

Contents lists available at [ScienceDirect](https://www.sciencedirect.com)

Digital Applications in Archaeology and Cultural Heritage

journal homepage: www.elsevier.com/locate/daach

From archives to museum and back: transcribing, digitizing, and enriching cultural heritage and manuscript legacy data of the Villa del Casale of Piazza Armerina

Giulia Marsili^{a,*} , Stephan N. Hassam^b ^a University of Bologna, Department of History and Cultures, Italy^b Randolph-Macon College, Department of Classics, USA

ARTICLE INFO

Keywords:

Archival documentation
 Manuscript legacy data
 Artificial intelligence
 Script analysis
 Web publishing
 Omeka

ABSTRACT

Legacy data from archaeological sites with long excavation histories present both challenges and opportunities for modern research. Such data - ranging from handwritten notes, excavation diaries, and photographs to artefacts and related inventories - often predate contemporary recording standards, yet they can hold invaluable information about archaeological sites that did not make it into the publication record. This research situates itself within the broader theoretical framework of “archive archaeology,” in the context of a digitization project at the Villa del Casale. Using existing AI technologies such as Transkribus and the Handwriting Analysis Tool, the project approaches archives not only as repositories of information but also as subjects of study. It explores the potential of these advanced digital technologies to transcribe and interpret unpublished handwritten legacy data - specifically, archival materials related to earlier excavations at the Villa del Casale in Piazza Armerina - in order to contribute to the broader analysis and understanding of legacy data and handwritten field notes. Additionally, the project discusses the creation of a Villa del Casale’s digital ecosystem to enhance the dissemination, accessibility, and reuse of both primary and secondary research data through the use of the open-source web publishing platform Omeka Classic, designed for the creation and management of digital collections and exhibits. The approach taken in this research seeks to integrate different up-to-date digital technologies to bridge the gap between historical archives, archaeological legacy data, and contemporary archaeological inquiry.

1. Introduction

Legacy data, particularly at sites with a long history of excavations, represents a significant challenge for archaeological investigation requiring careful management and integration into contemporary research frameworks. Handwritten notes, excavation diaries, unpublished manuscripts, field notes, photographs, drawings, as well as artefacts stored for decades with related inventories represent just some of key categories of evidence encompassed within this definition. These documents, many of which were created before current recording standards were established, represent an invaluable resource for modern research since they hold unique information about sites, artefacts, and contexts that may no longer be available due to environmental changes, poor preservation practices, or, more frequently, the loss of important primary data caused by the destructive nature of archaeological excavation. To unlock and grasp the potential of these documents, they must

be retrieved, systematised, deciphered and integrated into coherent systems. In this regard, the digitization of records and artefacts represents a transformative solution and a substantial advancement in archaeological publication and research through its ability to enhance the speed and accuracy of data collection, while improving its accessibility and useability in archaeological research.

In recent decades, new theoretical frameworks and improved access to advanced technologies have significantly increased interest in archaeological documentation and artefacts housed in museums and archives. A substantial body of theoretically driven research has sought to investigate this “Archive Archaeology” field, focusing on its potential and on developing effective strategies for engaging with archaeological archives (Baird, 2011; Swain, 2012; for the most recent attempts in the field: Raja, 2023; Frey and Raja, 2024, with references). This “archival turn” has shifted the focus from viewing archives merely as sources of understanding them (the archive-as-source) to subjects of study in their

* Corresponding author. University of Bologna, Piazza S. Giovanni in Monte, 2, 40124, Bologna, Italy.

E-mail addresses: giulia.marsili2@unibo.it (G. Marsili), stephanhassam@rmc.edu (S.N. Hassam).

<https://doi.org/10.1016/j.daach.2025.e00441>

Received 31 December 2024; Received in revised form 16 May 2025; Accepted 11 June 2025

Available online 17 June 2025

2212-0548/© 2025 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

own right (“the archive-as-subject”), which highlights the complexities of knowledge production (Stoler, 2009; Ward, 2022). By analyzing archival materials within their original contexts and considering their intended purposes, researchers can uncover alternative narratives and critically re-evaluate archival content, a responsibility shared by archaeological practices and museums.

In this context, specific challenges are posed by the very nature of handwritten documents. Their interpretation is often hindered by not only poor preservation conditions, but also peculiar handwriting styles, as well as archaic phrasing and vocabulary. To overcome these issues, extensive manual work and expertise is usually required. However, recent developments in AI are revolutionizing this field by offering enhanced tools to decipher manuscript documents and boost their accessibility and comprehensibility and thus render them useable for archaeological research.

This paper addresses digital strategies that were deployed to interpret and enhance the legacy data from the Villa del Casale of Piazza Armerina as well as integrate it into an up-to-date scientific research framework. Special focus is paid to handwritten documents discovered in the archives of the Soprintendenza per i Beni Culturali e Ambientali di Siracusa¹ that have been analysed, processed and interpreted by applying up-to-date HTR technologies that are usually employed for historical manuscripts and here experimented with to improve the speed and quality of transcription for the data mining of handwritten field notes. Moreover, a broad discussion about the creation of a digital ecosystem to expand the dissemination of and accessibility to primary and secondary research data to facilitate their improvement and reuse will be also presented.

2. Rationale

The archaeological site of Villa del Casale of Piazza Armerina has been ranked into the UNESCO World Heritage List since 1997. Largely appreciated for its elaborated architecture and mosaic decoration, it represents an exceptional context to investigate the development of a rural settlement over the *long durée*, from the so-called *villa rustica* (end of the 1st c. AD) to the Arab-Norman village (12th c.) (Gentili, 1999; Sfameni, 2006; Pensabene and Barresi (eds.), 2019; Bonanno, 2020; Baldini et al., 2025) (Fig. 1). This multi-phase settlement was unearthed over numerous excavation campaigns conducted with varying methodologies and objectives, spanning from the late 19th century to the present day (Nigrelli and Vitale, 2010, 24–25; Nigrelli and Vitale, 2014). Although knowledge of the existence of archaeological structures had been circulating among local antiquarians as early as the mid-17th century, the first documented discoveries and sporadic excavations did not occur until the late 19th century. Structural remains and mosaic floors were brought to light in 1820, while in 1881 L. Pappalardo uncovered the first figurative mosaics of the Trichoras hall (Chiarandà, 2025; Gentili, 1950, 291–296; Agnello, 1965). Later, between 1929 and 1940, P. Orsi, G. Cultrera and B. Pace supervised excavation works that enlarged Pappalardo’s excavations from the outer wall of the exedra of the elliptical courtyard to the Triapsidal room, the ambulatory of the Great Hunt and the room with the mosaic of the so-called “bikini girls,” mainly with the intent to uncover mosaics and walls (Orsi, 1934; Cultrera, 1936; Pace, 1951, 1955) (Fig. 2). These interventions were followed by the most impactful excavation campaign, which took place under the supervision of G. V. Gentili from 1950 to 1955 and brought to light most of the Late Antique period structures of the building, also supervising the restoration and conservation of the mosaics until 1963 (Gentili, 1999). Later, some trenches were carried out in 1970 by A. Carandini to verify the stratigraphical sequence of specific areas (Ampolo et al., 1971; Carandini et al., 1982). Moreover, between 1983

and 1988 excavation works under E. De Miro and G. Fiorentini’s supervision brought to light massive storage facilities in front of the entrance of the Villa (De Miro, 1988), while in 1996–1997 scattered medieval structures emerged during L. Guzzardi’s investigations. From 2004 to 2014, the team coordinated by P. Pensabene contributed to the discovery of a new bath complex to the south of the Villa and to the systematisation of a vast array of data from previous investigation seasons (Pensabene and Barresi, 2019).

In 2022, archaeological investigations resumed under the coordination of the Department of History and Cultures of the University of Bologna with the participation of the Institute of Cultural Heritage Sciences (ISPC) of the National Research Council and the University of Enna “Kore”, in agreement with the Archaeological Park of Morgantina and the Villa Romana del Casale in Piazza Armerina, and under the auspices of CISEM (Inter-University Centre for Studies on Late Antique Mediterranean Housing).² At the outset of the project, a geophysical investigations campaign was undertaken to explore the archaeological potential of the surrounding area of the Villa and define future research directions (Baldini et al., 2024). The University of South Florida’s Institute for Digital Exploration (USF IDEx) was included as an important collaborator with the project, responsible for the digitization in 3D activities in the field and of the legacy data, as well as the online dissemination of processed digital products. USF IDEx had in previous years carried out a far-reaching 3D digitization campaign of the Villa for documentation, conservation, and educational purposes (Fig. 3) (Gabellone et al., 2020; Vennarucci et al., 2021; Kingsland, 2023).³ The Villa del Casale’s closely intertwined relationship with the Museo di Palazzo Trigona is also at the core of the project “Digital strategies for enhancing cultural heritage: the Villa del Casale of Piazza Armerina, from the late antique building site to the Museum Collection” (Marsili, 2024). The project represents a case study of an EU-funded research initiative entitled “Virtual Technologies for Museums and Art Collections”, which acts as a thematic spoke of the project “CHANGES - Cultural Heritage Active Innovation for Next-Gen Sustainable Society”, prompting the experimentation of virtual technologies for the promotion, preservation, and enhancement of cultural heritage in Italian museums and art collections (Balzani et al., 2024).

The launch of the multidisciplinary project in Piazza Armerina offered the opportunity for new multipronged campaign of investigations on the abundant legacy data left over from previous excavations and research. Firstly, a campaign was undertaken to collect, systematise and digitize known legacy data in a GIS environment to verify the consistency and distribution patterns of different categories of findings in different settlement phases (Pizzi, 2023; Elena, 2023). Secondly, a substantial collection of artefacts from Gentili’s excavations, previously dispersed throughout the Villa’s storerooms, has been systematically recovered, catalogued, and 3D digitized. An important part of the project is represented by the publication of these materials on a Content Management System (CMC) in order to facilitate access to otherwise inaccessible resources and to create a foundation for advancing scientific research and enabling the virtual recontextualization of these elements (Hassam, forthcoming). Thus, a digital ecosystem has been set up via Omeka Classic, experimenting with the CMC’s potential in hosting and showcasing both 2D and 3D digital replicas of artefacts from past excavations and archival materials (Fig. 4), as discussed in Section 4.2. Lastly, a comprehensive review of the primary archival collections held by the regional Superintendencies, which serve as custodians of the historical record of past research activities, has led to the identification of largely unpublished graphic and photographic materials. Among these, numerous photographs document key moments

¹ <https://www.regione.sicilia.it/istituzioni/regione/strutture-regionali/soprintendenza-beni-culturali-ambientali-siracusa>

² <https://centri.unibo.it/cisem/en>

³ An example of the result of digital photogrammetry campaign carried out by USF IDEx of the Villa del Casale can be found here <https://youtu.be/3DYPrdYcy4?si=g0MIFi889EpWe6q5>.



