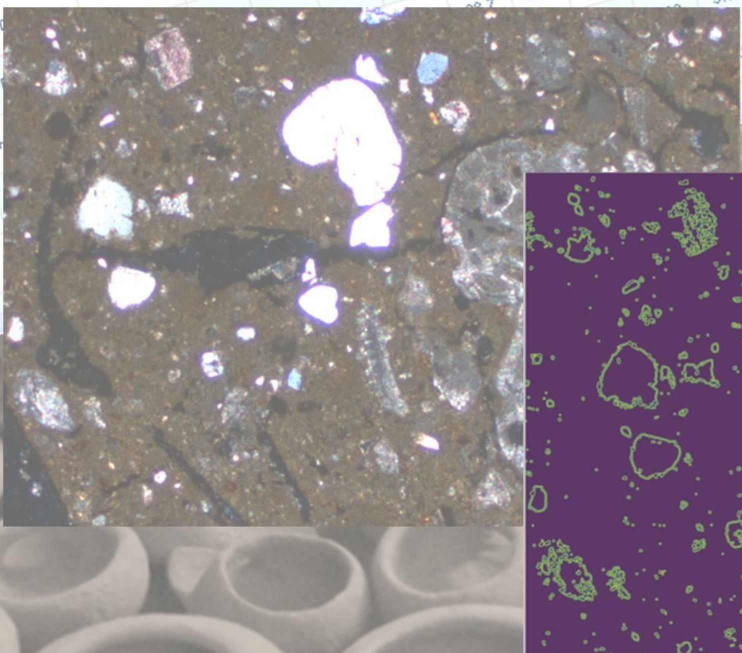
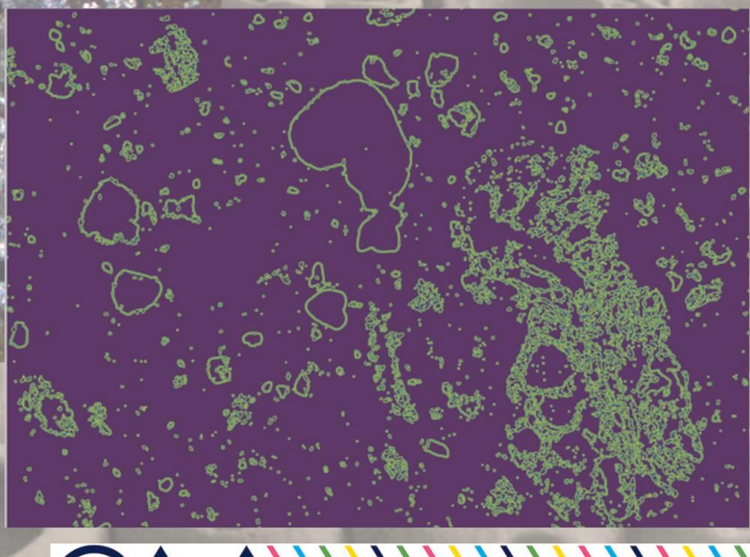
Big Data in Archaeology

Proceedings of the 4th Conference on
Computer Applications and Quantitative Methods in Archaeology
Greek Chapter (CAA-GR)
21-22 October 2021

Edited by

Anno Hein





NATIONAL CENTER FOR SCIENTIFIC RESEARCH "DEMOKRITOS"

CAA Greece Computer Applications & Quantitative Methods in Archaeology

Big Data in Archaeology

Proceedings of the4th Conference on Computer Applications and Quantitative Methods in Archaeology Greek Chapter (CAA-GR)

21-22 October 2021

Edited by

Anno Hein

Institute for Nanoscience and Nanotechnology – N.C.S.R. “Demokritos”

Computer Applications and Quantitative Methods in Archaeology – Greek Chapter (CAA-GR)

ISBN 000-000-00000-0-0

Athens, 2024

Table of Contents

Preface

The CAA-GR conference '*Computer Applications and Quantitative Methods in Archaeology, Greece*' is a biennial meeting with focus on research in the Eastern Mediterranean Region. After the previous conferences in Rethymno (2014), Athens (2016) and Limassol (2018) the 4th CAA-GR conference was initially planned to be held at N.C.S.R. "Demokritos" in October 2020. Under the circumstances of the beginning COVID-19 crisis, though, it was not possible to organize a conference hosting a larger number of participants at the N.C.S.R. "Demokritos" campus at that time. For this, we decided to postpone the conference for one year. Unfortunately, even though the situation had somewhat improved by 2021, we finally had to accept the fact that hosting even a limited number of participants was still not possible. As a further postponement appeared not to be reasonable the 4th CAA-GR conference with a particular focus on 'Big Data in Archaeology' was held as pure Web Event on 21st / 22nd October 2021 after all, gathering an audience of 58 registered participants. During the two day web conference 25 papers were presented covering subjects such as Data Management, Data Modeling, Statistics, 3D Modeling, Multimedia, GIS, Prospection Methods and Machine Learning.

The twelve papers included in the present conference proceedings provide an extract of the topics discussed during the CAA-GR 2021. Even though, we did not have the chance to meet in person then we would like to thank all participants and presenters for their contributions and support. We wish to thank as well the Scientific Committee of the 4th CAA-GR and the Steering Committee of the CAA-GR for helping us with reviewing the conference abstracts and the manuscripts submitted for the present proceedings.

Anno Hein

CAA-GR Steering Committee

Emeri Farinetti, *Chair*
George Pavlidis, *Deputy Chair*
N. Pinar Özgüner Gülhan, *Treasurer*
Theodora Moutsiou, *Secretary*
Dorina Moullou, *Publication Officer*
Stella Sylaiou, *Social Media, Communications and Outreach*

4th CAA-GR Conference 'Computer Applications and Quantitative Methods in Archaeology, Greece', 21st/22nd October 2021 (*Web Event*)

Organizing Committee

Anno Hein, Institute of Nanoscience and Nanotechnology, N.C.S.R. „Demokritos“
Vassilis Kilikoglou, Institute of Nanoscience and Nanotechnology, N.C.S.R. „Demokritos“
Ioannis Karatasios, Institute of Nanoscience and Nanotechnology, N.C.S.R. „Demokritos“
Vangelis Karkaletsis, Institute of Informatics and Telecommunications, N.C.S.R. „Demokritos“
George Petasis, Institute of Informatics and Telecommunications, N.C.S.R. „Demokritos“
Lida Arnellou, Institute of Informatics and Telecommunications, N.C.S.R. „Demokritos“

Scientific Committee

Athos Agapiou, *Dep. of Civil Engineering and Geomatics, Cyprus University of Technology*
Andreas Angourakis, *McDonald Institute for Archaeological Research, University of Cambridge*
Jacopo Bonetto, *Department of Cultural Heritage, University of Padova*
Despina Catapoti, *Dep. of Cultural Technology and Communication, University of the Aegean*
Panos Constantopoulos, *Dep. of Informatics, Athens University of Economic and Business*
Sylviane Déderix, *French School at Athens / AEGIS, UC Louvain*
Basilis Gatos, *Institute of Informatics and Telecommunications, N C.S R. "Demokritos"*
Dimitris Grigoropoulos, *German Archaeological Institute, Athens*
Avshalom Karasik, *The Institute of Archaeology, Hebrew University*
Markos Katsianis, *Dep. of Cultural Heritage Management & New Technologies, University of Patras*
Dimitrios Makris, *University of West Attica, Dep. of Conservation of Antiquities and Works of Art*
Maria Mertzani, *Directorate of Conservation of Ancient & Modern Monuments, General Directorate of Antiquities & Cultural Heritage*
Yiannis Papadatos, *Dep. of History and Archaeology, National and Kapodistrian University of Athens*
George Pavlidis, *Institute for Language and Speech Processing, 'Athena' Research Centre*
Maria Roussou, *Dep, of Informatics & Telecommunications, National and Kapodistrian University of Athens*

FIELD SURVEY OF A COLD-WAR LANDSCAPE IN SOUTHEAST BULGARIA

T. VALCHEV[1], A. SOBOTKOVA[2], P. HERMANKOVA[2] AND B. WEISSOVA[3]

[1] Yambol Regional History Museum, Bulgaria,
tvvulchev@gmail.com
[2] Department of History and Classical Studies, Aarhus University, Denmark
adela@cas.au.dk; petra.janouchova@gmail.com
[3] Institute of Archaeological Studies, Ruhr University, Bochum, Germany
bweissova@seznam.cz

Abstract

This article describes the reuse of a FAIMS2.6 digital recording module for mound monitoring and the results of pedestrian survey in the Bolyarovo municipality, southeast part of the Yambol Province in Bulgaria and a former Cold-War border. This deployment demonstrated the benefits of a generalised but customisable platform, which made digital capture of a wide range of archaeological features in the field possible despite a novice workforce and a brief campaign.

***Keywords**:* mobile computing, landscape archaeology, field data capture, pedestrian survey, burial mounds

Bolyarovo Campaign of 2018

The Bolyarovo municipality is situated in the south-eastern part of Yambol Province, Bulgaria, on the border with Turkey. The municipality covers 667 sq kms of rolling terrain leading up to the northern slopes of the Strandzha Mountain (Figure 1). The main rivulet winding through is the Popovska River.

While the Yambol Province has been the focus of intensive archaeological research since the end of the 19th century, Bolyarovo municipality has been a Cold-War border zone with restricted access and a "blank space" on the archaeological map of Bulgaria. The first excavation in the municipality took place in 1999.

Figure 1. Bolyarovo municipality landscape. (Photo: A. Sobotkova; CC-BY-4.0).

In the autumn of 2018, Todor Valchev and Dr. Adela Sobotkova organized an archaeological field survey in the Bolyarovo municipality. The project had two aims: 1) visit and register all potential burial mounds listed in old military maps, and 2) investigate the watershed of the Popovska River. To document the standing features (burial and settlement mounds, bunkers, and standing masonry) in the landscape, the team used a fully digital workflow. Specifically, we reused the 'Burial' module, a GIS-aware digital application for a systematic and offline collection of GPS points, photos, and structured data, developed on the FAIMS Mobile platform for 2017 fieldwork (Valchev and Sobotkova, 2019; Sobotkova and Weissova, 2020). The informal survey of the banks of the Popovska River was documented with long-hand notes and a handheld GPS. The 2018 record provides a foundation for future intensive surveys and for the reconstruction of past human activity in the area.

This report presents the used methodology and results from the field survey in Bolyarovo municipality. The development of computer technologies, especially mobile GIS applications in the last decades, allows us to record both the precise position of the archaeological sites as well as the status of these sites in a standardized manner. Systematic documentation coupled with precise geolocation of archaeological sites is essential for culture heritage protection during development and infrastructure projects. Reusing an Existing Data

Capture System

When selecting a data capture system, we prioritisedthe following criteria: 1) Can we easily merge data collected asynchronously across multiple devices and have mobile GIS integrated for different workflows in moble fieldwork? 2) Can we administer and run system offline? 3) Is it suitable for novise and FAIR data production? FAIMS v2.6 has been shown to facilitate all of the above (Sobotkova *et al.*, 2016; Valchev and Sobotkova, 2019; Sobotkova and Weissova, 2020). Alternative systems required interned or bilding with a desktop GIS (QField) which were a poor fir due to our degraded network and the use of R script for processing and visualization. Proprietary nature and the separation of structured and spatial data in systems such as ESRI ArcGIS Collector and Survey123 ran against the grain of FAIR data production and required reconciliation. In contrast, FAIMS supported our offline operation in 2017 and reduced administrative overhead: we were able to conduct intensive fieldwork, produce high-quality data with minimal post processing (file labeling and validation were automatic), and invest the time saving into secondary projects such as historical map digitization (Sobotkova *et al.*, 2023).

FAIMS Mobile is an open-source field data platform developed by Australian archaeologists from 2012 to 2014 (Ross *et al.*, 2013), and refined since. It generates customised Android applications for use offline (Sobotkova *et al.*, 2016; Ballsun-Stanton *et al.*, 2018), and has been deployed at 18 archaeological projects in Australia and abroad. Since 2017, FAIMS Mobile has also been used by researchers in other domains including geochemistry, ecology, and oral history; some 69 applications have been customised across all disciplines (for detailed discussion of a recent deployment, see Sobotkova *et al*., 2021).

The FAIMS team's involvement in the Yambol Province began in 2008, when Drs Shawn Ross and Adela Sobotkova started assisting the Yambol History Museum team members with digital registration of cultural heritage. Initial campaigns were dedicated to full-coverage survey, paleo-ecological sampling, and excavations (Ross *et al.*, 2010; Ross *et al.*, 2018). In 2017, Dr Sobotkova and Todor Vulchev started focusing on exclusively cataloguing burial mounds because of the mounting threat to this category of sites (Sobotkova and Weissova, 2019, 2020; Valchev and Sobotkova, 2019). This collaboration produced a module to record burial mounds (Nassif-Haynes *et al.* 2019, henceforth 'Burial'). The Burial module was designed to capture locations, morphology, and preservation status of burial mounds marked in the Soviet military maps, as well as artefact scatters and other features associated with the mounds or discovered en route (Valchev and Sobotkova, 2019; Sobotkova and Weissova, 2020). After the initial deployment in 2017, the module was improved and reused in 2018 in the Bolyarovo municipality campaign and later in 2020 at Perachora, Greece (Sobotkova *et al.* 2021).

FAIMS Platform

The 'core' FAIMS Mobile platform is an interpreter that generates a mobile application from definition files including Data Schema, User Interface (UI), and Logic (Sobotkova *et al.*, 2016, 2021; for more on FAIMS Mobile architecture and customisation see Ballsun-Stanton *et al.*, 2018). A single XML 'auto generator' file can generate all three definition files if they do not need to be manipulated separately (the Burial module was created in this manner). Users may also provide a text file to translate the interface into multiple languages. FAIMS uses an append-only datastore, intrinsically producing a data version history.

Data Schema

The Data Schema file defines entities, which contain attributes that can be assigned values. Entities can be joined by defined relationships (e.g., 'above', 'below', 'part of', etc.). Attributes can accept either a 'constrained' (i.e., a controlled vocabulary) or 'unconstrained' (i.e. free-text) value, and can include two pieces of metadata, free-text annotation and certainty on a 0–1 scale. Finally, each attribute can incorporate contextual 'help': text and images describing the attribute or how to record it.

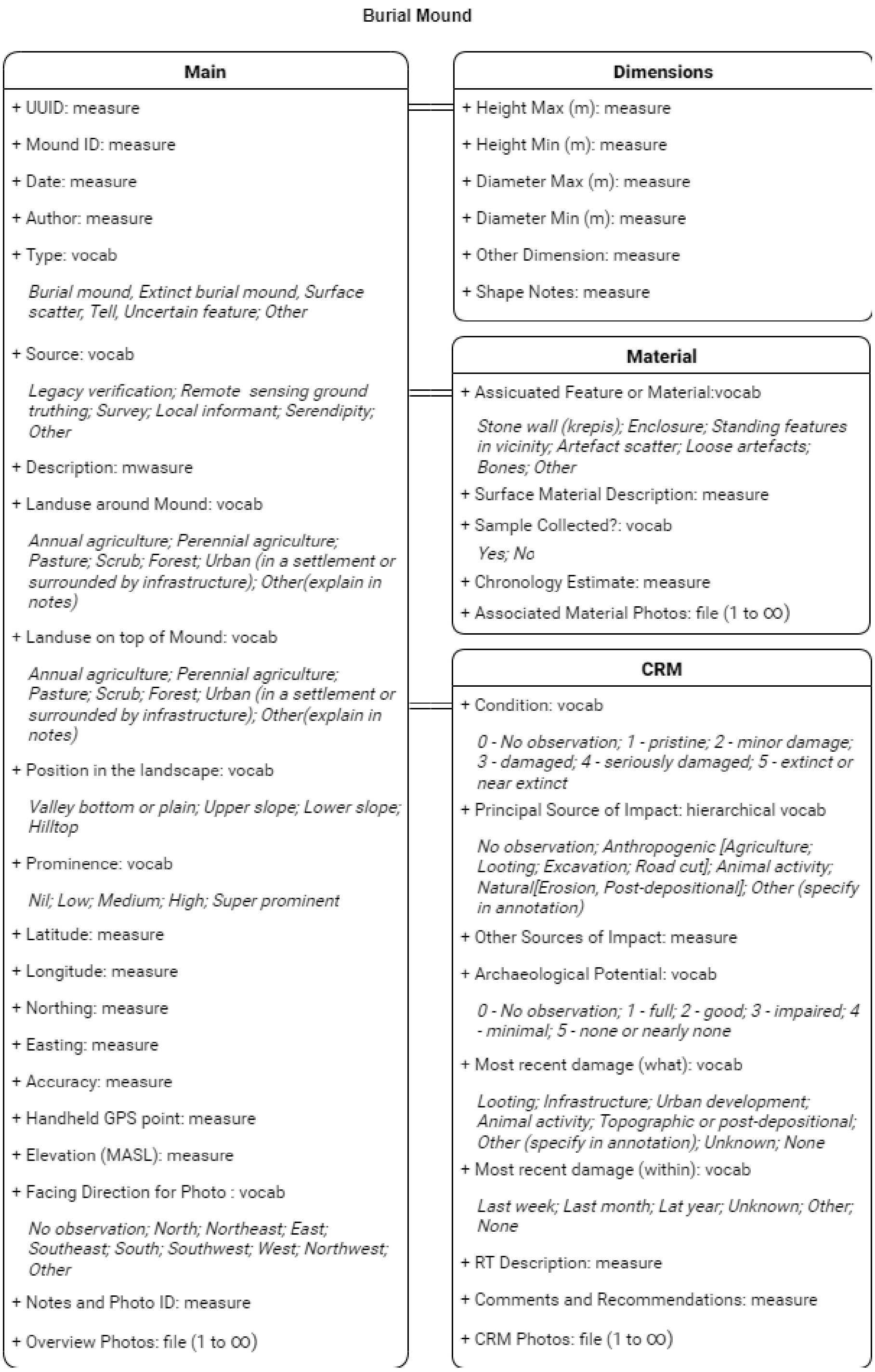


Figure 2. The Burial module data schema

The Burial module includes a single entity. A 'Burial Mound' entity records most feature attributes which are distributed across four tabs of screens, including 'Main', 'Dimensions', 'Material', 'CRM' (Figure 2). The attributes include a computer-generated universally unique identifier (UUID), record creation metadata (author and timestamp), geospatial data, description, dimensions, notes about associated finds, condition of the feature and main impacts, additional comments, and one or more device photos. The UUID remains unchanged during data manipulation and serves as the basis of relationships, but is not visible to the user. A human-readable identifier for each feature is generated by auto-incrementing a four-digit seed number set by the user.

User Interface

The module UI was designed to be simple. A user initially sees a 'Main' tab with 'Create New Record' button above a field showing the 'Next Mound ID'

(Figure 3). The Main tab allows users to reset the four-digit 'Next Mound ID' and view the GPS Diagnostics. The 'Search' tab offers free text and faceted search of existing records; 'Table' tab provides a tabular daily overview of records' essential attributes editable via 'Load' button; the 'Map' tab contains preloaded base maps and vectors and renders recorded features.

When a user taps the 'Create New Record' button, a Mound record is created, allocated a UUID, Mound ID, username, and timestamp. Users are then presented with a 'General' tab where they can capture a GPS location by tapping a button, select Type from a dropdown list, and provide Description, Land use on top and around mound, Prominence, Elevation and Comments. Buttons are also provided to take (multiple) overview photos and to photograph hand-drawn sketches. Notes and Photo ID text field allows the recording of start and end image numbers from digital SLR. An associated 'Dimensions' tab allows the entry of minimum and maximum Height and Diameter as well as other measurements and shape notes. A 'Material' tab allows specification of multiple feature or artefact material types associated with the mound (using checkboxes), sample collection checkbox, and a button to capture photographs of, e.g., artefacts. The 'CRM' tab contains dropdowns and free text fields to describe mound preservation status and sources of impact.

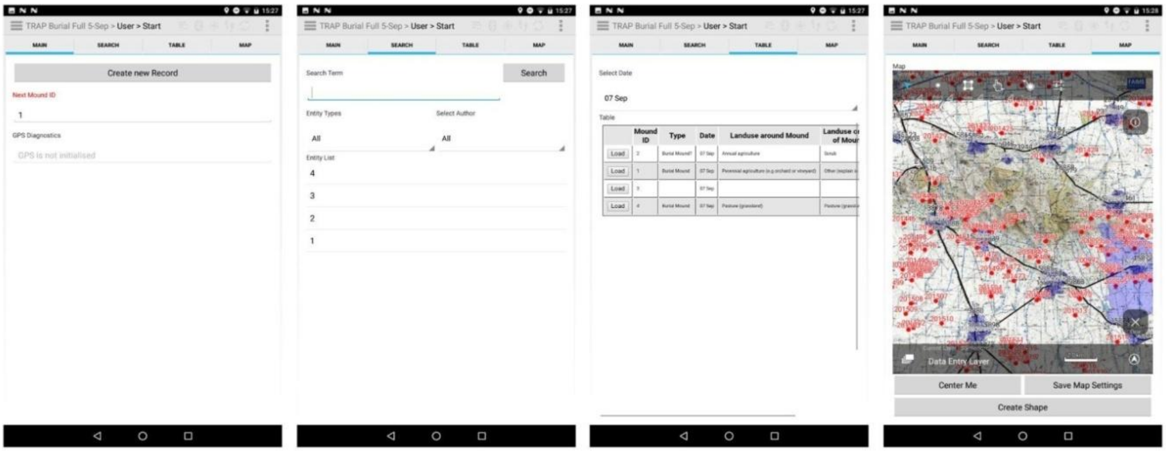


Figure 3. Screenshots from the module controls, showing navigation and search of the Burial module and the map display.

UI Logic

Tapping a 'Validate' button at the top of the General tab or the sidebar reports whether all required fields (coloured red in the UI) are complete, but does not prevent users from moving to the next record if they were not (Figure 4). Logic functions also allow the duplication of records except photos and coordinates, which proved useful when recording bunkers which varied little from one to the next.

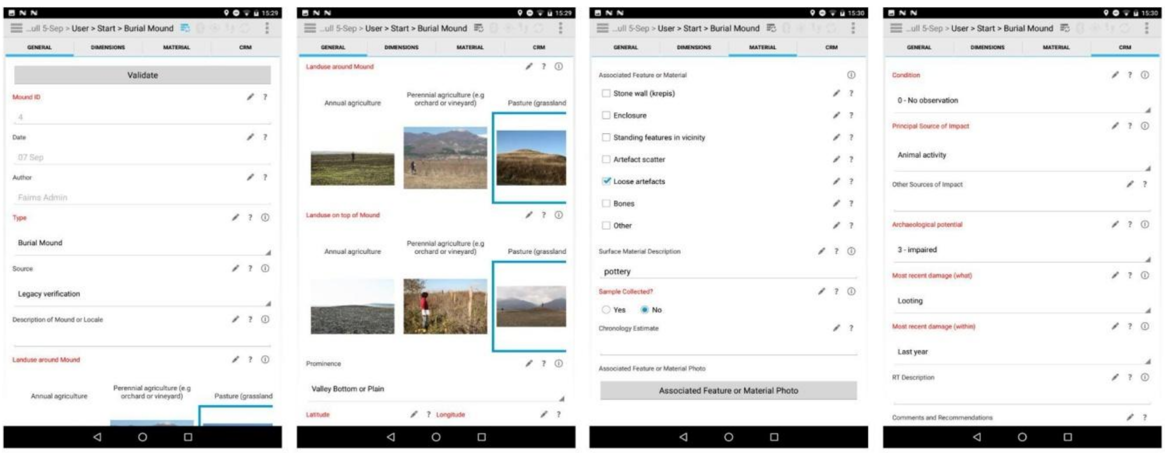


Figure 4. Screenshots from the General, Material and CRM recording tabs.

2017 to 2018 Improvements

After the 2017 season, we streamlined some of the vocabulary terms. We added missing terms such as 'Settlement mound' and 'Surface scatter' to the Type list and provided descriptions for them. We split the Prominence field into two dropdowns: 'Position in the landscape' and 'Mound prominence', with the former denoting a morphological unit such as valley, slope, or ridge and the latter being an assessment of how noticeable the mound was on approach (low, high, etc.). These changes followed in light of 2017 discussions and while not perfect (e.g. we were not expecting to find military installation and had no typology), we were happy we could modify the schema to better suit our field needs.

Data Collection

During the campaign, we worked in teams of two to three persons. Typically, there were three feature-recording teams and a fourth informal survey team. During the informal foot survey, team-members walked along the Popovska River, marking the locations of man-made concentrations with GPS, characterizing cultural material in a diary and documenting it with photographs. The Burial module was also used to make a rudimentary record. While the GPS points and photographs were downloaded upon return to base, the photos needed labelling and long-hand notes had to be transcribed and systematized upon return.

Figure 5. Recording mound-turned-into-bunker Feature 8707 in Bolyarovo municipality during September 2018 (Photo: A. Sobotkova; CC-BY-4.0).

Mound-monitoring teams visited locations of sunburst symbols, documenting their ground status. One team-member would enter data on a tablet and take photographs, while others held the scale and took measurements (Figure 5). Team-members using tablets carried spare power banks so that if batteries were low recording could continue. Each device was given a different range of Mound IDs (see Figure 3) to prevent duplication of human-readable identifiers. Whenever possible, devices were paired with Garmin Glo external Bluetooth GPS receivers to improve accuracy and time-to-fix under canopy.

The module's Map tab was furnished with topographic maps of the present area, vectors of points needing survey, and a new translation file that converted the module into Bulgarian upon loading. These changes and vocabulary descriptions led to more efficient recording of topographic prominence and improved consistency in feature types. While changes to prominence broke data compatibility with the 2017 prominence attribute, this change was exploratory and represented tweaking of a previously unusable attribute.

Data was recorded entirely offline. Working on the national border meant that the mobile network was seriously degraded. Large gaps in the coverage of both Bulgarian and Turkish mobile providers made phone communication unreliable, requiring advance planning and independent team operation until reaching the rendezvous spot.

Synchronisation and Postprocessing

Upon return to base, the module connected to a local server over the internet-free network that the server generated. Typically, all teams synchronised all tablets every day. The server also allowed users to edit data, view and revert to earlier versions of data, and export data in various formats (via a web interface). Post-fieldwork data correction was done on the server, so that all changes were captured in the data history, before data was exported as CSV and Shapefile for visualisation or analysis in other software. The remaining manual tasks were to download the backup GPSs and to label the publication-quality photos from DSLR cameras. To finalize the dataset, we would scan the manually drawn plans of burial mounds and finish diary entries and daily progress summaries for each team-day in the field. The final FAIMS-dataset was usable within a couple hours processing in OpenRefine, mostly to deaggregate (split into several columns) the structured data from the 'notes on the margins' captured in the Annotation or Certainty fields associated with each attribute.

Quality Assurance

As all FAIMS-collected data synchronisation and reference file labelling were done automatically upon export and proceeded smoothly from start, the team-leaders' tasks shifted from digitisation supervision to data quality checks. As soon as data was synchronized, team leaders reviewed validation flags and followed up any missing or incongruous data entries. We checked both the data values for consistency as well as the conceptual categories. Inter-team differences arose sometimes in the use of categorical attributes, such as mound type, condition, or archaeological potential. When teams encountered a feature, which was listed in a map as a mound, but in the field turned out to be a well-preserved military bunker, the opinion differed. Should its condition be marked as 'No observation', or 'Extinct' in relation to the mound that was supposed to be there, or 'Pristine' in relation to the 'new' military feature? While team-leaders understood the functional dependency between type and condition, the starting point (mound vs bunker) was sometimes ambiguous. Detailed data and photo review in the first week helped us streamline such procedures. Another inter-team difference that was harder to systematize was how teams gauged mound diameter. Mound boundaries cannot be always firmly established, especially if a mound is sprawling, ploughed over, and sits in rolling terrain. Walking estimates varied widely and the use of tape for measurement suffered from lacking bird's eye view. We decided to eye-ball mound diameter by standing at a distance and using a two-meter pole held by one team-member as a scale. Even these estimates varied by the observer's proximity, being more conservative (smaller) the closer one stood to a mound.

Discussions of rationale behind categories such as mound type and condition as well as dimension measurements had to happen at the start of fieldwork in order to be effective. We were able to reveal inconsistencies and address them thanks to the ease of aggregating and comparing all the structured data and the associated images within an afternoon every day from the very start of the season.

Dissemination

After the survey, visited site data were entered into the national online register of archaeological sites, the Archaeological Map of Bulgaria. Each registered site received a registration card with basic archaeological information about site type, size, location, chronology, cultural characteristics, associated materials and information about written sources. Geographic information on modern administrative borders, soils, orography, and hydrography was also provided. Finally, the database contains information about the protection status of the monument and recommendations for investigation, popularization, and socialization (Tzvetkova *et al.* 2012, 57-58; Kecheva, 2018; Valchev, Sobotkova 2019, 22).

Results

The campaign took place from 10 to 28 of September 2018. Fieldwork lasted a total of 15 recording days, during which we registered 240 features. These included one settlement mound, two ruined castles, a monastery, 29 surface concentrations, 123 burial mounds, 25 extinct burial mounds, and 59 other features (military bunkers, tank-emplacements, and uncertain features).

Field survey is a repeatable and large-scale approach to the investigation of regional past (Ammerman, 1981; Alcock *et al.*, 1994). Some of the results heavily depend on ground visibility and agricultural condition of the surveyed area. Building up a map of archaeological heritage through field surveys is a time consuming endeavour that is rarely complete. While returns increase with greater intensity, agricultural activity contributes to substantial inter-annual changes while regional interpretation is hampered by inconsistent methodologies and small sample sizes (Cherry, 1983; Alcock, 1993; Fentress, 2004; Terrenato, 2004). The main aim of the 2018 campaign was to revisit the burial mounds from the Soviet military topographic maps; however, the informal survey was productive enough to warrant a brief review of settlement evolution in the Bolyarovo municipality.

Surface Distributions Through Time

The 2018 campaign allowed us to capture the chronological distribution of sites in the municipality (Figure 6) and compare it with the regional development at large.

The earliest traces of human occupation in the watershed of the Popovska River date to the Late Neolithic (Karanovo IV culture, 5200 – 4900 BC). Material from this period was found at two artefact concentrations and one settlement mound. Late Neolithic was a period of unprecedented spread of early farmer settlements throughout the Yambol Province. 41 contemporary sites were documented in the Province including 16 settlement mounds and one pit sanctuary. This heyday of long-term farming communities owes to improved climatic conditions (specifically, increased moisture) which made agriculture possible (Connor *et al.*, 2013; Valchev, 2019, 139-140).

During the Chalcolithic period (4900 – 4300 BC), the farming communities fission and expand. Chalcolithic material is found at 91 habitations in the Tundzha River watershed (Valchev, 2019, 140), including a settlement mound near the town of Bolyarovo. This site, visited during the 2018 survey, was nearly destroyed in 1932/33 during the construction of a road and a gas-station. This development produced stone and flint tool finds, including one arrow-head, stored in the Yambol Museum archive. Other seven Chalcolithic sites are in the western part of the Bolyarovo municipality.

The Transition Period covers the span between the end of the Chalcolithic and the beginning of the Bronze Age in Bulgaria. This period of nearly 800 years is still very poorly understood (Николов and Петрова, 2013, 21-22; Valchev, 2019, 141).

The Bronze Age (3500 – 1100 BC) is a period when metal-working appears among Thracian communities. Bolyarovo residents use ore deposits from the slopes of the Strandzha Mountain to smelt metals. Two finds of oxhide ingots in the villages of Chernozem and Kirilovo in the western catchment of the Popovska River place this region within the international copper-ore trade network (Doncheva, 2012, 675-677, 686).

The Late Bronze Age (1500 – 1100 BC) is represented in the study area by one open-air settlement near the Gorska Polyana village. Material dating to the Early Iron Age (11th – 6th c. BC) was excavated in 2019 at a pit sanctuary near the Strandzha village (Агре and Дичев, 2020).

The urban settlement of Kabyle in the northern part of the Yambol Province was the regional capital during the Classical Period. It was conquered in 341 BC by Philip II of Macedon, but kept its leading position for the next centuries. Its zone of influence reached the northerly slopes of the Strandzha and Sakar Mountain (Драганов, 1993, 102-108), including the catchment of the Popovska River. The four Late Iron Age open-air settlements (5th – 1st c. BC) found here were likely under the control of Kabyle. Contemporary Thracian remains were excavated at several burial mounds in the Bolyarovo municipality, including the town of Bolyarovo (Попов, 1922, 175-177), Valchi Izvor (Агре, 2005a), Ruzhitsa (Агре, 2005b), and Zlatinitsa (Агре, 2006).

