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).

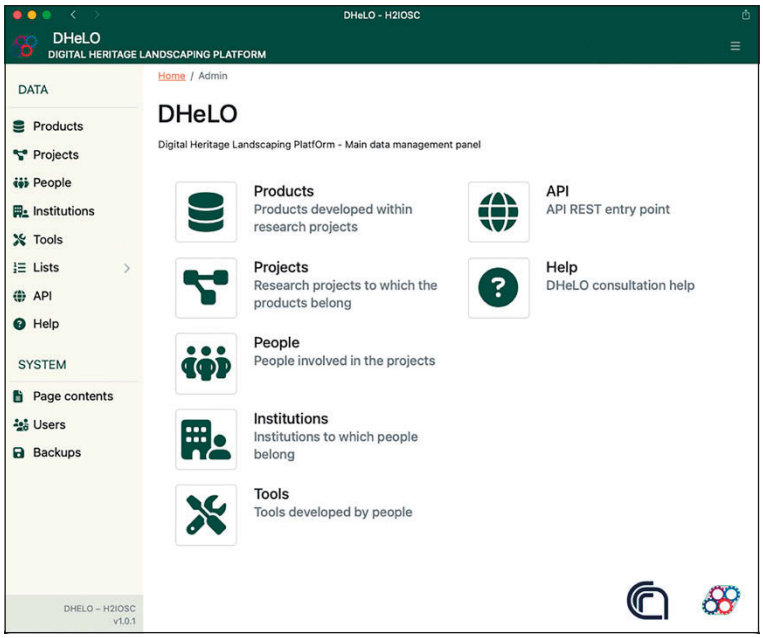

Fig. 1 – Screenshot of the DHeLO web app home page

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.

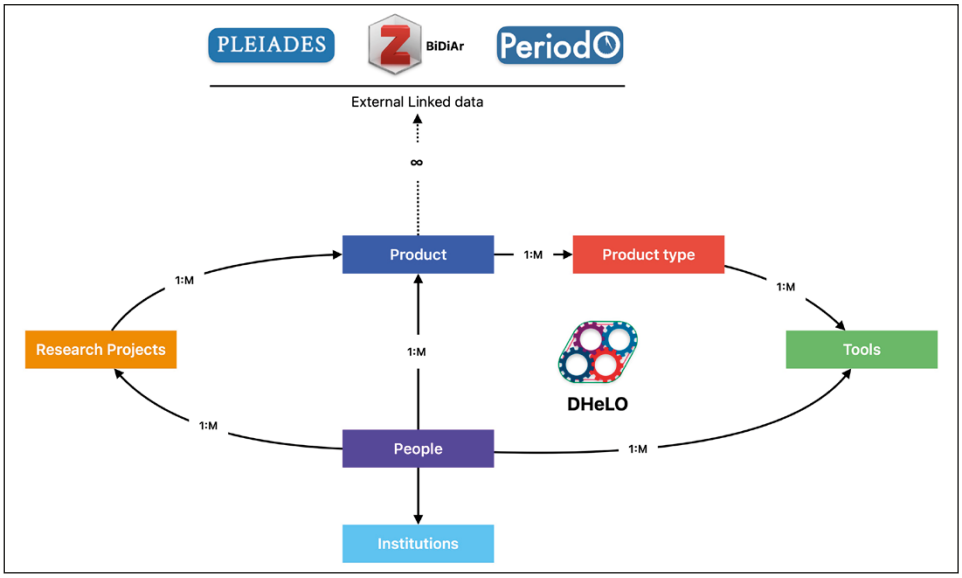

Fig. 2 – Overview of the DHeLO table schema structure.