Fig. 1. Aerial view of the Villa del Casale of Piazza Armerina. Image of the archaeological site taken using a drone (photo by C. Lamanna).



Fig. 2. Piazza Armerina, Villa del Casale, Triclinium, mosaic with the labour of Hercules during old excavations (Courtesy of Soprintendenza per i Beni Culturali e Ambientali di Siracusa, Proprietà della Regione Siciliana, Assessorato dei Beni Culturali e dell'Identità Siciliana - Dipartimento dei Beni Culturali e dell'Identità Siciliana - Soprintendenza per i Beni Culturali e Ambientali di Siracusa - reproduced with permission).

of old excavations or capture the monument's condition prior to the extensive restoration efforts undertaken in the last century. Administrative and epistolary documents provide valuable insights into the logistical and operational challenges of early archaeological work, including details surrounding significant discoveries such as the so-called colossal statue of Hercules (Fig. 5).

Of particular note is the discovery of an unpublished and unauthored manuscript from the archives of the Soprintendenza per i Beni Culturali e Ambientali di Siracusa, containing field notes likely associated with an early excavation phase. This notebook holds exceptional documentary value, especially given the absence of fieldnotes from the initial stages of exploration preceding the De Miro-Florentini's interventions. It consists of pocket style notebook with 64 handwritten pages, as well as a cover and front pages (Fig. 6). The support material consists of thin yellow

paper bound with brown cardboard written with a dip pen, featuring the title "Giornale degli Scavi in Contrada Casale Piazza Armerina" written on the cover. The diary documents a six-month excavation period, spanning from January 30th to August 1st, in which excavation activities are documented on a daily basis. Yet, it lacks any explicit reference to the year of the activity, either on the front page or within the text itself. The contents offer a rare glimpse into the daily operations of the excavation, which elucidate real-time methodological decisions, and reference structural elements (architecture, masonry) that no longer exist due to subsequent dismantling during later work.



Fig. 3. Image of 3D digitization activities at the Villa del Casale. Upper left, digital photogrammetry of the figural mosaics. Upper right, results of the digital photogrammetry of the interior of the Villa complex. Lower left, terrestrial laser scanning of the exterior of the complex. Lower right, overview map of the terrestrial laser scanning of the exterior of the Villa complex.

3. Materials and methods

3.1. Unlocking handwritten legacy data in archaeology with HTR and AI: methodologies, challenges and opportunities

As is common for historical excavation manuscripts, the Piazza Armerina notebook provides unique insights into early fieldwork techniques and methodologies, preserving important details about stratigraphy, artifact distribution, and excavation decisions that might otherwise remain undocumented. However, early archaeological practices often lacked standardized recording methods and terminologies, resulting in heterogeneity and genericity as prevailing trends. Furthermore, despite being complete, the manuscript is both unauthored and undated. These aspects significantly complicate the interpretation of this document and others of its kind, thereby requiring specific strategies to address these challenges. While digitization is crucial for preservation and accessibility, transcription remains indispensable for converting static images into searchable and analyzable text formats, ultimately enabling a more comprehensive understanding of the archaeological data, the potential identity of the author, and the broader historical context in which the manuscript was created.

In the present project, Transkribus was employed to accelerate and enhance the decoding of the handwritten excavation diary of the Villa del Casale of Piazza Armerina. Originating from the EU-funded READ project, Transkribus is an open-source, online platform for automated handwritten text recognition (HTR), with a focus on historical documents (Muehlberger et al., 2019). Fully aligned with Open Science

movement, Transkribus is available both as a Graphical User Interface (GUI) and as a web-based application. Relying on optical character recognition (OCR), it integrates cutting-edge machine learning and artificial intelligence to transcribe handwritten documents. At its core, Transkribus operates on a deep neural network architecture specifically tailored for sequence-to-sequence learning. Banking on examining and processing digitized images and transcriptions, this enables the software to identify and interpret a particular style of writing by recognizing patterns in the arrangement of letters, words, and lines. To train, validate, and test the HTR model on specific handwritten patterns, a properly curated set of manually transcribed training materials, defined as ground truth, must be provided. These materials serve as benchmarked data, and are essential for ensuring transcription accuracy (Muehlberger et al., 2019, 957; Romein et al., 2024, 3). The more complex the handwriting, the greater the amount of ground truth required to achieve reliable transcriptions.

Transkribus documentation recommends at least 15,000 words of diplomatic transcription to train an effective HTR model. Producing such a dataset from scratch for a corpus of 64 pages would require extensive manual transcription, negating the anticipated time-saving benefits. Prior experiments indicated that errors frequently occur in the recognition of accents, punctuation, or word tokenization, particularly in historical manuscripts with *scripta continua* writing styles. Given the relatively consistent handwriting of the diary, experiments were conducted to minimize manual effort while ensuring reliable transcription. As demonstrated in previous studies, optimal manuscript selection combined with transfer-learning techniques can reduce the

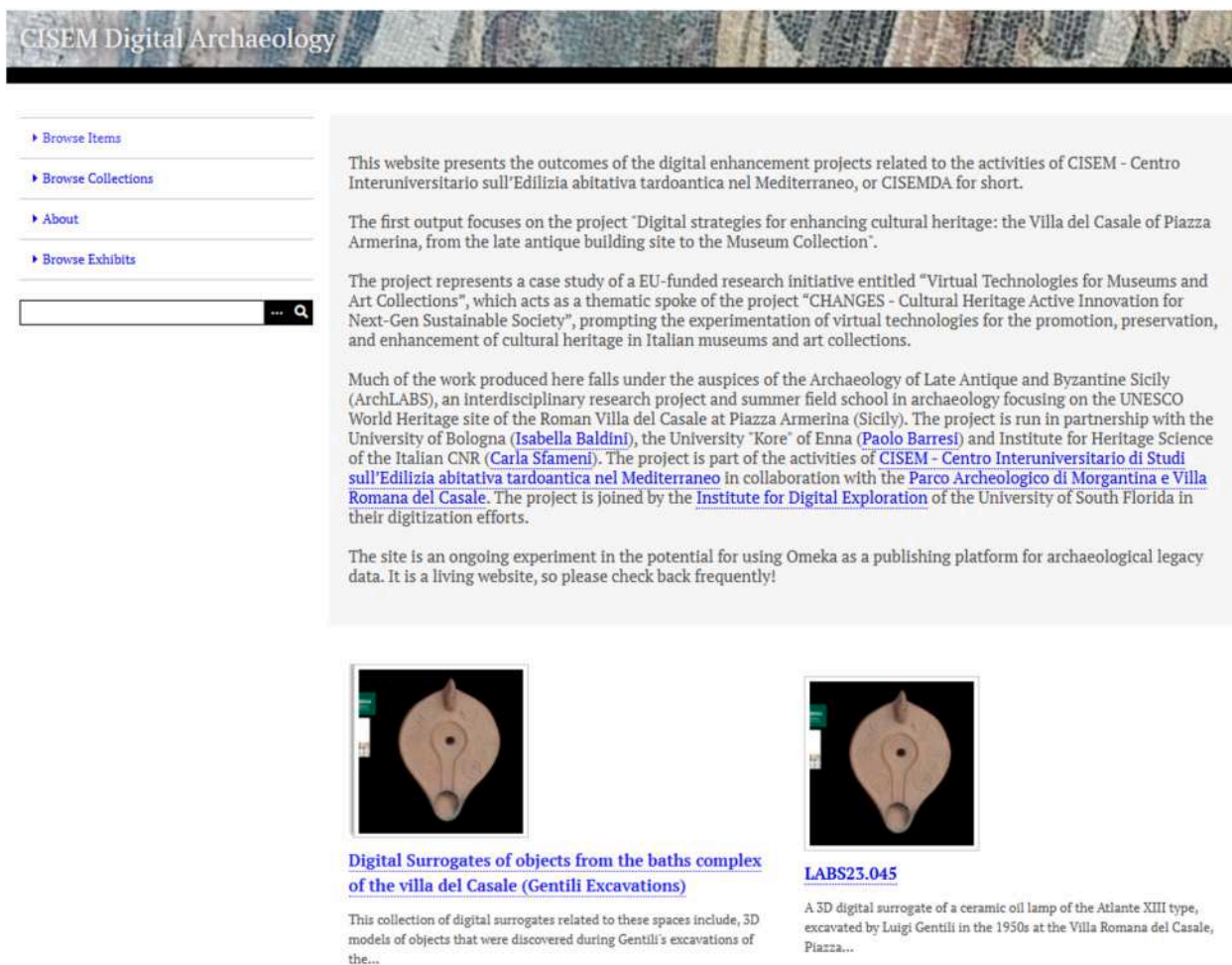


Fig. 4. Image of the home page of the Omeka Web Collection created for the project.

required training set size by up to 80 %, significantly decreasing manual transcription efforts (Perdiki, 2023). Given the digitized format supplied by the Superintendence, which consisted of double-page images, the first challenge addressed was text segmentation—dividing the document into regions and baselines. Major noise was created by the presence of portions of text inserted over the principal line, compelling manual adjustments. Later, the model has been trained on a relatively limited dataset of labelled textual input corresponding to the 25 % of the total (1133 words). Ground truth has been combined to existing public models with low Character Error Rate (CER) percentage sharing similar features (chiefly Italian language, 17th–20th c.). To this end, three different public models have been tested, namely the “Italian - 20th Century Minutes of Mediobanca’s Board of Directors & Executive Committee” (dataset: 90729 words; CER 2.38 %), the “Italian Administrative Hands 1550–1700” (dataset: 67361 words; CER 12.20 %) and Transkribus Italian Handwriting M1 (dataset 653630 words; CER 6.70 %). After some testing, the CER has been lowering significantly by cleaning major issues and obtaining text normalization (Fig. 7).

The selection of pretrained HTR models was guided by linguistic affinity (Italian language), chronological coherence (17th–20th century), and reported low Character Error Rate (CER) scores, used as indicators of transcription reliability. The models tested demonstrated varying degrees of effectiveness depending on the complexity of the manuscript. While the “Italian - 20th Century Minutes” model consistently produced the lowest CER (2.38 %), some inaccuracies persisted, particularly in the recognition of punctuation, diacritics, and word segmentation. Although transfer learning significantly reduced the required training set, human intervention remained necessary to

normalize outputs, correct semantic ambiguities, and fine-tune key archaeological terminologies. Overall, a hybrid approach combining automated transcription with targeted manual post-editing proved essential to ensure both accuracy and interpretability.