In AD 46, during the reign of Emperor Claudius, the Roman province of Thracia was created as a senatorial province. According to Lozanov, the territory of the town of Kabyle became the imperial domain and was administratively separated from the rest of the province (Лозанов, 2006,153-159). The present-day Bolyarovo municipality became a part of the territory of the city of Hadrianopolis/Adrianople (modern Edirne), and thus a neighbor to Byzantium, the later capital of the Eastern Roman Empire. The administrative and political changes at the beginning of the second century AD stimulated economic growth and stability in the province of Thracia. The peaceful and prosperous time saw the growth of new settlements, including smaller civilian vici as well as large villa estates (Ross *et al.*, 2012; Lozanov, 2015, 85; Sobotkova, 2018; Tušlová, 2022). This prosperity extended to Bolyarovo where 15 Roman-period artefact concentrations were documented.

In the second half of the third century, Gothic tribes crossed the Stara Planina Range into the province of Thracia. Their raids were followed by the Hunnic invasions at the very end of the fourth century. Their attacks strengthened in the middle of the fifth century. At this time, many cities in Thrace were abandoned or significantly reduced in size (Lozanov, 2015, 86; Tušlová, 2022).

In the first half of the sixth century, invasions continued with the incoming Bulgars and Slavs. By the end of the seventh century, the northern border of Byzantium ran through the Strandzha Mountain slopes. The Byzantine chronicler Cedrenus-Skylitzes called the area north of this border up to the Stara Planina Range "uninhabited in this time" as it was the no man's zone between the two political rivals, the Bulgarian kingdom and the Byzantine Empire (Бакърджиев and Вълчев, 2019,154).

Only after Emperor Basil II the Bulgar-Slayer conquered the Bulgarian kingdom, did Byzantium usher into a period of economic growth. Prosperity characterized the 11th and 12th centuries in the Balkan Peninsula. Popovska River watershed continued to be a part of the territory of Adrianople during these times. The field survey located five open-air settlements post-dating the 10th century. The latest archaeological investigations at a site of Malkoto Kale near the village of Voden showed that both the site, the monastery opposite it, as well as the open-air settlements found in Bolyarovo were all part of a territory ruled by the Vatatzes family (Бакърджиев and Русев, 2021). The Vatatzes family belonged to the elite of Adrianople and ascended the throne of the Empire of Nicaea in 1222 (Йорданов, 2005).

During the Third Crusade (1189-1192), the army of the German Emperor Frederik I Barbarossa conquered Adrianople. In the autumn of 1189, the fortress Malkoto kale near Voden village and the valley of Popovska River were raided by the crusaders (Бакърджиев and Вълчев, 2015).

The 13th and 14th centuries marked the period of massive changes across the Balkan Peninsula. The territories of the Bulgarian kingdom and the Byzantine Empire divided up into many smaller kingdoms. Ottoman expansion across the Balkan Peninsula also began in the 14th century, finishing with the conquest of the Tarnovo Kingdom in 1393, the Vidin Kingdom in 1396, and finally, Byzantium in 1453.

In the second half of the 14th century, the territory of present-day Bolyarovo municipality became a part of the sanjak (administrative division of the Ottoman Empire) of Adrianople. Three field concentrations dated to this period were noted during the 2018 survey, close to the present-day villages of Popovo, Dabovo, and Voden.

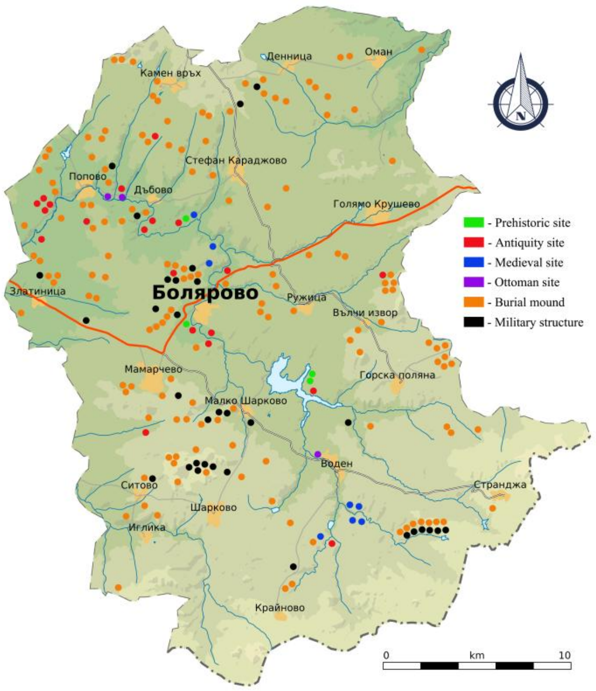


Figure 6. Map of results documented in the Bolyarovo municipality.

Bulgaria became a part of the Socialist Block on the 9th September 1944. Turkey, across the Bolyarovo border, was a US ally, and from 1952 a member of NATO and a part of the Western Alliance. For the next 45 years, Bolyarovo southern boundary became the Cold-War borderland, keeping the Warsaw Pact citizens in and Western Alliance forces out. Strategic locations with a good view of the surrounding countryside, often overlapping with mound locations which also favor prominent locations, have been repurposed into military bunkers, trenches, and tank emplacements. Among the 240 registered features, 44 had been converted into military structures. These structures were part of a complex defensive system that was supposed to delay the Turkish forces until the Soviets came to rescue. Built to endure, most bunkers still stand. Thirty years after the fall of the Iron Curtain, national defenses still lurk in the bushes and woods, waiting for the next discrepant moment in European history.

Burial Mounds

Burial mounds are conical earthen features, built as part of mortuary practices, that range from 0.5 to 10 m in height and up to 60 m diameter. In Bulgaria, their use spans from the Early Bronze Age to the beginning of the Late Antique period. Most burial mounds were built in relatively prominent locations, on isolated outcrops, ridges, and hilltops above habitation areas (Sobotkova and Weissova, 2019, 162-163, 170; Valchev and Sobotkova, 2019, 20). According to the Bulgarian law, burial mounds are protected cultural monuments of national significance.

During the 2018 campaign, our teams verified the condition of 143 burial mound markers - sunburst symbols - in the Soviet topographic maps. Additional 97 features were discovered as the teams walked through the landscape, reaching a total of 240 (148 mounds and extinct mounds, and 92 other features).

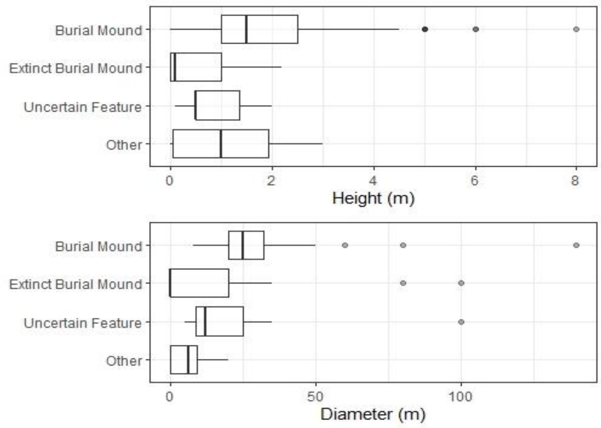


Figure 7. Feature height and diameter by type: burial mound, extinct burial mound, uncertain feature, and other

Boxplots in Figure 7 show that dimensions correlate with conceptual categories of mound-like features (Sobotkova and Weissova, 2020). Features classified as ‘burial mounds’ were the largest with 2.0 m average and 1.5 m median height, and an overall height range of 0 to 8 m. Burial mound diameter spanned from 8 to 140 m, with an average of 27.5 m. The ‘extinct’ mound measurements were lower: the height range was 0 to 2 m, with 0.5 m average and a 0 m median (assigned to the footprints of mounds). Despite ploughed-out outliers, the mean diameter was a mere 14 m because many extinct feature dimensions were unmeasurable or zero. The ‘uncertain’ mounds were smaller than burial mounds and bigger than the ‘extinct’ mounds. They ranged from 0.1 to 2.0 m in height, with an average of 0.9 and median of 0.5 m. Their diameter spanned from 5 to 100 m with 21.5 m average and 12 m median values. The measurements of uncertain mounds show tighter range in both height and diameter. Outliers with 100+m diameter appeared in all three groups as mound material was often found dispersed through ploughing. The category of 'other' mostly marks features that were affected by military activity, such as where map features led to bunkers and entrenchments rather than mounds. This category of 44 features ranged from 0 to 3 m in height (mean at 1.1m) and spanned 0 to 20 m in diameter with the mean of 6.6 m and no outliers due to standardized construction.

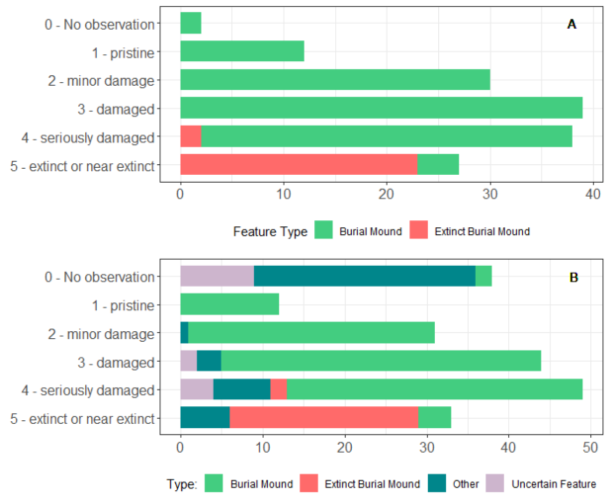


Figure 8. Condition of mound-like features. 'Other' denotes military installations.

The condition of 148 burial and extinct burial mounds was recorded on a Likert-style scale of one-to-five. One denoted a pristine mound and five denoted an 'extinct' mound, an essentially destroyed feature, which sometimes had a visible footprint (Sobotkova and Ross, 2018). Figure 8a shows 8% of mounds were in pristine condition and 20% of mounds suffered only minor damages. Damaged mounds comprised 26 %, seriously damaged mounds 26%, and the extinct or nearly extinct mounds represented 18% of total. The NA condition was assigned in two cases. Altogether, 28% (42) of all mound-like features were devoid of major damages, while 72% (104) were damaged or extinct. These percentages show more pristine and lightly damaged mounds than in the western regions of the Yambol province surveyed in 2009-2010, especially if we filter out the military installations and uncertain features (Sobotkova and Weissova, 2020).

Figure 8b shows that if we add the *ca* 50 military and uncertain features to the damaged mounds, the damaged category rises to nearly 80% (counting 0 - NA condition as 'extinct'). Minor damage drops to 21% of features. Military activity has a negative effect on immovable heritage in the border region.

Discussion

The Burial recording module was successfully expanded and reused in a short field season on the former Cold-War and now the EU-border between Turkey and Bulgaria. The module proved flexible enough to accommodate unexpected finds. By maintaining the data schema and carefully adjusting controlled vocabularies and the UI, we were able to fine-tune our recording system between seasons without breaking data compatibility in more than one field. This approach to balancing flexibility with data compatibility worked, but was entailed and fragile. Our experience informed design requirements for the next generation FAIMS platform (Ballsun-Stanton *et al.* 2021); 'FAIMS 3.0' will allow researchers to design and adjust their field data capture application via a graphic user interface, among other improvements such as cross-platform operation and more scalable and robust data collection.

The constraints on structured data entry (validation and automation) allowed fieldworkers, including novices, to record high-quality, analysis-ready data. Precise geo-location and standardized scales for mound condition facilitates future protection of these cultural monuments in the region. Digital annotations that accompanied each attribute allowed the teams to note uncertainties and mark discrepancies between the data schema and emergent field realities. When faced with the unexpected problem of assigning type and condition to military bunkers standing in place of a burial mound, or of documenting multiple land-uses where picture dictionaries allowed only one selection, field workers used the annotation to provide additional description. Digital annotations allowed for complete and flexible data entry, mitigating the inevitably rigid module structure. They facilitated automated post-processing and made emergent data accessible for statistical analysis via regular expressions.

The digital system administration and the volume of digital data placed additional demands on team-leader time. While the automated validation flagged any omissions, the content and consistency of data had to be checked manually. Integrated data view on tablets made it easy to engage novices in spot checks and corrections, which also led to improved in-field decision-making. After data review was complete, team-leaders engaged in module and data management, planning, and other manual tasks such as the transfer of data to AKB (batch upload not being available). Team-members had spare time and plunged into other digital projects. They produced a Bulgarian translation of the Burial module and successfully deployed an older digital tool to measure mound volume from mound photos. A number of students also digitized more mounds from Soviet military maps using a Map-Digitisation FAIMS2.6 module (Nassif-Haynes *et al.*, 2021). The relative abundance of free time after fieldwork made it possible to extend the scope of the project by supporting individual student initiatives.

Conclusion

Pedestrian survey of the Popovska River watershed in the Bolyarovo municipality revealed traces of past human activity spanning from Prehistory to the 20th century. The reuse of an existing FAIMS module made it possible for us to visit 240 features in 15 days, systematically documenting archaeological heritage in a former Cold-War zone for the first time.

Acknowledgements

Development of the Burial Mound Monitoring Module was funded by the Research Attraction and Acceleration Program (RAAP) NSW Department of Industry (16805/00) between 2016 and 2018. The most recent survey work was conducted with the support of the grant no.19686 from the Endeavour Short Term Mobility Program of the Australian Department of Education and the Ministry of Culture of Bulgaria: РД 11-00-230/31.07.2018.

References

Alcock, S. E., 1993, 'Graecia Capta.' The Landscapes of Roman Greece. Cambridge: Cambridge University Press.

Alcock, S. E. et al., 1994, 'Intensive Survey, Agricultural Practice and the Classical Landscape of Greece', in Morris, I. (ed.) Classical Greece: ancient histories and modern archaeologies. Cambridge: Cambridge University Press, pp. 137–170.

Ammerman, A. J., 1981, 'Surveys and archaeological research', Annual review of anthropology, 10, pp. 63–88.

Ballsun-Stanton, B. et al., 2018, 'FAIMS Mobile: Flexible, open-source software for field research', SoftwareX, 7, pp. 47–52. https://doi: 10.1016/j.softx.2017.12.006.

Cherry, J., 1983, 'Frogs around the Pond: Perspectives on Current Archaeological Survey Projects in the Mediterranean Region', in Keller, D. and Rupp, D. (eds) Archaeological Survey in the Mediterranean. Oxford: British Archaeological Reports (International Series 155), pp. 375–416.

Connor, S. E. et al., 2013, 'Environmental conditions in the SE Balkans since the Last Glacial Maximum and their influence on the spread of agriculture into Europe', Quaternary science reviews, 68, pp. 200–215.

Doncheva, D., 2012, 'The northern "journey" of Late Bronze Age copper ingots', in Паунов, Е. and Филипова, С. (eds) ΗΡΑΚΛΕΟΥΣ ΣΩΤΗΡΟΣ ΘΑΣΙΩΝ Изследвания в чест на Иля Прокопов от приятелите и учениците му по случай неговата 60-годишнина. Велико Търново, pp. 671–714.

Fentress, E., 2004, 'What are we counting for?', in Barker, G. and Mattingly, D. (eds) The Archaeology of Mediterranean Landscapes. Cambridge: Cambridge University Press, pp. 44–52.

Kecheva, N., 2018, 'Archaeological Map of Bulgaria – Transport and Pipeline Infrastructure Projects', Internet Archaeology 51. https://doi.org/10.11141/ia.51.2

Lozanov, I., 2015, 'Roman Thrace', in Valeva, J., Nankov, E., and Graninger, D. (eds) A Companion to Ancient Thrace. Oxford: Wiley, pp. 75–90.

Nassif-Haynes, Ch. et al., 2019, TRAP Burial Mound Monitoring Module (version 89333d). Github. https://github.com/FAIMS/burial/releases/tag/1.0.0.

Nassif-Haynes, Ch. et al., 2021, Map-Digitisation Module (version fc0b21c). Github. https://github.com/FAIMS/map-digitisation/releases/tag/map-dig-2018.

Ross, S. A. et al., 2010, 'An Interdisciplinary Pilot Project in the Environs of Kabyle, Bulgaria', Archaeologia Bulgarica, 14(2), pp. 69–85.

Ross, S. A. et al., 2012, The Tundzha Regional Archaeological Project: Elhovo 2009 Preliminary Report. Yambol.

Ross, S. A. et al., 2013, 'Creating eResearch Tools for Archaeologists: The Federated Archaeological Information Management Systems Project', Australian Archaeology, 77, pp. 107–119.

Ross, S. A. et al., 2018, 'The Tundzha Regional Archaeology Project: History, Aims, and Outcomes', in Ross, S. A. et al. (eds) The Tundzha Regional Archaeological Project: Surface Survey, Palaeoecology, and Associated Studies in Central and Southeast Bulgaria, 2009-2015 Final Report. Oxford: Oxbow Books, Limited, pp. 2–10.

Sobotkova, A. et al., 2016, 'Measure twice, cut once: cooperative deployment of a generalised, archaeology-specific field data collection system', in Averett, E. W., Gordon, J. M., and Counts, D. B. (eds) Mobilizing the Past: Recent Approaches to Archaeological Fieldwork in the Digital Age. University of North Dakota Digital Press in conjunction with Mukurtu 2.0 (Center for Digital Archaeology), pp. 334–370.

Sobotkova, A., 2018, 'Spatial variability in surface artefact distributions in the Yambol study areas', in Ross, S. A. et al. (eds) The Tundzha Regional Archaeological Project: Surface Survey, Palaeoecology, and Associated Studies in Central and Southeast Bulgaria, 2009-2015 Final Report. Oxford: Oxbow Books, Limited, pp. 164–179.

Sobotkova, A. et al., 2021, 'Deploying an Offline, Multi-User, Mobile System for Digital Recording in the Perachora Peninsula, Greece', Journal of Field Archaeology, pp. 571–594. https://doi: 10.1080/00934690.2021.1969837.

Sobotkova, A. et al., 2023, 'Creating Large, High-Quality Geospatial Datasets from Historical Maps Using Novice Voluntiers', Applied Geography, 155 (june): 102967. https://doi.org/10.1016/j.apgeog.2023.102967.

Sobotkova, A. and Ross, S. A., 2018, 'Kazanlak Survey Results', in Ross, S. A. et al. (eds) The Tundzha Regional Archaeological Project: Surface Survey, Palaeoecology, and Associated Studies in Central and Southeast Bulgaria, 2009-2015 Final Report. Oxford: Oxbow Books, Limited, pp. 66–81.

Sobotkova, A. and Weissova, B. (2019) 'Locational Analysis of Burial Mounds in the Middle Tundzha River Watershed. Combining Historical Maps with Field Survey and Satellite Image Analysis Data', in Studia in Honorem Iliae Iliev. Yambol, pp. 161–175.

Sobotkova, A. and Weissova, B., 2020, 'Soviet topographic maps and burial mounds of the Yambol province: digital workflow for mortuary landscape verification', Archaeological Prospection, 33, p. 233. https://doi: 10.1002/arp.1769.

Terrenato, N., 2004, 'Sample Size Matters!The Paradox of Global Trends and Local Surveys', in Alcock, S. E. and Cherry, J. (eds) Side-by-Side Survey: Comparative Regional Studies in the Mediterranean. Oxford: Oxbow Books, pp. 36–48.

Tušlová, P., 2022, 'The Historical and Topographical Setting of Yurta-Stroyno', in Tušlová, P., Weissová, B., and Bakardzhiev, S. (eds) The Yurta-Stroyno Archaeological Project. Studies on the Roman Rural Settlement in Thrace. Praha.

Valchev, T., 2019, 'Yambol region and its place in the Prehistoric map of Bulgaria Some notes about the settlements during Neolithic, Chalcolithic and Bronze Age', in Studia in Honorem Iliae Iliev. Yambol, pp. 135–158.

Valchev, T. and Sobotkova, A., 2019, 'Monitoring burial mounds in the Yambol province: deploying mobile technology to improve cultural heritage protection', in Kyriakidis, P., Agapiou, A., and Lysandrou, V. (eds) Spreading Excellence in Computer Applications for Archaeology and Cultural Heritage. Proceedings of the 3rd Conference on Computer Applications and Quantitative Methods in Archaeology Greek Chapter (CAA-GR) Limassol, 18-20 June 2018. Limassol-Cyprus: Cyprus University of Technology, pp. 19–23.

Агре, Д., 2005a, 'Археологически разкопки на могила в землището на с. Вълчи извор, община Болярово', Археологически открития и разкопки през 2004 г. София, pp. 145–146.

Агре, Д., 2005b, 'Археологически разкопки на надгробна могила в землището на с. Ружица, община Болярово, през 2004', Археологически открития и разкопки през 2004 г. София, pp. 146–148.

Агре, Д., 2006, 'Владетелски гроб от могила между селата Маломирово и Златиница, Елховско (Ямболска област)', Поселищен живот в Тракия, IV, pp. 139–144.

Агре, Д. and Дичев, Д., 2020, 'Спасителни археологически проучвания на обект от ранната желязна епоха, землище на с. Странджа, общ. Боляровом обл. Ямбол (I и II етап)', Археологически открития и разкопки през 2019 г. София, pp. 425–428.

Бакърджиев, С. and Вълчев, Т., 2015, 'Археологически проучвания на средновековна крепост „Малкото кале", с. Воден, община Болярово, област Ямбол', Археологически открития и разкопки през 2014 г. София, pp. 703–706.

Бакърджиев, С. and Вълчев, Т., 2019, 'Ранносредновеконо селище до с. Лозенец, област Ямбол', Acta Musei Varnaensis, XIII(1). Варна, pp. 153–174.

Бакърджиев, С. and Русев., Я., 2021, 'Редовни археологически проучвания на средновековната крепост „Малкото кале" при с. Воден, общ.

Болярово’, Археологически открития и разкопки през 2020 г. София.

Драганов, Д., 1993, Монетосеченето на Кабиле. София.

Йорданов, И., 2005, ‘Печати на византийската фамилия Ватаци’, Нумизматика, сфрагистика и епиграфика, 2, pp. 145–168.

Лозанов, И., 2006, ‘Кабиле през късноенелистическата и римската епоха: проблеми на селищното развитие – II в. пр. Хр. – IV в. сл. Хр’, Поселищен живот в Тракия IV, pp. 145–176.

Николов, В. and Петрова, В., 2013, ‘Проучвания на тел Караново в контекста на проблема за продължителността на хиатуса между късния халколит и ранната бронзова епоха в Тракия’, Археология, (2), pp. 7–24.

Попов, Р., 1922, ‘Материали за проучване на халщатската и ла-тенска култури в Бѫлгария и Македония’, Годишник на Народния музей за 1921 год., pp. 153–182.

GIS AND 3D DIGITAL MODELLING FOR TRACING AND FIGURING OUT URBAN CULTURAL COMPONENTS: IOANNINA CITY CASE STUDY

ATHINA CHRONI[1] & ANDREAS GEORGOPOULOS[2]

[1] Archaeologist-Art Historian N.K.U.A.-MSc UNI.PI.-PhD N.T.U.A.- Post-doctorate N.T.U.A., Hellenic Ministry of Education, Religious Affairs and Sports 20, Paramithias st., 10435 Athens, Greece, *athina.chroni@gmail.com*

[2] Professor, Laboratory of Photogrammetry-National Technical University of Athens 9, Iroon Polytechniou st., 15780 Athens, Greece, *drag@central.ntua.gr*

Abstract

Proposing innovative techniques for tracing and figuring out urban cultural components, safeguarding and monitoring cultural heritage, maintaining the thread of society's communication with culture even in extreme situations, like the COVID-19 pandemic, rebuilding our societies and economies and, finally, conveying a message of solidarity, hope, and unity have been the great challenges of IASIS Postdoctoral Research Project designed and implemented by Athina Chroni, supervised by Andreas Georgopoulos.

Ioannina city, at the north-western corner of Greece, a city which, already until the year 1430, the starting year of its Post-Byzantine/Ottoman period, had gained a leading role, being at the crossway of trade routes, and known, consequently, an economic and spiritual flourishment, thus "belonging to the category of cities that were developed after the expansion of the Ottoman Empire in the Balkan countries, cities with a significant urban past, in which the Ottoman influence was partial, but clear", (Kanetakis 1994) has been chosen to form the urban environment case-study for implementing an integrated system for remotely managing, documenting, protecting and highlighting cultural heritage by applying new technologies, under the ultimate goal of rehabilitating and recovering collective memory, as well as activating, primarily, the local community in the direction of participation in matters of culture and cultural heritage, via innovative, open-access software and methodologies, through a user-friendly interface and an open-culture mentality, thus reinforcing and restoring the democratic dimension of culture.

Keywords*: Cultural Heritage, GIS, 3-D Modelling, Open Sources, Ioannina*

1. Introduction

"Space is not static, material as a whole and empty, as is often defined in the sciences. It is artificial, as an imprint in space, and social, as a field of human life, absolute or abstract." (Lefebvre 1974)

The imprint of time sometimes imposes terrible consequences on cultural heritage, it often becomes necessary to not only recover the memory of original features of historical buildings, urban and landscape environments, but also understand its likely evolution. (Rodríguez-Gonzálvez et al. 2017) In this framework, *IASIS* Postdoctoral Research Project has focused on tracing and 3D digitally approaching specific landmarks of the city, nowadays forever lost, thus studying and pointing out the alterations of the city's urban web during the Byzantine and the Post-Byzantine/Ottoman period, consequently figuring out Ioannina's pluralistic profile, as it was shaped, mainly, after the year 1430, as a result of the city's three cultures' osmosis, i.e., the Christian, the Jewish and the Muslim, under the reasoning that the ever-evolving dynamics of the city have been principally depicted in buildings, public or private, religious or secular, conventional or more elaborate, each one of a unique historical and architectural profile.

Furthermore, on laying the foundation stone for the registration and documentation of the edifices coming from the city's Post-Byzantine period as well as for the regular status-updating of those few still standing, categorized as "Buildings at Risk" in the afore-mentioned framework, by developing a related digital platform.

Engagement of the local community has formed one more challenge of the project: classified crowdsourcing is thus also anticipated for the future updating of the web-based platform.

2. The acronym/term *I.A.S.I.S.*

IASIS[1] is an *open-to-the-people* project, the term *IASIS*, implying the concept of treatment/rehabilitation and constituting an indirect reference both to the material as well as to the mental/emotional/psychological dimension of the procedure, attaching thus equal significance to tangible and intangible cultural heritage.

Quoting the etymology of the word *IASIS*: originally, it is an ancient Greek word, meaning the "cure"[2] and maintains the same spelling in English (*IASIS*) as well as in Greek (*ΙΑΣΙΣ*), thus allowing it to be read by a wide audience, either English-speaking or Greek-speaking, also keeping the same sounding and pronunciation in both languages.

3. Deconstructing the cultural palimpsest

All places have hidden secrets, revealed when the cultural palimpsest goes under a decomposition procedure by tracing and revealing its structural elements. (Chroni 2019) Ioannina, being a city with a long Byzantine past, should have the chance to reveal its hidden secrets, given that "each place is a distinctive autonomous entity, as a unique idea, which has spiritual power and emotional content, expressed through the collective, overall and perpetual consciousness of its people". (Stefanou 2005).

The coexistence and cohabitation of people of diverse religions, ethnicities and cultural identities diachronically, have shaped a unique, highly fruitful and creative multicultural character for the city. Taking into consideration that the scales and meanings of the architectural work concerning public and private buildings are directly connected to the urban phenomenon, due to conception, position and role, tracing the city's physiognomy in its architectural expressions and decomposing the cultural palimpsest has been the great challenge of the project. Recovering and enhancing the collective memory of the place as well as involving the local community in matters of cultural heritage, have been the final objectives.

In this framework, the landmarks chosen to be 3D digitally approached come from all the three cultural components of the city[3], often stratigraphically and chronologically succeeding one another, thus creating an interesting intangible palimpsest, reflecting the fluidity of human reality. (Chroni & Georgopoulos 2020) Unfortunately, all of the landmarks, have been destroyed due to natural disasters and the unbridled, often uncontrolled modern constructions, leaving either no traces at all, as is the case of Taxiarchis Archangel Michael Christian Byzantine Cathedral, Pantocrator Christian Byzantine Church, Thomas Preljubović Tomb, Hagia Paraskevi Christian Byzantine Monastery, Namaz Giyah Muslim Mosque, the Ottoman Government House, the Barracks, Kahal Kadosh Hadash Jewish Synagogue, Epifaneios Educational Institution, the Moat and swampy area located next to the Castle's western arm, or just a few traces, as is the case of Ali Pasha Sarai and Chairedin Pasha Sarai. Fethiye Muslim Mosque, the Castle's western arm and the Army Headquarters are the exceptions, still standing, almost intact, although representing their final constructional phase. (Fig. 1)

Priority was given to landmarks which, meeting all the previous criteria, run through the longitudinal axis of the urban web in the north-south direction, thus providing the possibility of further studying the wider area around each landmark and its dialogue (buffer zone) with the surrounding urban space, an element fruitfully contributing to the interpretation of the urban and the social fabric of Ioannina mainly in the Post-Byzantine/Ottoman period.

4. The city of Ioannina-Historical framework

The first turning point for the city is the year 1430, when surrendered to the Ottomans; the second one is the year 1611, when a failed Christian uprising[4] would lead to the abolition of the privileges the Ioannites[5] had secured with the *Decree by Sinan Pasha* in 1430, thus triggering undisguised disaster of the Christian religious buildings. After that year, 35 Christian churches and monasteries were destroyed, of which 18 were located inside the Castle of Ioannina. Any attempt to identify the sites of the Castle's extinct temples, other than the *Cathedral of Taxiarchis Archangel Michael* and the adjacent *Church of Pantocrator*, in the *Inner Acropolis* (*Its Kale*) of the Castle's southeastern citadel, would be in vain, as no relevant tradition has been passed over from generation to generation. From 1430 to 1913, 17 mosques were built inside and outside the Castle, two metzites inside the Castle and three *tekes*, each one at each entrance of the city. Christian churches and monasteries existed in most of the

[1] ***I***oannina ***A***rchitectural ***S***ocietal ***I***nfrastructure ***S***tratification.

[2] From the Greek verb *ἰάομαι -ῶμαι* (to cure), *ἴασις, -εως* (the cure). (Hofmann 1950)

[3] The Christians, the Jews and the Muslims.

[4] Organised by Dionisius, Metropolitan of Larissa-Trikki.

[5] *Ioannites* are called the inhabitants of Ioannina city.

metzitia standing, according to Aravantinos. In the years 1920s-1950s, 15 of the 19 mosques and the metzites in the city of Ioannina have been destroyed. The Jewish synagogue inside the Castle had already been founded since the 9th century, as estimated, while in 1540 a second synagogue had been founded outside the Castle. The building of the last one has been destroyed in the years 1960s. In 1913 the city of Ioannina was handed over by the Ottomans to the Hellenic State. (Chroni & Georgopoulos 2020).