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

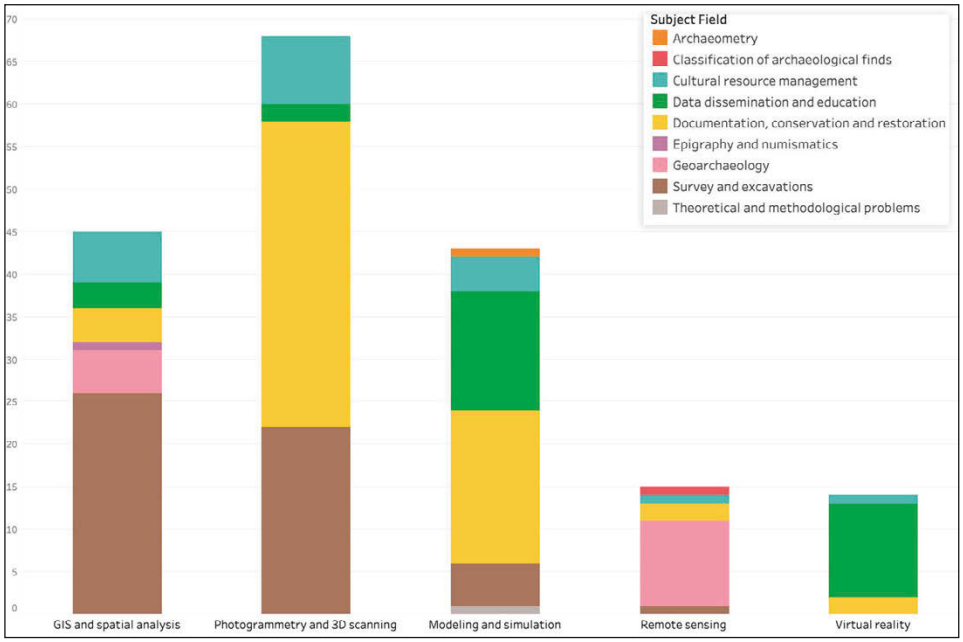

Fig. 3 – Bar chart of a selection of Products related to the most popular categories of technological applications in the digital CH/HS sectors, colored according to the archaeological research fields primarily concerned.

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

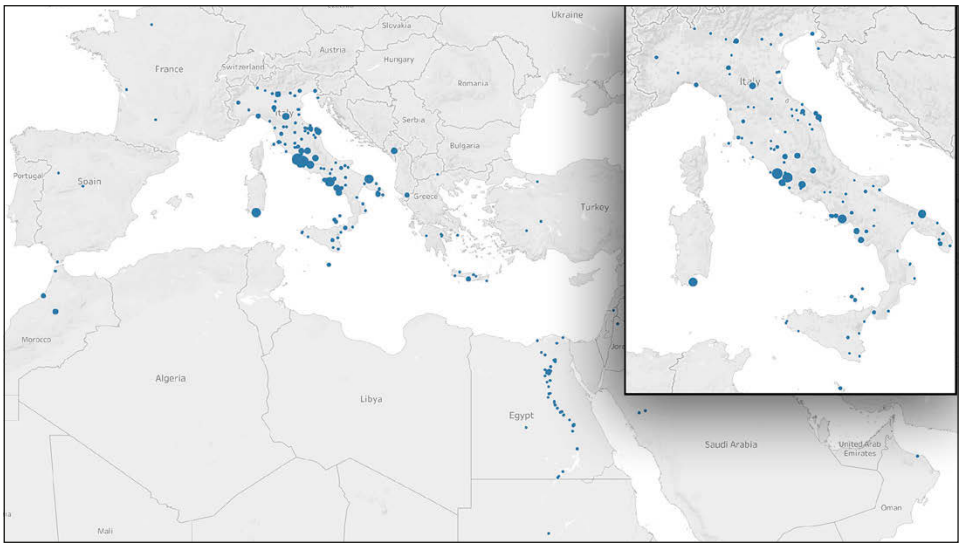

Fig. 4 – Map of the Products in the DHeLO database, geotagged through the Pleiades references.