Further research task consists in the semantic enrichment of the digital manuscript, indexing document images to enable semantic searches. As stated in literature, this process, largely performed on machine-printed documents, poses some challenges for handwritten document images (Tüselmann and Fink, 2022). Named Entity Linking has been performed to recognize entities and links them to their corresponding entries in a Knowledge Base. To this end, Transkribus tagging functionality has been implemented to annotate the text and provide a knowledge base for data mining. In addition to Transkribus’s default structural, textual, and relational tags, new textual tags have been introduced to reflect archaeological concepts such as layers, structures, and materials (e.g., stone, pottery, mosaic) (Fig. 8). Following the standard entity linking architecture, for all these entities named by entity recognition, candidate generation from Wikidata (that are automatically implemented in the Transkribus tag section) and disambiguation were implemented to produce a semantically enriched text. A subsequent phase of the research will focus on searching, reusing, and integrating the extracted information - geolocated with precision - into existing GIS environments. This integration will enable the visualization of the extent of the excavations and the spatial distribution patterns of artifact categories, thus improving the user experience.

Internal references within the manuscript allow for a relatively precise localization of excavation activities in the area north of the Triconch hall. Multiple mentions of prior excavations and the revelation

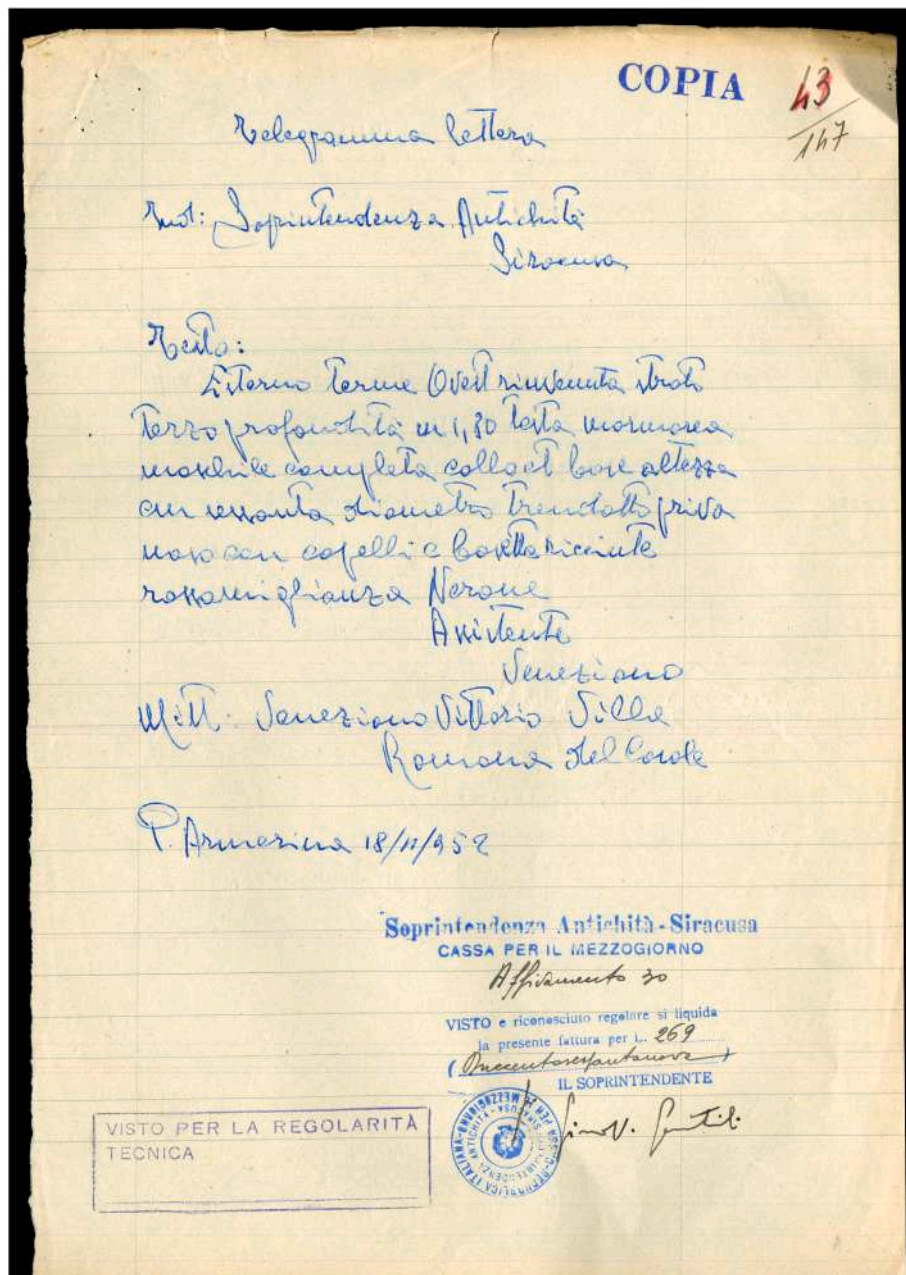


Fig. 5. Letter of the local attendant V. Veneziano to G.V. Gentili announcing the discovery of a marble statue in 1952 (Courtesy of Soprintendenza per i Beni Culturali e Ambientali di Siracusa, Proprietà della Regione Siciliana, Assessorato dei Beni Culturali e dell'Identità Siciliana - Dipartimento dei Beni Culturali e dell'Identità Siciliana - Soprintendenza per i Beni Culturali e Ambientali di Siracusa - reproduced with permission).

of mosaic floors appear to correspond to the mythological cycle depicted in the apses of the Triclinium. References to different modern cadastral properties (Ciancio, Milazzo, Pergola) at the boundaries of the expropriated zone further support this hypothesis. Therefore, the sector under investigation as described in the notebook likely corresponds to the area between the northern and western apses of the Triclinium, and the southern apse of the Great Hunt corridor, where external walls featured remnants of plastered and painted decorations. Frequent references to piles of marble and mosaic tesserae imply a gradual dismantling of the lavish decorations of the late antique residence, with these materials stockpiled in open areas for reuse or storage.

Moreover, the field notes hold significant potential for providing insights into medieval structures that were later dismantled during subsequent phases of the excavation. These structures are occasionally interpreted as small residential units and, in other instances, as segments

of the aqueduct. Their later chronology is suggested by both the described masonry's quality and typology, and the associated material culture. These data contribute to enriching the understanding of the settlement patterns of the Medieval village established over the ruins of the late antique building, adding new structures located in the area between the Triclinium and Great Hunt ambulatory to the picture provided by recent research (Baldini et al., 2025).

3.2. An excavation diary in search of an author: overcoming anonymity issue matching AI with historical data

As already mentioned, one of the major issues in analysing Piazza Armerina field diary consists in the lack of data about authorship and year of composition. The extensive history of research conducted at the Villa, involving a succession of prominent figures, including eminent



Fig. 6. Cover and first page of the Piazza Armerina notebook (Courtesy of Soprintendenza per i Beni Culturali e Ambientali di Siracusa, Proprietà della Regione Siciliana, Assessorato dei Beni Culturali e dell'Identità Siciliana - Dipartimento dei Beni Culturali e dell'Identità Siciliana - Soprintendenza per i Beni Culturali e Ambientali di Siracusa - reproduced with permission).



Fig. 7. Screen captures of Transkribus web-app: left, Piazza Armerina notebook collection with baselines and regions; right, training and validation statistical data of Character Error Rate (CER).



Fig. 8. Screen capture of Transkribus web-app: tag function.

archaeologists and their respective assistants, hampers a secure attribution of the excavation diary to a specific research season. To overcome these challenges, a detailed internal analysis of the information contained in the text was combined with artificial intelligence tools for analysing, sorting, and identifying handwriting styles of different

authors.

As mentioned, the field notes describe in detail the daily activities carried out between January 30th and August 1st. Both working days and holidays are noted down, along with their respective activities. Digging works, practicalities of the excavation, number of workers

involved are the usual contents of ferial days, while excursions in Sicilian cities, visits of public officials (e.g. the Superintendent, the Prefect), and small restoration activities usually take place in festive days. Relevant for setting the chronological framework is the mention of Easter holidays, from Holy Friday to Monday, which occupy the days 19-20-21 and 22 of April. According to the Gregorian calendar, April 21st corresponds to the Easter day in the years 1867, 1878, 1889, 1935, 1946, 1957. Comparing these dates to what we know about the history of the research at Villa del Casale, potential matching dates correspond to Cultrera's excavations in 1935 and Gentili's restoration season in 1957. These dates refer to very different scenarios with reference to the advancement of discoveries, the portions of structures already uncovered, as well as the different excavation methodologies employed. Therefore, useful insights in attributing the diary may come from a thorough evaluation of its contents.

Firstly, relevant logistical details regard the mention of Decauville wagons to remove the backfill. The railway line is known to have been installed among the structures of the Villa since the campaigns of the 1930s (Nigrelli and Vitale, 2010, 106) and then extensively utilized and renewed during the Gentili's excavations, as also documented by the correspondence preserved in the archives of the Superintendency of Syracuse (Fig. 9). Secondly, the connection to previous excavation campaigns, seemingly aimed at bringing to light some mosaics, is taken as a benchmark to direct the new excavating operations. In the first days 14 workers were engaged in digging trenches to verify the depth of the archaeological basin, until they reach the undisturbed soil at 2.9 m below the ground level. In the next weeks, further trenches were opened

to verify the existence and position of mosaics previously uncovered.

The description of the excavation activities repeatedly refers to operations such as the removal of earth and clearing of various material. Regarding findings, sherds from different periods were found and mentioned in brief terms. The documents demonstrate a fair amount of knowledge about prehistoric pottery, cited in relation to the productive *facies* (e.g., early Sicilian pottery "ceramica del primo periodo siculo"; "ceramica del terzo periodo siculo" etc.). The description of later materials is more generic, among which the presence of glazed ware ("un vaso in terracotta stagnato di colore verde mare, molto probabilmente arabo lavorato con arabeschi a bassorilievi") can sporadically be inferred along with other abundant medieval materials, which are sometimes illustrated by sketches (Fig. 10). Therefore, the general framework gives the impression of an early excavation season, conducted out of a stratigraphic framework, and directed by an archaeologist with a broad background, well rooted also in the knowledge of prehistoric chronologies.