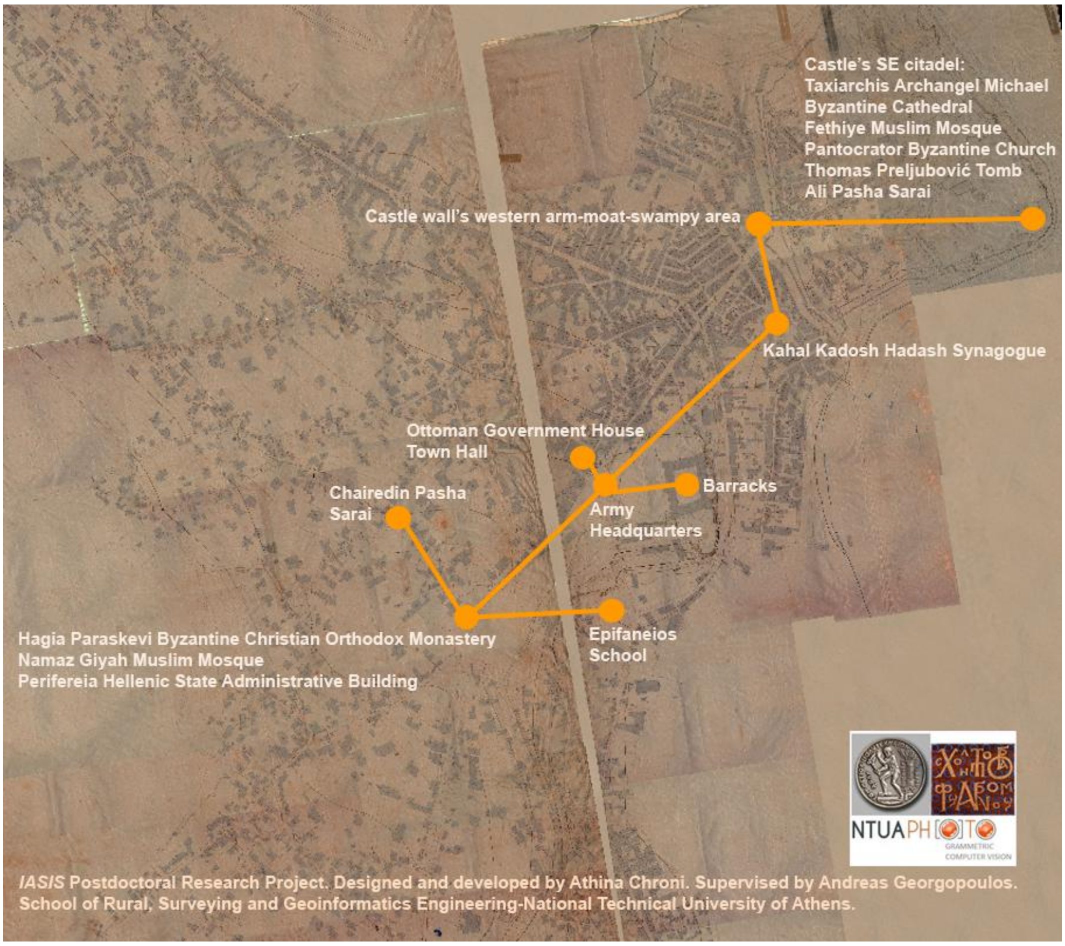


Fig. 1. The lost landmarks to be 3D digitally developed in the framework of *IASIS* Postdoctoral Research Project. Cartographic datum: Urban plan of Ioannina/1916-1918/Implemented by Melirrytos (& Christides)-Signed by Melirrytos. Digital processing designed and developed by Athina Chroni.

5. Application method-Tools of the Postdoctoral Research Project

Nowadays, the achievement of good practices and sustainable workflows to manage heterogeneous data for cultural heritage has been made possible thanks to the exponential growth of digital tools for data registration, acquisition and management, mandatory at each level and essential to preserve and protect cultural assets. (Pierdicca et al. 2013)

In the context of the specific Postdoctoral Research Project, glimpses of collective memory have formed the starting point for the decomposition of the cultural urban palimpsest.

The medium for finding the path towards Ioannina's Byzantine and Post-Byzantine urban context has been extensive documentation based on historiographic, bibliographic and archaeological data, as well as typological architectural data on Byzantine Christian churches and monasteries as well as on Muslim mosques; also on secular architecture of the respective periods and cultures. (Chroni & Georgopoulos 2020) All of the data collected have been analyzed and cross-examined: the outcome of the related scientific findings has been digitally processed and furthermore combined with the study of geospatial data, terrestrial and aerial photographs, as well as with artistic depictions such as engravings, paintings, postcards, the afore-mentioned comparatively studied in a GIS Environment (Chroni 2012) georeferenced to the city's ortho-photo of the year 2015.

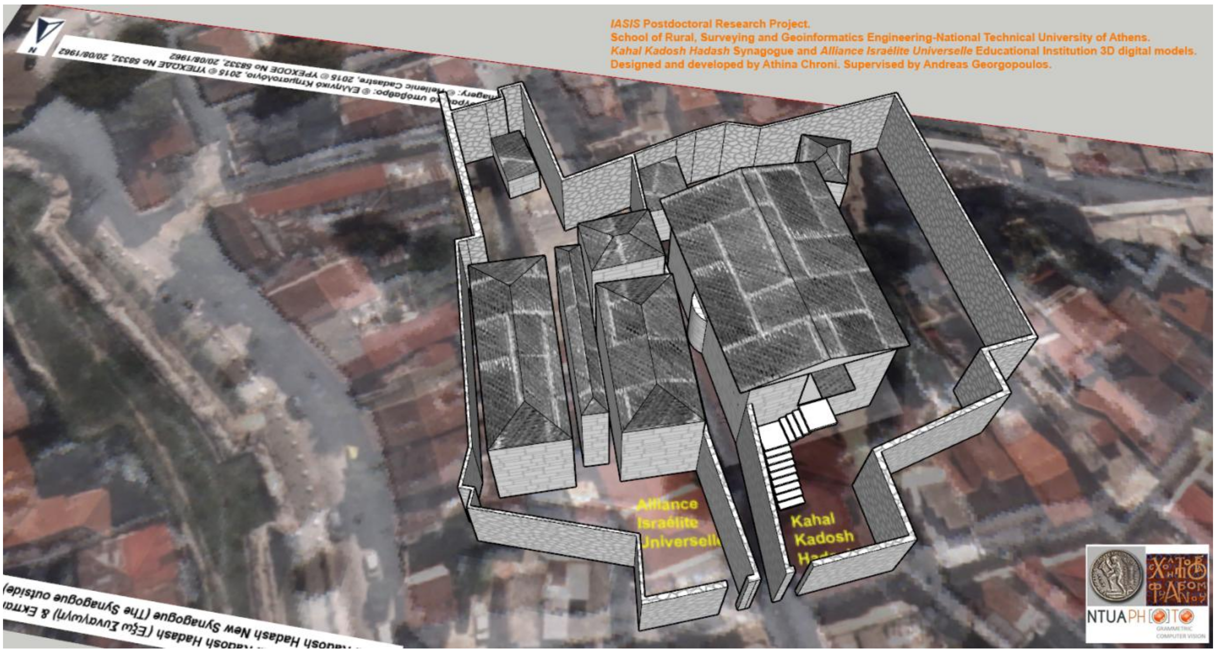


Fig. 2. *Kahal Kadosh Hadash* Synagogue and *Alliance Israélite Universelle* Educational Institution 3D digital models overlayed on the GIS digital image processing product. Designed and developed by Athina Chroni.

Consequently, the dating, as well as the location of Byzantine and Post-Byzantine landmarks within the modern city's urban web and their subsequent incorporation in an interactive GIS has been possible, thus allowing a flexible, multi-layered management of the varied data and, also, laying the foundations for a related web database. Many advantages arise from the GIS approach: at a first level, merging quantitative and qualitative data, renders possible their effective combination, while at a second level, it provides an easy and intuitive access to the information for non-expert users. (Pierdicca 2013) Under this perspective, multiple topographic, cartographic and architectural data, of various dates, have been georeferenced according to the physical color ortho-image of the city of Ioannina dated in the year 2015, thus rendering possible the location of the lost landmarks in the urban web, as well as their form and size, allowing the transition to the following stage of work, the 3D digital model development of the landmark already documented as afore-mentioned. The ortho-image combining the characteristics of the image, i.e., the optical realistic representation of the city, with those of a map, i.e., the georeferencing and metrics info, has been considered to be the optimal choice for the georeferencing of each landmark's 3D digital model. (Chroni & Georgopoulos 2020)

The ortho-image has been provided by the *Hellenic Cadastre* (Hellenic Cadastre 2022) for exclusive use in the framework of the specific Postdoctoral Research Project implementation. It consists of 132 sub-ortho-images, type LS025, dating in 2015, at a resolution of 25cm, georeferenced at *EPSG:2100-GGRS87/Greek Grid-Projected* georeference system.

The 3D digital abstractive representation (Stathopoulou et al. 2015; Georgopoulos 2015; Georgopoulos 2018) of selected landmarks has been overlayed on the GIS digital image processing product including always the aerial view of the modern city, in order to render the sense of each landmark's interaction with the urban web and figure out the alterations of the urban texture, aiming also at a more recognizable view of the city from the part of the end users. (Fig. 2) Given that each landmark's documentation has formed the basic axis of the research project, a web data base has been also developed; its consequent 3D digital rendering has formed the medium for visualizing cultural data and achieving its communication to the public. (Chroni 2021) The fragmentary character of the documentation available, as well as the intention to highlight the alterations of the texture in the urban fabric, gave the lead to an abstractive approach of the landmarks' 3D digital representation, in order not to impose the researcher's point of view but just imply the form of the building. Concluding, a minimalistic optical approach of the past has been the fundamental concept of the research project. (Chroni & Georgopoulos 2020) The landmarks' 3D digital representation proved out to contribute to a flexible study of a building's structural complex, as well as further comprehending the pluralistic physiognomy of the city, while constituting also a digital tool and motive for further research in the future.

At a final stage of the 3D approach, a virtual walk-through-the-3D-digital-model has become possible by producing an mp.4 digital file under the perspective of offering a more vivid experience to the visitor of the website which has been developed as a portal to the specific research study (Chroni & Georgopoulos 2020).

Fig. 3. Screenshot of *IASIS* website. Designed and developed by Athina Chroni. *IASIS* Website. https://athinachroni.wixsite.com/my-site-1

Fig. 4. Screenshot of *IOANNINA 1430-1913* virtual museum. Designed and developed by Athina Chroni. *IASIS* Virtual Museum. https://www.artsteps.com/view/5feca5aafe659e68d58a48c8

Additional tools, the development of an online website,[6] involving also the public though crowdsourcing, an online virtual museum[7] (Kazanis et al. 2017) and an online routes platform proposing cultural walks in the city,[8] based on the research's cultural and scientific axis, aiming at rendering *IASIS* project user-friendly and opening up to the people's community.

[6] *IASIS* Website. https://athinachroni.wixsite.com/my-site-1; Designed and developed by Athina Chroni.
[7] *IASIS* Virtual Museum-*IOANNINA 1430-1913*. https://www.artsteps.com/view/5feca5aafe659e68d58a48c8; Designed and developed by Athina Chroni.

[8] *IASIS* Cultural Routes. https://izi.travel/en/browse/d23654bc-d0bf-4b11-ab4a-184cdd0cf456#75019ea5-3313-42f5-aeae-d0359c792691; Designed and developed by Athina Chroni.

6. *IASIS* Website, *IOANNINA 1430-1913* Virtual Museum

IASIS website (Fig. 3), intending to be interactive with the people's community and furnishing cultural information data on the city, fulfills the role of the portal to the project, giving also way out to a virtual museum (Fig. 4) (Chroni & Georgopoulos 2021: Ref., No 3)

Consequently, a virtual museum, considered as the most user-friendly medium to communicate the digital product to the public, (Chroni & Georgopoulos 2021: Ref., No 4) has been also developed, where the visitor will have the opportunity to navigate in Ioannina city's cultural past. Text, image, sound and virtual walk-through-the-3D-digital-models mp.4 files, produced in the framework of this research project, are the virtual items of the exhibition, aiming to attract the viewer's attention, thus becoming an "open call" for a visit in the physical space of the modern city, as well. (Chroni & Georgopoulos 2021: Ref No 3; Chroni & Georgopoulos 2021: Ref., No 4)

7. Social inclusion and QR tagging

Culture is a *property*, the element that connects people to each other, to their past and present, ensuring a solid step towards the future. (Chroni 2019) Cultural heritage might be considered as the cultural imprint of the long-running course of man; being formed under the influence of diverse time and location components, it crystallizes and expresses the *collective memory* on local terms, as well as on global terms, since the local is the minimum piece of the broader mosaic of mankind's cultural activity and production; overcoming local cultural diversifications, it becomes the "common place" that brings people closer. (Chroni, 2018)

In the afore-mentioned framework, an *open access* profile for the project has been adopted under the perspective of reinforcing and restoring the democratic dimension of culture: consequently, priority has been given to the implementation of open data sources and software, thus complying with the *values-based approach model* of cultural heritage management, emphasizing the values attributed to cultural heritage by different interest groups of people.[9]

Under the intention to achieve more interaction of the project with the local community, QR coded informative labels are set at the specific sites, where the landmark buildings used to stand and have been now 3D digitally developed,[10] reaching out *IASIS* project to the public, thus fulfilling the role of *interactive portals* to *IASIS* virtual environment, and, finally connecting the *intangible* (digital) with the *tangible* (physical space).

8. Conclusions

The preservation of cultural heritage requires documentation, supervision, management. Cultural heritage management is composed of the field of protection, conservation and highlighting. (Chroni 2021) In the last two years, COVID-19 crisis has profoundly affected societies around the world, plunging the global economy into a deep recession. The situation is detailed in a formal *UNESCO Report* following a survey in May 2020. Moreover, according to the *United Nations Development Programme-U.N.D.P. 2020*, the catastrophic consequences of COVID-19 crisis, climate emergency and growing inequalities are not evenly distributed. They are, however, interconnected, and so should be treated. (Chroni 2021)

A survey of the *NEMO*[11] (Chroni 2021) network reveals that the majority of larger museums (81%) increased their digital capacities as a result of the COVID-19 pandemic. It is also interesting to note that social media posts have the lion's share. Among them, *Facebook* is dominant with a percentage of 75%, second is *Instagram*, with a percentage of 21.4%. (Chroni 2021) "Museums have no borders, they have a network", according to the *International Council of Museums-ICOM*. (ICOM 2022)

The aim of the specific Postdoctoral Research Project is to provide good practices to be used as guideline to manage heterogeneous data and communicate it to the public overpassing extreme situations, like the *lockdowns*. The integrated application of GIS and 3D visualization allowed flexible identification, localization, mapping and morphological documentation of various types of cultural heritage features. Additionally, the short-term target is to share cultural heritage information, while the long-term goal is to keep on backfilling and updating the web database.

[9] The *values-based approach* cultural heritage management model emerging in the western world from the 1980s onwards; it was accompanied by the development of the *post-processual archaeology theory*. (*The Burra Charter* 1999)

[10] With the support of the Municipality of Ioannina.

[11] *NEMO*: The Network of European Museum Organisations. (NEMO 2022)

Concerning cultural heritage, the year 1830 is a turning point in the Greek legislative framework. According to the Greek Law 3028/2002, Art. No 2: "(aa) As ancient monuments or ancient relics are considered all cultural goods belonging to Prehistoric, Ancient, Byzantine and Post-Byzantine times and are dated up to the year 1830, subject to the provisions of Article 20. … bb) As newer monuments are considered cultural goods that are dated later than the year 1830 and the protection of which is imposed due to their historical, artistic or scientific significance, under the regulations of Articles 6 and 20." (Hellenic State-Law No 3028/2002, The Greek Government Gazette No 153/A/28-6-2002; Hellenic State-Law No 4858/2021, The Greek Government Gazette No 220/A/19-11-2021)

The case of the successive destruction of multiple landmarks in Ioannina city case study, as well as around the world, even in modern times, due to the expansion of the structured landscape and the consequent over-urbanized environments, and, also due to the vandalisms, the wars and the physical disasters, calls for an urgent plan, globally, for monitoring-protecting-preserving-highlighting cultural heritage and strengthening collective memory. (Chroni & Georgopoulos 2021: Ref Nr 4)

"The laws of nature are taking on a new meaning today, they are no longer dealing with certainties but with possibilities and probabilities, they confirm the procedure of becoming not only of being. The distinction between the humanities and the sciences is therefore neither given nor absolute. They have approached each other; they have become entangled within each other." (Themelis 2018) In this framework, we firmly believe in the potential of interdisciplinary approaches for the study, restoration, preservation and highlighting of cultural heritage. As such, the specific Postdoctoral Research Project combines the disciplines of Archaeology-Cultural Heritage Management and New Technologies, fulfilling the role of a *virtual collective memory box*, while maintaining a scientific, educational, social and developmental role, largely reminiscent of the concept of a museum. (Chroni 2019).

References

Chroni, A., 2012, Applications of Photointerpretation and Remote Sensing in Archaeology. Doctoral Thesis, National Technical University of Athens-School of Rural and Surveying Engineering-Department of Topography-Laboratory of Remote Sensing, Athens.

Chroni A., 2018, Cultural Heritage Digitization & Copyright Issues. Springer Nature, Switzerland 2018, ISSN 0302-9743, ISSN 1611-3349, Lecture Notes in Computer Science ISBN 978-3-030-01761-3, ISBN 978-3-030-01762-0, Library of Congress Control Number: 2018956722, Chapter 34, pp. 396-407. https://doi.org/10.1007/978-3-030-01762-0

Chroni, A., 2019, *Syros Virtual Museum*: the tangible and the intangible element-Integrated cultural heritage management and new museological approaches-Standard data models and controlled vocabularies for cultural goods documentation. MSc Thesis, University of Piraeus-Division of IT and Technical Services-Faculty of IT, Athens.

Chroni, A., 2021, Museums: New technologies and museological approaches in post-COVID times. Proceedings of The Centre for Ancient Cultural Heritage and Environment (CACHE), Macquarie University, Living Digital Heritage Conference "Integrating the Past into the Present and Future", 5-7 November 2021, Sydney, Australia.

Chroni, A. and Georgopoulos, A., 2020, Documentation and 3D Digital Modelling: The Case of a Byzantine Christian Temple and an Ottoman Muslim Mosque in Ioannina City, Greece. Proceedings of the 8th International Euro-Mediterranean Conference-EuroMed 2020 on Digital Heritage: Progress in Cultural Heritage. Documentation, Preservation, and Protection, November 2-5, 2020, Nicosia, Cyprus, Editors M. Ioannides et al., Springer Nature Switzerland.

Chroni, A. and Georgopoulos, A., 2021, Documentation of Historic Buildings for Their 3D Reconstruction in a Digital Cultural Heritage Management System. 2nd International Conference TMM-CH, Transdisciplinary Multispectral Modelling and Cooperation for the Preservation of Cultural Heritage Rebranding the World in Crisis through Culture, December, 12-15, 2021, Eugenides Foundation Athens, Greece.

Chroni, A. and Georgopoulos, A., 2021, *IASIS*-Integrated Cultural Heritage Management System: A Christian Monastery, a Muslim Mosque and A Contemporary State Administrative Building, Ioannina: Digital Approach. 2nd International Conference TMM-CH, Transdisciplinary Multispectral Modelling and Cooperation for the Preservation of Cultural Heritage Rebranding the World in Crisis through Culture, December, 12-15, 2021 Eugenides Foundation Athens, Greece.

Georgopoulos, A., 2015, Digital Technologies in the Service of Cultural Heritage: The Case of the Byzantine Churches of Troodos. Panhellenic Conference on Digitization of Cultural Heritage, Volos,

Georgopoulos, A., 2018, Modern Technologies for Heritage Documentation. Proceedings of 1st International Conference on the Documentation of Natural and Cultural Heritage, 7-8 March 2018, Tehran, Iran.

Georgopoulos, A., Tapinaki, S., Ioannidis, C., Frentzos, E., Stampolidis, N., Maragoudakis, N., 2016, Development of a Georeferenced Archaeological Information Data Base for Eleutherna in Crete. Proceedings of 8th International Congress on Archaeology, Computer Graphics, Cultural Heritage and Innovation, pp. 333-337. 5-7 September 2016, Valencia, Spain.

Hofmann, J.B., 1974, Etymologisches Wörterbuch Des Griechischen. Verlag Von R. Oldenbourg, München, (1950), Translated by Papanikolaou A., Athens.

Kanetakis, G., 1994, The Castle: Contribution to the urban history of Ioannina. Published by the Technical Chamber of Greece, Athens.

Kazanis, S., Kontogianni, G., Chliverou, R., Georgopoulos, A., 2017, Developing a Virtual Museum for the Ancient Wine Trade in Eastern Mediterranean. Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci., XLII-2/W5, 399-405, https://doi.org/10.5194/isprs-archives-XLII-2-W5-399-2017.

Lefebvre, H., 1991, The Production of Space. Donald Nicholson-Smith trans., Oxford: Basil Blackwell.

Pierdicca R., Malinverni E.S., Tassetti A.N., Mancini A., Bozzi C.A., Clini P. and Nespeca R., 2013, Development of a GIS environment for archaeological multipurpose applications: The Fano Historic Centre, Aversa, Capri.

Rodríguez-Gonzálvez, P. et al., 2017, 4D reconstruction and visualization of cultural heritage: analyzing our legacy through time. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume XLII-2/W3, Nafplio.

Stathopoulou, E. K., Georgopoulos, A., Panagiotopoulos, G., Kaliampakos, D., 2015, 3D Visualisation of Lost Cultural Heritage Objects Using Crowdsourcing. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume XL-5/W7, 2015 25th International CIPA Symposium 2015, 31 August – 04 September 2015, Taipei, Taiwan.

Stefanou, J., 2005, About site and landscape. ARCHITECTS SADAS-PEA Magazine, Volume 49, Jan.-Feb., Athens.

Themelis, P., 2018, Excavation. Papazisis Editions, Athens.

Hellenic Cadastre. www.ktimatologio.gr (Last accessed: January 21, 2022).

Hellenic State-Law No 3028/2002, The Greek Government Gazette No 153/A/28-6-2002. https://www.kodiko.gr/nomologia/download_fek?f=fek/2002/a/fek_a_153_2002.pdf&t=3bb1eb7cb8692b8bd76bd3250104103e (Last accessed: January 21, 2022).

Hellenic State-Law No 4858/2021, The Greek Government Gazette No 220/A/19-11-2021. https://www.kodiko.gr/nomothesia/document/759457/nomos-4858-2021 (Last accessed: August 11, 2023).

IASIS Cultural Routes. https://izi.travel/en/browse/d23654bc-d0bf-4b11-ab4a-184cdd0cf456#75019ea5-3313-42f5-aeae-d0359c792691

IASIS Virtual Museum. https://www.artsteps.com/view/5feca5aafe659e68d58a48c8

IASIS Website. https://athinachroni.wixsite.com/my-site-1

ICOM. https://icom.museum/en/ (Last accessed: January 21, 2022).

NEMO: The Network of European Museum Organisations. https://www.ne-mo.org/

The Burra Charter, Articles 24 & 25 - The Australia ICOMOS Charter for Places of Cultural Significance, (1999). http://australia.icomos.org/wp-content/uploads/BURRA-CHARTER-1999_charter-only.pdf (Last accessed: January 21, 2022).

DEPLOYING LEGACY DATA THROUGH DIGITAL TECHNOLOGIES FOR THE RECONSTRUCTION OF THE BUILDING HISTORY OF HOUSE B AT THESSALONIKI TOUMBA

K. EFKLEIDOU[1], M. KARANTONI[1], S. TRIANTAPHYLLOU[1], S. ANDREOU[1], D. KAIMARIS[1], M. TASSOPOULOU[1], A. STAMNAS[1] AND N. KOUIDIS[81]

[1]Aristotle University of Thessaloniki, University Campus, 54124, Thessaloniki, Greece
kalefkleidou@hotmail.com, marwkarantoni@hotmail.com, strianta@hist.auth.gr, andrest@hist.auth.gr, kaimaris@auth.gr, tstamnas@yahoo.gr, mtasso@live.com

[2] Infodim, 21 25th Martiou Str., 54645 Thessaloniki, Greece,
n.kouidis@infodim.gr

Abstract

Thessaloniki Toumba is a multi-period mound settlement in Central Macedonia, Greece, founded during the Bronze Age and inhabited until the Hellenistic period. Continued re-building, occasional clearance episodes, and, finally, the establishment of a burial ground on the top during the Byzantine era resulted in a highly complex stratigraphy. In the context of a project aimed at presenting the site effectively to the public, different episodes of the history of House B were reconstructed and presented in 3D based exclusively on legacy data compiled over 25 years of excavations. The need for more extensive use of legacy data in archaeology, especially in cases of presenting old excavations to the public, has been repeatedly highlighted because of the huge amounts of information amassing underexploited in archaeological archives since the early days of scientific excavations. The issue persists primarily because of the variable quality of the data in terms of extent and accuracy and of the obsolete form they are preserved in. This paper presents the workflow implemented for the operationalization of the legacy data collection available from the excavations conducted at the site of Thessaloniki Toumba between 1988 and 2013 with the aim to reconstruct and present to the public the biography of House B.

Keywords: Thessaloniki Toumba, Late Bronze Age, Central Macedonia Greece, 3D representation, excavation documentation

1. Introduction

Rendering archaeological sites and artifacts in 3D has become a trend in public archaeology nowadays because they offer the public more detailed and immersive visualisations of past material culture. With the rapid growth of technologies (such as laser scanners and photogrammetry) and 3D modelling software, the production of 3D models has become a quick, efficient and low-cost method for rendering effectively anything from excavated sites to heritage monuments and individual objects (Figure 1).

The procedure is straightforward when it involves capturing data still standing or visible in the field during excavation. It is not, however, straightforward for the bulk of archaeological sites whose investigation has already been concluded and which, for various reasons, were not intended to be made accessible to the public or the scientific community. These sites are commonly backfilled, some have degraded, or have even been destroyed leaving no room for experts and general public to visit. Similar problems are presented by multi-phase sites, where remains of different occupation levels are found in close spatial association and/or deposits and architectural features are overlayed in deep stratigraphic sequences requiring the selective removal of deposits, artifacts and structures before proceeding to the excavation of deeper deposits.

These sites could be made accessible to the public not only through the publication of textual descriptions, images of their excavation and finds, or architectural plans, but also by rendering them in 3D. 3D models allow archaeologists and the public to directly encounter and actively engage with an excavated space irrespective of physical accessibility. This procedure would involve, however, the incorporation of legacy data, meaning documentation data produced, collected or captured without the technologies that are available nowadays and outside the scopes of contemporary excavation methods and means of dissemination (Allison, 2008). As a result, these data may often be insufficient in terms of accuracy or completeness, and, in 2D form, which do not "translate" well into 3D.

This paper presents the workflow regarding the operationalization and use of legacy data in effectively disseminating the archaeological remains of past excavation sites. In our case study, we used the legacy data (reports, drawings, 2D plans, and photographs) collected between 1988 and 2013 during the excavation of House B, a multi-phase building complex excavated at the archaeological site of Thessaloniki Toumba. Due to the fragile state of the architectural remains unearthed, this archaeological site is closed to the public and remains largely unknown to the wider audiences in the city. The dissemination of the biography of House B followed two phases: the reconstruction of the building's history based on the study of the stratigraphic and architectural data collected during its excavation; and the visual rendering of this house's history through a sequence of 2D plans and 3D models, whereby each plan or model presents the excavated remains associated with a different phase in the house's history.

Figure 1. A 3D model of the excavated area using structure-from-motion photogrammetry produced during the 2021 field season at Thessaloniki Toumba (Model by SmartEye Project, ©Thessaloniki Toumba Excavation Archive & SmartEye Project).

2. The archaeological site of Thessaloniki Toumba

The site of Thessaloniki Toumba lies at the northeastern part of the modern-day city of Thessaloniki, in Northern Greece (Figure 2). The site is a mound-settlement, formed over hundreds of years of continuous habitation on the same restricted area of what originally was a low natural hill. Habitation at the site was probably established at the beginning of the Middle Bronze Age (MBA, 2050-1650 B.C), but the best preserved and most extensively investigated occupation levels belong to the Late Bronze Age (1650-1050 B.C.). The excavated remains on the top of the mound belong to several multi-room free-standing buildings, which were separated by narrow streets. Most of the buildings have only been partially investigated, but the remains and archaeological deposits of three (Houses A, B, and E) provide us with important information on the size, architecture, building practices and activities that took place in their interior (Figure 3). Houses were built with mudbricks on stone socles with wooden supports. The floors were predominantly earthen, and the roofs made of clay and reeds. Over the course of the settlement's history, houses were repeatedly repaired and rebuilt on top of their predecessors' remains after the latter had been partially or entirely backfilled. This practice resulted in a highly complex stratigraphic sequence both at the scale of the settlement and at the scale of individual buildings (Andreou et al., 2022, Andreou, 2009, Andreou and Kotsakis, 1997).

Figure 2. Aerial view of the archaeological site of Thessaloniki Toumba (view from west) (Photo by K. Kotsakis ©Thessaloniki Toumba Excavation Archive).

House B, which forms our case study, is a building complex that was partially exposed over an area of ca. 80 m^2 and at a depth of almost 3m. Preliminary study of the stratigraphic sequence and architecture of the building (Efkleidou et al., 2018) revealed that the building was entirely rebuilt at least four times over a period of more than 200 years (1210-950 B.C.). In-between these rebuilding events, the building saw multiple episodes of minor (renewals of earthen floors and internal furnishings) and major (reconstruction of individual walls) repairs. The evidence for Iron Age and Archaic period habitation at this part of the excavation has been heavily disturbed or destroyed. However, during the Classical period a large building with strong foundations was built above the remains of the Bronze Age House B.

The settlement was probably abandoned during the synoecism that led to the foundation of the Hellenistic town of Thessaloniki in 315 B.C. by King Cassander of Macedon. However, during the Middle Byzantine period the area above House B was transformed into a burial ground (Kotsakis and Andreou, 1989, Andreou et al., 2022). The opening of several of these Byzantine period graves caused the destruction of Prehistoric and Classical period deposits and architectural features of House B.

3. The data and their limitations

The excavation of House B took place between 1988 and 2013 and the remains extend over an area of ca. 110 m^2. Excavation at the site follows the single context recording system which respects the integrity of the stratigraphic context and adheres to a deposit-removal strategy that follows from the reverse the deposition sequence and life-history of the building (Kotsakis et al., 1995). The documentation procedures were rigid: each context was drawn at 1:50 scale, contexts considered complicated or important for their finds or attributes were also sketched in the excavation diary and in sequenced plans at 1:20 scale; multiple contexts of interest were also drawn in top plans at 1:20 scale. Single or groups of contexts and their associations were routinely photographed. Finally, stratigraphic sections of at least two sides of each trench were drawn at 1:10 scale. The collection, thus, of documentary data from House B includes 72 top plans, several sketches drawn mostly without scale, 11 stratigraphic sections, and 1532 photographs. Added to these, there are handwritten excavation diaries, catalogues of features, finds etc. and reports from each excavation trench for every field season.

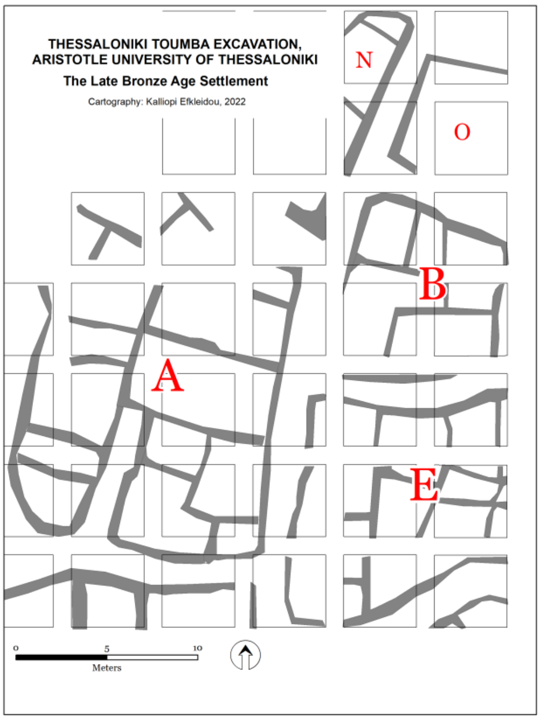


Figure 3. Simplified plan of the Late Bronze Age settlement at the top of Thessaloniki Toumba (Map by K. Efkleidou, ©Thessaloniki Toumba Excavation Archive).

While the documentation strategy at Thessaloniki Toumba excavation (initiated in the 1980s) followed pioneering protocols for its time, the accumulated data, particularly from the earlier periods of the excavation, present limitations which constrain the detail of the final publication of the archaeological remains. These limitations became further pronounced in the face of our intention to produce 3D models of the excavated space as part of our ongoing commitment to make the site accessible and effectively engage the public with its historic environment (see Kotsakis et al., 1995 for early attempts to reconstruct the site in 3D).

These constrains are linked to three types of problems affecting the usability of the legacy data at variable degrees: problems related to the documentation protocols; to the technical equipment used, and to human error.