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

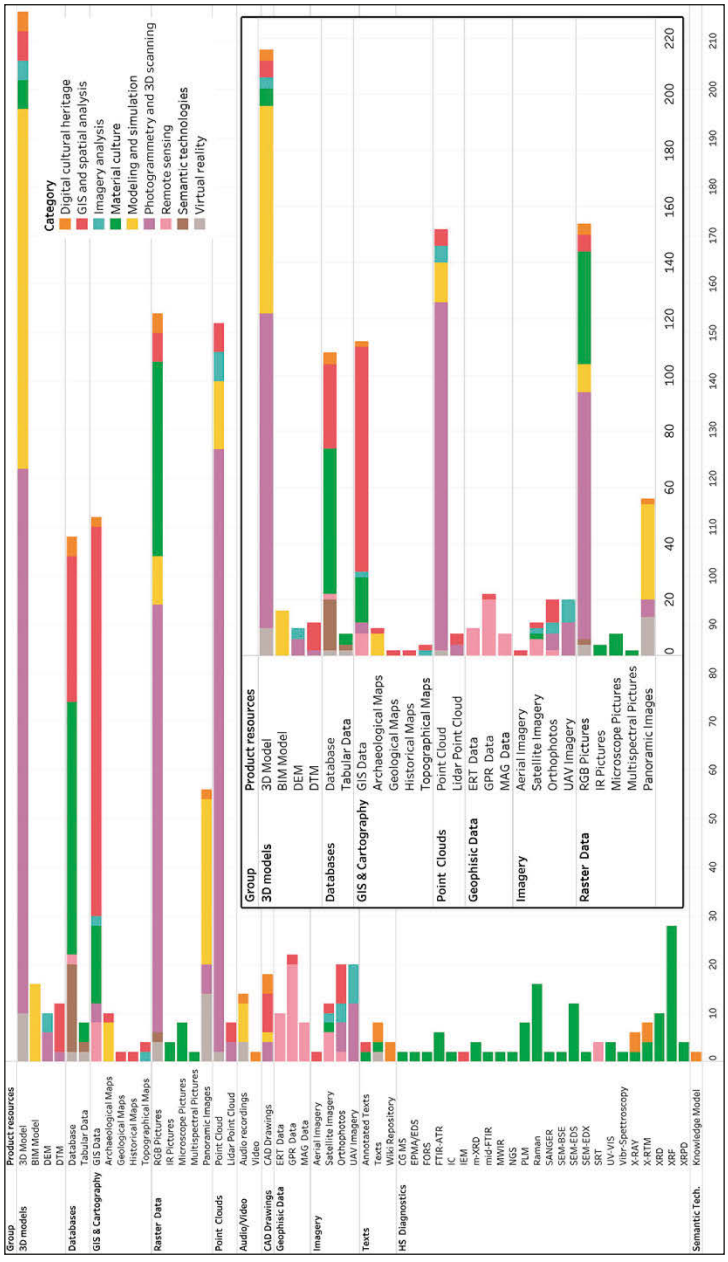

Fig. 5 – Data types recorded in the DHeLO database thematically grouped and color-coded by category. On the right, within the black box, a selection of the most popular for the CH sector.

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'

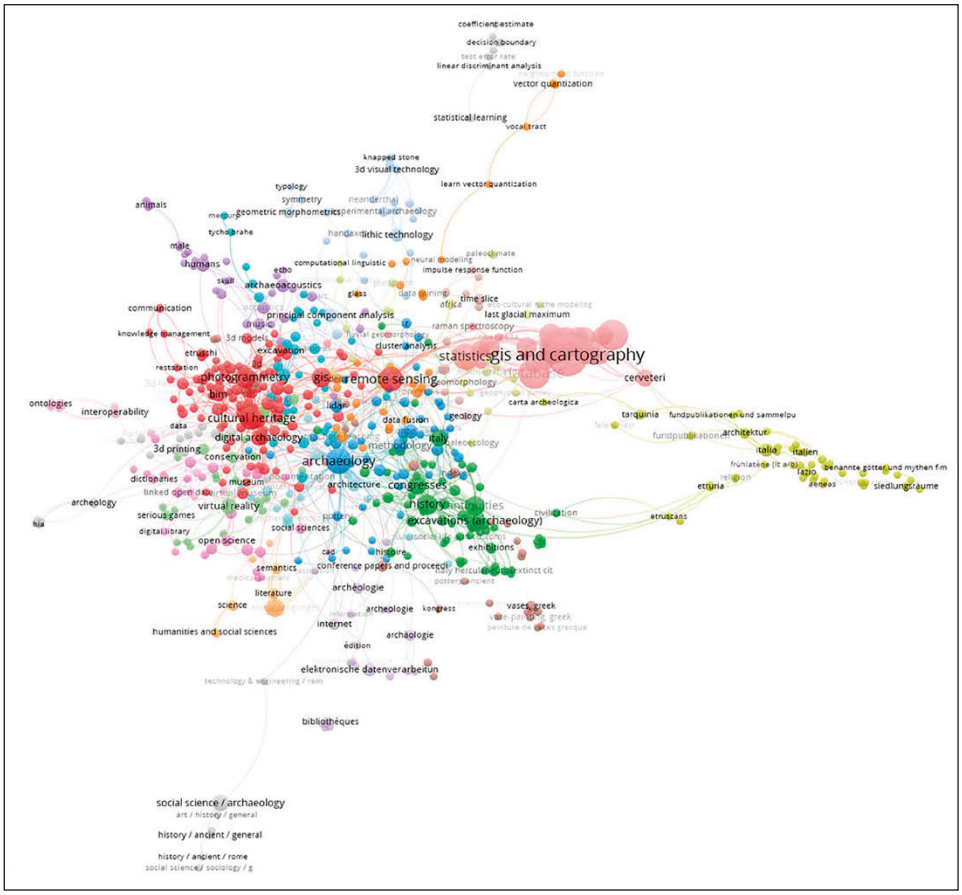

Fig. 6 – Network visualizations of the keywords associated to bibliographic references in BiDiAr. Worth noticing 4 main clusters to explore with further analysis: GIS and cartography (pink), archaeology (blue), photogrammetry (red), excavations (green).