These elements likely suggest an attribution of the excavation diary to the earliest period of G. Cultrera excavations, which, thanks to funding from the Municipality of Piazza Armerina and on the occasion of the celebrations of the Augustan Bimillennium, in 1935 resumed the excavations in the Triclinium area, previously investigated by P. Orsi (Muscolino, 2017, 176). This attribution seems to be reinforced by internal references to mosaics discovered "seven years earlier", which would refer to the investigations conducted by Orsi in 1929. In the brief notes given to the press, Cultrera reported that the work was limited to a vast and deep removal of the backfill soil, with the collection of

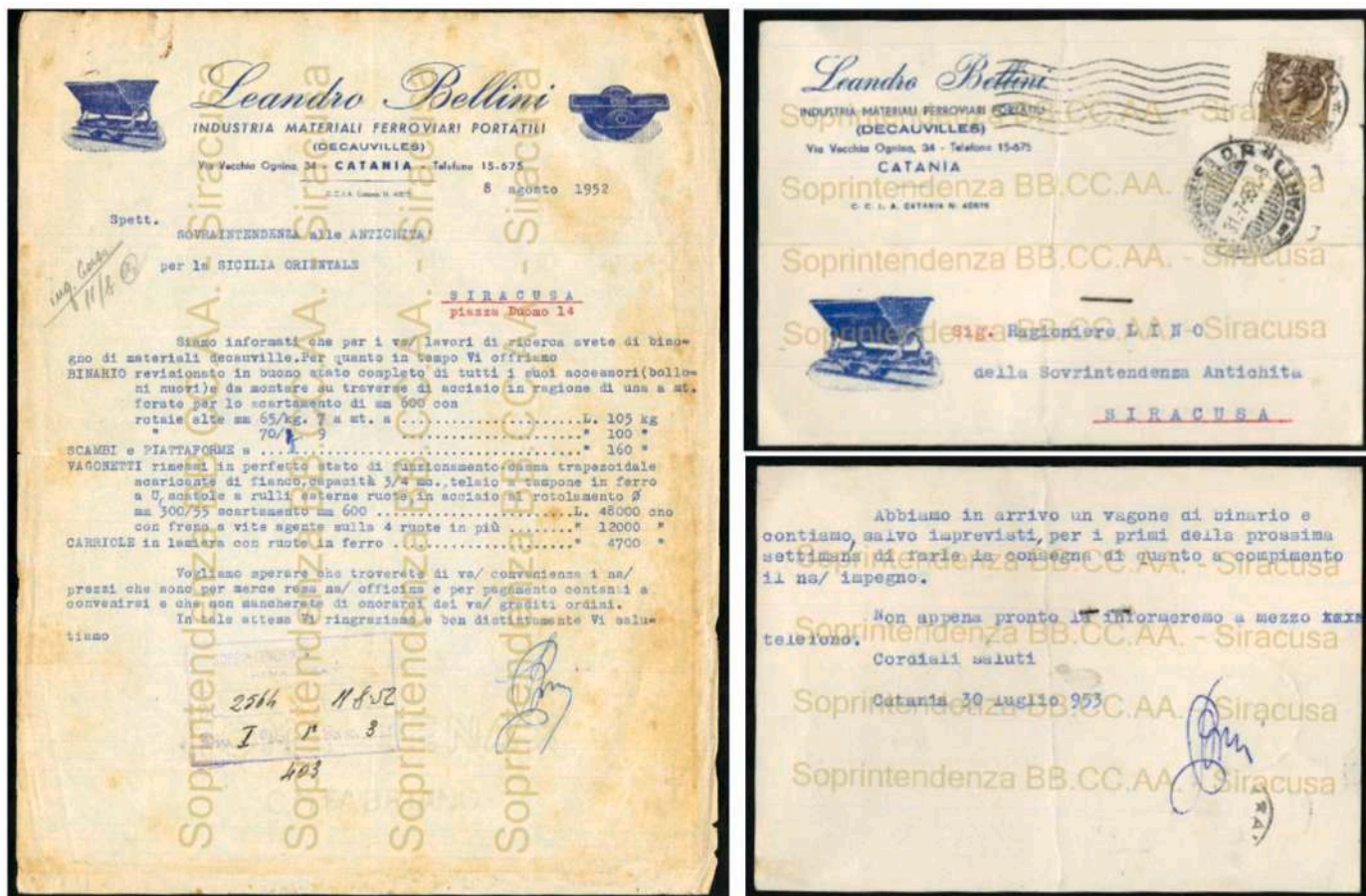


Fig. 9. Quotation from the company Leandro Bellini of Catania for the supply of Decauville wagons and rails (August 1952) and delivery letter for a spare part (July 1953) during Gentili's excavations (Courtesy of Soprintendenza per i Beni Culturali e Ambientali di Siracusa, Proprietà della Regione Siciliana, Assessorato dei Beni Culturali e dell'Identità Siciliana - Dipartimento dei Beni Culturali e dell'Identità Siciliana - Soprintendenza per i Beni Culturali e Ambientali di Siracusa - reproduced with permission).

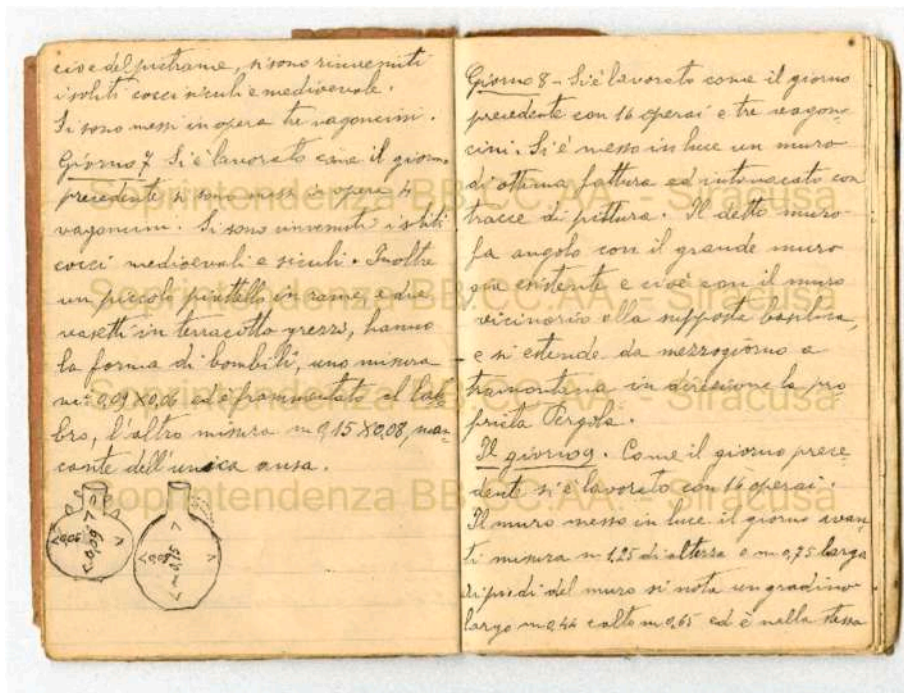


Fig. 10. Piazza Armerina notebook, page corresponding to May, 7th with a sketch of Medieval pottery (Courtesy of Soprintendenza per i Beni Culturali e Ambientali di Siracusa, Proprietà della Regione Siciliana, Assessorato dei Beni Culturali e dell'Identità Siciliana - Dipartimento dei Beni Culturali e dell'Identità Siciliana - Soprintendenza per i Beni Culturali e Ambientali di Siracusa - reproduced with permission).

intercepted archaeological material (Cultrera, 1936, 12–13; Cultrera, 1940). The management of excavation works was entrusted to Domenico Inglieri, P. Orsi's former collaborator, who was appointed site manager on the excavation site in place of the Superintendent (Bonifazio, 1950, 17–19; Alagna, 2008, 126; Inglieri, 2017, 21–23; Puglisi et al., 2017, 54–56). A brief mention of the accounts provided by Inglieri is contained in one of Gentili's first published reports, in relation to the excavation of the Triclinium (Gentili, 1950, 305).

To further verify this research hypothesis, a Handwriting Analysis Tool (HAT) was tested to attribute the handwriting to an individual style. The HAT is a software tool embedded in a set of Pattern Analysis Software Tools (PAST) developed by the Universität Hamburg that target the automatic analysis of visual and tabular patterns in the research data from the study of ancient written artefacts (Mohammed, et al., 2018, 2022). The HAT is an easy-to-use off-line software application of pattern recognition methods used for analyzing images of handwriting. It can simultaneously scan and compare multiple handwriting styles, sorting them according to their correspondence to a questioned or unknown style. The tool employs pattern recognition and statistical analysis techniques to generate similarity scores, facilitating the study of handwritten manuscripts and aiding in the identification of scribes in historical documents.

Five pages were selected from the notebook that shared the same handwriting style because they were written by the same hand and were used as "unlabelled handwriting style" to be analysed. Four labelled writing styles were then added to compare to the unlabelled one. They belong to four letters written with ink nib by four archaeologists involved in Villa del Casale excavations at different stages. The first three are preserved and digitized by the manuscripts section of the Library of the Civic Museum of Rovereto,⁴ while the fourth belongs to the

Document Archive of the Superintendence of Syracuse. They consist of: 1. a minute of a letter from P. Orsi to E. Salerno, dated to May 1932⁵; 2. a letter of G. Cultrera to P. Orsi, dated to August 1932⁶; 3. a letter of D. Inglieri to P. Orsi, dated to December 1927,⁷ and 4. a document with handwritten notes about Piazza Armerina excavation by G.V. Gentili, included in the section "Attività della Cassa per il Mezzogiorno", dated to 1952.

The unlabelled document (Piazza Armerina notebook) and D. Inglieri's handwriting obtained the highest similarity values (64,1 %), followed by Gentili (18,2 %) and Cultrera's (17,7) (Fig. 11). We might well assume that the fieldnotes of the Piazza Armerina manuscript and D. Inglieri's letter are written by the same person, thus concluding that Inglieri was the author of Piazza Armerina notebook. This conclusion aligns well with what was surmised in the analysis of the contents of the manuscript: considering the involvement of Inglieri at the very beginning of Cultrera's excavations, we might propose 1935 as a date for the manuscript composition.

4. Digitization and web-publishing

The discovery of new sources of information regarding legacy excavations at the Villa del Casale is an important part of the new campaign of research that was inaugurated through an agreement between the Parco Archeologico di Morgantina e della Villa Romana del Casale di Piazza Armerina and CISEM. An important part of the project is to divulge the new information of the analysis of the legacy data, as well

⁵ https://www.fondazionemcr.it/patrimonio_abbonamenti.jsp?area=279&ID_LINK=114702&page=1961&IDCTX=446006&id_context=446006 (last access: 18 December 2024).