One of the most important problems arising from the documentation protocols followed at Thessaloniki Toumba lies with the decision to draw top plans of the excavated space whenever a context of interest appeared. Such a context would involve the detection of a structure or an artifact or a group of artifacts (i.e., a hearth, a pit, a wall, a group of pottery sherds etc.). These top plans depicted only the context of interest, in an effort to limit any unintentional association of contexts that might have belonged to non-associated events. Such associations were to be considered and determined during the post-excavation study of the stratigraphy, after which the top plans would be combined. While this protocol intended to remove the pressure of interpreting complex stratigraphic and contextual associations from the excavator’s shoulders while in the field, it often resulted in decontextualizing the documented features and artifacts. This practice impeded the post-excavation process of conceptualizing or reconstructing the excavated space both mentally and in 3D.

Furthermore, a limitation arose by the number of elevations recorded in the plans. While the interpretation of a feature and its contextual associations can be achieved even with a small number of measurements, it was soon highlighted that the number of elevations noted in the plans was too small to allow the accurate reconstruction in 3D of the 2D features in the top plans.

Photographic documentation was also focused on capturing the subtleties of archaeological features and artifacts as they were exposed in the field, but the process involved both close-ups and trench overviews that allow one to gain an understanding of the wider context of the features or artifacts documented. The process also involved photographing from different angles and directions to limit the effect of shadows or distortions. Still the limited number, angles, and quality of photographs taken on any occasion, although probably sufficient for an archaeological publication, have proved insufficient for the reconstruction of the excavated context as a 3D model with the use of photogrammetric software, such as Agisoft Metashape.

Issues pertaining to the technical equipment used over the years affected the accuracy of the measurement data recorded in the top plans, sections, and the diaries. Until the end of the 1990s and before the time of high-accuracy total stations in archaeological excavations, depths were measured with a water level and for a brief period with a dumpy level from temporary benchmarks. The coordinates of contexts, features and artifacts were measured in the field with tape measures from the sides of each trench within the established site grid and subsequently transformed into the site's local reference system. As a result of the technical equipment available and on-site problems that commonly occur in the fieldwork (i.e., contraction of trench limits as excavation proceeds) some accuracy loss was observed both on the horizontal and on the vertical plain.

Substantial limitations, however, were also brought about by human error. Under the pressure of time or due to the training character of the fieldwork campaigns involving archaeology students, the guidelines regarding the symbology used in the top plans or the accuracy of the measurements noted were not rigidly followed.

4. The workflow for the operationalization of the legacy data

The first step taken was to scan all legacy data and translate them into digital images. Diaries and reports were scanned into PDF files, whereas all plans and sketches were digitized in high-resolution grayscale TIFF files (600dpi) (Figure 4, left).

A GIS system was setup in which all plans were imported and georeferenced. These plans were ordered by date produced to create sequenced planar snapshots of the excavation procedure. The sequences of overlayed plans allowed us to determine: features that were present in multiple plans (a preliminary indication of the close spatial association of permanently standing features with deposits and structures of more than one occupational phase); features whose morphology, attributes, or interpretation evolved as excavation progressed; and deposits or features at different levels that were removed as excavation proceeded.

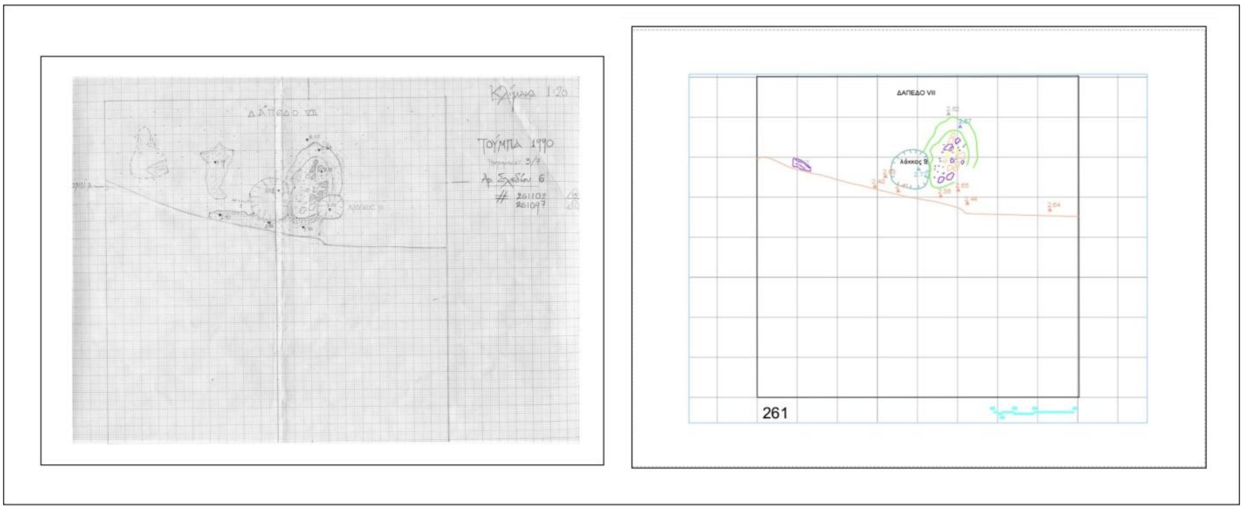


Figure 4. Left: top plan of features in trench 261 produced in the field. Right: the same plan digitized in Autocad (Original plan by M. Dosi (1990), digitized plan by G. Vlahodimos (2000-2001) and M. Karantoni (2020), ©Thessaloniki Toumba Excavation Archive).

Digitization of the top-plans could not be automated as they were drawn on millimeter paper, the grid of which created a very dense web of lines that impaired the distinction of individual features on the plans. A heads-up manual digitization method was followed instead, which was carried out in Autodesk Autocad software (Figure 4, right). Even though the digitization process is a technical job, an increased level of interpretation was involved in those cases where the symbology did not conform to standards and feature type, material or even shape could not be directly recognized from the drawing. During this process, the classification of features in the plans became standardized, so that all digital plans would share a common set of layers and pen attributes. The layers followed a classification system for features based on type ("Architecture", "Structures", and "Portable finds"), function (e.g., Architecture was subdivided into socle, wall, surface), and material (e.g., stone, mudbrick, clay, gravel, etc.) (Table 1).

Once the digitization of the plans had finished, a systematic effort was made to address the issue of the small number of elevations noted in the plans. The missing information was retrieved from the descriptions and measurements noted in the excavation diaries. Where such information could not be found in the diaries, it was reconstructed from the photographs.

The diary descriptions and photographs of features were also scrutinized in those cases where mistakes were observed in the shape or size of the features drawn. In many cases, this involved the orthorectification of photographs of features to ascertain the correct location or size of the features.

The final step before the production of the 3D models of House B involved the production of phase plans. The phase plans of a building or a wider archaeological site combine all the structures, deposits, and other features that are interpreted as contemporary or as falling within the timespan of the same occupation period. In cases of multi-period sites, such as Thessaloniki Toumba, where houses were repeatedly rebuilt, several phase plans are expected to be produced.

The stratigraphic analysis of the deposits with the use of the well-known Harris-Matrices (Figure 5) and the 3D modelling of the excavation units (following Tsipidis et al., 2011, Katsianis et al., 2015) (Figure 6) allowed us to determine more than 100 different events in the life of House B. These events involved re-buildings of the house, repairs in walls or interior furnishings, multiple floor layings and repairs, destructions, infilling episodes, burials, insertions or removals of pithoi or other features, and, finally, the usages of the spaces of the house itself. The sequence of events in mound settlements in general, and at Thessaloniki Toumba in particular, followed a roughly cyclical order: construction, use, destruction/ abandonment, infilling, and rebuilding. Each time the house was rebuilt, a new cycle and a new occupation phase begun.

TYPE	FUNCTION	MATERIAL
A (ARCHITECTURE)	WALL	STONE
		MUDBRICK
		CLAY
	FLOOR	GRAVEL
		EARTH
S (STRUCTURE)	HEARTH	
	OVEN	
	PIT	
	BASKET	
	POST-HOLE	
	PLATFORM	
F (PORTABLE FINDS)	POTTERY	PITHOS
		OTHER
	SEASHELLS	
	STONES	
	GRAVEL	
	SKELETAL (REMAINS)	HUMAN
		ANIMAL
	OTHER ARTIFACTS	STONE
		CLAY
		BONE

Table 1. The organization and naming conventions of the layers used in the digitized top plans.

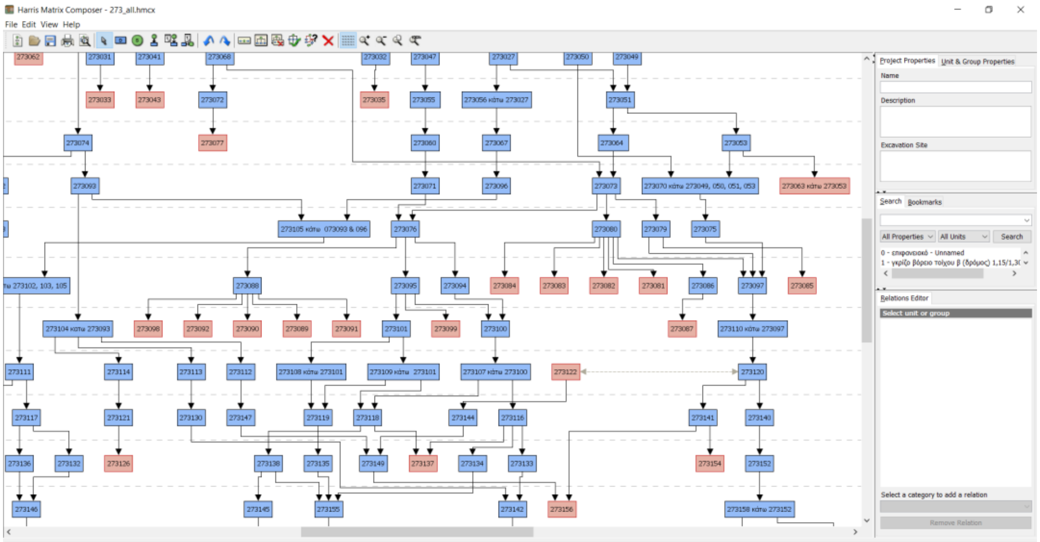


Figure 5. Part of the Harris matrix produced during the stratigraphic analysis of the deposits of House B (Figure by K. Efkleidou).

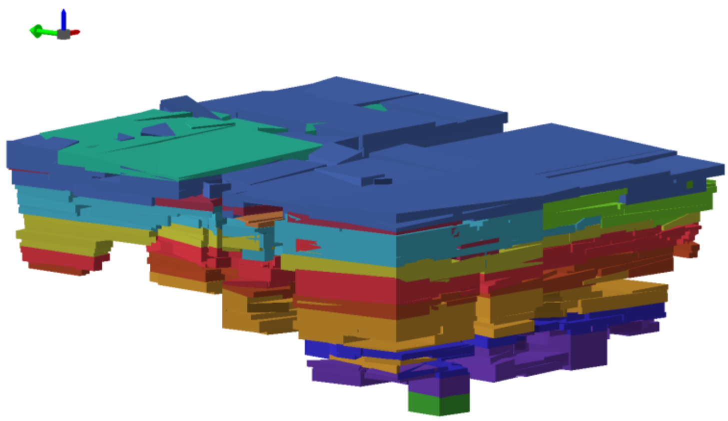


Figure 6. The 3D modelled excavation units from House B at Thessaloniki Toumba classified by occupation phase in the ESRI ArcScene environment (Model by M. Karantoni).

As part of the stratigraphic sequencing procedure, the archaeological deposits were associated with built-in features (i.e., walls, platforms, hearths etc), and most importantly with the floors or living horizons of each occupation phase of the building. Portable artifacts (i.e., pottery, stone tools etc.) are depicted on the phase plans only when they have been found lying on these floors or on built-in features and structures.

Seven occupation phases which chronologically extend from prehistory (Bronze Age) to the Classical/ Hellenistic period, were, thus, distinguished in the history of House B at Thessaloniki Toumba. The final phase of occupation involved the conversion of this part of the settlement into a Christian burial ground dating to the Byzantine times. The respective phase plans presenting the excavated remains of each occupation phase were finally modelled into 3D with the use of appropriate software, such as Autocad 3D and 3DS max (Figure 7).

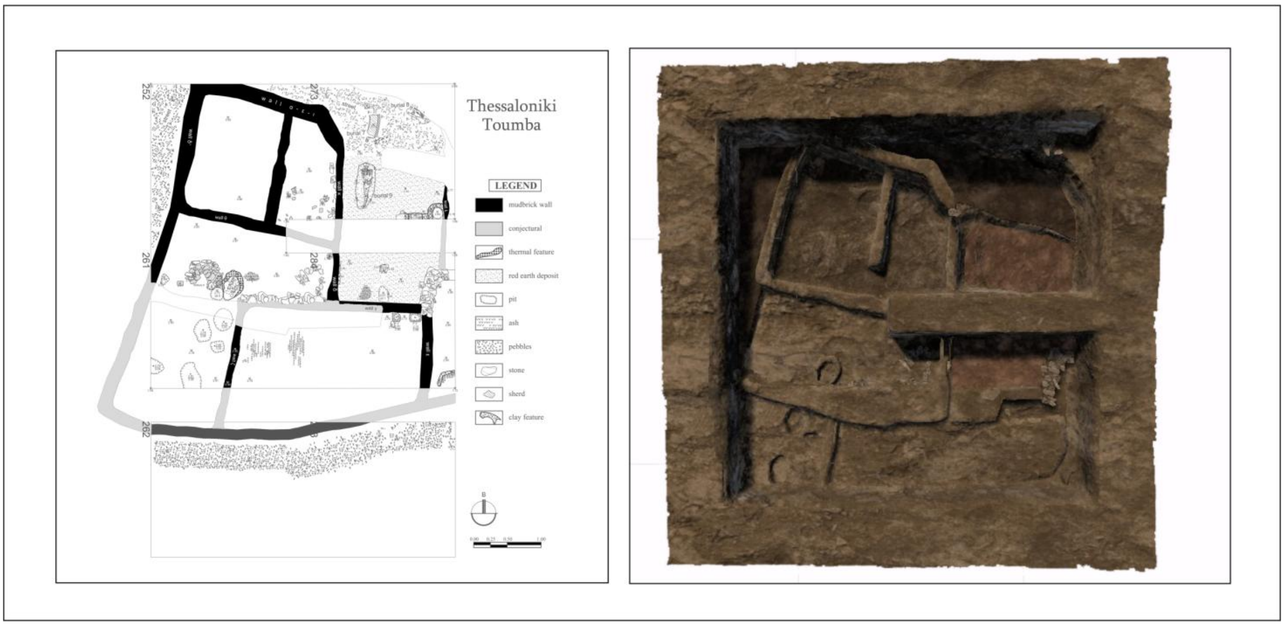


Figure 7. The (aggregate) top plan (left) and the 3D model (right) of the excavated remains of phase 2B at Thessaloniki Toumba House B (Plan by K. Efkleidou and M. Karantoni, model by SmartEye Project, © Thessaloniki Toumba Excavation Archive & SmartEye Project).

Absolute Chronology	Thessaloniki Toumba House B Occupation Phases	Relative Chronology Southern Mainland Greece
ca. 6th-12th c. AD	1A	Byzantine Period
500 - 316 B.C.	1B	Classical & Hellenistic period
1070/1040-1030/1000 BC	2A	Early Protogeometric
1110/1100-1070/1040 BC	2B	Late Helladic IIIC LATE
1140/1130-1110/1100 BC	3	Late Helladic IIIC ADVANCED
1210-1140/1130 BC	4	Late Helladic IIIC DEVELOPED
		Late Helladic IIIC EARLY

Table 2. The dating of the occupation phases identified in the area of House B at Thessaloniki Toumba.

5. Results: the biography of House B

The history of House B clearly demonstrates the conscious efforts of the respective household to maintain the external form of the house identical for a period of over 200 hundred years (Efkleidou et al., 2022) (Table 2).[1] The earliest excavated remains are of limited spatial extent but are securely dated on relative stratigraphic basis to settlement phase 4 (1210-1130 B.C.) (Figure 8). The remains were covered by a destruction/fill deposit that preserved them in considerably good condition. In fact, at the west side of the investigated room, a low clay-plated platform was found with charred and fossilized organic food remains preserved inside a clay bowl and on the surface of a stone grinder. At a small distance, the partial remains of a thermal feature, probably a domed oven, testify to the use of this space as primarily a food preparation and cooking area (Efkleidou et al., 2018, Andreou et al., 2022). During subsequent occupation phases,

[1] *The occupation phases of House B have been stratigraphically correlated with the settlement-wide phases at Thessaloniki Toumba, the absolute chronology of which is based on a series of radiocarbon dating samples discussed in Andreou (2009; see also discussion in Jung et al. 2009; Jung and Wenninger 2002; 2004).*

however, the functions of the different spaces in the house changed.

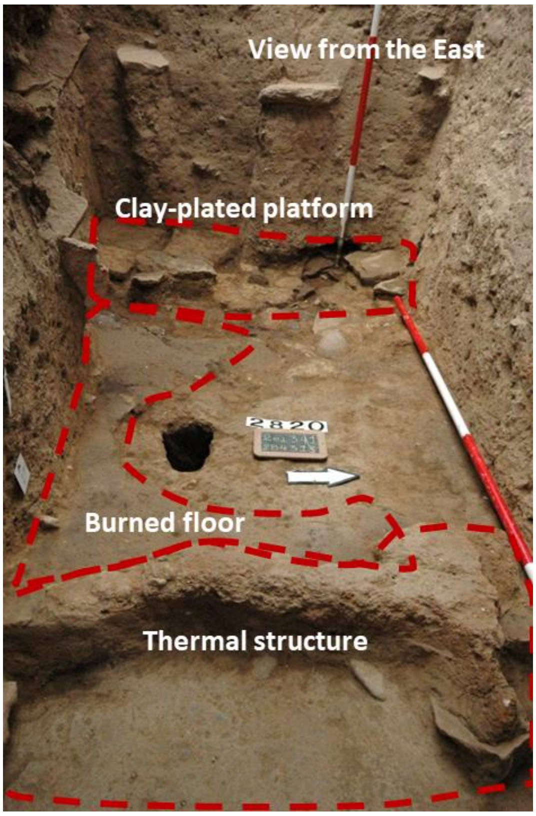


Figure 8. View of the excavated remains of occupation phase 4 (©Thessaloniki Toumba Excavation Archive).

During phase 3, when the house was rebuilt by replicating the exact same plan that it had during the earlier occupation phase, two large storerooms were functioning on either side of a space that was partly used as a purple-dye production workshop. The storerooms contained pithoi, numerous large and medium-sized jars, as well as very large woven baskets with a capacity of ca. 600 liters or more, that could be used as granaries. One of the rooms in the northeast corner was infilled with a particular type of red clay soil and then abandoned (Efkleidou et al., 2018, Andreou et al., 2022) (Figure 9).

After another violent destruction, the house was rebuilt (occupation phase 2B). Once again, it replicated the plan already familiar to the household. During this period, house floors were repeatedly repaired, as they were worn out by everyday use. One space, however, presents increased interest as it was transformed into a semi-public courtyard space open to the street. The area was used for routine activities, such as spinning or the production of household equipment (tools), and the consumption of meals. However, it is particularly interesting that this space, together with the area of the street right outside this part of the house, was also considered the designated area for the burial of four individuals (three children and a male adult) (Andreou et al., 2014). After each burial event, the area was restored to its routine use. It is only after the last of the burials, that of a seven-year-old child which was exceptionally laid face-down, that the house was rebuilt once again (occupation phase 2A) (Figure 10).

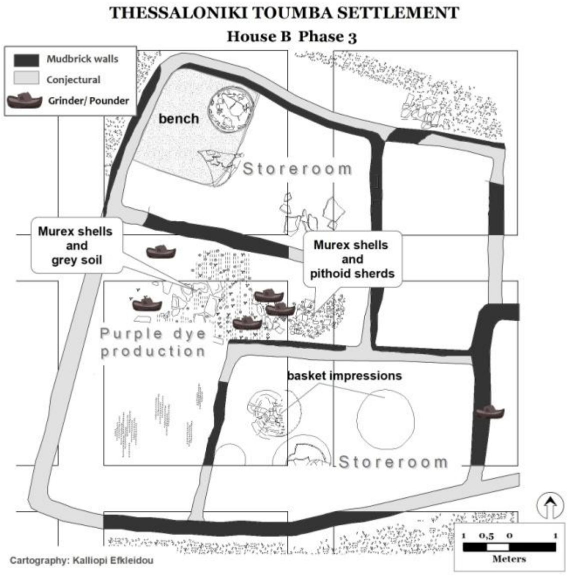


Figure 9. The (aggregate) top plan of occupation phase 3 at House B (Cartography by Kalliopi Efkleidou).

Over the next period and until Classical times habitation expanded outside the confines of the mound top. A settlement was established around the foothills of the mound (known as *trapeza* or table) that continued to be occupied until the abandonment of the site in general as part of the synoecism for the foundation of the city of Thessaloniki by Kassandros in 316 B.C. (Andreou, 2019, Andreou and Kotsakis, 1997).

While habitation on the mound top did not cease, as recent investigations have demonstrated (Andreou et al., 2022), all such evidence was obliterated from the area of House B by an organized clearance event that prepared the area for the construction of a large Classical-period house. The size of this house (covering an area of at least 100 m^2) exceeded that of the prehistoric one. Some of its walls were built on top of the stubs of the prehistoric house's walls, but interior walls especially were built on high rubble-stone socles laid inside foundation trenches cut through the prehistoric deposits. The house comprised at least four spaces, but the fragmentary nature of the many architectural remains in its vicinity does not allow us to reconstruct its complete plan, at least in the excavated area of the site. The house suffered a major destruction during the Classical period, that caused parts of the roof to collapse inside the house. Yet, the structure was rebuilt again, as the very partially preserved stone walls document.

The final episode in the biography of House B was its conversion to a Christian burial ground. Five burials of three infants and two adults were found with the bodies placed in an extended position, with their arms crossed over the chest, oriented west to east. The burials were neatly ordered in two rows, one for the adults and one for the infants, and regularly spaced (Andreou and Kotsakis, 1994, Andreou et al., 2022).

Figure 10 3D model of occupation phase 2B and the last burial performed during this phase in the northern part of House B (Model by SmartEye Project, © Thessaloniki Toumba Excavation Archive & SmartEye Project).

6. Discussion

The growing need to make past archaeological projects and their findings more accessible to the public integrating means such as immersive 3D models has become widely acknowledged among archaeologists nowadays. 3D models provide direct interaction with heritage without damaging it and without hindering its continued investigation. This understanding has led to the growing incorporation in archaeological excavation practice of new methods (i.e., structure-from-motion photogrammetry, point clouds) and technologies (i.e., drones, laser scanners) that improve and accelerate the creation of high-quality and high-resolution 3D models of excavations, archaeological deposits, built-in features and portable artifacts in 3D.

However, when it comes to presenting excavations that have already been backfilled or sealed off for reasons of preservation, the use of these methods and technologies is impossible. Turning to the use of legacy data then becomes a one-way street with all the problems and restrictions posed by the aims, methods, and conditions under which the legacy data were compiled. The methodology presented here demonstrates that the processing of legacy data (plans, sketches, diaries, and reports) can achieve their operationalization and allow the production of 3D models of the excavated landscape.

In the case study presented, many of the problems identified in the legacy data available were the result of excavation protocols and technologies that were continuously developed, evaluated and adapted or upgraded over a period of almost 40 years. Most importantly, however, they were the result of their time and collected without ever having in mind the use in which we aim to put them today. However, considering the irreversibly destructive effect of all archaeological excavations, whereby artifacts and occasionally entire occupation phases are removed and obliterated, it is imperative that we continue to seek new and enhanced ways to reconstruct the archaeological remains in a way which represents faithfully and accurately their condition during excavation. The present paper presents an attempt to achieve this aim by making use, to the best possible extent, of the available legacy data, even when they present substantial limitations by today’s standards.

In fact, the operationalization of the visual data available (plans, sketches, sections and photographs) together with the analysis of the stratigraphic and architectural data from the excavations at the archaeological site of Thessaloniki Toumba, allowed us to unlock and effectively present the complex building history of House B. The evolution of the continuous habitation in this area was thus successfully presented in the traditional form of phase-plan sequences as well as in the form of 3D-model sequences enabling the direct visualization of the site's ongoing excavation and the archaeological remains' materials, hue, and form. As a final comment, we would like to note that 3D visualization did not provide any useful insights during the stratigraphic sequencing and phasing of the archaeological deposits because the latter is an analytical step that needs to precede the building of the 3D model. However, the models finally produced have provided both expert and wider audiences with a unique way to envision and gain a better understanding of how past spaces looked and were used by the people who lived in them. In the end, these 3D models and our ability to "walk" through and explore them is worth a thousand words!

Acknowledgements

This research was carried out in the context of the project SMART EYE: Development of a prototype for smart visualization of invisible antiquities (T6YBΠ-00023, MIS 5056205). The project is co-financed by Greece and the European Union (European Regional Development Fund) through the Operational Program Competitiveness, Entrepreneurship and Innovation 2014-2020, Special Actions "Aquaculture" - "Industrial Materials" - "Open Innovation in Culture".

The Thessaloniki Toumba excavation is funded by the Aristotle University of Thessaloniki and the Research Committee of the Aristotle University of Thessaloniki.

References

Allison, P., 2008, Dealing with Legacy Data - an introduction. *Internet Archaeology,* 24, DOI 10.11141/ia.24.8.

Andreou, S., 2009, Stratified Wheel made Pottery Deposits and Absolute Chronology of the LBA to the EIA Transition at Thessaloniki Toumba. *In:* Deger-Jalkotzy, S. & Bächle, A. E. (eds.) *LH III C Chronology and Synchronisms III: LH III C Late and the Transition to the Early Iron Age. Proceedings of the international workshop held at the Austrian Academy of Sciences at Vienna, February 23rd and 24th, 2007.* Wien: Verlag der Österreichischen Akademie der Wissenschaften.

Andreou, S., 2019, Μετά την Τούμπα: παράδοση και νεωτερισμοί στις κοινωνίες της Κεντρικής Μακεδονίας στο τέλος της 2ης και στις αρχές της 1ης χιλιετίας π.Χ. *In:* Kotsakis, K. (ed.) *Η αντίπερα όχθη. Κοινωνικός χώρος και ιδεολογία στις προϊστορικές κοινότητες. Μελέτες αφιερωμένες στη μνήμη του καθηγητή Γ. Χ. Χουρμουζιάδη.* Θεσσαλονίκη: Αριστοτέλειο Πανεπιστήμιο Θεσσαλονίκης.

Andreou, S., Efkleidou, K. & Triantaphyllou, S., 2014, Η Πανεπιστημιακή Ανασκαφή στην Τούμπα Θεσσαλονίκης κατά το 2009. *Το Αρχαιολογικό Έργο στη Μακεδονία και Θράκη,* 24 (2010)**,** 359-364.

Andreou, S. & Kotsakis, K., 1994, Η ανασκαφή στην Τούμπα της Θεσσαλονίκης το 1991. *Το Αρχαιολογικό Έργο στη Μακεδονία και Θράκη,* 5**,** 210-18.

Andreou, S. & Kotsakis, K., 1997, Η προϊστορική τούμπα της Θεσσαλονίκης. Παλιά και νέα ερωτήματα. *Το Αρχαιολογικό Έργο στη Μακεδονία και Θράκη,* 10Α (1996)**,** 369-87.

Andreou, S., Triantaphyllou, S. & Efkleidou, K., 2022, Ανασκαφή Τούμπας Θεσσαλονίκης 2006-2016. *Το Αρχαιολογικό Έργο στη Μακεδονία και Θράκη,* 30 (2016), 281-290.

Efkleidou, K, Karantoni, M, Triantaphyllou, S & Andreou, S., 2022, Unlocking building biographies during the Late Bronze Age in Central Macedonia: the case of the Thessaloniki Toumba mound settlement. *Journal of Greek Archaeology,* 7: 94-121.

Efkleidou, K., Valamoti, S. M., Vliora, E., Karathanou, A., Kyrillidou, S., Papadias, Y., Triantaphyllou, S. & Andreou, S., 2018, Η πανεπιστημιακή ανασκαφή στην Τούμπα Θεσσαλονίκης κατά το 2013. *Το Αρχαιολογικό Έργο στη Μακεδονία και Θράκη,* 27 (2014)**,** 307-315.

Jung, R., Andreou, S. & Weninger, B., 2009, Synchronisation of Kastanas and Thessaloniki Toumba at the end of the Bronze and the beginning of the Iron Age. *In:* Deger-Jakotzy, S. & Baehle, A. (eds.) *LH III C Chronology and Synchronisms III. LH III C Late and the Transition to the Early Iron Age. International Workshop held at the Austrian Academy of Sciences at Vienna, 23-24.2.2007.* Wien: Verlag der Osterreichischen Akademie der Wissenschaften.

Jung, R. & Weninger, B., 2002, Zur Realitaet der Discrepanz zwieschen den kalibrierten 14C-Daten und der historischaerchaeologishen Datierung in Kastanas.

Kastanas. Ausgrabungen in einem Siedlugshuegel der Bronze- und Eisenzeit Makedoniens 1975-1979. Die Drehscheibenkeramik der Schichten 19 bis 11. Teil 1 & 2. Kiel: Verlag Oetker/Voges.

Jung, R. & Weninger, B., 2004, Kastanas and the chronology of the Aegean Late Bronze Age and Early Iron Age. *In:* Higham, T., Bronk Ramsey, C. & Owen, C. (eds.) *Radiocarbon and Archaeology. Proceedings of the 4th symposium, Oxford 2002.* Oxford: Oxford University School of Archaeology.

Katsianis, M., Tsipidis, S. & Kalisperakis, I., 2015, Enhancing excavation archives using 3D spatial technologies. *In:* Papadopoulos, C., Paliou, E., Chrysanthi, A., Kotoula, E. & Sarris, A. (eds.) *Archaeological Research in the Digital Age. Proceedings of the 1st Conference on Computer Applications and Quantitative Methods in Archaeology Greek Chapter (CAA-GR). Rethymno, Crete, 6-8 March 2014.* Rethymno: Institute for Mediterranean Studies – Foundation of Research and Technology (IMS-FORTH).

Kotsakis, K., Andreou, S., Vargas, A. & Papoudas, D., 1995, Reconstructing a Bronze Age settlement by CAD. *In:* Huggett, J. & Ryan, N. (eds.) *Computer applications and quantitative methods in archaeology 1994: Proceedings of the 22nd CAA conference 1994.* Oxford: Tempus Reparatum, Archaeological and Historical Associates Limited.

Kotsakis, K. & Andreou, S., 1989, Προκαταρκτικές παρατηρήσεις στις αρχιτεκτονικές φάσεις του οικισμού της Τούμπας Θεσσαλονίκης. *Το Αρχαιολογικό Έργο στη Μακεδονία και Θράκη,* 2**,** 201-212.

