoudi, G. Cheng, J. Li (eds.), *The Semantic Web – ISWC 2023*, Cham, Springer Nature Switzerland, 38-58 (https://doi.org/10.1007/978-3-031-47243-5_3).

Kennedy C.J. 2015, *The role of Heritage Science in conservation philosophy and practice*, «The Historic Environment: Policy & Practice», 6, 3, 214-228 (https://doi.org/10.1080/17567505.2015.1099925).

Kennedy C.J., Penman M., Watkinson D., Emmerson N., Thickett D., Bosché F., Forster A.M., Grau-Bové J., Cassar M. 2024, *Beyond Heritage Science: A review*, «Heritage», 7, 3, 1510-1538 (https://doi.org/10.3390/heritage7030073).

Le Franc Y., Parland-von Essen J., Bonino L., Lehväslaiho H., Coen G., Staiger C. 2020, *D2.2 FAIR Semantics: First recommendations* (https://zenodo.org/records/3707985; last accessed 29/05/2024).

Moraitou E., Aliprantis J., Christodoulou Y., Teneketzis A., Caridakis G. 2019, *Semantic bridging of Cultural Heritage disciplines and tasks*, «Heritage», 2, 1, 611-630 (https://doi.org/10.3390/heritage2010040).

Narula G.S., Wason R., Jain V., Baliyan A. 2018, *Ontology mapping and merging aspects in Semantic web*, «International Robotics & Automation Journal», 4, 1 (https://doi.org/10.15406/iratj.2018.04.00087).

Sartini B., Baroncini S., Van Erp M., Tomasi F., Gangemi A. 2023, *ICON: An ontology for comprehensive artistic interpretations*, «Journal on Computing and Cultural Heritage», 16, 3, 1-38 (https://doi.org/10.1145/3594724).

Skublewska-Paszkowska M., Milosz M., Powroznik P., Lukasik E. 2022, *3D technologies for intangible cultural heritage preservation. Literature review for selected databases*, «Heritage Science», 10, 1, 3 (https://doi.org/10.1186/s40494-021-00633-x).

Souza R.R., Tudhope D., Almeida A.M.B. 2012, *Towards a taxonomy of KOS: Dimensions for classifying knowledge organization systems*, «Knowledge Organization», 39, 3, 179-192 (https://doi.org/10.5771/0943-7444-2012-3-179).

Strlič M. 2018, *Heritage Science: A future-oriented cross-disciplinary field*, «Angewandte Chemie International Edition», 57, 25, 7260-7261 (https://doi.org/10.1002/anie.201804246).

van Eck N.J., Waltman L. 2010, *Software survey: VOSviewer, a computer program for bibliometric mapping*, «Scientometrics», 84, 2, 523-538 (https://doi.org/10.1007/s11192-009-0146-3).

Vandenbussche P.-Y., Atemezing G.A., Poveda-Villalón M., Vatant B. 2017, *Linked Open Vocabularies (LOV): A gateway to reusable semantic vocabularies on the web*, «Semantic Web», 8.3, 437-452 (https://doi.org/10.3233/SW-160213).

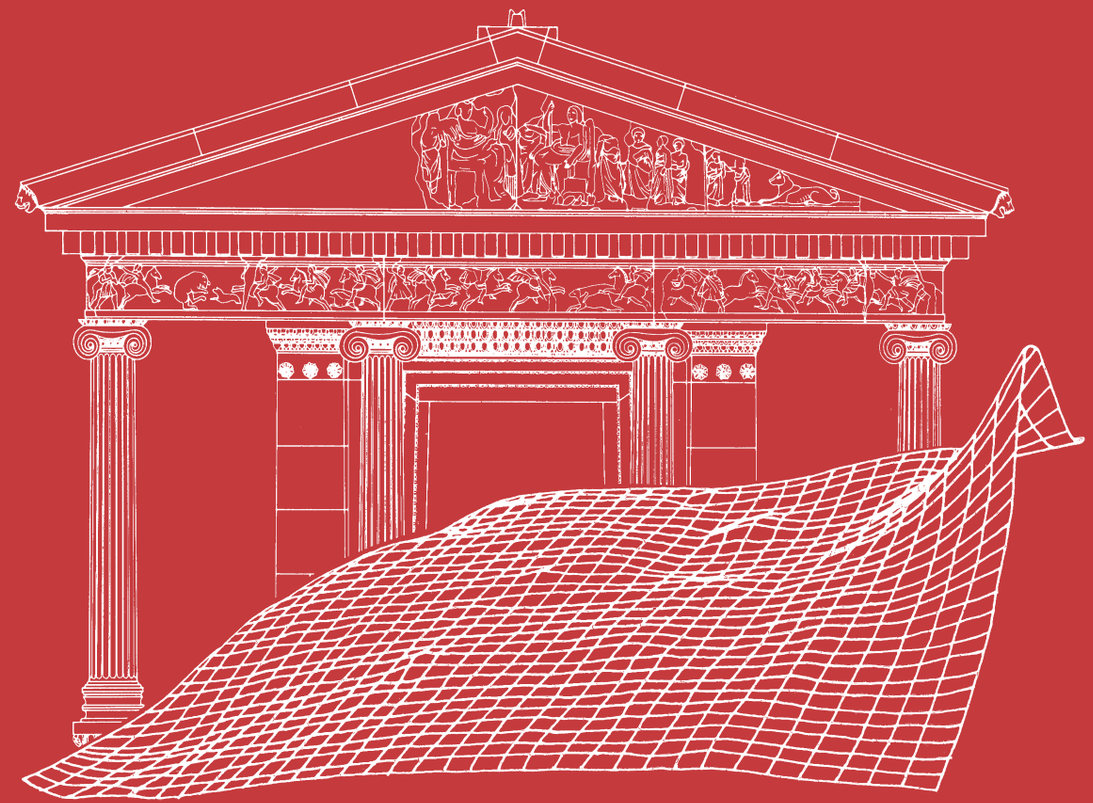
Wilkinson M.D., Dumontier M., Aalbersberg Ij.J. *et al*. 2016, *The FAIR Guiding Principles for scientific data management and stewardship*, «Scientific Data», 3, 1, 160018 (https://doi.org/10.1038/sdata.2016.18).

ABSTRACT

This article explores the contributions of the Milan branch of CNR-ISPC to the Humanities and Cultural Heritage Italian Open Science Cloud (H2IOSC) Project, focusing on facilitating data integration within Heritage Science. Its primary objective is to ensure seamless interoperability between resources from multiple institutions by establishing a shared semantic framework. The multidisciplinary nature of Heritage Science underscores the necessity for shared data repositories and effective management tools. Recent literature highlights the importance of semantic technologies in improving data integration and interoperability. To this end, the H-SeTIS database is currently under development. H-SeTIS will function as a hub for the systematic surveying and description of various semantic tools relevant to the Heritage domain. Interestingly, a preliminary analysis of data within H-SeTIS reveals that many semantic resources specifically designed to address the unique requirements of the Heritage domain do not meet the minimum quality requirements of accessibility and reusability. This finding underscores a potential area for future development: the creation of H-SeTIS aims to support the ongoing development of a comprehensive ontology for Cultural Heritage, enhancing data FAIRness and the discipline's overall impact.

35.1

2025



€ 60,00
ISSN 1120-6861
e-ISSN 2385-1953
ISBN 978-88-9285-278-5
e-ISBN 978-88-9285-279-2
AC-35-1

ARCHEOLOGIA
E CALCOLATORI

ARCHEOLOGIA E CALCOLATORI

35.1

2024

All'Insegna del Giglio