⁶ https://www.fondazionemcr.it/patrimonio_abbonamenti.jsp?area=279&ID_LINK=114702&page=2167&IDCTX=446769&id_context=446769 (last access: 18 December 2024).

⁷ https://www.fondazionemcr.it/patrimonio_abbonamenti.jsp?area=279&ID_LINK=114292&page=182&IDCTX=417919&id_context=417919. (last access: 18 December 2024).

⁴ <https://www.fondazionemcr.it/archivi>

Analyse Selected Handwriting Styles

Unlabelled Handwriting Styles		Labelled Handwriting Styles	
Name of Handwriting Style	Number of Uploaded Files	Name of Handwriting Style	Number of Uploaded Files
Unlabelled 1_Piazza Armerina notebook	1	Cultrera	1
		Gentili	1
		D. Inglieri	1

Parameter Selection

Currently Selected settings: Keypoints Detection Algorithm: FAST
Main Keypoint Parameter: 10

Analysis Results

Results for Unlabelled 1_Piazza Armerina notebook:

Style Name	Score (Relative Similarity)
D. Inglieri	64,1%
Gentili	18,2 %
Cultrera	17,7 %

Fig. 11. Screen capture of the HAT system webpage: processing section and analysis results with keypoints.

as the legacy data itself, in its myriad forms, in a format that approaches the principles of FAIR scientific publishing. USF IDEx became a partner of the project to aid in the capture, analysis, and dissemination of the Villa itself, the excavations, and, most pertinently to this contribution, the legacy data from past excavations at the Villa del Casale. The Villa del Casale, being such an important site to the Roman archaeology of Sicily, is relatively poor accessible from a digital perspective. While the initial publication of the excavations cited above have covered the importance of certain architectural aspects of the Villa and much of its mosaics, there has been a paucity of investigation into the Villa's legacy documentation and archaeological materials. Much of this is due to relative inaccessibility to the materials as the Cultural Heritage Institutions (CHIs) responsibility to preserve the data conflicts with the ability to make the data accessible and requires a great deal of investment from outside sources. Following a trend that can be identified in Sicily and elsewhere in Roman archaeology, the relative lack of accessibility to data regarding the excavations from Villa del Casale, both in the sense of a relative paucity of published materials from earlier excavations and lack of accessibility to documentation and objects, has thus stifled the Villa's central importance to Roman archaeology of Sicily as well as reduced its impact on the wider field of Roman archaeology (Hassam, 2023). However, CHIs are largely preoccupied with the preservation of the materials under their authority and often do not have the resources to engage in digitization efforts, or if they are able to digitize their materials, effectively disseminate them to the public in an accessible way. It becomes the responsibility of projects such as the international Summer School "ARCHLABS. Archaeological Heritage in Late Antique and Byzantine Sicily"⁸ to utilize resources and expertise

available to improve accessibility to the public and academic communities on behalf of CHIs.

The realization of the vital importance of making Cultural Heritage Objects (CHOs) available has been made abundantly clear in the past few years, as CHIs and tourism economies were devastated by a global pandemic and continue to be threatened geopolitical conflicts such as those in Ukraine and Syria, sparking an increased interest and urgency to digitization projects in museums across the world (Carandini et al., 1982; Christiansen, 2021; Giannakouloupolos et al., 2021). Digitization itself is not enough, however, as the contents generated need to be disseminated in an appropriate way so as to be properly accessible to the public and the scholarly community. Thus, one of the principal goals of the project is to make the legacy data of the Villa del Casale available to scholars and the public with an eye towards FAIR principles in scientific publishing (Wilkinson et al., 2016). As the digital humanities have expanded, there has been a large number of Content Management Systems (CMSs) have been developed to accommodate various project and disciplinary needs, but there is no "one size fits all" solution (Toscano et al., 2023).

4.1. Digitization of the Villa and its legacy data

With the increase of the availability of technologies and infrastructures that permit researchers to create born digital research projects, and the increasing understanding by CHIs to engage in digitization of collections for digital forms of accessibility for the public, digitization of archaeological artefacts for museums, excavations, and the documentation of legacy data is well developed and constantly being improved (Barzaghi et al., 2024). The University of South Florida's Institute for Digital Exploration (USF IDEx) has had as its mission since its inception the priorities of providing Virtual Accessibility to

⁸ <https://site.unibo.it/piazza-armerina-cisem/it/summer-school>

archaeological sites and museum collections that are not physically accessible or have a limited digital presence, as well as creating digital replicas or surrogates for cultural heritage artefacts and collections that are in danger or not physically accessible.⁹ Beginning in 2017, USF IDEX engage in a campaign of digitization of 3D of the Villa del Casale to create digital accessibility to its mosaics. As sharing the entirety of the model was much too large to visualize online, each individual room was made into its own model to be uploaded to Sketchfab. Unfortunately, the decimation of the model reduced the metric quality of the scans, and the textures are not as photorealistic as those in the original models (Gabellone et al., 2020). This activity resulted in the production of digital twins of 38 rooms with mosaic floors, currently accessible on a collection dedicated to ARCHLabs on Sketchfab (Fig. 12). As a partner in the new excavations and legacy data project, USF IDEX aims to achieve three objectives, critically reusing the 3D visualizations produced in past campaigns for (Tanasi and Hassam, forthcoming), producing new digital assets of the Villa's exterior aimed at documenting and completing the technical documentation of the Villa, and finalising the digital documentation in 3D of the legacy data in order to disseminate them in light of new information gleaned from archival research. The latter objective will be discussed in this section.

There are many digital tools using various web protocols and data structures, but there are no ideal solutions for the dissemination of digitized information in 3D format or otherwise (Champion and Rahman, 2020; Toscano et al., 2023). While there are a number of viable 3D repositories and visualization tools used by CHIs and archaeological projects, such as 3DHop (Galeazzi et al., 2016) and ATON (Fanini et al., 2021), to this day, the most accessible 3D repository for cultural heritage is the private company Sketchfab (Champion and Rahman, 2020; Garstki et al., 2020). For this reason, USF IDEX continues to use it as the primary repository and visualizer for its digital products in 3D. As the goal of the project is to increase the digital footprint of the excavations of the Villa and its legacy data, we needed to produce a way to store large quantities of data as and link that data with its metadata. While the digitization efforts carried out indeed have a great deal of potential to aid in research, public outreach, and conservation efforts related to the Villa, the products of the Sketchfab collections do not themselves meet the criteria to be used reused in research according to FAIR principles. One of the primary roadblocks to the effective dissemination of archaeological information is that it is separated into data silos that are not accessible across disciplinary boundaries, nor can they be reused in a manner apart from their original use, thus reducing the utility of the data. The development of web-sourced publishing has the potential to overcome many of these boundaries, but the method in which it is done is of central importance (Harrison, 2022). Thus, it was decided to experiment with a CMS that would have the capability to share a wide range of material culture types in a malleable system that would follow a standard metadata schema while also allowing for the creation of collections that would suit the needs of public outreach.

4.2. Experimentation with Omeka Classic

Given the limited resources typically available to CHIs, we also propose a modular and interoperable framework aimed at facilitating the long-term sustainability and incremental expansion of this digital ecosystem, with minimal infrastructural demands. After a survey of viable solutions, it was decided to experiment with the use of the Omeka open-source CMS. An important tool that has been increasingly used by GLAM (Galleries, Libraries, Archives, and Museums) institutions, the software's name comes from the Swahili word meaning "to display" or "lay out for discussion". The Omeka web publishing platform is a CMS created by the Center of History and New Media at George Mason

University designed to create an exhibition and archival space (Scheinfeldt, 2008). Since its inception in 2008, it has been continuously developed and has been used increasingly frequently for the dissemination of archival materials and cultural heritage objects. Omeka, now known as the Omeka Classic, was a free to use and low-cost alternative CMS that required significantly less investment in than creating one's own and was designed for use by small institutions to create an open platform for digital collections. In 2010, a hosting service called Omeka Classic was released to allow users to use the software without the necessity of installing the software and launching one's own webserver. In 2017, the CMS Omeka S was released for use by large scale organizations, while allowing the management of multiple websites based on a single database. However, this last version of the software, while being the most versatile, is also the costliest in terms of time investment and required expertise.

In all cases, the Omeka CMS has the potential to cost-effectively create a collection of many different material types, as long as their metadata fit into the Dublin Core metadata schema (Weibel and Koch, 2000). Being a well-established metadata schema that can easily be adapted to material culture (Mi and Pollock, 2018; Marsili and Orlandi, 2019; Giannakoulopoulos et al., 2021) has contributed to its potential for use as a repository for archaeological information. In the educational sphere, the open-source web platform made it a popular tool, allowing academics to create teaching tools to be used in teaching courses on Roman archaeology (Webber, 2016)¹⁰ The ability to allow user submissions and comments also adds a participatory factor to Omeka. For example, the Museum of Looted Antiquities offers the ability for users to submit collections of looted objects that have been repatriated in order to preserve information about their ownership history that can be used to track illicit trade.¹¹

While designed for use by GLAMs, it has increasingly been experimented with for the publication of archaeological datasets. The Omeka CMS is continuously updates plugins and utilities designed with an eye towards the Open Science movement, including the systematization of metadata and the mass import of images and bibliographies (Rueff, 2024). Though there has not been full integration of web-based 3D viewers into Omeka itself, it is possible to link content in Omeka with outside 3D web viewers (Vagts and Athassapoulos, 2016), though this has not been widely practiced in the CMS. While there have been recent advances in CMSs focusing on archaeological excavations, such as AIR and Heurist (Derudas et al., 2023), the metadata standards needed by the project for the publication of 2D excavation journals and relevant photographic information and the ease with which Omeka can be used to create digital exhibitions in the future is better suited to needs of the project. With this in mind, an Omeka page was created to host the legacy data being assessed within the scope of the project, including both archival materials analysed in sections 3.1, 3.2, and the digitized legacy data discussed in section 4.1 (Fig. 13).¹² It contains collections referring to a selection of the products of the digitization of legacy data in 3D in the form of lamps related to the Gentili excavations of the 1950s, as well as items related to the new information gleaned from the Piazza Armerina notebook.