Tsipidis, S., Koussoulakou, A. & Kotsakis, K., 2011, Geovisualization and Archaeology: supporting excavation site research. *In:* Ruas, A. (ed.) *Advances in Cartography and GIScience. Volume 2: Selection from ICC 2011*. Berlin, Heidelberg: Springer, DOI 10.1007/978-3-642-19214-2_6

ΠΡΟΣΕΓΓΙΣΕΙΣ ΑΝΟΙΚΤΗΣ ΕΠΙΣΤΗΜΗΣ ΣΤΗΝ ΨΗΦΙΑΚΗ ΑΡΧΑΙΟΛΟΓΙΚΗ ΈΡΕΥΝΑ

Μ. ΚΑΤΣΙΑΝΗΣ[1] ΚΑΙ Δ. ΤΣΙΑΦΑΚΗ[2]

[1] Τμήμα Ιστορίας-Αρχαιολογίας, Πανεπιστήμιο Πατρών, Κτίριο Β, Πανεπιστημιούπολη Πάτρας, Ρίο, 26504, *mkatsianis@upatras.gr*

[2]Ερευνητικό Κέντρο «Αθηνά», Ινστιτούτο Επεξεργασίας του Λόγου, Πανεπιστημιούπολη Κιμμερίων, Ξάνθη, Τ.Θ. 159, 67100,
tsiafaki@athenarc.gr

Περίληψη

Ως μέρος μιας παγκόσμιας τάσης για την εφαρμογή πρακτικών της Ανοικτής Επιστήμης, κρίνεται ιδιαιτέρως χρήσιμη η παρουσίαση των πρωτοβουλιών σε διεθνές επίπεδο και η σύγκριση με το ελληνικό παράδειγμα και την προσωπική μας εμπειρία, όπου παρά την αυξανόμενη χρήση της ψηφιακής τεχνολογίας στην αρχαιολογική έρευνα, οι στρατηγικές συγκρότησης, επιμέλειας και ανοικτής διάθεσης των ψηφιακών συνόλων δεδομένων σπάνια επαρκούν. Οι ολοένα αυξανόμενες εκκλήσεις για την ενθάρρυνση πρακτικών Ανοικτής Επιστήμης τονίζουν την ανάγκη για τη διαφάνεια της ερευνητικής διαδικασίας και την υπολογιστική αναπαραγωγιμότητα κατά τη δημιουργία αρχαιολογικών συνόλων δεδομένων, με ό,τι αυτό συνεπάγεται. Στην Ελλάδα η σχετική συζήτηση στον χώρο της Αρχαιολογίας χρήζει εντατικοποίησης προκειμένου να λάβει υπόψη το τοπίο των διαθέσιμων υποδομών αποθετηρίων, καθώς και την ανάγκη για πρακτικές οδηγίες σε οργανισμούς και ερευνητές που ασχολούνται με την παραγωγή, την οργάνωση και τη διατήρηση συλλογών αρχαιολογικών δεδομένων. Στο παρόν άρθρο σκιαγραφούνται η πρόοδος και τα εμπόδια ως προς την επιμέλεια και τη διάθεση των αρχαιολογικών δεδομένων τα τελευταία χρόνια, ενώ περιγράφονται ενδεικτικά οφέλη από τη διατήρηση και τη δυνατότητα επαναχρησιμοποίησης της ψηφιακής αρχαιολογικής παραγωγής που αντιστοιχεί στο έργο μια ολόκληρης γενιάς. Παρουσιάζονται ακόμη ολοκληρωμένες και τρέχουσες διεθνείς προσπάθειες για την ενίσχυση των πρακτικών Ανοικτής Επιστήμης στην αρχαιολογική έρευνα μέσα από ερευνητικά προγράμματα και συναφή δίκτυα. Τέλος, συνοψίζονται τα στοιχειώδη βήματα που απαιτούνται προς μια “Ανοικτή Αρχαιολογία”.

Keywords*: επιμέλεια ψηφιακών δεδομένων, μακροπρόθεσμη διατήρηση, σύνολα αρχαιολογικών δεδομένων, Ανοικτή Επιστήμη, Ανοικτή Αρχαιολογία*

1. Εισαγωγή

Η παρούσα εργασία επιχειρεί να σκιαγραφήσει το τοπίο της *Ανοικτής Επιστήμης* (Open Science) στην Αρχαιολογία καταγράφοντας τις εγχώριες και τις διεθνείς εξελίξεις. Αρχικά προσδιορίζεται η γενικότερη σημασία και οι στόχοι της Ανοικτής Επιστήμης, περιγράφονται οι σχετικές ευρωπαϊκές υποδομές και πολιτικές που την ενισχύουν, ενώ γίνεται αναφορά και στη σταδιακή εκκίνηση της σχετικής συζήτησης στην Ελλάδα. Στη συνέχεια, συνοψίζονται οι δυνατότητες που ανοίγει η Ανοικτή Επιστήμη στην αρχαιολογική έρευνα, καθώς και οι ιδιαιτερότητες του Ελληνικού παραδείγματος. Παρουσιάζονται ακόμη ενδεικτικά διεθνή αρχαιολογικά ερευνητικά έργα που έχουν στην ατζέντα τους την Ανοικτή Επιστήμη και επισημαίνονται οι παράγοντες που μπορούν να τοποθετήσουν την εφαρμογή των σχετικών προσεγγίσεων σε ένα καλύτερα διαμορφωμένο πλαίσιο, ανοίγοντας το δρόμο προς μια πραγματικά *Ανοικτή Αρχαιολογία* και στη χώρα μας.

2. Τί περιλαμβάνει η Ανοικτή Επιστήμη;

Είναι γεγονός ότι δεν υπάρχει ένας καθολικά αποδεκτός και μεμονωμένος ορισμός της Ανοικτής Επιστήμης. Οι Fecher και Friesike (2013) διακρίνουν αρκετές επιμέρους χροιές που περιλαμβάνουν έννοιες όπως τη διαφάνεια και τη διαθεσιμότητα των ερευνητικών υποδομών, τη διάχυση των ερευνητικών διαδικασιών στο ευρύτερο κοινό, την ελεύθερη προσβασιμότητα στα ερευνητικά αποτελέσματα, τη διαμόρφωση εναλλακτικών τρόπων υπολογισμού του ερευνητικού αντίκτυπου, καθώς και τη συνεργατική ή πληθοποριστική έρευνα. Καθώς όλες οι παραπάνω κατευθύνσεις είναι ακόμη υπό διαμόρφωση, έχει πολύ ενδιαφέρον η παρακολούθηση της συζήτησης λόγω της σημαντικής επίδρασης που δύναται να έχει συνολικά στους τρόπους άσκησης του επιστημονικού έργου κατά τον 21ο αιώνα.

Ωστόσο, σε ένα πρώτο επίπεδο, η Ανοικτή Επιστήμη, ορίζει μια νέα προσέγγιση στην επιστημονική διαδικασία προκειμένου «…*να κάνει τα πρωτεύοντα αποτελέσματα των δημοσίως χρηματοδοτούμενων ερευνητικών αποτελεσμάτων - δημοσιεύσεις και ερευνητικά δεδομένα - δημόσια προσβάσιμα σε ψηφιακή μορφή χωρίς κανέναν ή ελάχιστους περιορισμούς…*» (OECD 2015:7). Η Ανοικτή Επιστήμη αφορά σε όλους ανεξαρτήτως τους επιστημονικούς κλάδους περιγράφοντας, ουσιαστικά, τις διαδικασίες που καθιστούν την επιστημονική παραγωγή και τα αποτελέσματά της ανοιχτά και προσβάσιμα προς όλους. Καίριας σημασίας είναι το 'όλους', καθώς αφορά τόσο στο ακαδημαϊκό ακροατήριο όσο και στο ευρύτερο κοινό. Επιπλέον, η Ανοικτή Επιστήμη στοχεύει στη σύνδεση των επιστημονικών αποτελεσμάτων με τα δεδομένα και τις διαδικασίες που χρησιμοποιούνται για την παραγωγή τους, υποστηρίζοντας την επαλήθευση και την αναπαραγωγιμότητα των ερευνητικών αποτελεσμάτων και προάγοντας τη διαφάνεια της έρευνας. Αυτό επιτυγχάνεται με τη διαθεσιμότητα συνόλων ερευνητικών δεδομένων παράλληλα με τα δημοσιευμένα αποτελέσματα μέσω σχετικών υπηρεσιών (όπως τα αποθετήρια δεδομένων) και άλλων συνεργατικών υποδομών διεξαγωγής έρευνας (όπως τα Εικονικά Εργαστήρια). Με την υιοθέτηση των πρακτικών Ανοικτής Επιστήμης προωθούνται η ανταλλαγή δεδομένων και η ερευνητική συνεργασία, ενώ ο κατά βάση χρηματοδότης και τελικός αποδέκτης της έρευνας, δηλαδή το ευρύτερο κοινωνικό σύνολο, μπορεί να έχει πρόσβαση, να επαναχρησιμοποιήσει και να επωφεληθεί από την επένδυσή του (βλ. Αθανασίου *et al.* 2020:5-9, Geser 2019: 25-27)[1]. Εν κατακλείδι, η Ανοικτή Επιστήμη διευρύνει την ίδια την έρευνα, τα αποτελέσματα και τους αποδέκτες της, μεταστρέφοντας την μέχρι πρότινος εσωστρέφειά της σε εξωστρέφεια.

3. Ανοικτή Επιστήμη: Ευρωπαϊκή πολιτική και υποδομές

Η Ευρωπαϊκή Επιτροπή έχει τοποθετήσει την έννοια της Ανοικτής Επιστήμης στην κορυφή της τρέχουσας και μελλοντικής ερευνητικής ατζέντας. Ο αναμενόμενος αντίκτυπος του ανοίγματος της έρευνας είναι ο μετασχηματισμός της επιστήμης που οδηγεί στην καινοτομία και σε γενικότερα κοινωνικά οφέλη. Εκτός από τις συστάσεις και τις πολιτικές για τα *Ανοικτά Δεδομένα* (Open Data) (ΕΕ 2018) και την *επαναχρησιμοποίηση* (reuse) των πληροφοριών του Δημόσιου Τομέα (ΕΚΒ 2019), ίσως η πιο φιλόδοξη πρωτοβουλία για την ενθάρρυνση της Ανοικτής Επιστήμης είναι η ανάπτυξη του *Ευρωπαϊκού Νέφους Ανοικτής Επιστήμης* (European Open Science Cloud – EOSC)[2]. Το EOSC είναι άμεσα συνδεδεμένο με την Ευρωπαϊκή (ψηφιακή) Στρατηγική και κατέχει σημαντική θέση στο πλαίσιο του Horizon Europe[3]. Στόχος του είναι να υποστηρίξει πρακτικές Ανοικτής Επιστήμης που περιλαμβάνουν την ελεύθερη χρήση και επαναχρησιμοποίηση ενός εκτεταμένου εύρους ψηφιακών ερευνητικών υποδομών. Το νέφος έχει περιγραφεί ως «*μια πανευρωπαϊκή ομοσπονδία υποδομών δεδομένων*» (Jones & Abramatic 2019:14) που θα προσφέρει σε ερευνητές και επαγγελματίες «*ένα εικονικό περιβάλλον με δωρεάν πρόσβαση, καθώς και ανοιχτές και ενιαίες υπηρεσίες για την αποθήκευση, τη διαχείριση, την ανάλυση και την επαναχρησιμοποίηση ερευνητικών δεδομένων, σε διεθνές και διεπιστημονικό επίπεδο*» (Jones & Abramatic 2019:7). Ωστόσο, η επιτυχία του δε σχετίζεται απλώς με τη δημιουργία μιας υποδομής για την κατάθεση και διάθεση δεδομένων, αλλά με τη γενικότερη ανάπτυξη και καλλιέργεια μιας ευρύτερης νοοτροπίας ανταλλαγής δεδομένων μεταξύ των παραγωγών ερευνητικών δεδομένων που θα τροφοδοτεί την πλατφόρμα με περιεχόμενο και θα (επανα)χρησιμοποιεί αυτό και τα εργαλεία της.

Προς αυτή την κατεύθυνση, διαμορφώθηκε η ιδέα των *Δίκαιων* δεδομένων, ευρύτερα γνωστών με το αγγλικό ακρώνυμο *FAIR* (Findable / Accessible / Interoperable / Reusable). Πρόκειται για ερευνητικά δεδομένα που είναι *ευρέσιμα, προσβάσιμα, διαλειτουργικά* και *επαναχρησιμοποιήσιμα*, αντικαθιστώντας τη γενική έννοια των Ανοιχτών Δεδομένων με μια πιο συγκεκριμένη και μετρήσιμη στη βάση διαμορφωμένων κριτηρίων[4]. Χωρίς να επεκταθούμε στις ιδιαιτερότητές τους, σημειώνουμε ότι οι αρχές FAIR καθορίζουν «*συγκεκριμένες προϋποθέσεις για τα σύγχρονα περιβάλλοντα δημοσίευσης δεδομένων, όσον αφορά την υποστήριξη τόσο της μη αυτόματης όσο και της αυτοματοποιημένης εναπόθεσης, εξερεύνησης, κοινής χρήσης και επαναχρησιμοποίησης*» των ερευνητικών δεδομένων (Wilkinson *et al.* 2016). Αυτές

[1] Φυσικά δεν λείπουν και φωνές που τονίζουν ότι οι ανοικτές ερευνητικές πρακτικές, όταν εφαρμόζονται μόνο στη δημόσια και όχι στην ιδιωτικά χρηματοδοτούμενη έρευνα, συνδράμουν τελικά στην αύξηση της ασυμμετρίας των σχετικών πρακτικών (π.χ. Fernández Pinto 2020).

[2] https://eosc-portal.eu/.

[3] https://digital-strategy.ec.europa.eu/en/policies/open-science-cloud.

[4] https://www.go-fair.org/fair-principles/.

οι αρχές λειτουργούν ως οδηγός για τους εκδότες και τους επιμελητές δεδομένων, βοηθώντας τους να αξιολογήσουν, αν τα δεδομένα τους καθίστανται συμβατά με την Ανοικτή Επιστήμη. Προφανώς, πολλές από αυτές τις αρχές μπορεί να είναι αρκετά τεχνικές, απαιτώντας αρκετή προσπάθεια στην εφαρμογή τους. Ως εκ τούτου, παρά την τρέχουσα αύξηση των δεδομένων FAIR, διακρίνεται έλλειψη εφαρμογής, κατανόησης ή ακόμη και γνώσης αυτών των αρχών μεταξύ των ερευνητών, υπογραμμίζοντας την ανάγκη για ενημέρωση και κατάρτιση (βλ. και Geser 2019:27-31).

4. Ανοικτή Επιστήμη: Το παράδειγμα της Ελλάδας

Μέσα στο πλαίσιο αυτό και εστιάζοντας στον χώρο των ανθρωπιστικών επιστημών και του πολιτισμού, θα πρέπει να σημειωθεί πως η Ελλάδα έχει πραγματοποιήσει βήματα για την προώθηση των αρχών της Ανοικτής Επιστήμης. Αυτό βέβαια δεν σημαίνει απαραίτητα ότι η σχετική πρόοδος μπορεί να χαρακτηριστεί απολύτως ικανοποιητική. Ως ορόσημα της συζήτησης στον χώρο των ανθρωπιστικών επιστημών και του πολιτισμού μπορούν να αναφερθούν το 4ο Διεθνές Συνέδριο με τίτλο "*Η Ανοικτή Πρόσβαση στα Ερευνητικά Δεδομένα ως μοχλός για την Ανοικτή Επιστήμη*" (Αθήνα, 15-16/1/2015) και η ημερίδα με τίτλο "*Ανοικτή Επιστήμη: Ζητήματα και Προοπτικές*" (Αθήνα, 15/6/2017) που διοργανώθηκαν από το Εθνικό Κέντρο Τεκμηρίωσης (ΕΚΤ) (ΕΚΤ 2015, Μάλλιου *et al.* 2017). Οι δράσεις αυτές συνέβησαν παράλληλα με την προσπάθεια διαμόρφωσης υποδομών αποθετηρίων, ως ψηφιακή υπηρεσία (Software As A Service – SAAS) από το ΕΚΤ και την ανάπτυξη των εθνικών συσσωρευτών πολιτιστικού και ερευνητικού ψηφιακού περιεχομένου, *SearchCulture* και *OpenArchives* αντίστοιχα[5]. Στο πλαίσιο της συνολικής δράσης του ΕΚΤ, συμπληρώθηκαν με την έκδοση σχετικών οδηγών για την τεκμηρίωση της πολιτιστικής πληροφορίας (Αγγελάκη 2014, Βασιλογαμβράκης & Μπάρτζη 2015, Σταθόπουλος *et al.* 2013), αλλά και βοηθητικών εφαρμογών για την προτυποποίηση μεταδεδομένων, όπως το *Semantics.gr*[6] (Georgiadis *et al.* 2016).

Ακολούθησαν το συμπόσιο "*Ανοικτή Επιστήμη στον ελληνικό ερευνητικό ιστό: Ερευνητικές Διαδικασίες, Ερευνητικά Δεδομένα, Συνεργασίες*" (Αθήνα, 29-30/11/2018) που συνοδεύτηκε από τη διημερίδα "*Ημέρες Ανοικτής Επιστήμης*" (Αθήνα, 21-22/10/2021)[7]. Οι σχετικές δράσεις πραγματοποιήθηκαν με πρωτοβουλία του Ερευνητικού Κέντρου Αθηνά στο πλαίσιο της συμμετοχής του ως εθνικού κόμβου στην ευρωπαϊκή υποδομή OpenAIRE[8], συνοδεύοντας τη συντονισμένη προσπάθεια για τη διαμόρφωση μιας γενικότερης στρατηγικής για την Ανοικτή Επιστήμη, η οποία αποτυπώθηκε στο «*Εθνικό Σχέδιο για την Ανοικτή Επιστήμη*» (Αθανασίου *et al.* 2020) και συνέβαλε στη σύσταση στις 28/2/2022 της Ελληνικής Πρωτοβουλίας για την Ανοικτή Επιστήμη (ΕΠΑΕ) / Hellenic Open Science Initiative (HOSI)[9]. Στον χώρο της Αρχαιολογίας ειδικά αξίζει ακόμη να σημειωθεί η διαδικτυακή συνεδρία του Ελληνικού παραρτήματος του οργανισμού *Εφαρμογές Υπολογιστών και Ποσοτικές Μέθοδοι στην Αρχαιολογία* (CAA-GR) με τίτλο "Open Digital Archaeological Content in the Connected World: Curation and Stewardship"[10].

Όπως και στο εξωτερικό, η συζήτηση στην Ελλάδα είχε ως αφετηρία την ανάγκη για ανοικτή πρόσβαση στις δημοσιεύσεις και την επονομαζόμενη *γκρίζα* βιβλιογραφία (grey literature), ωστόσο σταδιακά επεκτάθηκε προς συνολικότερες θεματικές υπό το ευρύτερο πρίσμα της Ανοικτής Επιστήμης. Σημαντικό ρόλο στην εξέλιξη αυτή έπαιξαν και οι συνολικές κατευθύνσεις των ερευνητικών πολιτικών της Ευρωπαϊκής Ένωσης, που προώθησαν τη διαφάνεια και την προσβασιμότητα των δεδομένων του δημόσιου τομέα. Στην πρόσφατη έκδοση της εγχώριας *Βίβλου του Ψηφιακού μετασχηματισμού*[11], αναφέρονται τόσο οι στόχοι όσο και οι τρέχουσες και προγραμματιζόμενες

[5] https://www.searchculture.gr & https://www.openarchives.gr
[6] https://www.semantics.gr/.

[7] Οι παρουσιάσεις και το οπτικοακουστικό υλικό από τις εκδηλώσεις είναι διαθέσιμα στο διαδίκτυο. Βλ. https://www.athenarc.gr/el/events/open-science-symposium & https://www.athenarc.gr/el/events/imeres-anoiktis-epistimis-stin-ellada αντίστοιχα.
[8] Το OpenAIRE λειτουργεί υποστηρικτικά στην προσπάθεια ανάπτυξης του EOSC μέσω της ευθυγράμμισης θεσμικών και εθνικών πολιτικών εντός της ΕΕ για την Ανοικτή Επιστήμη, τον συνδυασμό διαθέσιμων υποδομών και την ανάπτυξη νέων, καθώς και τη διεύρυνση της εκπαίδευσης και της ευαισθητοποίησης της ελληνικής ακαδημαϊκής και ερευνητικής κοινότητας. Βλ. https://www.openaire.eu/ & https://www.openaire.eu/os-greece, όπου παρουσιάζεται και το σύνολο των δράσεων που πραγματοποιήθηκαν μαζί με το αντίστοιχο πληροφοριακό υλικό.
[9] https://www.hellenicopenscience.gr/
[10] Η συνεδρία πραγματοποιήθηκε στις 21/4/2021 (https://gr.caa-international.org/2021/03/24/caa-gr-2021-sessions-spring-edition/). Οι παρουσιάσεις είναι διαθέσιμες μέσω του δικτύου CARARE (Connecting ARchaeology and ARchitecture in Europeana) στην πλατφόρμα Vimeo (https://vimeo.com/user124611809).
[11] https://digitalstrategy.gov.gr/.

δράσεις ανοικτής διακυβέρνησης για το διάστημα 2020-2025, περιλαμβάνοντας αναφορές στις στοχεύσεις των σχετικών με τον πολιτισμό και την επιστημονική έρευνα υπουργείων (ΥΨΔ 2021).

5. Γιατί χρειάζεται η Ανοικτή Επιστήμη στην Αρχαιολογία;

Σε διεθνές επίπεδο η έννοια της Ανοικτής Επιστήμης έχει χαρακτηριστεί ως μία από τις μεγάλες προκλήσεις της αρχαιολογικής πρακτικής δίνοντας ιδιαίτερη έμφαση στην ανοικτή πρόσβαση, τα ανοικτά δεδομένα και τις ανοικτές μεθόδους (Marwick *et al.* 2017). Άλλωστε, οι αρχαιολογικές επεμβάσεις, όπως η ανασκαφή θεωρείται ότι «καταστρέφουν» ή έστω «μετουσιώνουν» τα πρωτογενή αρχαιολογικά υλικά σε ερευνητικά τεκμήρια (βλ. Lucas 2001), ενώ όλες οι εργασίες που σχετίζονται με την αρχαιολογία και την πολιτιστική κληρονομιά εκτελούνται στο όνομα του δημόσιου συμφέροντος. Ωστόσο, η αρχαιολογική έρευνα μέχρι σήμερα περιλαμβάνει πολλά *μεμονωμένα* σύνολα δεδομένων (data silos) με τη μορφή ερευνητικών αρχείων, αδημοσίευτων εκθέσεων ή ακαδημαϊκών προϊόντων (η λεγόμενη *γκρίζα* βιβλιογραφία) που συχνά είναι απροσπέλαστα σε προσωπικές ή κλειστές υποδομές αποθήκευσης (Huggett *et al.* 2018, Moore & Richards 2015). Οι ερευνητικές δημοσιεύσεις, ακόμη και όταν περιέχουν εκτεταμένα παραρτήματα με πίνακες, σχέδια και φωτογραφίες που τεκμηριώνουν το ερμηνευτικό αποτέλεσμα, σπάνια παρέχουν αναλυτικά τα *πρωτογενή* σύνολα δεδομένων τους, ενώ συχνά πολλές από τις διαδικασίες που εμπλέκονται στην αρχαιολογική ερμηνεία (π.χ. δειγματοληπτικές μεθοδολογίες) τεκμηριώνονται με διαφορετικούς βαθμούς ανάλυσης και επιλεκτικά.

Τα ψηφιακά σύνολα δεδομένων, παρά την αυξανόμενη χρήση τους ως υποστηρικτικά στοιχεία μιας έρευνας, είναι εξίσου δυσεύρετα. Θα πρέπει να διευκρινιστεί εδώ πως δεν πρέπει να συγχέουμε τα ερευνητικά σύνολα δεδομένων, με τη δημόσια διαθεσιμότητα ψηφιακών συλλογών (όπως π.χ. η συλλογή εκθεμάτων ενός μουσείου ή το αρχειακό απόθεμα ενός οργανισμού). Αναφερόμαστε συγκεκριμένα σε σύνολα δεδομένων της αρχαιολογικής έρευνας και για τον τρόπο με τον οποίο μπορούν να είναι διαθέσιμα και τεκμηριωμένα ως προς την προέλευσή τους και τις επακόλουθες τροποποιήσεις κατά τη διάρκεια της ερευνητικής διαδικασίας. Κάτι τέτοιο συνεπάγεται προσοχή στον τρόπο με τον οποίο τα αρχαιολογικά δεδομένα συγκεντρώνονται εξαρχής, οργανώνονται, μετασχηματίζονται και τελικά αρχειοθετούνται, με σκοπό τη μακροπρόθεσμη διατήρησή τους και την περαιτέρω διαθεσιμότητά τους (Beck & Neylon 2012, Costa *et al.* 2014).

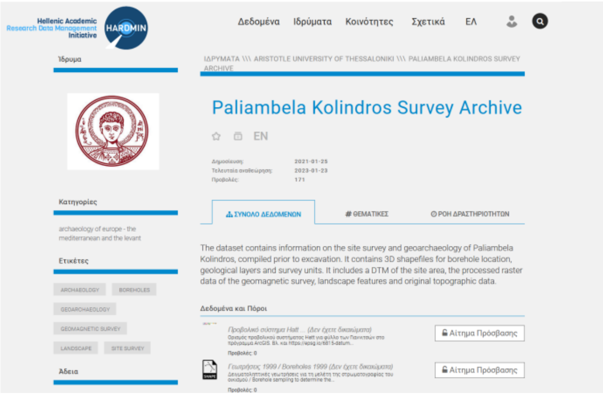
Υπάρχουν αρκετά πλεονεκτήματα στην ανοικτή διάθεση αρχαιολογικών ερευνητικών συλλογών δεδομένων. Καταρχάς, το τρέχον περιβάλλον ψηφιακών εκδόσεων έχει υπερκεράσει τα παραδοσιακά εμπόδια στη συμπερίληψη δεδομένων και υποστηρίζει τη συμπλήρωση των ερευνητικών αποτελεσμάτων με τα αρχικά σύνολα δεδομένων τους (βλ. Kansa 2012, Moore & Richards 2015, Opitz 2018). Οι πρακτικές ψηφιακής διάδοσης επιχειρούν μάλιστα να είναι πιο προσιτές προκειμένου τα ερευνητικά αποτελέσματα να είναι κατανοητά από οποιονδήποτε θέλει να τα μελετήσει και να τα αναθεωρήσει. Η διαθεσιμότητα των δεδομένων αποτρέπει τη επαναληπτική συλλογή δεδομένων και διοχετεύει τη διαθέσιμη χρηματοδότηση στην ανάλυση δεδομένων αντί για την εκ νέου συλλογή τους. Καθώς η αρχαιολογική έρευνα εξαρτάται όλο και περισσότερο από ψηφιακά δεδομένα και υπολογιστικές διαδικασίες που είναι επιρρεπείς σε αλλαγές, τα απροσπέλαστα δεδομένα αυξάνουν τον κίνδυνο απώλειας. Ο πιο αποτελεσματικός τρόπος για να διατηρεί κανείς τα δεδομένα του είναι να τα χρησιμοποιεί, να κάνει γνωστή την ύπαρξη και τη διαθεσιμότητά τους, καθώς και να ενθαρρύνει περαιτέρω και άλλους να τα χρησιμοποιήσουν. Τα υπάρχοντα σύνολα δεδομένων είναι σημαντικό να είναι προσβάσιμα και να έχουν τη δυνατότητα της πολλαπλής χρήσης για τη διερεύνηση νέων ερευνητικών ερωτημάτων, επιτρέποντας την αξιοποίηση προηγούμενων επενδύσεων σε εργασία και πόρους. Ταυτόχρονα, ενισχύονται και οι μέθοδοι επεξεργασίας *Μεγάλων Δεδομένων* (Big Data) που βασίζονται σε σημαντικό βαθμό στη συγκέντρωση και την ανοικτή διαθεσιμότητα εκτεταμένων ερευνητικών πληροφοριακών συνόλων (Gattiglia 2015, βλ. και Geser 2019:75-76, 119-120).

6. Ανοικτή επιστήμη και Αρχαιολογία στην Ελλάδα

Εστιάζοντας τελικά στον χώρο της αρχαιολογίας, σε μια πρόσφατη επισκόπηση των τρεχουσών πρακτικών διαχείρισης ερευνητικών δεδομένων στην Αρχαιολογία στην Ελλάδα, συνειδητοποιήσαμε αρκετά πράγματα που σχετίζονται με αυτήν τη συζήτηση (Tsiafaki & Katsianis 2021). Αρχικά, πρέπει να σταθούμε στην παρατηρούμενη επιφυλακτικότητα των αρχαιολόγων-ερευνητών να υιοθετήσουν προσεγγίσεις της Ανοικτής Επιστήμης, λόγω των κινδύνων που σχετίζονται για παράδειγμα, με την κλοπή δεδομένων ή τη διαφάνεια πιθανών ερευνητικών σφαλμάτων. Υπάρχουν επίσης ανησυχίες για την προστασία της πολιτιστικής κληρονομιάς και ζητήματα που αφορούν στην προστασία των προσωπικών πνευματικών δικαιωμάτων, ενώ συγκεκριμένοι περιορισμοί του υφιστάμενου αρχαιολογικού νόμου συχνά λειτουργούν αποτρεπτικά ως προς το άνοιγμα της έρευνας και του διαμοιρασμού αρχαιολογικών τεκμηρίων (βλ. Κανελλοπούλου-Μπότη *et al.* 2013, Πάντος 2013, Tsiavos 2020).

Ειδικά στην Ελλάδα φαίνεται ότι οι παράγοντες αυτοί επιτείνονται από μια εδραιωμένη νοοτροπία ενάντια στην ανοικτή διάθεση των δεδομένων της αρχαιολογικής έρευνας (βλ. και Sitara & Vouligea 2014). Αυτή η νοοτροπία ωστόσο, φαίνεται πως έχει αρχίσει να αλλάζει και μπορεί να ενισχυθεί προς την κατεύθυνση αυτή με την εισαγωγή νέων πολιτικών που ενθαρρύνουν και υποστηρίζουν τη διατήρηση ψηφιακών δεδομένων και την ανοικτή διάδοση. Προς το παρόν, δεν υπάρχουν προδιαγεγραμμένες κατευθυντήριες γραμμές για τη διαχείριση δεδομένων ανοικτής πρόσβασης, όπως σε άλλες χώρες, ενώ οι φορείς χρηματοδότησης συνήθως δεν απαιτούν *Σχέδια Διαχείρισης Δεδομένων* (Data Management Plans - DMP). Ακόμα κι αν υπήρχαν ρυθμιστικά κίνητρα, ωστόσο, δεν υπάρχει επαρκής τεχνογνωσία από την πλευρά των ερευνητών για να ακολουθήσουν μια ανοικτή επιστημονική προσέγγιση. Τα δεδομένα FAIR σε οποιοδήποτε επίπεδο (προσωπικό ή θεσμικό) απαιτούν ψηφιακές δεξιότητες και εκπαίδευση, καθώς και διεπιστημονική συνεργασία για την υλοποίηση τους. Εξίσου σημαντικό ρόλο, έχει παίξει και η σχετική καθυστέρηση στην ανάπτυξη υποδομών αποθήκευσης ή ακόμη και στη διαμόρφωση διαδικασιών δομημένης εναπόθεσης. Όλοι αυτοί οι παράγοντες συμβάλλουν σε αυτό που περιγράψαμε ως ένα *τοπίο κατακερματισμού*, όταν πρόκειται για πρακτικές διαχείρισης δεδομένων που αποτελούν κατά μία έννοια τη ραχοκοκαλιά μιας Ανοικτής Επιστήμης στον χώρο της Αρχαιολογίας.

Μέσα σε αυτό το τοπίο, όσον αφορά στις υποδομές αποθετηρίων, μια πρόσφατη προσθήκη είναι η *Ελληνική Υπηρεσία Δεδομένων* ή αλλιώς *HELIX*. Η ανάπτυξή του πραγματοποιείται στο πλαίσιο του OpenAIRE από το Ερευνητικό Κέντρο Αθηνά και το ΕΔΥΤΕ (Εθνικό Δίκτυο Υποδομών Τεχνολογίας και Έρευνας) με στόχο την παροχή μίας ψηφιακής υποδομής νέφους για την οριζόντια ερευνητική υποστήριξη, που περιλαμβάνει επεκτάσιμες ψηφιακές υπηρεσίες για τη διαχείριση και το διαμοιρασμό επιστημονικών δημοσιεύσεων, συνόλων δεδομένων και εικονικών εργαστηρίων[12]. Μέσα σε αυτή την υποδομή εντάσσεται το *HARDMIN*, το εθνικό καταθετήριο ερευνητικών δεδομένων που αναπτύχθηκε σε συνεργασία με τον *Σύνδεσμο Ελληνικών Ακαδημαϊκών Βιβλιοθηκών* (ΣΕΑΒ) (Pispiringas *et al.* 2019). Το HARDMIN απευθύνεται κυρίως στην ακαδημαϊκή κοινότητα, ωστόσο δύναται να αποτελέσει μια λειτουργική λύση συνολικά για τα αρχαιολογικά ερευνητικά δεδομένα που παράγονται από κρατικούς ερευνητικούς φορείς, καθώς και άλλους ιδιωτικούς ή μη κερδοσκοπικούς οργανισμούς πολιτιστικής κληρονομιάς (Εικ. 1). Κατά μία έννοια, λοιπόν, το πρόβλημα της έλλειψης κατάλληλων υποδομών αποθήκευσης δεδομένων φαίνεται σταδιακά να ξεπερνιέται.



Εικόνα 1 Συλλογή αρχαιολογικών δεδομένων στην πλατφόρμα HARDMIN από το αρχαιολογικό πρόγραμμα Παλιαμπέλων Κολινδρού.