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.

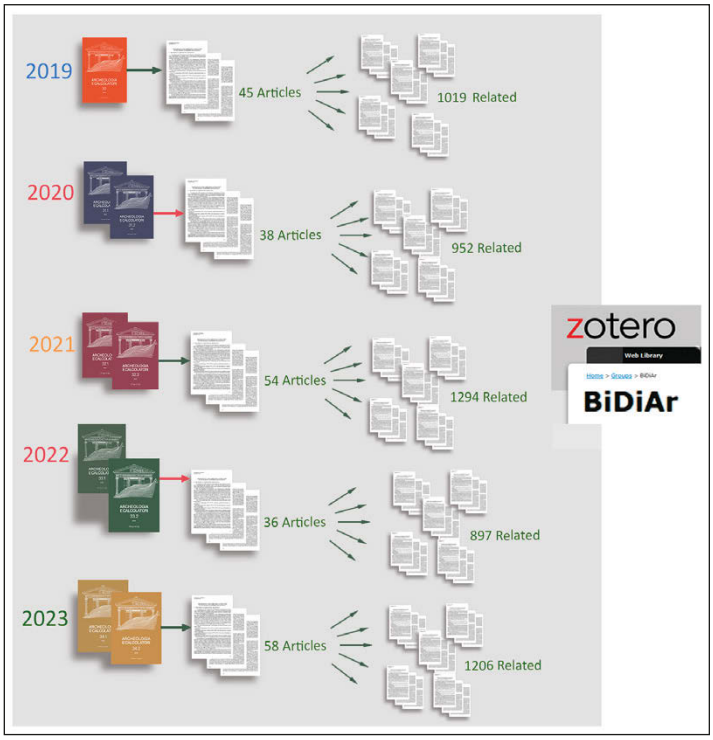

Fig. 7 – Graph illustrating the relationships between A&C papers and their correlated records within BiDiAr.

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

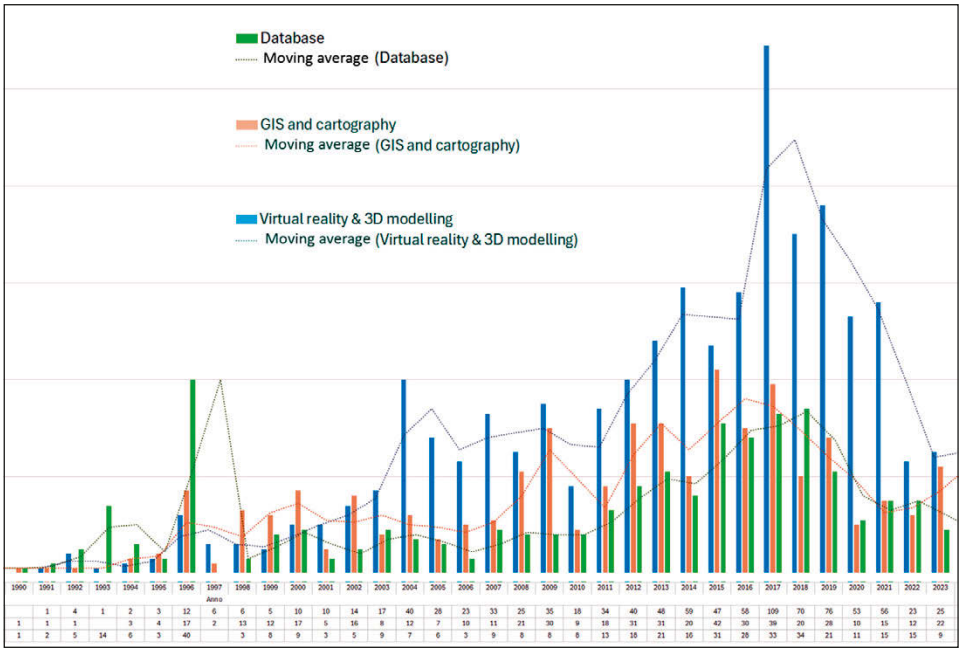

Fig. 8 – Graph illustrating the distribution over time of the number of publications concerning: Virtual reality and 3D modelling, GIS and cartography and Database.

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.