5. Conclusions

This study presents a part of renewed research activities at the Villa del Casale of Piazza Armerina and the integration of interdisciplinary and multimodal documentation, analysis, and publication of legacy data originating from one of the premier Roman villas in the Mediterranean region. The integration of advanced AI tools such as Transkribus and HAT for the transcription and enrichment of the excavation diary of

⁹ <https://www.usf.edu/arts-sciences/institutes/idex/about/mission-and-goals.aspx>

¹⁰ <http://omeka.wellesley.edu/piranesi-rome/>

¹¹ <https://mola.omeka.net/>.

¹² <https://www.cisemda.com/>.

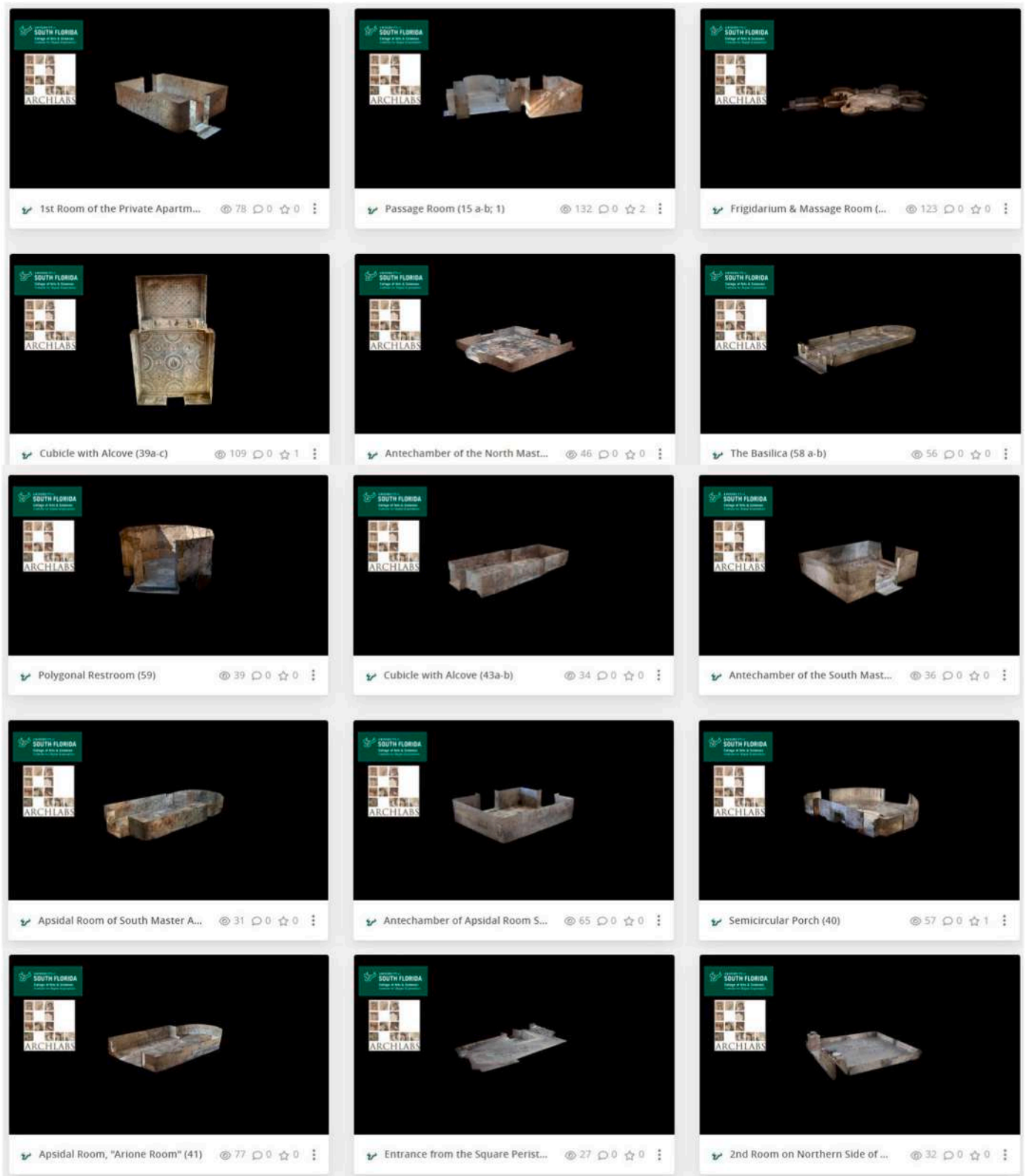


Fig. 12a. Screen capture of the digital collection of the decorated rooms of the Villa del Casale on Sketchfab, created by the Institute for Digital Exploration at the University of South Florida.

Piazza Armerina successfully demonstrate how these technologies can overcome challenges related to variability, genericity, and incomplete authorship information when confronted with the analysis of handwritten archaeological documents. By leveraging machine learning models and pattern recognition software, researchers can significantly

reduce manual effort, enhance transcription accuracy, and extract semantic data for further analysis. Additionally, the identification of the diary's author and its chronological placement provide a valuable context for understanding early excavation methodologies and the historical evolution of the site. These findings reinforce the role of AI in

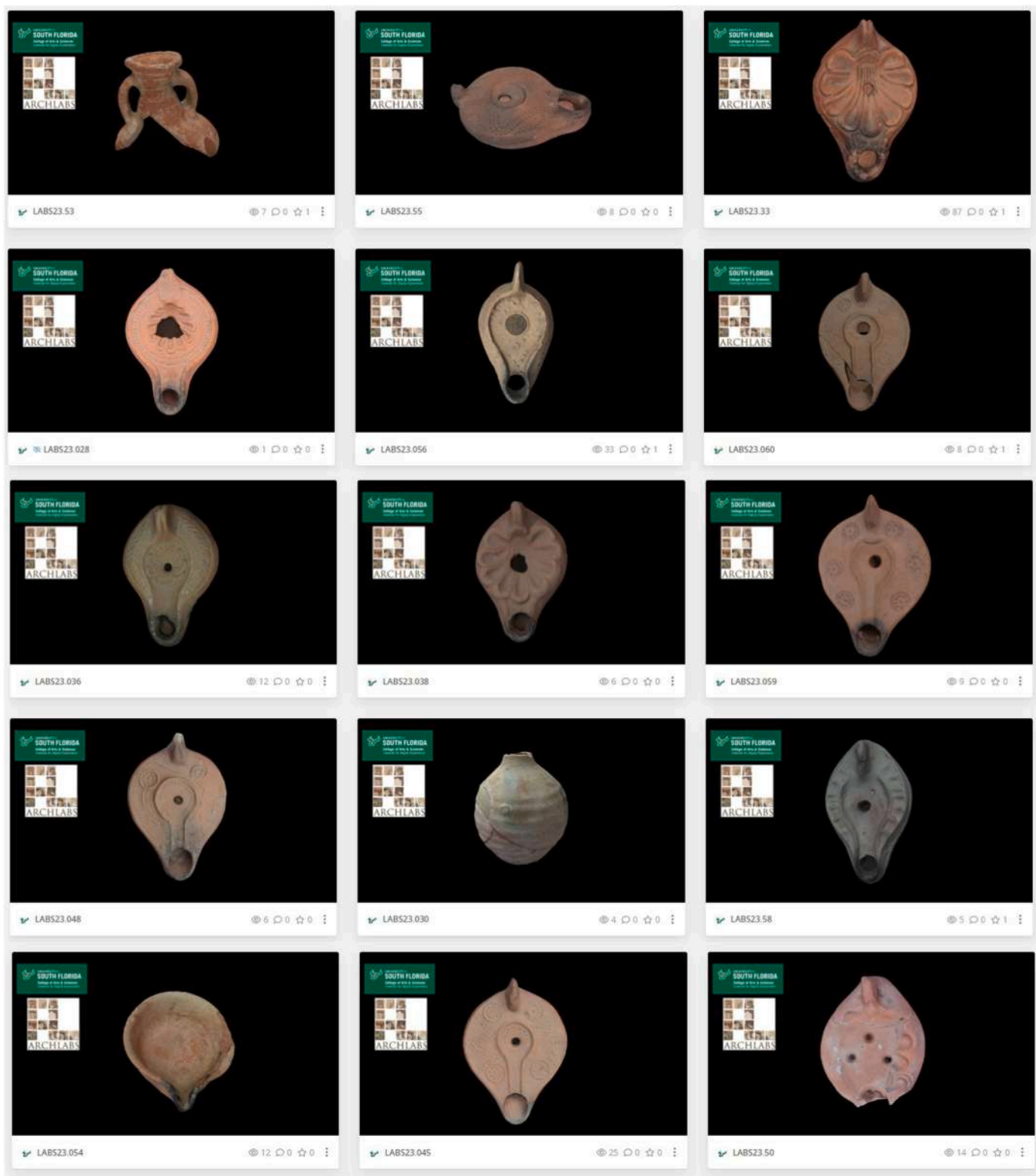


Fig. 12b. Screen capture of the digital collection of legacy data from Gentili’s excavations of the western bath complex of the villa del Casale on Sketchfab, created by the Institute for Digital Exploration at the University of South Florida.

advancing archaeological research, offering innovative approaches to preserving and interpreting the cultural heritage embedded in handwritten documents.

The collection, digitization, and analysis of legacy data is a crucial step forward for understanding the Villa and the project aims to maximize the utility of these activities by following principles of the FAIR

publishing. Thus, the project has outlined the rationale for choosing and experimenting with the CMS Omeka Classic to present legacy data to the public and scholarly community. By using a platform geared towards GLAM style CHIs, in which documents and objects can be created using a standardized metadata schema, the project can subsequently produce exhibits for the public, as well as engage in a new form of sharing

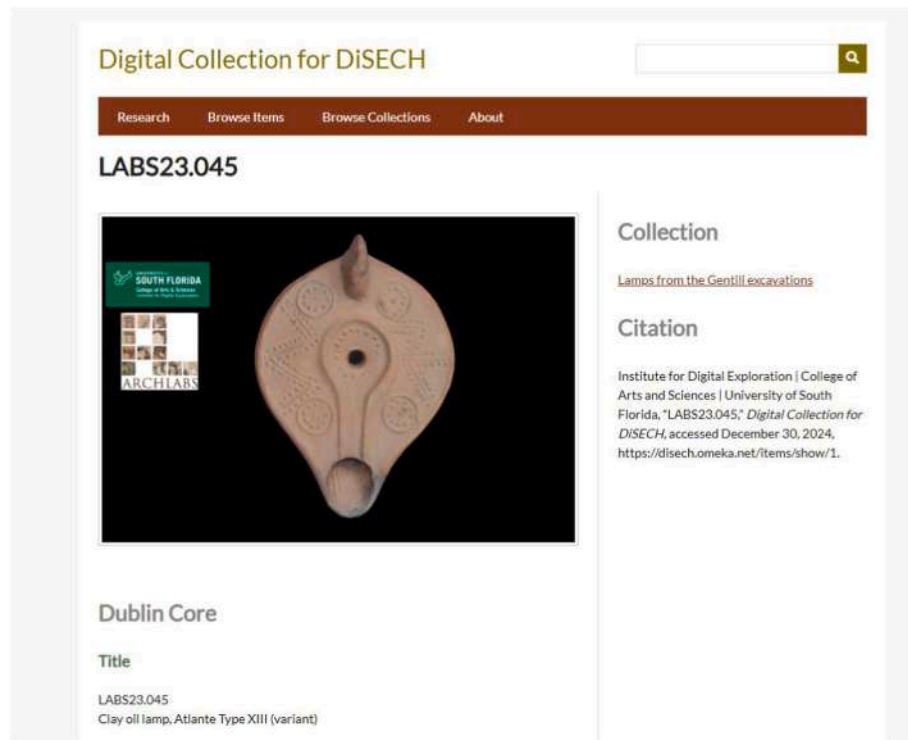


Fig. 13. Example of a landing page for one of the items in the digital collection of objects on the Omeka Platform.

academic research. As a free software, Omeka Classic was explicitly designed to be a low-cost CMS aimed at small institutions. The relatively limited cost in resources and expertise of hosting a server for the content makes it a sustainable platform for the project that can be used by local CHIs for their own purposes.