7. Ανοικτή επιστήμη και διεθνείς προσπάθειες στο χώρο της Αρχαιολογίας

Πολλά από τα προβλήματα που επισημάνθηκαν για την Ελλάδα αναδεικνύονται και στο ευρωπαϊκό πλαίσιο, καθώς κάθε χώρα έχει τη δική της στρατηγική για την επίτευξη του στόχου της Ανοικτής Επιστήμης. Ωστόσο, παρατηρούμε ότι ήδη υπάρχουν πολλά διεπιστημονικά ή θεματικά αποθετήρια, ενώ σημαντική προσπάθεια έχει ήδη καταβληθεί στην ανάπτυξη κατευθυντήριων γραμμών και πολιτικών για την Ανοικτή Επιστήμη στην Αρχαιολογία (βλ. και Fernández Cacho 2021). Τα τελευταία χρόνια έχουν διευρυνθεί οι επιλογές για δικτύωση, με σκοπό την απόκτηση και ανταλλαγή σχετικής εμπειρίας μέσω της ενεργούς συμμετοχής ή της απλής ενημέρωσης για τα αποτελέσματα από αρκετά διεθνή έργα. Σε αυτό το πλαίσιο καταβάλλεται ιδιαίτερη προσπάθεια, ώστε να γίνουν πιο διαλειτουργικές οι διαφορετικές προσεγγίσεις για τη διαχείριση δεδομένων και να ενταχθούν δεδομένα από διαφορετικούς χώρους αποθήκευσης σε μεγάλες διεθνείς υποδομές. Τέλος, η ευαισθητοποίηση και η εκπαίδευση για την Ανοικτή Επιστήμη αρχίζει να κερδίζει την προσοχή, όπως

[12] https://hellenicdataservice.gr/

υποδηλώνουν οι στόχοι μεγάλων ευρωπαϊκών χρηματοδοτούμενων έργων, από τα οποία αναφέρονται ενδεικτικά τα παρακάτω:

Το έργο *PARTHENOS* (2015-2019) είχε ως στόχο την ενίσχυση της συνοχής της έρευνας στην Ψηφιακή Κληρονομιά, ενσωματώνοντας επιμέρους πρωτοβουλίες και φέρνοντας σε επαφή επαγγελματίες από συναφείς τομείς[13]. Επικεντρώθηκε στον καθορισμό και την υποστήριξη κοινών λύσεων και κοινών προτύπων σε ψηφιακές επιστημονικές πρακτικές στον τομέα της πολιτιστικής κληρονομιάς, καθώς και στην εναρμόνιση των σχετικών πολιτικών και την εφαρμογή τους (Uiterwaal *et al.* 2021). Μεταξύ των παραδοτέων του έργου, περιλαμβάνονται ένα πρότυπο σχέδιο διαχείρισης αρχαιολογικών ερευνητικών δεδομένων (Giorgio & Ronzino 2018), καθώς και κατευθυντήριες γραμμές για την εφαρμογή των αρχών δεδομένων FAIR. Το σχετικό έγγραφο έχει μεταφραστεί σε πολλές γλώσσες και διατίθεται και στα ελληνικά (PARTHENOS *et al.* 2019).

Λίγο πριν την πανδημία διαμορφώθηκε διεθνές ερευνητικό δίκτυο με το ακρώνυμο *SEADDA* (2019-2023), έχοντας ως στόχο τη "*διάσωση της Ευρωπαϊκής Αρχαιολογίας από τον Ψηφιακό Μεσαίωνα*"[14]. Ουσιαστικά στο πλαίσιο του προγράμματος επιχειρείται η δημιουργία ενός συνδέσμου ερευνητών που επιχειρούν να καταγράψουν την τρέχουσα κατάσταση αιχμής σε όλη την Ευρώπη ως προς τις πρακτικές διαχείρισης αρχαιολογικών δεδομένων, να χαρτογραφήσουν υπάρχουσες υποδομές και απαιτήσεις ψηφιακής αρχειοθέτησης και να προωθήσουν προτάσεις για τον μετριασμό των προβλημάτων που προκύπτουν από την έλλειψη προδιαγραφών τεκμηρίωσης ψηφιακών δεδομένων και στρατηγικών διατήρησης, μέσα από βέλτιστες πρακτικές για τη βιωσιμότητα των συνόλων δεδομένων αρχαιολογικής έρευνας. Η ιδέα δεν περιορίζεται στο θέμα της διατήρησης, αλλά εξετάζει τις ανάγκες του τρέχοντος περιβάλλοντος της Ανοικτής Επιστήμης και των δεδομένων FAIR. Στο πλαίσιο των εργασιών του δικτύου, δημοσιεύτηκε ειδικό τεύχος στο περιοδικό *Internet Archaeology* με θέμα τη χαρτογράφηση της τρέχουσας κατάστασης ως προς την επιμέλεια αρχαιολογικών δεδομένων ανά χώρα σε διεθνές επίπεδο (Richards *et al.* 2021, βλ. και Tsiafaki & Katsianis 2021 για την Ελλάδα).

Έχοντας συναφή στόχευση, το ευρωπαϊκό πρόγραμμα *ARIADNEplus*[15] (2019-2022) προσπάθησε να συνδυάσει περιεχόμενο από πολλαπλά αρχαιολογικά αποθετήρια και εμπειρία από επαγγελματίες του χώρου, προκειμένου να ευθυγραμμίσει τις πολιτικές επιμέλειας ψηφιακών αρχαιολογικών δεδομένων σε διεθνές επίπεδο (Niccolucci & Richard 2019). Το πρόγραμμα αποτέλεσε διάδοχο του προγράμματος ARIADNE που έλαβε χώρα μεταξύ 2013-2017. Η προσέγγιση συνδέεται με τη σταδιακή ανάπτυξη μιας υποδομής *μετα*-αναζήτησης για σύνολα δεδομένων που έχουν παραχθεί σε παγκόσμια γεωγραφική κλίμακα με ποικίλο επίπεδο ανάλυσης, τεκμηρίωσης, εμβέλειας και σημασιολογικής διαλειτουργικότητας. Η κοινοπραξία συμπεριέλαβε 41 εταίρους από 23 χώρες και 4 διεθνείς οργανισμούς συντονίζοντας την προσπάθεια ενοποίησης και ομογενοποίησης σε ενιαία ψηφιακή υποδομή άνω των δύο εκατομμυρίων ψηφιακών αρχαιολογικών δεδομένων κάθε είδους (από δεδομένα ανασκαφών και επιφανειακών ερευνών, έως εθνικά αρχεία μνημείων ή αποτελέσματα επιστημονικών αναλύσεων). Μετά το πέρας της περιόδου χρηματοδότησης η κοινοπραξία ίδρυσε τη μη κερδοσκοπική ερευνητική υποδομή ARIADNE RI[16] με στόχο τη βιωσιμότητα του ενιαίου μετα-αποθετηρίου της πύλης ARIADNE, σε συνδυασμό με τη διατήρηση αλλά και την περαιτέρω επέκταση της αρχαιολογικής κοινότητας που δημιουργήθηκε στη διάρκεια των δύο προγραμμάτων (2013-2022).

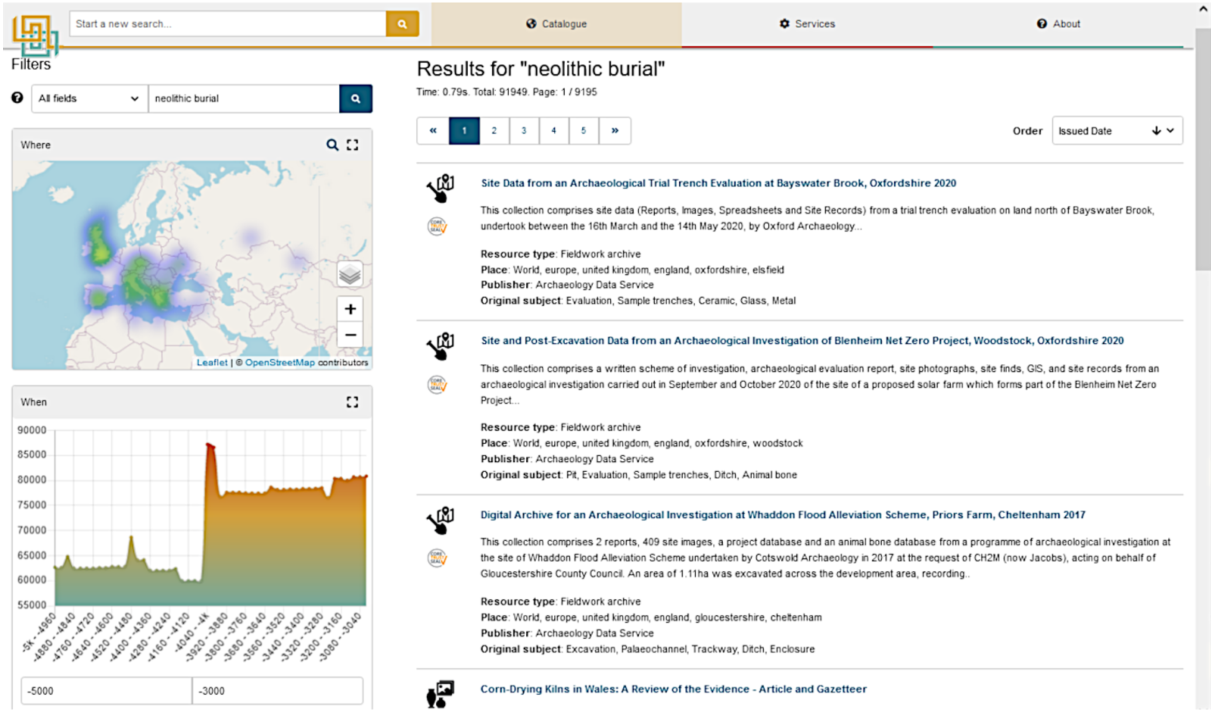
Η πύλη ARIADNE λειτουργεί ως πλατφόρμα διάθεσης ψηφιακών αρχαιολογικών δεδομένων από την Ευρώπη και όχι μόνο. Η πύλη αναζήτησης χρησιμοποιεί την υποκείμενη κοινή σημασιολογική χαρτογράφηση σε επίπεδο συλλογής των επιμέρους ψηφιακών πόρων και περιλαμβάνει εργαλεία για την συνδυαστική τους αναζήτηση και το φιλτράρισμα των αποτελεσμάτων (Εικ. 2). Το περιεχόμενο και τα εργαλεία της πλατφόρμας δύνανται να διασυνδεθούν με το EOSC επιτρέποντας τη διαθεσιμότητά τους μέσω και της υποδομής του τελευταίου.

[13] *Pooling Activities, Resources and Tools for Heritage E-research Networking, Optimization and Synergies* – Horizon 2020 PARTHENOS (GA654119). https://www.parthenos-project.eu.

[14] *Saving European Archaeology from the Digital Dark Age* – COST Action SEADDA (CA18128). https://www.seadda.eu/.

[15] *Advanced Research Infrastructure for Archaeological Dataset Networking in Europe plus* – Horizon 2020 ARIADNE+ (GA823914). https://ariadne-infrastructure.eu/. Η σχετική πύλη αναζήτησης δεδομένων είναι προσβάσιμη στη διεύθυνση https://portal.ariadne-infrastructure.eu/.

[16] ARIADNE Research Infrastructure AISBL. https://www.ariadne-research-infrastructure.eu/.

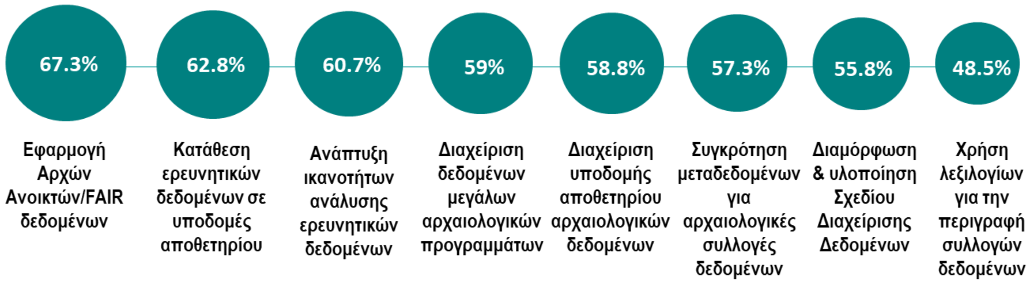


Εικόνα 2 Αναζήτηση στην πύλη του ARIADNEplus (στις 27/01/2022) με τον όρο “neolithic burial” και χρησιμοποιώντας χρονολογικό φίλτρο (5000-3000 π.Χ.). Εμφανίζονται 91949 αποτελέσματα, τα οποία παρατίθενται σε λίστα και εμφανίζονται ως προς τη χωρο-χρονική τους συγκέντρωση στα συνοδευτικά παράθυρα.

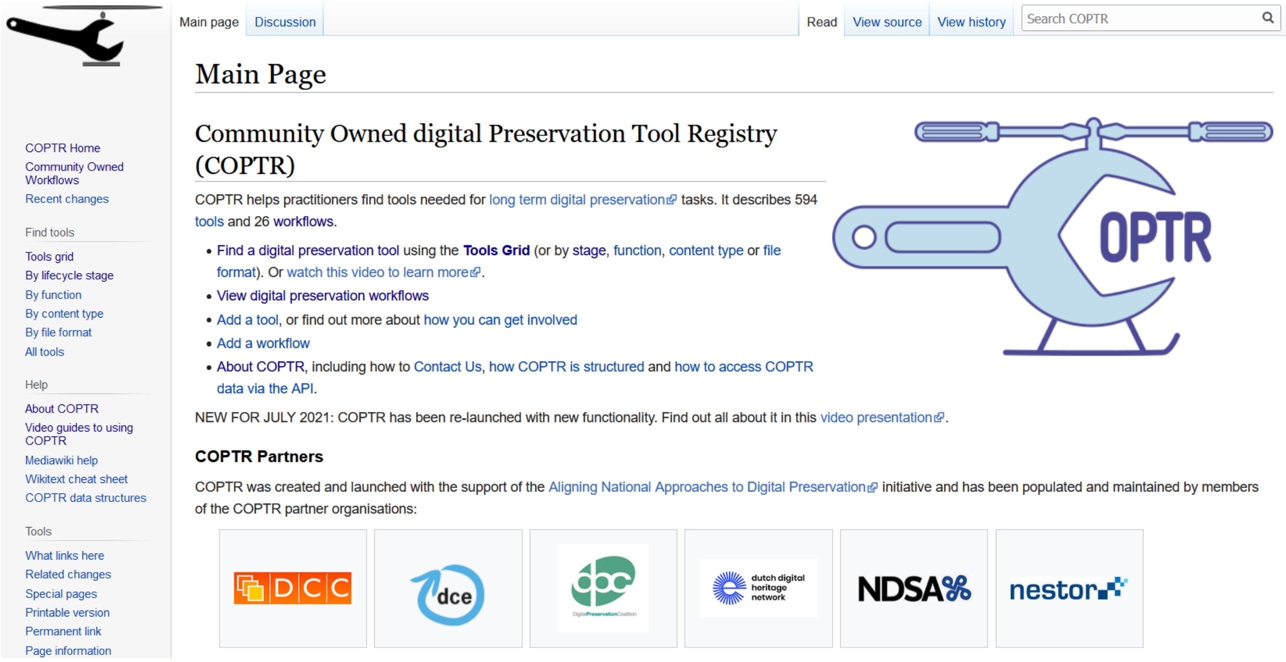
Στο πλαίσιο του ίδιου προγράμματος διεξήχθη έρευνα για τις ανάγκες της αρχαιολογικής επιστημονικής κοινότητας εστιάζοντας – μεταξύ άλλων – στο τρέχον επίπεδο υιοθέτησης της Ανοικτής Επιστήμης και στις αντίστοιχες ανάγκες κατάρτισης (Geser 2019). Όσον αφορά στα εμπόδια ως προς την κατάθεση ερευνητικών δεδομένων σε ψηφιακά αποθετήρια (Εικ. 3), κρίθηκαν ως ιδιαίτερα σημαντικά: η έλλειψη επαγγελματικής αναγνώρισης και ανταμοιβής ως προς την ανοικτή διάθεση των δεδομένων μιας έρευνας, τα πιθανά ζητήματα πνευματικής ιδιοκτησίας επί της παραγόμενης πληροφορίας, καθώς και η επένδυση σε εργασία που απαιτεί η προσαρμογή των δεδομένων σύμφωνα με τα πρότυπα των αποθετηρίων. Ακολουθούν οι ελλείψεις ως προς την επιβολή της σχετικής υποχρέωσης από τους φορείς ερευνητικής χρηματοδότησης ή/και την αρχαιολογική νομοθεσία, καθώς και οι περιορισμοί στις διαθέσιμες υποδομές καταθετηρίων και τα σχετικά κόστη (Geser 2019: 65-68, 73-74). Όλες οι επισημάνσεις αναδεικνύουν την ανάγκη για ρύθμιση των ερευνητικών υποδομών και συνθηκών διεξαγωγής ερευνητικού έργου τόσο ως προς τις υποχρεώσεις των ερευνητών όσο και ως προς τις ανάγκες χρηματοδοτικής κάλυψης των σχετικών υποχρεώσεων.

Από την άλλη σημαντικές είναι και οι ανάγκες κατάρτισης (Εικ. 4), για τις οποίες η ίδια έρευνα έδειξε ότι πρέπει να δοθεί ιδιαίτερη έμφαση σε μια σειρά από θεματικές, όπως: η ενημέρωση και η καθοδήγηση για την εφαρμογή των αρχών δεδομένων FAIR, η επεξήγηση της ροής εργασιών κατάθεσης δεδομένων, η ανάπτυξη σχετικών δεξιοτήτων για την ανάλυση επιστημονικών δεδομένων, η διαμόρφωση και υλοποίηση Σχεδίων Διαχείρισης Δεδομένων, η ανάπτυξη και υποστήριξη υποδομών διαχείρισης δεδομένων, καθώς και η εκπαίδευση ως προς την εφαρμογή προτύπων μεταδεδομένων, ορολογίας και σημασιολογίας (Geser 2019:120-122). Είναι γεγονός ότι η ενθάρρυνση της Ανοικτής Επιστήμης απαιτεί εξίσου σημαντική επένδυση τόσο σε υποδομές και σχετικές χρηματοδοτικές προβλέψεις όσο και στην εκπαίδευση των νέων ερευνητών στα σχετικά ψηφιακά εργαλεία και τις ροές ψηφιακής επιμέλειας.

Εικόνα. 3 Ιεράρχηση εμποδίων ως προς την κατάθεση ερευνητικών δεδομένων σε ψηφιακά αποθετήρια με βάση τις απαντήσεις «πολύ» ή «αρκετά» σημαντικό [398 απαντήσεις επί συνόλου 415]. Προσαρμοσμένο στα ελληνικά από Geser 2020[17].



Εικόνα 4 Διαδικασίες για τις οποίες η πρακτική εξάσκηση και εκπαίδευση θα ήταν «πολύ βοηθητικές» [328 απαντήσεις επί συνόλου 330]. Προσαρμοσμένο στα ελληνικά από Geser 2020.



Εικόνα 5 Η σελίδα της κοινότητας COPTR.

[17] Οι φωτογραφίες 2 & 3 αποτελούν επεξεργασμένες εκδοχές των γραφικών που περιλαμβάνονται στο κείμενο της σύνοψης των βασικών ευρημάτων της έρευνας, διαθέσιμο στη σελίδα https://ariadne-infrastructure.eu/key-results-of-the-community-needs-survey/ (Ημ/νία πρόσβασης 28/01/2022).

Πληροφοριακό υλικό για την κάλυψη πολλών αναγκών σε εκπαίδευση είναι ήδη διαθέσιμο στο πλαίσιο των προαναφερθέντων έργων και άλλων πρωτοβουλιών[18]. Πολλά αποθετήρια παρέχουν κατευθυντήριες γραμμές σχετικά με τον τρόπο προετοιμασίας των δεδομένων για κατάθεση σύμφωνα με τις αρχές FAIR, ενώ σχετικές προβλέψεις έχουν αρχίσει να ενσωματώνονται και από σημαντικούς φορείς χρηματοδότησης (π.χ. ERC 2021). Επιπλέον, διατίθενται συλλογές εργαλείων για τη διατήρηση δεδομένων και την ανοικτή επιστημονική πρακτική, κατά προτίμηση χρησιμοποιώντας εργαλεία ανοιχτού κώδικα. Ένα πρόσφατο παράδειγμα συνιστά η συμμετοχική πλατφόρμα *COPTR*, (Εικ. 5) που συνιστά ένα μητρώο εργαλείων ψηφιακής επιμέλειας το οποίο ανήκει στην ερευνητική κοινότητα, πράγμα που σημαίνει ότι οι εργαλειοθήκες που περιλαμβάνονται αναπτύσσονται και διατηρούνται από την ερευνητική κοινότητα που τις χρησιμοποιεί (COPTR contributors 2021).

8. Προς μία Ανοικτή Αρχαιολογία στην Ελλάδα

Ήδη από το 2012 η έννοια της *Ανοικτής Αρχαιολογίας* (Open Archaeology) έκανε την εμφάνισή της περιλαμβάνοντας αρκετές ακόμη επιπλέον εκφάνσεις που σχετίζονται με τον εκδημοκρατισμό της αρχαιολογικής γνώσης και τη συμμετοχή του κοινού (Aspöck 2019, Beck & Neylon 2012, Lake 2012). Στο παρόν άρθρο επικεντρωθήκαμε στις σχέσεις μεταξύ αρχαιολογικής πρακτικής και ψηφιακών ροών εργασιών, πληροφοριακού αποθέματος και ερευνητικού αποτελέσματος. Συνοψίζοντας, σημειώνονται κάποιες προκαταρκτικές σκέψεις.

Συνολικά, όπως έχει σημειωθεί από πολλούς ερευνητές, η υιοθέτηση των πρακτικών της Ανοικτής Επιστήμης απαιτεί συντονισμένη προσπάθεια σε πολλαπλά επίπεδα, από τον ίδιο τον ερευνητή ως το ευρύτερο πλαίσιο λειτουργίας της ερευνητικής διαδικασίας. Σε πρώτο επίπεδο, είναι αναγκαία η αλλαγή νοοτροπίας που στερεί την επιστημονική κοινότητα από πολύτιμα δεδομένα λόγω ανησυχιών για πιθανή κακόβουλη χρήση τους ή παράλειψη αναφοράς, καθώς και στενά ιδιοτελή κίνητρα για συσσώρευση δεδομένων.

Θεμελιακή κρίνεται η σημασία της θέσπισης ενός κεντρικού πλαισίου και γενικών κατευθυντηρίων γραμμών που θα ακολουθούνται στην κατάρτιση ενός Σχεδίου Διαχείρισης Δεδομένων (DMP) κατά την έναρξη ενός ερευνητικού έργου. Αυτό θεωρείται αναγκαίο προκειμένου να γίνουν κατανοητές οι πιθανές απαιτήσεις σε υποδομές ψηφιακών δεδομένων και να εξορθολογιστεί η ψηφιακή ροή παραγωγής δεδομένων, ώστε να αποφεύγονται φαινόμενα συμφόρησης, τόσο κατά την παραγωγή όσο και κατά την τελική αποθήκευση των παραγόμενων δεδομένων (βλ. Strupler & Wilkinson 2017). Ας μην ξεχνάμε ότι πολλές ψηφιακές μεθοδολογίες (π.χ. 3Δ φωτογραμμετρία) περιλαμβάνουν πολλά ενδιάμεσα στάδια επεξεργασίας με επιμέρους προϊόντα, τα οποία συνιστούν κομμάτια μιας ψηφιακής αλυσίδας παραγωγής και πρέπει να συνοδεύουν το τελικό προϊόν, προκειμένου να είναι δυνατή η πιθανή μελλοντική επανεπεξεργασία τους με διαφορετικές μεθόδους (βλ. και Κατσιάνης 2013).

Πέρα από τις προβλέψεις και τα συστήματα ασφάλειας, διαχείρισης και διάθεσης δεδομένων κάθε οργανισμού, υπάρχουν πρόσθετες επιλογές για την αποθήκευση, διατήρηση και ανοικτή διάθεση δεδομένων, τόσο στο πλαίσιο προσβάσιμων υπηρεσιών αποθετηρίου (π.χ. GitHub) όσο και σχετικών διεθνών αρχαιολογικών υποδομών (π.χ. ADS, tDAR). Σε πρόσφατη έρευνα για τις στρατηγικές διαχείρισης δεδομένων και την παρούσα κατάσταση ως προς τη λειτουργία ψηφιακών αρχαιολογικών αποθετηρίων οι Geser et al. (2022) σημειώνουν σημαντικές διαφοροποιήσεις ως προς τις πρακτικές διαχείρισης δεδομένων και τις υφιστάμενες υπηρεσίες αρχαιολογικών αποθετηρίων σε διεθνές επίπεδο. Σε κάθε περίπτωση, ωστόσο, διακρίνεται η συνειδητοποίηση από σημαντικό ποσοστό των επαγγελματιών της πολιτιστικής κληρονομιάς της ανάγκης για βελτίωση των υποδομών και για ρύθμιση του σχετικού περιβάλλοντος της επιμέλειας των ψηφιακών αρχαιολογικών δεδομένων. Η πρόσφατη ανάπτυξη της Ελληνικής Υπηρεσίας Δεδομένων (HELIX) διαμορφώνει μια ακόμη επιλογή σε εγχώριο επίπεδο, η επιβίωση και η εξέλιξη της οποίας θα εξαρτηθεί σε σημαντικό βαθμό από τη υιοθέτηση και χρήση της από την ελληνική ερευνητική κοινότητα εν γένει.

[18] Βλέπε https://portal.ariadne-infrastructure.eu/services, όπου περιλαμβάνονται πρότυπα διαμόρφωσης Σχεδίων Διαχείρισης Δεδομένων, κατευθυντήριες γραμμές για τη διαχείριση δεδομένων από αρχαιολογικά προγράμματα, οδηγίες για την εφαρμογή των αρχών FAIR, καθώς και υπηρεσίες που αναπτύχθηκαν με τη μορφή εικονικών εργαστηρίων ή μεμονωμένων εφαρμογών για την επεξεργασία, ομογενοποίηση, σημασιολογική χαρτογράφηση, ανάλυση και δημοσίευση αρχαιολογικών δεδομένων. Επίσης, αρκετές πληροφορίες για τον πλήρη κύκλο ζωής της αρχαιολογικής πληροφορίας προσφέρει και η ανανεωμένη έκδοση των οδηγών καλής πρακτικής του Archaeological Data Service (ADS), οι οποίοι είναι προσβάσιμοι στο https://archaeologydataservice.ac.uk/help-guidance/guides-to-good-practice/.

Παράλληλα με τις προβλέψεις για τα ίδια τα ερευνητικά δεδομένα, ο τρόπος με τον οποίο τα πρωτογενή δεδομένα μετατρέπονται σε τελικά ψηφιακά προϊόντα θα πρέπει να τεκμηριώνεται για την ενημέρωση των μελλοντικών χρηστών. Αναφερόμαστε στη *γενεολογία* (lineage) και στην τεκμηρίωση της *προέλευσης* (provenance) των δεδομένων, οι οποίες χρήζουν διαφάνειας και είναι άρρηκτα συνδεδεμένες με τις υπολογιστικές διαδικασίες επεξεργασίας των δεδομένων. Ερευνητικά έργα που τεκμηριώνουν τη διαδικασία δημιουργίας δεδομένων τους και διατηρούν τα ενδιάμεσα αποτελέσματα επεξεργασίας ενισχύουν τις πιθανότητες της επαναχρησιμοποίησης και της επαύξησης των δεδομένων τους με νέες τεχνικές επεξεργασίας (π.χ. Katsianis *et al.* 2021). Προς αυτή την κατεύθυνση ενθαρρύνεται η κατάθεση των ψηφιακών ροών επεξεργασίας δεδομένων στη μορφή ανοικτού προγραμματιστικού κώδικα ή ψηφιακών εργαλείων προκειμένου να διευκολύνει την επανάχρηση και την περαιτέρω εξέλιξη των αρχαιολογικών υπολογιστικών μεθόδων (Huggett 2014, Cobb *et al.* 2019).

Για να ενσωματωθούν αυτές οι διατάξεις στην αρχαιολογική ερευνητική διαδικασία, είναι σημαντική η ευρύτερη γνωστοποίηση των πρωτοβουλιών της Ανοικτής Επιστήμης. Τα διεθνή έργα που παρουσιάζονται παρέχουν πολλές ευκαιρίες για συμμετοχή ή/και αξιοποίηση ευκαιριών κατάρτισης ή ακόμη και χρηματοδότησης σε επαγγελματίες του χώρου που ασχολούνται με την επιμέλεια ψηφιακών δεδομένων. Ωστόσο, προσοχή πρέπει επίσης να δοθεί στην εκπαίδευση και κατάρτιση των νέων αρχαιολόγων μέσω της συμπερίληψης μαθημάτων ανάπτυξης ψηφιακών δεξιοτήτων και *γραμματισμού δεδομένων* (data literacy) και στα σχετικά προγράμματα σπουδών των πανεπιστημιακών τμημάτων (βλ. Kansa & Kansa 2021). Προς αυτή την κατεύθυνση το έργο *DELTA* επιχείρησε να χαρτογραφήσει τον ψηφιακό γραμματισμό στα αρχαιολογικά προγράμματα σπουδών (Polymeropoulou *et al.* 2020) και να καλύψει μέρος του κενού σχεδιάζοντας ένα υβριδικό μάθημα επάνω στη χρήση ψηφιακών μεθόδων στο ανασκαφικό έργο που συνδυάζει διαδικτυακό περιεχόμενο με τη μορφή ενός Μαζικού Ανοιχτού Διαδικτυακού Μαθήματος (MOOC - Massive Open Online Course) και τη συμμετοχή στο φυσικό χώρο της αρχαιολογικής ανασκαφής στο Πλάσι του Μαραθώνα που διεξάγεται από το Τμήμα Ιστορίας και Αρχαιολογίας του ΕΚΠΑ[19].

Τέλος, αξίζει να υπογραμμίσουμε και τη σημασία της διαφάνειας και στην κρατική πολιτιστική πολιτική. Ακόμη και το *4ο Εθνικό σχέδιο δράσης για την Ανοικτή διακυβέρνηση 2019-2021* (ΤΔΑΔ-ΥΔΑ 2020:35-40) τόσο στην ανεξάρτητη αξιολόγηση του σχεδιασμού (Γκούσκος *et al.* 2021: 65-67) όσο και στην τελική έκθεση (OGP 2023: 15-17) σημειώνεται ότι χωλαίνει ως προς τη συμπερίληψη των άμεσα ενδιαφερόμενων στις σχετικές αποφάσεις, κάτι που μπορεί να αλλάξει εφόσον υιοθετηθούν με καταστατικό τρόπο ευρύτερες πτυχές της Ανοικτής Επιστήμης. Βήματα προς την κατεύθυνση αυτή πραγματοποιήθηκαν με την Ανοιχτή Πρόσκληση συνεργατικού σχεδιασμού του *5ου Εθνικού σχεδίου δράσης 2022-2025* (ΥΨΔ 2023: 3-10), από όπου όμως απουσιάζουν δεσμεύσεις που αφορούν στα δεδομένα και τους υφιστάμενους πόρους του Υπουργείου Πολιτισμού και Αθλητισμού[20].

9. Προτάσεις - Συμπεράσματα

Η παρούσα εργασία επιχειρεί να αποτελέσει έναυσμα, ώστε να ανοίξει περαιτέρω η συζήτηση για τα πλεονεκτήματα, αλλά και τα μειονεκτήματα ή τις δυσκολίες που εισάγει η Ανοικτή Επιστήμη στην ψηφιακή αρχαιολογική έρευνα. Η πανδημία του κορονοϊού (COVID-19) άλλαξε ίσως για πάντα τον τρόπο της ερευνητικής εργασίας και συνέβαλε στην ακόμη μεγαλύτερη διείσδυση των ψηφιακών μεθόδων στην ερευνητική διαδικασία (βλ. Geser 2021).