The possibility of freely accessing online digital twins of archival documents and artefacts preserved in storage will provide users with enriched cultural experiences. This approach can be adapted to other archaeological or museal projects with legacy data challenges, promoting the wider the use, and perhaps more importantly, wider dissemination, of legacy data pertaining to old excavations. Furthermore, the proposal to place a QR code within the Palazzo Trigona Museum - a venue housing numerous artefacts excavated from the Villa - will not only enhance the visitor experience but also serve as an ideal connection between the archaeological site, the Museum and the archives, that is geographically distant sites that preserve the millennia-old history of the Villa del Casale, a cultural crossroads for centuries.

CRedit author statement

Stephan Hassam: Writing – review & editing, Writing – original draft (Sections 4.1, 4.2, 5), Visualization, Validation, Software, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. Giulia Marsili: Writing – review & editing, Writing – original draft (Sections 1, 2, 3.1, 3.2, 5), Validation, Supervision, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was partly funded by Project PE 000020 CHANGES - CUP

B53C22003780006, NRP Mission 4 Component 2 Investment 1.3, Funded by the European Union – NextGenerationEU.

The authors sincerely thank the directors of the archaeological mission at Villa del Casale of Piazza Armerina, prof. I. Baldini, P. Barresi, and C. Sfameni, as well as prof. D. Tanasi, responsible for the 3D digitization and enhancement project, for granting permission to carry out this research.

References

- Agnello, S.L., 1965. La Villa romana di Piazza Armerina ai primi dell'Ottocento. *Arch. Stor. Siracus.* 11, 57–77.
- Alagna, A., 2008. Franco Minissi. Restauro e musealizzazione dei siti archeologici in Sicilia, Tesi di Dottorato. Università Degli Studi Di Napoli Federico II.
- Ampolo, C., Carandini, A., Pucci, G., Pensabene, P., 1971. La villa del Casale a Piazza Armerina. *Problemi, saggi stratigrafici ed altre ricerche. Mélanges de l'école française de Rome* 83 (1), 141–281.
- Baird, J.A., 2011. Photographing dura-Europos, 1928-1937. *An archaeology of the archive. Am. J. Archaeol.* 115 (3), 427–446.
- Baldini, I., Sfameni, C., De Giorgi, L., Barbolla, D.C., Ferrari, I., Giuri, F., Torre, C., Leucci, G., 2024. Villa del Casale Piazza Armerina (EN): new data related to geophysical investigations. *J. Appl. Geophys.* 227. <https://doi.org/10.1016/j.jappgeo.2024.105439>.
- Baldini, I., Barresi, P., Sfameni, C., Tanasi, D., 2025. La ripresa delle ricerche alla villa del Casale di Piazza Armerina: nuovi dati e prospettive per la storia dell'insediamento "post villan". In: Cavalieri, M., Castrorao Barba, A., Sfameni, C. (Eds.), *La Villa Dopo la Villa*, 3. Trasformazione di un Sistema Insediativo ed Economico nell'Italia Meridionale E Nelle Isole Maggiori tra Tarda Antichità e Medioevo, Louvain, pp. 181–206.
- Balzani, R., Barzaghi, S., Bitelli, G., Bonifazi, F., Bordignon, A., Cipriani, L., Colitti, S., Collina, F., Daquino, M., Fabbri, F., Fanini, B., Fantini, F., Ferdani, D., Fiorini, G., Formia, E., Forte, A., Giacomini, F., Girelli, V.A., Gualandi, B., Heibi, I., Iannucci, A., Manganelli Del Fà, R., Massari, A., Moretti, A., Peroni, S., Pescarin, S., Renda, G., Ronchi, D., Sullini Tini, M.A., Tomasi, F., Travaglini, L., Vittuari, L., 2024. Saving temporary exhibitions in virtual environments: the digital renaissance of ulisse aldrovandi – acquisition and digitisation of cultural heritage objects. *Digit. Appl. Archaeol. Cult. Herit.* 32. <https://doi.org/10.1016/j.daach.2023.e00309>.
- Bonanno, C., 2020. *Piazza Armerina. L'area nord dell'insediamento medievale presso la Villa del Casale. Indagini Archeologiche 2013-2014.* ArchaeoPress, Oxford.
- Bonifazio, A., 1950. *La Città Di Piazza Armerina E Suoi Dintorni Con Speciale Riguardo Agli Scavi Archeologici Del Casale.* Ferrara.
- Carandini, A., Ricci, A., De Vos, M., 1982. *Filosofiana, la Villa di Piazza Armerina. Immagine Di Un Aristocratico Romano Al Tempo Di Costantino.* Flaccovio ed. Palermo.

- Champion, Erik, Rahaman, Hafizur, 2020. Survey of 3D digital heritage repositories and platforms. *Virtual Archaeol. Rev.* 11 (23), 1–15. <https://doi.org/10.4995/var.2020.13226>.
- Chiarandà, G.P., 2025. Piazza città di Sicilia antica, nuova, sacra, e nobile. In: Messina: per Gl'Heredi Di Pietro Brea.
- Cultrera, G., 1936. Scavi, Scoperte E Restauri Di Monumenti Antichi in Sicilia Nel Quinquennio 1931-1935, 24. E. F., ASIPS, Siracusa, pp. 1–5, 11-13.
- Cultrera, G., 1940. Piazza armerina. *Bollettino Comunale* 68, 129–130 notiziario.
- De Miro, E., 1988. La Villa del Casale di Piazza Armerina: nuove ricerche. In: Garaffo, S. (Ed.), *La Villa Romana Del Casale Di Piazza Armerina, Atti Della IV Riunione Scientifica Della Scuola Di Perfezionamento in Archeologia Classica Dell'Università Di Catania (Piazza Armerina, 28 settembre-1 Ottobre 1983)*, Cronache Di Archeologia Suppl. 23, pp. 58–73.
- Derudas, P., Nurra, F., Svensson, A., 2023. New AIR for the archaeological process? The use of 3D web semantic for publishing archaeological reports. *J. Comput. Cult. Herit.* 16 (3), 1–23. <https://doi.org/10.1145/3594722>.
- Ellena, S., 2023. Gli scavi De Miro a Piazza Armerina. *Campagne Di Scavo Degli Anni 1983, 1986, 1988, Tesi Di Laurea, Università Di Bologna*.
- Fanini, B., Ferdani, D., Demetrescu, E., Berto, S., d'Annibale, Enzo, 2021. ATON: an open-source framework for creating immersive, collaborative and liquid web-apps for cultural heritage. *Appl. Sci.* 11 (22), 11062. <https://doi.org/10.3390/app112211062>.
- Frey, J., Raja, R. (Eds.), 2024. *Trends in Archive Archaeology. Current Research on Archival Material from Fieldwork and its Implications for Archaeological Practice*. Brepols, Turnhout.
- Garstki, K., Counts, D.B., Averett, E.W., Kansa, S.W., Kansa, E.C., 2020. A square peg in a round hole: rethinking archaeological publication. *Near E. Archaeol.* 83 (4), 264–269. <https://doi.org/10.1086/712391>.
- Gabellone, F., Chiffi, M., Tanasi, D., Decker, M., 2020. Integrated technologies for indirect documentation, conservation and engagement of the roman mosaics of Piazza Armerina (Enna, Italy). In: Cicalò, Enrico (Ed.), *Proceedings of the 2nd International and Interdisciplinary Conference on Image and Imagination, Advances in Intelligent Systems and Computing*, 1140. Springer International Publishing, Cham, pp. 1016–1028. https://doi.org/10.1007/978-3-030-41018-6_83.
- Galeazzi, F., Callieri, M., Dellepiane, M., Charno, M., Richards, J., Scopigno, R., 2016. Web-Based visualization for 3D data in archaeology: the ADS 3D viewer. *J. Archaeol. Sci.: Report* 9, 1–11. <https://doi.org/10.1016/j.jasrep.2016.06.045>.
- Gentili, G.V., 1950. Piazza Armerina, grandiosa Villa romana in contrada «Casale. *Atti dell'Accademia Nazionale dei Lincei, Notizie degli scavi di antichità.* S 291–335. VIII, IV, fasc. 1-6.
- Gentili, G.V., 1999. *La Villa Romana di Piazza Armerina Palazzo Erculio*. Fondazione Don Carlo, Osimo.
- Giannakouloupolous, A., Pergantis, M., Poulimenou, S.M., Deliyannis, I., 2021. Good practices for web-based cultural heritage information management for Europeana. *Information* 12, 1–12.
- Hassam, S., 2023. *Visualizing Empire: Advancing Roman Archaeology in Sicily and Malta through 3D Digitization and Spatial Technologies*. FL: University of South Florida, Tampa. Dissertation.
- Harrison, L.K., 2022. Closing the loop on the digital data lifecycle: reviving a salvage archaeology dataset. In: Garstki, K. (Ed.), *Critical Archaeology in the Digital Age: Proceedings of the 12th IEMA Visting Scholar Conference*, Cotsen Digital Archaeology Series. The Cotsen Institute of Archaeology Press, UCLA, pp. 79–96.
- Inglieri, D., 2017. L'uomo e la sua vita. In: Inglieri, D., Puglisi, D., Schembari, M.R., D'Amato, T.V. (Eds.), *Raffaello Umberto Inglieri. Vita e Opera di un Archeologo Comisano*. Pro Loco Comiso, Comiso, pp. 21–38.
- Kingsland, K., 2023. *Archaeogaming and the Re-Use of Digital Archaeological Materials: Generating Serious Games for the Villas of Roman Sicily*, PhD diss. University of South Florida.
- Marsili, G., Orlandi, L.M., 2019. Digital humanities and cultural Heritage preservation: the case of the BYZART (Byzantine Art and Archaeology on Europeana) Project. *Studies in Digital Heritage* 3 (2), 144–155.
- Marsili, G., 2024. Digital strategies for enhancing cultural heritage: the Villa del Casale of Piazza Armerina project, from legacy data to digital ecosystem. *Archeologia e Calcolatori* 35.2, 445–454. <https://doi.org/10.19282/ac.35.2.2024.46>.
- Mi, X., Pollock, B.M., 2018. Metadata schema to facilitate linked data for 3D digital models of cultural heritage collections A university of South Florida libraries case Study. *Cataloging Classif. Q.* 56, 273–286.
- Mohammed, H., Märgner, V., Stiehl, H.S., 2018. Writer identification for historical manuscripts: analysis and optimisation of a classifier as an easy-to-use tool for scholars from the humanities. In: *16th International Conference on Frontiers in Handwriting Recognition (ICFHR)*. Niagara Falls, NY, pp. 534–539.
- Mohammed, H., Helman-Wazny, A., Colini, C., Beyer, W., Bosch, S., 2022. Pattern Analysis Software Tools (PAST) for written artefacts. In: Uchida, S., Barney, E., Eglin, V. (Eds.), *Document Analysis Systems. DAS. Lecture Notes in Computer Science*, 13237. Springer, Cham. https://doi.org/10.1007/978-3-031-06555-2_15.
- Muehlberger, G., Seaward, L., Terras, M., Ares Oliveira, S., Bosch, V., Bryan, M., Colutto, S., Déjean, H., Diem, M., Fiel, S., Gatos, B., Greinoecker, A., Grüning, T., Hackl, G., Haukovaara, V., Heyer, G., Hirvonen, L., Hodel, T., Jokinen, M., Kahle, P., Kallio, M., Kaplan, F., Kleber, F., Labahn, R., Lang, E.M., Laube, S., Leifert, G., Louloudis, G., McNicholl, R., Meunier, J.-L., Michael, J., Mühlbauer, E., Philipp, N., Pratikakis, I., Puigcerver Pérez, J., Putz, H., Retsinas, G., Romero, S., Sablatnig, R., Sánchez, J.A., Schofield, P., Sfikas, G., Sieber, C., Stamatopoulos, N., Straub, T., Terbul, T., Toselli, A.H., Ulreich, B., Villegas, M., Vidal, E., Walcher, J., Weidemann, M., Wurster, H., Zagoris, K., 2019. Transforming scholarship in the archives through handwritten text recognition: Transkribus as a case study. *J. Doc.* 75 (5), 954–976. <https://doi.org/10.1108/JD-07-2018-0114>.
- Muscolino, F., 2017. Giuseppe Cultrera Soprintendente in Sicilia: politica, scavi e restauri (1931-1941). In: Capaldi, C., Dally, O., Gasparri, C. (Eds.), *Archeologia E Politica Nella Prima Metà Del XX Secolo Incontri, Protagonisti E Percorsi Dell'Archeologia Italiana E Tedesca Nel Mediterraneo*, Atti Delle Giornate Internazionali Di Studio Napoli, 24-26 Febbraio 2016. Naus Editoria, Pozzuoli, pp. 167–180.
- Nigrelli, F.C., Vitale, M.R., 2010. *Piazza Armerina: Dalla Villa Al Parco*. Studi E Ricerche Sulla Villa Romana Del Casale E Il Fiume Gela, Biblioteca Del Cenide, Reggio Calabria.
- Nigrelli, F.C., Vitale, M.R., 2014. Dalla Villa al paesaggio. Il tema della protezione e della musealizzazione del sito archeologico di Piazza Armerina fra esigenze conservative, concezione del paesaggio e pianificazione del territorio. In: *La Villa Restaurata E I Nuovi Studi Sull'Edilizia Residenziale Tardoantica*, Atti Del Convegno Internazionale Del CISEM (Piazza Armerina, 7-10 Novembre 2012), pp. 651–662. Edipuglia, Bari.
- Orsi, P., 1934. *Romanità e avanzi romani di Sicilia*. Roma, rivista di studi e di vita romana 12 (6), 253–260.
- Pace, B., 1951. Note su di una Villa romana presso Piazza Armerina. *Atti Dell'Accademia Nazionale Dei Lincei, Rendiconti Della Classe Di Scienze Morali, Storiche E Filologiche*. S. VIII. VI, pp. 454–476 fasc. 11-12.
- Pace, B., 1955. *I Mosaici Di Piazza Armerina*. Casini, Roma.
- Pensabene, P., Barresi, P., 2019. *Piazza Armerina. Villa Del Casale: Scavi E Studi Nel Decennio 2004-2014. L'Erma di Bretschneider*, Roma.
- Perdiki, E., 2023. Preparing big manuscript data for hierarchical clustering with minimal HTR training. *J. Data Min. Digit. Human.* 1–13. <https://hal.science/hal-03880102v3>.
- Pizzi, M., 2023. Legacy Data e Sistemi Informativi Geografici: proposti di progetto GIS sui dati d'archivio degli scavi Gentili nella Villa del Casale di Piazza Armerina (EN). Tesi di Specializzazione, UNIBO.
- Puglisi, 2017. Raffaele Umberto Inglieri in Grecia: un percorso esemplare fra archeologia e guerra. In: Inglieri, D., Puglisi, D., Schembari, M.R., D'Amato, T.V. (Eds.), *Raffaele Umberto Inglieri. Vita E Opera Di Un Archeologo Comisano*. Pro Loco Comiso, Comiso, pp. 53–74.
- Raja, 2023. *Shaping archaeological archives. Dialogues Between Fieldwork, Museum Collections, and Private Archives*. Brepols, Turnhout.
- Romein, C.A., Hodel, T., Gordijn, F., Zundert, J. J. van, Chagué, A., Lange, M. van, Jensen, H.S., Stauder, A., Purcell, J., Terras, M.M., Heuvel, P., Keijzer, C., Rabus, A., Sitaram, C., Bhatia, A., Depuydt, K., Afolabi-Adeolu, M.A., Anikina, A., Bastianello, E., et al., 2024. Exploring data provenance in handwritten text recognition infrastructure: sharing and reusing ground truth data, referencing models, and acknowledging contributions. Starting the conversation on how we could get it done. *Journal of Data Mining & Digital Humanities. Documents histories et reconnaissance automatique de texte*. <https://doi.org/10.46298/jdmhd.10403>.
- Rueff, Bastien, 2024. Dealing with post-excitation data: the omeka S TIMMA web-database. *Peer Community J.* 4 (January), e5. <https://doi.org/10.24072/pcjournal.366>.
- Sfameni, C., 2006. L'insediamento medievale: la documentazione degli scavi precedenti. In: Pensabene, P., Sfameni, C. (Eds.), *Iblatasah Placea Piazza. L'Insediamento Medievale Sulla Villa Del Casale: Nuovi E Vecchi Scavi, Mostra Archeologica*. Piazza Armerina, pp. 81–90.
- Stoler, A.L., 2009. *Along the Archival Grain: Epistemic Anxieties and Colonial Common Sense*. Princeton University Press, Princeton.
- Swain, H., 2012. *Archive archaeology*. In: Skeates, R., McDavid, C., Carman, J. (Eds.), *The Oxford Handbook of Public Archaeology*. Oxford University Press, Oxford, pp. 351–372.
- Tüselmann, O., Fink, G.A., 2022. Named entity linking on handwritten document images. In: Uchida, S., Barney, E., Eglin, V. (Eds.), *Document Analysis Systems, Proceedings of 15th IAPR International Workshop, DAS 2022 (La Rochelle, France, May 22-25, 2022)*. Springer, Cham, pp. 199–213.
- Vennarucci, Rhodora, Fredrick, David, Tanasi, Davide, Reynolds, Nicholas, Kingsland, Kaitlyn, Jenkins, Brianna, Hassam, Stephan, 2021. In Ersilia's Footsteps: toward an Interactive WebGL Application for Exploring the Villa Romana Del Casale at Piazza Armerina, Sicily. In: Ganovelli, Fabio, McDonald, Carol, Banterle, Francesco, Potenziani, Marco, Callieri, Marco, Jung, Yvonne (Eds.), *The 26th International Conference on 3D Web Technology*. Association for Computing Machinery, New York, pp. 1–7. <https://doi.org/10.1145/3485444.3487646>. Pisa, Italy - November 8 -12, 2021.
- Ward, C., 2022. Excavating the archive/archiving the excavation: archival processes and contexts in archaeology. *Adv. Archaeol. Pract.* 10 (2), 160–176.
- Webber, Monique, 2016. *Art of Empire: forum of augustus*, 2016. <https://moniquewebber.omeka.net/exhibits/show/forumofaugustus/starthere>.
- Weibel, S.L., Koch, T., 2000. The Dublin core Metadata initiative: mission, current activities, and future directions. *D-Lib Mag.* 6. <https://doi.org/10.1045/december2000-weibel>.
- Wilkinson, M.D., Dumontier, M., Aalbersberg, I.J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., Da Silva Santos, L.B., Bourne, P.E., Bouwman, J., Brookes, A.J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C.T., Finkers, R., Gonzalez-Beltran, A., Gray, A.J.G., Groth, P., Goble, C., Grethe, J.S., Heringa, J., 'T Hoen, P.A.C., Hooft, R., Kuhn, T., Kok, R., Kok, J., Lusher, S.J., Martone, M.E., Mons, A., Packer, A.L., Persson, B., Rocca-Serra, P., Roos, M., Van Schaik, R., Sansone, S.-A., Schultes, E., Sengst, T., Slater, T., Strawn, G., Swertz, M. A., Thompson, M., Van Der Lei, J., Van Mulligen, E., Velterop, J., Waagmeester, A., Wittenburg, P., Wolstencroft, K., Zhao, J., Mons, B., 2016. The FAIR guiding principles for scientific data management and stewardship. *Sci. Data* 3. <https://doi.org/10.1038/sdata.2016.18>.