Έχει έρθει η στιγμή, ώστε τα πρώτα βήματα με τη μορφή συνεδρίων και τη θέσπιση ομάδων εργασίας για την υιοθέτηση της Ανοικτής Επιστήμης στον ευρύτερο χώρο των Ψηφιακών Ανθρωπιστικών επιστημών τα τελευταία χρόνια, να εξειδικευτούν στο επίπεδο της αρχαιολογικής έρευνας με τη συμμετοχή όλων των επιμέρους φορέων που δραστηριοποιούνται στην παραγωγή αρχαιολογικών δεδομένων (κρατικές υπηρεσίες, πανεπιστήμια, ερευνητικά ινστιτούτα, ξένες αρχαιολογικές σχολές, αρχαιολογικές εταιρείες και άλλες ιδιωτικές πρωτοβουλίες ή κινήσεις πολιτών). Μέσω της ανταλλαγής εμπειρίας για τον ρόλο της ψηφιακής τεχνολογίας στην ερευνητική δραστηριότητα, του καθορισμού της σειράς των

[19] *Digital Excavation through Learning and Training in Archaeology* – DELTA (Erasmus+/KA2 Project No: 2019-1-EL01-KA203-062875). http://www.project-delta.eu/. Η πλατφόρμα του μαθήματος είναι προσβάσιμη στη διεύθυνση https://mooc.cti.gr/delta.html.

[20] Εξαίρεση αποτελεί το σημαντικό πολιτιστικό και επιστημονικό απόθεμα της Ακαδημίας Αθηνών, το οποίο και εντάσσεται σε δράσεις ψηφιοποίησης και ψηφιακής επιμέλειας με απώτερο στόχο την αξιοποίηση και ευρύτερη προβολή του με όρους ανοικτής πρόσβασης (ΥΨΔ 2023: 23-28).

βημάτων για την προσαρμογή του γενικότερου κλάδου στο *άνοιγμα* της Αρχαιολογίας, της ουσιαστικής διαβούλευσης για την χάραξη πολιτικών ενθάρρυνσης ανοικτών πρακτικών και για διαμόρφωση του κατάλληλου νομικού, εκπαιδευτικού και χρηματοδοτικού πλαισίου, ίσως γίνει εφικτό να ξεπεραστούν πολλά από τα εμπόδια που αποθαρρύνουν στην πράξη τις ανοικτές πρακτικές στη χώρα μας. Προς αυτή την κατεύθυνση είναι χρήσιμη, έως αναγκαία, μια ανοικτή εκδήλωση με τη μορφή ενός συνεδρίου και τη συμμετοχή εκπροσώπων από το σύνολο των φορέων που δραστηριοποιούνται στην εγχώρια αρχαιολογική έρευνα.

Σε κάθε περίπτωση, ήδη χρησιμοποιούμε δεδομένα που διατίθενται σε ανοικτή πρόσβαση και ένα πρώτο απλό βήμα που μπορούμε να κάνουμε καθεμία και καθένας ξεχωριστά είναι να αναγνωρίζουμε τη χρήση τους στην έρευνά μας παραπέμποντας σωστά σε αυτά. Μόνο με αυτό τον τρόπο μπορεί να αναγνωριστεί η προσπάθεια των ερευνητών να διαθέσουν τα δεδομένα της δικής τους έρευνας προς όφελος της ερευνητικής κοινότητας και του ευρύτερου κοινωνικού συνόλου.

Ευχαριστίες

Η έρευνα έχει υποστηριχθεί από το πρόγραμμα H2020 ARIADNEplus project (H2020-INFRAIA-2018-1-823914) και από το πρόγραμμα SEADDA στο πλαίσιο της Δράσης COST (CA18128). Οι γνώμες των συγγραφέων δεν αντικατοπτρίζουν απαραίτητα εκείνες της Ευρωπαϊκής Επιτροπής. Το παρόν κείμενο αποτελεί επεξεργασμένη και ενημερωμένη (08/2023) εκδοχή της αρχικής υποβολής (02/2022) για τα πρακτικά του συνεδρίου του CAA-GR 2020. Ευχαριστούμε τους αξιολογητές για τα χρήσιμα σχόλιά τους. Θα θέλαμε ακόμη να ευχαριστήσουμε την κ. Έλλη Παπαδοπούλου, Επιστημονική Συνεργάτιδα του ΕΚ Αθηνά, για την παροχή πληροφοριών σχετικών με τις δράσεις του OpenAIRE στην Ελλάδα.

Βιβλιογραφία

Αγγελάκη, Γ., 2014, *Υπηρεσίες SaaS από το ΕΚΤ: Ανοικτά και Διασυνδεδεμένα Πολιτιστικά Δεδομένα*. Αθήνα: ΕΚΤ. https://hdl.handle.net/20.500.12776/14420.

Αθανασίου, Σ., Αμοιρίδης, Β., Γαβριηλίδου, Μ., Γερασόπουλος, Ε., Δημόπουλος, Α., Κακλαμάνη, Γ., Καραγιάννης, Φ., Κλαμπάνος, Ι., Κονδύλη, Δ., Κουμάνταρος. Κ., Κωνσταντόπουλος, Π., Λενάκη, Κ., Λυκιαρδόπουλος, Α., Μανωλά, Ν., Μητροπούλου, Δ., Μπενάρδου, Α., Μπούκος, Ν., Νούσιας, Α., Νταουντάκη, Μ., Οίκουτα, Σ., Ολλανδέζου, Λ., Παπαδοπούλου, Ε., Παπασταματίου, Η., Πιπερίδης, Σ., Σπάϊερ, Ο., Τσιαβός, Π., Χρυσίνα, Ε. & Ψωμόπουλος, Φ. 2020, *Εθνικό Σχέδιο για την Ανοικτή Επιστήμη*. Zenodo. https://doi.org/10.5281/zenodo.3908953.

Aspöck, E., 2019, Moving towards an open archaeology: projects, opportunities and challenges. *Mitteilungen der VÖB*, 72 (2): 538-554. https://doi.org/10.31263/voebm.v72i2.3249.

Βασιλογαμβράκης, Ν. & Μπάρτζη, Κ., 2015, *Καλές πρακτικές τεκμηρίωσης*. Αθήνα: ΕΚΤ. https://hdl.handle.net/10442/15292.

Beck, A. & Neylon, C., 2012, A vision for Open Archaeology. *WorldArchaeology*, 44 (4): 479-497. https://doi.org/10.1080/00438243.2012.737581.

Cobb, P. J., Sigmier, J. H., Creamer, P. M. & French, E. R., 2019, Collaborative Approaches to Archaeology Programming and the Increase of Digital Literacy Among Archaeology Students. *Open Archaeology*, 5 (1): 137-154. https://doi.org/10.1515/opar-2019-0010.

COPTR contributors, 2021, *COPTR*, https://coptr.digipres.org. Ημερομηνία πρόσβασης: 8/8/2023.

Costa, S., Beck, A., Bevan, A. & Ogden, J., 2014, Defining and Advocating Open Data in Archaeology. Στο: Earl, G., Sly, T., Chrysanthi, A., Murrieta Flores, P., Papadopoulos, C., Romanowska, I. & Wheatley, D. (eds) *Archaeology in the Digital Era: Proceedings of the 40th Conference in Computer Applications and Quantitative Methods in Archaeology, Southampton, United Kingdom, 26-30 March 2012*. Amsterdam University Press, Amsterdam, 449-456. https://discovery.ucl.ac.uk/id/eprint/1379503/4/Bevan_Costaetal14.pdf. Ημερομηνία πρόσβασης: 8/8/2023.

Γκούσκος, Δ., Καραμαγκιώλη, Ε., Στάιου, Ε.Ρ. & Παπαδοπούλου, Η.Η., 2021, *Ανεξάρτητος Μηχανισμός Αξιολόγησης (IRM): Αναφορά Σχεδιασμού 2019-2021 για την Ελλάδα.* https://www.opengovpartnership.org/wp-content/uploads/2020/11/Greece_Design_Report_2019-2021_GR_for-public-comment.pdf. Ημερομηνία πρόσβασης: 8/8/2023.

OGP, 2023, *Ανεξάρτητος Μηχανισμός Αξιολόγησης. Έκθεση Αποτελεσμάτων: Ελλάδα 2019-2022*. Open Government Partnership. https://www.opengovpartnership.org/wp-content/uploads/2023/06/Greece_Results-Report_2019-2022_EL.pdf.

ΕΚΤ (Εθνικό Κέντρο Τεκμηρίωσης), 2015, Αφιέρωμα: Η Ανοικτή Πρόσβαση στα Ερευνητικά Δεδομένα ως Μοχλός για την Ανοικτή Επιστήμη. *Καινοτομία. Έρευνα και Τεχνολογία*, 98: 20-27. http://kainotomia.ekt.gr/issue/2015/98/files/assets/common/downloads/Kainotomia_98.pdf. Ημερομηνία πρόσβασης: 8/8/2023.

ΕΕ (Ευρωπαϊκή Επιτροπή), 2018, *Σύσταση (ΕΕ) 2018/790 της Επιτροπής της 25ης Απριλίου 2018 για την πρόσβαση στις επιστημονικές πληροφορίες και τη διαφύλαξή τους*. https://eur-lex.europa.eu/legal-content/EL/TXT/PDF/?uri=CELEX:32018H0790&from=EN. Ημερομηνία πρόσβασης: 8/8/2023.

ΕΚΒ (Ευρωπαϊκό Κοινοβούλιο και Συμβούλιο), 2019, Οδηγία (ΕΕ) 2019/1024 του Ευρωπαϊκού Κοινοβουλίου και Συμβουλίου της 20ής Ιουνίου 2019 για τα ανοικτά δεδομένα και την περαιτέρω χρήση πληροφοριών του δημόσιου τομέα. *Επίσημη Εφημερίδα της Ευρωπαϊκής Ένωσης*, L 172: 56-83. https://eur-lex.europa.eu/legal-content/EL/TXT/PDF/?uri=CELEX:32019L1024&from=En. Ημερομηνία πρόσβασης: 8/8/2023.

ERC (European Research Council), 2021, *Open Research Data and Data Management Plans. Information for ERC grantees*. Version 4.1 – 20/04/2022. https://erc.europa.eu/sites/default/files/document/file/ERC_info_document-Open_Research_Data_and_Data_Management_Plans.pdf. Ημερομηνία πρόσβασης: 8/8/2023.

Fecher, B. & Friesike, S., 2013, Open Science: One Term, Five Schools of Thought. *RatSWD_WP* (German Council for Social and Economic Data - Working Paper Series), 218. http://dx.doi.org/10.2139/ssrn.2272036.

Fernández Pinto, M., 2020, Open Science for private Interests? How the Logic of Open Science Contributes to the Commercialization of Research. *Frontiers in Research Metrics and Analytics*, 5: 588331. https://doi.org/10.3389/frma.2020.588331.

Fernández Cacho, S., 2021, Open archaeology. Advances and challenges in the management of archaeological contents. *Complutum*, 32 (2): 443-457. https://dx.doi.org/10.5209/cmpl.78570.

Gattiglia, G., 2015, Think big about data: Archaeology and the Big Data challenge. *Archäologische Informationen*, 38: 113-124. https://doi.org/10.11588/ai.2015.1.26155.

Georgiadis, H., Papanoti, A., Paschou, M. Roubani, A. Pelekanou, D. Chardouveli, D & Sachini, E., 2016, Semantics.gr: a self-improving service to repositories and aggregators for massively enriching their content. Στο: *Proceedings of the 10th International Conference on Metadata and Semantics Research, Digital Humanities and Digital Curation (DHC) workshop, Göttingen, Germany, 22-25/11/2016*. http://hdl.handle.net/10442/15420.

Geser, G., 2019, *ARIADNEplus: D2.1 Initial Report on Community Needs*. Zenodo. https://doi.org/10.5281/zenodo.4916190.

Geser, G., 2021., *ARIADNEplus:Impact of COVID-19 on Archaeology and Cutlural Heritage.* https://ariadne-infrastructure.eu/wp-content/uploads/2021/11/COVID-19_impact-archaeology-and-cultural-heritage_29Oct2021.pdf. Ημερομηνία πρόσβασης: 8/8/2023.

Geser, G., Richards, J.D., Massara, F. & Wright, H., 2022, Data Management Policies and Practices of Digital Archaeological Repositories, *Internet Archaeology*, 59. https://doi.org/10.11141/ia.59.2.

Giorgio, S. D. & Ronzino, P., 2018, PARTHENOS Data Management Plan template for Open Research in Archaeology. Στο: *3rd Digital Heritage International Congress (DigitalHERITAGE) held jointly with 24th International Conference on Virtual Systems & Multimedia (VSMM 2018)*, 1-4. https://doi.org/10.1109/DigitalHeritage.2018.8810039.

Huggett, J., 2014, Promise and Paradox: Accessing Open Data in Archaeology. Στο: Mills, C., Pidd, M. & Ward, E. (eds). *Proceedings of the Digital Humanities Congress 2012*. Studies in the Digital Humanities. Sheffield: HRI Online Publications. http://dx.doi.org/10.17613/qc5d-9x46.

Huggett, J., Reilly, P. & Lock, G., 2018, Whither Digital Archaeological Knowledge? The Challenge of Unstable Futures. *Journal of Computer Applications in Archaeology*, 1 (1): 42-54. https://doi.org/10.5334/jcaa.7.

Jones, S. & Abramatic, J.(eds.), 2019, *European Open Science Cloud (EOSC) strategic implementation plan*. European Commission, Directorate-General for Research and Innovation, Publications Office. https://doi.org/10.2777/202370.

Κανελλοπούλου-Μπότη, Μ., Βουλιγέα, Ε. & Σιταρά, Μ., 2013, Ανοικτή Πρόσβαση και αρχαιολογικά Δεδομένα. Στο: *Διεθνές συνέδριο: Ανοικτή πρόσβαση @ ΕΚΤ: Αναπτύσσοντας κοινές ευρωπαϊκές πολιτικές* *για την καινοτόμο χρήση της δημόσιας και επιστημονικής πληροφορίας. Αθήνα, ΕΚΤ, 16-18/10/2013*. https://ihrc.gr/download.php?fgr=announcement_open_access.pdf. Ημερομηνία πρόσβασης: 8/8/2023.

Kansa, E., 2012, Openness and Archaeology's Information Ecosystem. *World Archaeology* 44 (4): 498-520. http://doi.org/10.1080/00438243.2012.737575.

Kansa, E., & Kansa, S., 2021, Digital Data and Data Literacy in Archaeology Now and in the New Decade. *Advances in Archaeological Practice*, 9 (1): 81-85. http://doi.org/10.1017/aap.2020.55.

Katsianis, M., Kotsakis, K., & Stefanou, F., 2021, Reconfiguring the 3D excavation archive. Technological shift and data remix in the archaeological project of Paliambela Kolindros, Greece. *Journal of Archaeological Science: Reports*, 36: 102857. https://doi.org/10.1016/j.jasrep.2021.102857.

Κατσιάνης, Μ., 2013, Σύγχρονες προκλήσεις στην τεκμηρίωση των αναστηλωτικών έργων της αθηναϊκής Ακρόπολης, *Ενημερωτικές ειδήσεις από την αναστήλωση των μνημείων της Ακρόπολης* 13: 22-26. https://workfun.openabekt.gr/files/items/558993/P91.25.pdf. Ημερομηνία πρόσβασης: 8/8/2023.

Lake, M., 2012, Open archaeology. *World Archaeology*, 44 (4): 471-478. https://doi.org/10.1080/00438243.2012.748521.

Lucas, G., 2001, Destruction and the Rhetoric of Excavation. *Norwegian Archaeological Review*, 34 (1): 35-46. https://doi.org/10.1080/00293650119347.

Μάλλιου, Ν., Τσουκαλά, Β., Μανωλά, Ν., Αγγελάκη, Μ., Σαραντοπούλου, Ι., Νταουντάκη, Μ., Διομήδη-Παρπούνα, Έ., Franck, G., Carvalho, J. & Grootveld, M., 2017, *Ανοικτή Επιστήμη: Ζητήματα και Προοπτικές, Αθήνα, 15/6/2017.* Αθήνα: ΕΚΤ. https://hdl.handle.net/10442/15536.

Marwick, B, d'Alpoim Guedes, J., Barton, C. M., Bates, L. A., Baxter, M., Bevan, A., Bollwerk, E. A., Bocinsky, R. K., Brughmans, T., Carter, A. K., Conrad, C., Contreras, D. A., Costa, S., Crema, E. R., Daggett, A., Davies, B., Drake, B. L., Dye, T. S., France, P., Fullagar, R., Giusti, D., Graham, S., Harris, M. D., Hawks, J., Heath, S., Huffer, D., Kansa, E. C., Kansa, S. W., Madsen, M. E., Melcher, J., Negre, J., Neiman, F. D., Opitz, R., Orton, D. C., Przstupa, P., Raviele, M., Riel-Savatore, J., Riris, P., Romanowska, I., Smith, J., Strupler, N., Ullah, I. I., Van Vlack, H. G., VanValkenburgh, N., Watrall, E. C., Webster, C., Wells, J., Winters, J. & Wren, C. D., 2017, Open science in archaeology. *SAA Archaeological Record*, 17 (4): 8-14.

Moore, R. & Richards, J., 2015, Here Today, Gone Tomorrow: Open Access, Open Data and Digital Preservation. Στο: Wilson, A.T. & Edwards, B. (eds) *Open Source Archaeology*. Warsaw, Poland: De Gruyter Open Poland, 30-43. https://doi.org/10.1515/9783110440171-004.

Niccolucci, F. & Richards, J., 2019, ARIADNE and ARIADNEplus. Στο: Niccolucci, F. & Richards, J. (eds). *The ARIADNE Impact*. Budapest: Archaeolingua, 7-26. https://doi.org/10.5281/zenodo.3476712.

OECD (Organisation for Economic Co-operation and Development), 2015, Making Open Science a Reality. *OECD Science, Technology and Industry Policy Papers*, No. 25. Paris: OECD Publishing. http://dx.doi.org/10.1787/5jrs2f963zs1-en.

Opitz, R., 2018, Publishing Archaeological Excavations at the Digital Turn. *Journal of Field Archaeology*, 43: S68-S82. https://doi.org/10.1080/00934690.2018.1505409.

Πάντος, Π., 2013, Το θεσμικό πλαίσιο των ανασκαφών στην Ελλάδα από τον 20ο στον 21ο αιώνα. *Ανάσκαμμα*, 6: 93-130. https://anaskamma.files.wordpress.com/2013/06/06_pantos.pdf. Ημερομηνία πρόσβασης: 8/8/2023.

PARTHENOS, Hollander, H., Morselli, F., Uiterwaal, F., Admiraal, F., Trippel, T. & Di Giorgio, S., 2019, *PARTHENOS Οδηγίες για την εφαρμογή των αρχών FAIR στη διαχείριση και επανάχρηση δεδομένων*. Zenodo. https://doi.org/10.5281/zenodo.3363386.

Pispiringas, L., Athanasiou, S., Simaioforidis, Z. & Skoutas, D., 2019, HEAL-Link and HELIX open collaboration to facilitate and promote scholarly communication through Open Access and engage Research Infrastructures in Open Science. Στο: *85th IFLA General Conference and Assembly, World Library and Information Congress* (IFLA WLIC 2019). http://library.ifla.org/id/eprint/2498. Ημερομηνία πρόσβασης: 8/8/2023.

Polymeropoulou, P., Kameas, A., Papadatos, Y., Kalara, A., Sogliani, F., Roubis, D., Vullo, M., Tóth, P., Hons, D., Malíšková, J., Nosek, V. & Pokutta, D., 2020, *Training students in the use of digital technologies in archaeology. Desk and field research final report*. Erasmus+/KA2 project. DELTA: Digital

Excavation through Learning and Training in Archaeology. http://www.project-delta.eu/wp-content/uploads/2022/11/DELTA-IO1-Desk-and-Field-Research-Report.pdf. Ημερομηνία πρόσβασης: 8/8/2023.

Richards, J.D., Jakobsson, U., Novák, D., Štular, B. & Wright, H., 2021, Digital Archiving in Archaeology: The State of the Art. Introduction. *Internet Archaeology*, 58. https://doi.org/10.11141/ia.58.23.

Sitara, M., & Vouligea, E., 2014, Open Access to Archeological Data and the Greek Law. Στο Sideridis, A., Kardasiadou, Z., Yialouris, C. & Zorkadis, V. (eds) *E-democracy, security, privacy and trust in a digital world*. e-Democracy 2013. Communications in Computer and Information Science, 441. Cham: Springer. https://doi.org/10.1007/978-3-319-11710-2_16.

Σταθόπουλος, Π., Χούσος, Ν. & Σταύρου, Γ., 2013, *Προδιαγραφές και χαρακτηριστικά διαλειτουργικότητας για ανοικτό ψηφιακό περιεχόμενο*. Αθήνα: ΕΚΤ. http://hdl.handle.net/10442/8887.

Strupler, N., & Wilkinson, T. C., 2017, Reproducibility in the Field: Transparency, Version Control and Collaboration on the Project Panormos Survey. *Open Archaeology*, 3 (1): 279-304. https://doi.org/10.1515/opar-2017-0019.

ΤΔΑΔ-ΥΔΑ, 2020, *4ο Εθνικό Σχέδιο Δράσης για την Ανοικτή Διακυβέρνηση 2019-2022. Επικαιροποιημένη έκδοση (Δεκέμβριος 2020)*. Τμήμα Διαφάνειας και Ανοικτής Διακυβέρνησης (ΤΔΑΔ-ΥΔΑ), Γενική Διεύθυνση Δημοσίων Οργανώσεων, Υπουργείο Διοικητικής Ανασυγκρότησης. https://www.opengovpartnership.org/wp-content/uploads/2019/05/Greece_Action-Plan_2019-2022_Update_GR.pdf. Ημερομηνία πρόσβασης: 8/8/2023.

Tsiafaki, D. & Katsianis, M., 2021, Stewardship of Digital Archaeological Data in Greece: a landscape of fragmentation. *Internet Archaeology*, 58. https://doi.org/10.11141/ia.58.8.

Tsiavos, P., 2020, State, Netocrats and the Commons: Developing a Cultural Policy in the Era of Platforms. Στο: Lekakis, S. (ed) *Cultural heritage in the realm of the commons*. London: Ubiquity Press, 109-126. https://doi.org/10.5334/bcj.g.

Uiterwaal, F., Niccolucci, F., Bassett, S., Krauwer, S., Hollander, H., Admiraal, F., Romary, L., Bruseker, G., Meghini, C., Edmond, J. & Hedges, M., 2021, From disparate disciplines to unity in diversity How the PARTHENOS project has brought European humanities Research Infrastructures together. *International Journal of Humanities and Arts Computing*, 15 (1-2): 101-116. https://doi.org/10.3366/ijhac.2021.0264, https://hal.inria.fr/hal-03402145.

ΥΨΔ, 2021, *Βίβλος Ψηφιακού Μετασχηματισμού 2020-2025*. Υπουργείο Ψηφιακής Διακυβέρνησης. https://digitalstrategy.gov.gr/vivlos_pdf. Ημερομηνία πρόσβασης: 8/8/2023.

ΥΨΔ, 2023, *5ο Εθνικό Σχέδιο Δράσης για την Ανοικτή Διακυβέρνηση 2023-2025*. Υπουργείο Ψηφιακής Διακυβέρνησης. https://www.opengovpartnership.org/wp-content/uploads/2023/01/Greece_Action-Plan_2022-2024_December_GR.pdf. Ημερομηνία πρόσβασης: 8/8/2023.

Wilkinson, M.D., Dumontier, M., Aalbersberg, I.J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.W., Bonino da Silva, L. S., 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., 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., Sengstag, T., Slater, T., Strawn, G., Swertz, M.A., Thompson, M., Van der Lei, J., van Mulligen, E., Velterop, J., Waagmeester, A., Wittenberg, P., Wolstencroft, K., Zhao, J. & Mons, B., 2016, The FAIR guiding principles for scientific data management and stewardship. *Sci Data*, 3: 160018. https://doi.org/10.1038/sdata.2016.18.

EXPLORING NEANDERTHAL HANDEDNESS THE CONTRIBUTION OF DIGITAL APPLICATIONS

S. LIGKOVANLIS[1], A. AIDONIS[2], N. GALANIDOU[3]& C. PAPAGEORGOPOULOU[4]

[1] University of Crete-Department of History and Archaeology-Laboratory of Educational Collections, Casts and Originals, University Campus of Gallos, Rethymno 74100
stligkov@gmail.com
[2] Democritus University of Thrace - Department of History and Ethnology-Laboratory of Physical Anthropology, Campus DUTh, Komotini 69100
saidonis@gmail.com
[3] University of Crete, Department of History and Archaeology, University Campus of Gallos, Rethymno 74100
galanidou@uoc.gr
[4] Democritus University of Thrace- Department of History and Ethnology, Campus DUTh, Komotini 69100
cpapage@he.duth.gr

Abstract

The *NeandLang* project investigates the Neanderthal ability of speech articulation and comprehension. This is based on the neurophysiological evidence that these capacities are directly related to the formation of cerebral lateralization reflected also in the right- over left-handedness dominance. Research explores the rate of hand lateralization on these hominins through the examination of their manual specialization imprinted in the technical procedures of lithic production. To this end, observations deriving from both experimental (stone tools produced by modern knappers) and archaeological data (Neanderthals' stone tools from Kalamakia Cave-Peloponnesus) are compared. A series of digital applications are incorporated in our methodological protocol in order the research objectives to be met: Software for the analysis of the experimental sessions' video recordings contributes to the understanding of the kinematics characterizing the knapping procedures manual specialization and their consequences imprinted on the lithic products. 3D scanning, photogrammetry, GIS analytical tools and digital image processing of the experimentally produced lithics and archaeological objects enable targeted technical observations and calculations to be made. The construction of special databases permits an evaluation of the collected data and leads to the identification of handedness patterns. In this paper, we present the digital methodology of our study along with the results obtained. The use of innovative digital approaches to the research of interplay between knapped stone technology and hominin cognition are also discussed.

Keywords: *Cognitive archaeology, speech articulation origins, hominin handedness, prehistoric lithic technology, digital applications.*

1. Introduction

Neandlang project explores aspects of the Neanderthals' cognitive capabilities and especially their ability of language articulation and comprehension. This is a long-lasting debate, with some researchers claiming that those hominins were capable of structured oral communication (e.g. Dedieu& Levinson 2013; Mellars 1996) and others disputing such a proposal. The latter argue that the lack of language was a major reason for Neanderthals' extinction, since the absence of such a trait, would be a disadvantage concerning natural competition and evolution (e.g. Berwick et al. 2013; Lieberman 2007).

Based on the neurophysiological evidence that linguistic capacity on modern humans is directly related to the formation of cerebral lateralization and reflected also in the right- over left-handedness dominance (e.g. Gentilucci & Corballis 2006; Knecht et al. 2000), the current research explores the rates of handedness on Neanderthal populations, through the examination of their manual specialization imprinted in the technical procedures of lithic production. This is undertaken by the idea that chiropractic procedures of the artifacts' construction, can leave characteristic traces on them, reflecting the hand preference of their manufactures. Following such an approach, if it could be verified that Neanderthals had a handedness rate similar to that observed in modern populations, this would be an indirect evidence that they had the neurophysiological prerequisites, to develop linguistic communication. To this end, observations deriving from both experimental (stone tools produced by modern knappers) and archaeological data (Neanderthals' stone tools from Kalamakia Cave-Peloponnesus) are compared.

A distinct stigma, which is more or less observable on the lithic flakes' ventral surfaces and it is caused by the mechanics of stone flaking, was chosen to be closely investigated. This is the so-called cone of percussion, cone crack or Hertzian cone. It is a triangular shaped imprint, bordered by distinct flaws, the cone crack paths. This component is caused by Hertzian fracture (e.g. Frank & Lawn 1967), a physical mechanism of brittle materials breakage, such as glass but also flint, a major lithic raw material for prehistoric stone tool construction **(**Figure 1).

Laboratory experiments on glass and ceramics have shown that during Hertzian fracture, and when the impact loading is perpendicular to a striking surface, then the cone crack resamples an isosceles triangle. Consequently, when contact is not perpendicular but inclined, below or above 90°, the triangular imprint of the cone crack presents a distinct skewness (e.g. Aydelotte et al. 2016; Chaudri & Lianghui 1989) (Figure 2).

Earlier efforts used this mechanism, in order to explore prehistoric handedness through lithics. It was *a priori* considered, that due to human physiology, during prehistoric stone tool production, left- and right-handers would bear mirroring blows on the core surfaces, resulting to a distinctive skewness of the cone cracks, according to the knappers' manual laterality (Figure 3). By experimentally monitoring this hypothesis through blind tests, those studies have come to a dead-end, offering contradictory results (e.g. Bargallo & Mosquera 2013; Ruck et al. 2015; Rugg & Mullane 2001;Uomini 2001). Therefore, such an approach was never applied to the archaeological record.

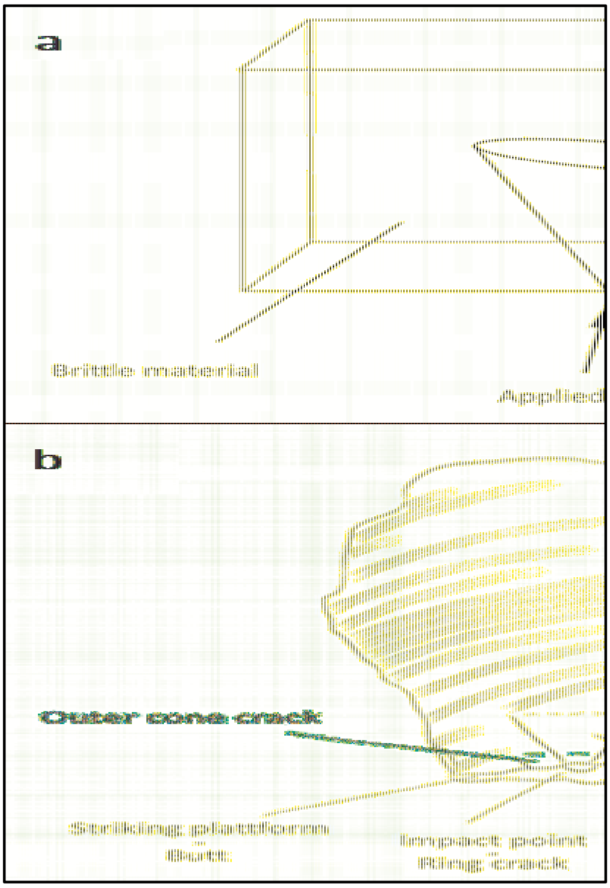


Figure 1. a)Formation of a Hertzian cone on a brittle material. b) Basic features of a Hertzian fracture imprinted on the ventral side of a lithic flake (b).

That failure was partly attributed (e.g. Ruck et al. 2020) to a low determinability of the method, but also a lack of objectivity, during the examinations of the lithics, which in all cases were made macroscopically.

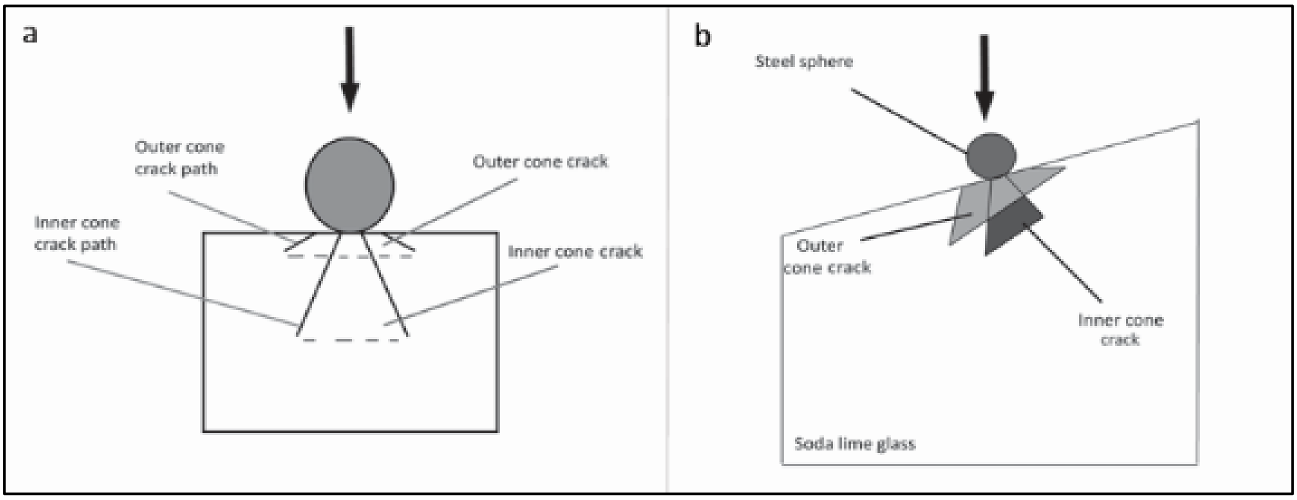


Figure 2.a) Schematic diagram of inner- and outer-cone crack geometries according to a perpendicular blow. b) formation of cone cracks on soda lime glass when a blow is inclined according to Chaudri&Liangyi 1989.

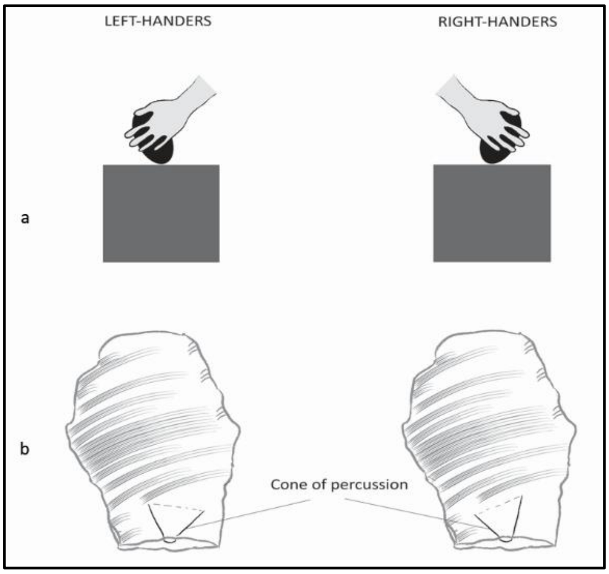


Figure 3.a)Hypothesized distinct human kinematics during flaking according to knappers hand preference. b) handedness indicators based on the cone of percussion features on a flake considered by Rugg & Mullane 2001, Bargallo & Mosquera 2013, Ruck et al. 2015 (e-ii).

2. Research methodological protocol and the contribution of digital applications

Within the context of the current research, the limitations faced by previous work was attempted to be surpassed with the insertion of a methodological protocol combining the conduction of controlled knapping experiments, and a close, multifactor, study of their results. In all these procedures, digital applications played a decisive role.

Two major experimental knapping sessions, an 'open observation' and a blind one, have been conducted, in which 7 right- and 6 left-handers modern knappers, have been participated. During these controlled procedures our empirical material, of over 400 flint flakes, have been generated for further examination (286 derived from the 'open observation' experiment and 172 from the blind one). The biggest part of the 'open observation' experiment, has been recorded using two high frame digital cameras (1080p-120fps), placed in front of the knappers and on the side of their non-dominant hand.

The video recordings have been analyzed with the use of Kinovea® video software, permitting frame per frame observations, offering at the same time tools for the conduction of digital measurements. With such an analysis, except the evaluation of individual knapping styles, we targeted the understanding of every knapper's blows inclination, in order to comprehend how these, affect the cone crack geometry of each extracted flake. In this way, we examined thevalidity of the *a priori* hypothesis of previous studies, that left- and right-handers, as a rule, bring mirroring blows on the cores.

Figure 4. Knapping experiments' video analysis using Kinovea® software and examples of blows' inclination during the experimental flaking sessions from left and right-handers: a: 'expected', b: 'perpendicular', c: 'invert'.

The inclination of the knappers' blows, has been evaluated at the time that the hammerstone hits the core surface, and it was divided into 3 broad categories: 'expected', if it was in accordance with the hypothesis of their relation to the knapper's hand preference, 'invert', if they were contrary to the expected one, and 'perpendicular', if a blow inclination was nearly vertical to a striking surface **(Figure 4)**.

Observations on the cone crack geometry have been also conducted using digital tools. In this manner, we attempted to achieve a maximum reduction of subjectivity, concerning the understudy features' evaluation. The ventral sides of the flakes, produced during the 'open observation' experiment, have been documented with close-up photography techniques. 3D models have been also generated by the application of laser scanning and photogrammetry. The goal of these processes was to clearly highlight the cone crack paths, bordering the flakes' cone cracks. Consequently, appropriate measurements of the cone crack paths geometry could objectively reveal the potential skew of the cone cracks.

In practice, and after exhaustive tests, it was made clear that the most 'profitable' technique for the cone crack paths' highlighting is, for the moment, close-up photography combined with the appropriate digital processing of the images produced. It should be also noticed that the creation of 3D models and automatic analyses, had in some cases spectacular results concerning the crack paths' highlighting. Yet, the universal application of these techniques showed that for the majority of the flakes, the results were not very clear, prohibiting objective evaluations. This is due to the inability of high-resolution zoom analysis provision by today's available 3D recording tools. Nevertheless, it is also certain that technological development, will eliminate such issues in the future **(Figure 5)**.

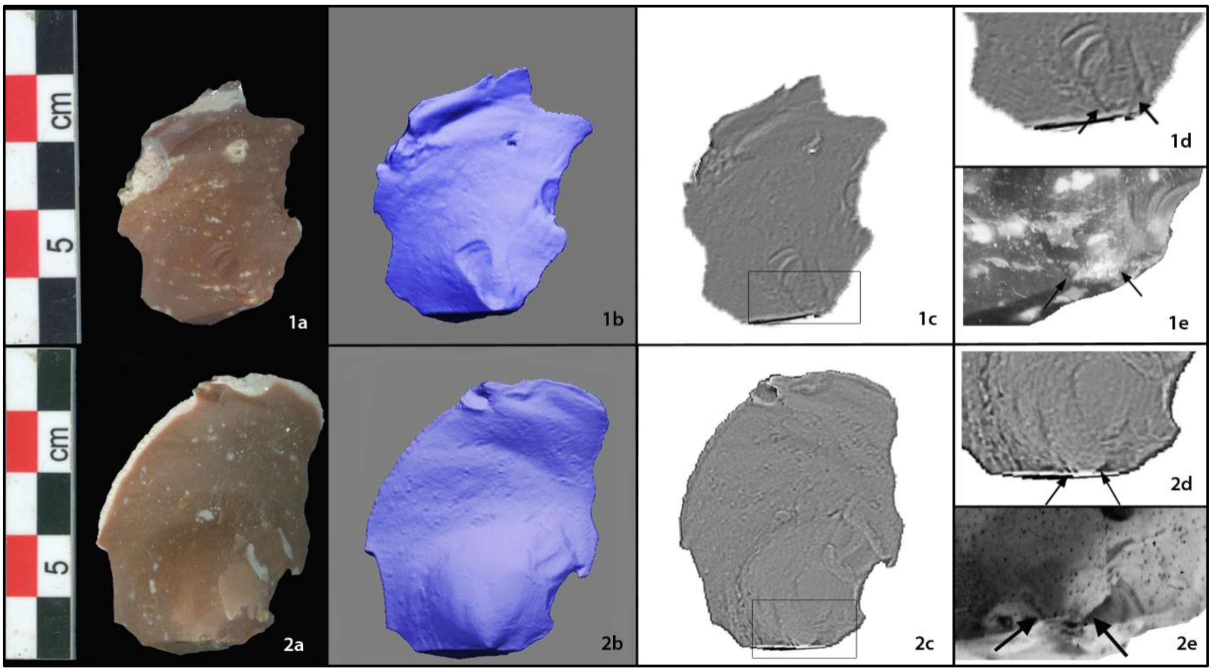


Figure 5. Examples of lithic flakes' cone crack paths highlighting on solid 3D models (1b-2b), using curvature analysis on raster models (1c-d, 2c-d) and after digital image processing (1e-2e).

Digital measurements have been executed using the Adobe Photoshop® rule tool and Kinovea® angle tool. Measurements focused on the exterior angles of the cone crack paths, in relation to a specific reference point. This is represented by a straight line, passing through the origins of the crack paths at a flake's striking platform. That area, according to contact mechanics, represents the impact surface of the hammerstone into the striking surface.

In practice, the difference occurring from the subtraction of the left to the right angle measured, could define, as a specific value, the skewness of a cone crack, but also the inclination of the impact blow. Theoretically, quotients with negative signs should be in agreement with an expected blow inclination of a left-handed, whereas quotients with positive signs, with that of a right-hander. For the needs of the research, in order to eliminate potential mistakes on measurements, quotients on the -

5/+5 region, considered to reveal unskewed cones, corresponding, in theory, to perpendicular blows on a striking surface **(Figure 6)**.

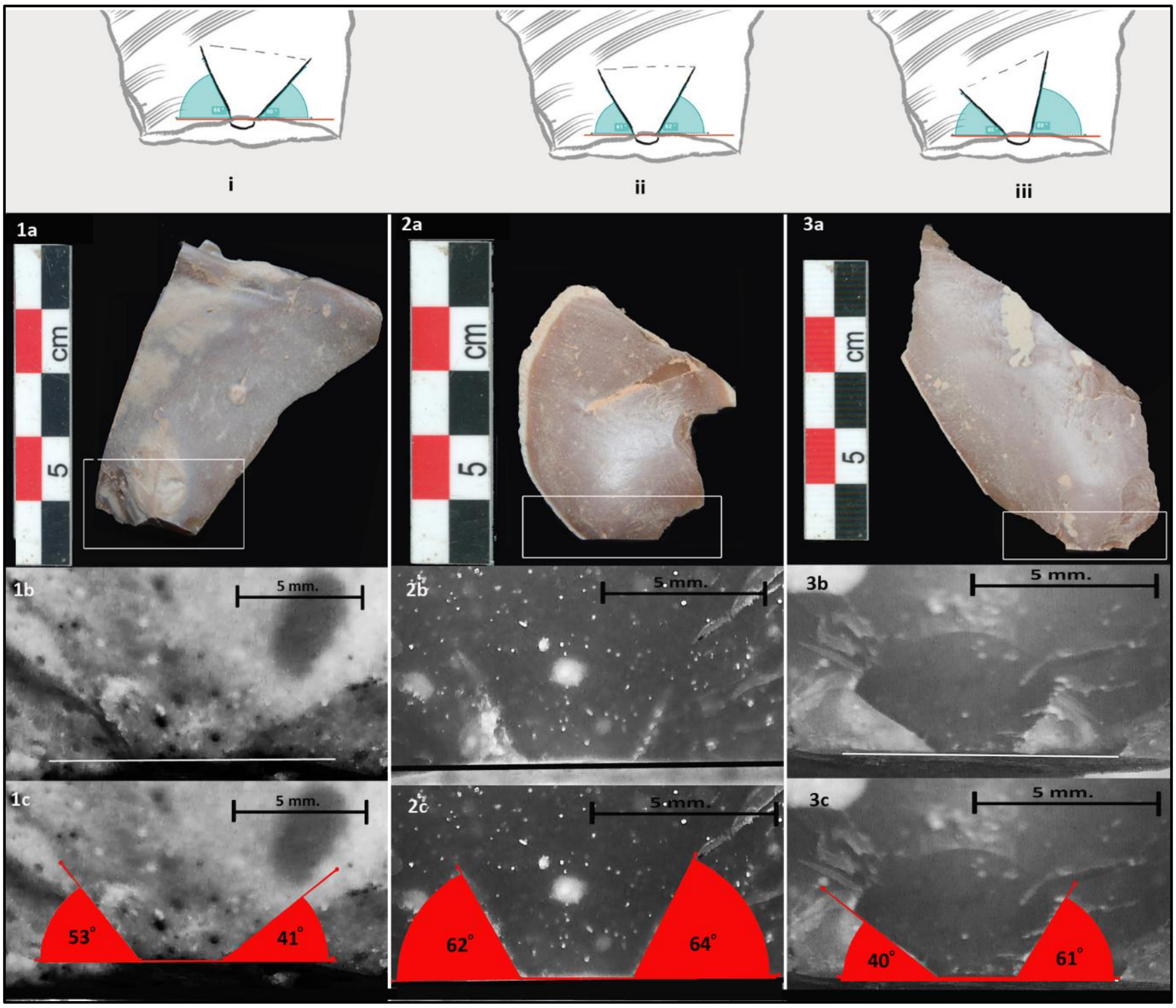


Figure 6. Application of digital measurements on experimentally produced flakes: 1: flake produced by a left-hander. 2: flake with an unskewed cone of percussion. 3: flake produced by a right-hander

The same practices were followed during the evaluation of the flakes produced, during the blind experiment. Yet, in this case independent, macroscopical, judgments preceded, and they were remained unknown, till the evaluation of the digital measurements.

Finally, concerning the contribution of the digital applications in the implementation of our research, the construction of the appropriate databases, facilitated the gathering and management of the empirical data gained.

3. Results

With regards to the results of the study's experimental part, a strong positive correlation between the cone crack geometry and the subjects' handedness is evident, as the values of the digital measurements indicate. The agreement rates approach an average of 78% and the determinability of the method is exceeding 90% **(Figure 7)**.

Considering the knappers' kinetic behavior, blow inclination data show that both left- and right-handers tend to bring on flaking angles, which confirm in an extended degree the *a priori* hypothesis of mirroring blows. Yet, the percentage of opposite than expected, and less or more perpendicular oriented blows, is not negligible **(Figure 8)**.

The results of the blind experiment confirm, at first, the general picture of the analyses carried out through 'open observations', and indicate that the methodology of the determination of the cone crack skewness, through its quantified digital evaluation, can offer high objectivity. Again, determinability ranges at a high level, and rates of successful judgments of knappers' handedness are over 72%. Of particular interest, however, are the results of the macroscopic observations, showing that successful judgments are just over 50%, a rate which would render the proposed method invalid, and it can reveal the inherent subjectivity featuring macroscopic examinations **(Figure 9)**.

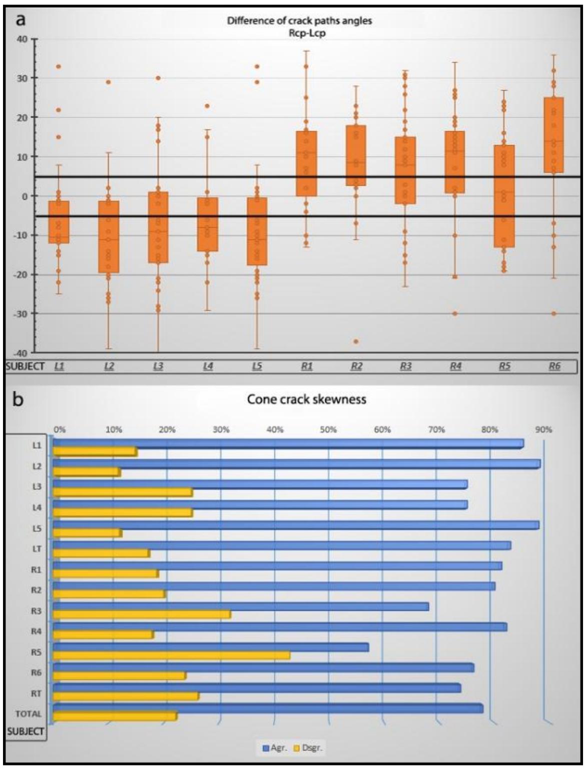


Figure 7. (a)Difference of cone crack path angles and(b)rates of cone crack skewness, on the experimentally produced lithic flakes according to subjects' hand-preference. L1, L2: Left-hander 1, 2…R1, R2: Right-hander 1,2…LT: Left-handers total. RT: Right-*handers total. Agr.: In agreement with hypothesis. Dsgr.: in disagreement with hypothesis*

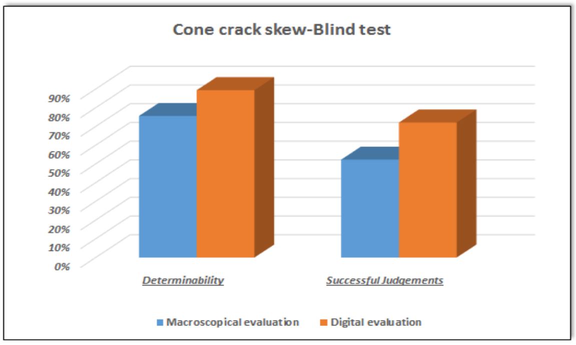


Figure 8. Correlation between blow inclination and cone crack skewness on the experimentally produced flakes among left- and right-handers. LH: Left-handers. RH: Right-handers. Agr.: In agreement with hypothesis. Dsgr.: in disagreement with hypothesis. Exp.: Expected, Perp.: Perpendicular, Inv.: Invert.

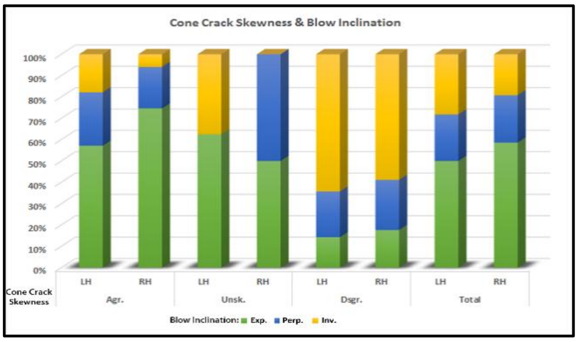


Figure 9. Results of handedness correlationwithin the context of the research blind test.

4. Discussion

The analysis of the experimental data indicates that during knapping tasks, the inclination of blows delivered, seems to an important extent be consequently connected with the geometry of the cone cracks, created on the flakes' ventral faces. It is also understood that observations made on artifacts using digital techniques, can reveal many more useful information about lithic technology's unknown aspects, than traditional macroscopic approaches.

A main issue arising from our experimental study is how handedness can be clearly associated with specific blow inclinations among left- and right-hand knappers, since our empirical data, show that during knapping procedures a deviation on this handedness expression is observed: even the 'high-scoring' left- and right-handers produce in some degree flakes, the features of which refer to knappers with an inverse hand preference.

Such an observation leads us to think that the non-absolute agreement rates between the knappers' handedness and their produced flakes' cone crack geometry, may not constitute a misjudging of the method represented, but a real fact of deviation from the manifestation of a 'normal' hand preference. Although such a phenomenon could well be associated with a series of technical causes, it could be also related, with the degree of hand preference that every individual brings, and the expression of his ambidexterity in various expressions of life (e.g. Bryden & Steenhuis 1987; Schachter 2000). Thus, this ratio of 'personal handedness' is very likely to be imprinted also in lithic production (for this issue see also Ligkovanlis 2022).

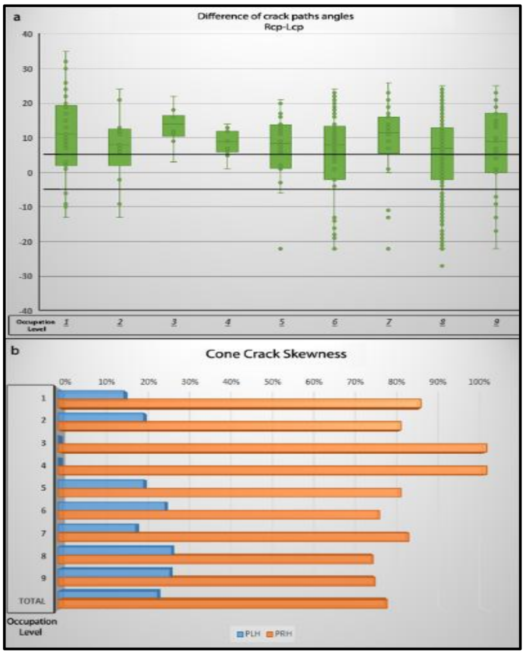


Figure 10.a) Difference of cone crack path angles of the flint flakes from Kalamakia cave, in relation to the distinct occupation levels. b) Potential hand-preference attribution of the artifacts based on the rates of determinable cone crack skewness. PLH: Possible left-handers. PRH: Possible right-handers.

The above discussion is inevitably involved with the issue of how in practice we could approach prehistoric handedness rates, among the archaeological record. Data from the current study showed us that at this step, we may be unable to approach the exact rates of handedness among hominins, but we could at least detect a predominance of right- or left-handers.

Following such a consideration, the pilot implementation of the study's methodology in the archaeological record has been decided, in order how the rates of prehistoric handedness could be approximated in real conditions, to be investigated. In total, we examined 468 flint flakes, derived from 9 different occupation levels of Kalamakia cave, which is located at Mani peninsula, southern Greece. That material is clearly associated with Neanderthals (Darlas 2007;Harvatiet al. 2013).

The results of this study demonstrate that these hominins in the cave, in a percentage more than 78%, created their flint stone products, bringing blowing angles that could be mostly connected with right-handers **(Figure 10)**. A statistical comparison with our experimental data shows that most probably the stone artefacts studied, constitute in their great majority the products of individuals with a right-hand preference **(Table 1)**.

	Experimental data-Left handers (6)	*Experimetal data-Right handers (7)*	*Kalamakia cave-Flint flakes*
Mean	**-5.97**	**7.06**	**5.85**
Standard Error	0.90	0.99	0.56
Median	**-8**	**9**	8
Mode	-11	8	8
Standard Deviation	13.10	14.72	11.56
Sample Variance	171.77	216.90	133.81
Kurtosis	0.69	-0.26	-0.10
Skewness	0.60	-0.44	-0.56
Range	73	74	62
Minimum	-40	-37	-27
Maximum	33	37	35
Count	208	220	421

Table 1. Descriptive statistics of the cone crack path angles' difference on the experimentally produced flakes and the flint flakes from Kalamakia cave.

5. Conclusions

The current research indicates that Neanderthal flintknappers, at least at Kalamakia cave, represented a strongly formatted 'right' manual lateralization and, thus, they probably had the neurophysiological prerequisites for linguistic communication. More research at other Middle Palaeolithic assemblages is needed, in order such an argument to be confirmed. Moreover, comparisons between assemblages created within a distance of several millennia by different hominin types might also indicate evolutional trends, concerning the 'dominant hands' of prehistoric societies.

Another crucial conclusion of the current research is the decisive contribution of modern technologies towards the understanding of our past. The current issue was a puzzle for over 2 decades, and it seems that its solution is promoted through technological development, which manages to reveal former invisible aspects of both the material culture and behavior of its carriers.

In such a manner, the current effort comes to be added to a series of other approaches (e.g. Stout et al. 2015; Li et al. 2017) using digital tools, in order, through stone tools, to explore the past and evolution of human cognition, contributing, with this manner, to the understanding of our present conscious or unconscious minds.

It is sure that such efforts will be intensified, proving that a kind of a computer assisted archaeology, does not come to simply switch the analogical into digital, but to answer on matters of substance, helping us to promote our knowledge, and fulfill our inherent curiosity. In such a research future, interdisciplinarity and co-operation are emerging as an essential component, in order to combine our 'left and right brains' for the benefit of science, but also our life.

Acknowledgments

We are grateful to the volunteers participated to the experimental knapping sessions. Special thanks are due to Dr. Andreas Darlas, Dr. Ourania Palli, Dr. Peny Tsakanikou, Dr. Vasiliki Boura, Dr. George Stergiou and Dr. Katerina Ligkovanli for their help during the study implementation.

This research is co-financed by Greece and the European Union (European Social Fund- ESF) through the Operational Programme «Human Resources Development, Education and Lifelong Learning» in the context of the project "Reinforcement of Postdoctoral Researchers - 2nd Cycle" (MIS-5033021), implemented by the State Scholarships Foundation (IKY).

References

Aydelotte, B., Jannotti, P., Andrews, M., Schuster, B., 2016, A Comparison of Damage in Glass and Ceramic Targets. *Advances in Ceramic Armor, Bioceramics, and Porous Materials* 608, 3-12.

Bargalló, A., Mosquera, M., 2013, Can hand laterality be identified through lithic technology? *Laterality: Asymmetries of Body, Brain and Cognition* 19(1), 37-63. DOI: https://doi.org/10.1080/1357650X.2013.769559

Berwick, R.C., Hauser, M., Tattersall, I., 2013, Neanderthal language? Just-so stories take center stage. *Frontiers in Psychology* 4, 671. DOI: https://doi.org/10.3389/fpsyg.2013.00671

Bryden, M.P., Steenhuis, R.E., 1987, Handedness is a matter of degree. *Behavioral and Brain Sciences* 10(2), 266-267. DOI: https://doi:10.1017/S0140525X00047737

Chaudhri, M.M., Liangyi, C., 1989, The orientation of the Hertzian cone crack in soda-lime glass formed by oblique dynamic and quasi-static loading with a hard sphere. *Journal of Materials Science* 24(10), 3441-3448. DOI: https://doi.org/10.1007/BF02385722

Darlas, A., 2007, Le Moustérien de Grèce a la lumière des récentesrecherches. *L'Anthropologie*111-3, 346-366.

Dediu, D., Levinson, S.C., 2013, On the antiquity of language: the reinterpretation of Neandertal linguistic capacities and its consequences. *Frontiers in psychology* 4, 397. DOI: https://doi.org/10.3389/fpsyg.2013.00397

Frank, F.C., Lawn, B., 1967, On the theory of Hertzian fracture. *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences* 299(1458), 291-306. DOI: https://doi.org/10.1098/rspa.1967.0137

Gentilucci, M., Corballis, M.C., 2006, From manual gesture to speech: A gradual transition. *Neuroscience & Biobehavioral Reviews* 30(7), 949-960. DOI: 10.1016/j.neubiorev.2006.02.004

Harvati, K., Darlas, A., Bailey, S.E., Rein, T.R., El Zaatari, S., Fiorenza, L., Kullmer, O., Psathi, E., 2013, New Neanderthal remains from Mani peninsula, Southern Greece: the Kalamakia Middle Paleolithic cave site. *Journal of Human Evolution* 64(6), 486-99. DOI: https://doi.org/10.1016/j.jhevol.2013.02.002

Knecht, S., Dräger, B., Deppe, M., Bobe L., Lohmann, H., Flöel, A.S., Ringelstein, E.B., Henningsen, H., 2000, Handedness and hemispheric language dominance in healthy humans. *Brain* 1 23(12), 2512-2518. DOI:10.1093/brain/123.12.2512

Li, H., Kuman, K., Lotter, M.G., Leader, G.M., Gibbon, R.J., 2017, The Victoria West: earliest prepared core technology in the Acheulean at Canteen Kopje and implications for the cognitive evolution of early hominids. *Royal Society Open Science* 4(6), 170288. DOI: https://doi.org/10.1098/rsos.170288

Ligkovanlis S., 2022, Hominin Hand-Preference and Lithic Production: Exploring Neanderthal Handedness Rates through the Study of Hertzian Fracture Features on Lithic Blanks. *Open Quaternary* 8(4), 1-21. DOI: 10.5334/oq.111

Mellars, P., 1996, *The Neanderthal Legacy*. Princepton University Press, New Jersey.

Lieberman, P., 2007, Current views on Neanderthal speech capabilities: A reply to Boë et al. (2002). *Journal of Phonetics* 35(4), 552-563. DOI: https://doi.org/10.1016/j.wocn.2005.07.002

Ruck, L., Broadfield, D.C., Brown, C.T., 2015, Determining hominid handedness in lithic debitage: a review of current methodologies. *Lithic Technology* 40(3), 171-188.

Ruck, L., Holden, C., Putt, S., Shick, K., Toth, N., 2020, Inter- and Intra-rater Reliability in Lithic Analysis: a Case Study in Handedness Determination Methodologies.

Journal of Archaeological Method and Theory 27, 220–244.
DOI: https://doi.org/10.1179/2051618515Y.0000000009

Rugg, G. & Mullane, M., 2001, Inferring handedness from lithic evidence. *Laterality: Asymmetries of Body, Brain and Cognition* 6(3), 247-259.1
DOI: https://doi.org/10.1080/713754411

Schachter, S.C., 2000, The Quantification and Definition of Handedness: Implications for Handedness Research. In Mandal MK., Bulman-Fleming, M.B. and Tiwari, G. (eds.), *Side Bias: A Neuropsychological Perspective*, 155-175. Dordrecht: Springer.
DOI: https://doi.org/10.1007/0-306-46884-0_6

Stout, D., Hecht, E, Khreisheh, N., Bradley, B. & Chaminade, T., 2015, Cognitive Demands of Lower Paleolithic Toolmaking. *PLoS ONE* 10(4), e0121804.
DOI: https://doi.org/10.1371/journal.pone.0121804

Uomini, N., 2001, Lithic indicators of handedness: Assessment of methodologies and the evolution of laterality in hominids. Unpublished thesis (M. Sc.), University of Durham.

USING DIGITAL APPROACHES TO REVEAL THE HISTORY OF A RARE PORTABLE MOSAIC ICON CREATION, USE, ALTERATIONS, INTERVENTIONS

D. MAKRIS[1], S. CHLOUVERAKI[1] AND S. AKPEK[2]

[1]*Department of Conservation of Antiquities and Works of Art, University of West Attica, demak@uniwa.gr, schlouve@uniwa.gr*

[2]*Department of Radiology, VKF American Hospital, Istanbul, Turkey*

Abstract

The mosaic icon of 'Pammakaristos', the All-blessed Mother of God, an emblematic icon of the Ecumenical Patriarchate of Constantinople,is one of the finest examples of Byzantine art, dated to the early 12th century. A documentation framework which involves an interdisciplinary team of conservators, computer graphic and medical radiology experts, was pursued in order to investigate the condition of the icon and trace its history through time. Three-dimensional digital imaging methodologies were employed to address conservation/preservation. Furthermore, an assessment framework was defined aiming to evaluate their contribution towards objective diagnostic and documentation procedures. A full three-dimensional digital imaging survey was carried out as part of the icon's comprehensive study. The aim of the study was to provide baseline information for a thorough examination and documentation from structural, geometrical and morphological perspective. A three tiers acquisition protocol based on visible (active and passive radiance capturing methods) and non-visible spectra (dual-energy computed tomography) acquisition methods was undertaken, guided by the conservators' requirements. The integrated digitization workflow and the subsequent analysis were instrumental in the conservation decision-making process prior to intervention. These studies have so far provided a comprehensive and detailed record of the current condition of the icon and will be further applied to guide, monitor and document the current interventions and the resulting changes in the icons' 'geomorphometry'.

Keywords*: optical-laser-scan, structure-from-motion, computed tomography, conservation, Byzantine icon, cultural heritage documentation*

1. Introduction

The portable mosaic icon of 'Pammakaristos', is an emblematic ecclesiastical artwork of the Ecumenical Patriarchate of Constantinople, renowned for its outstanding religious and artistic value. It is a masterpiece of the early 12thCentury, made of minute glass, gold leaf and stone tesserae set on wooden panel with wax-resin paste, a rather rare technique that demonstrates the meticulousness and brilliance of Byzantine art (Fig.1). The icon was subjected to various conservation treatments in the past, however, its condition today is still critical, owing to the severe damages related to its use and to physical deterioration processes (Chlouveraki 2020). Macroscopic observation, structure-from-motion, optical laser scanning and computed tomography were employed in the examination, documentation and assessment of its condition. Moreover, the technology employed in its creation as well as the various interventions applied over the centuries were investigated

Mosaic surfaces can be rather challenging as they exhibit irregular concavities and convexities related to the constituent materials, material joints, and the form and synthesis of tesserae. Portable mosaics present special features owing to their three-dimensional nature, which parallels panel icons. Intrinsic qualities of tesserae such as their color, degree of diffraction, refraction and reflection present challenges on any three-dimensional digital acquisition approach. In the case of the mosaic icon under study, additional limitations were met as the icon had to be examined within the premises of the Patriarchate.

The implementation of three-dimensional (3D) modeling in conservation has widened the geometric and topographic survey processes (Dellepiane et al., 2011; Eros et al., 2017; Ioannidis et al. 2019; Makris et al., 2021). Models of high resolution and accuracy which derive from visible and non-visible methods can be mutually merged in a holistic model that can be further connected with different media resources and archival material. 3D models allow the deduction of both metric and semantic knowledge. Structure-from-Motion (Koutsoudis et al. 2014; Remondino et al., 2014; Aicardi et al., 2018), optical three-dimensional laser scanning (Bornaz et al., 2007; Kersten et al. 2018; Makris et al., 2018), and dual-energy computed tomography (Re et al., 2016; Zhao et al., 2018; Montainaet al., 2021) are the non-invasive digital investigation approaches employed. These technologies are complementary and their role is widely recognized and appreciated by cultural heritage conservation and preservation professionals (Hess 2015; Palma et al., 2019; Montaina et al., 2021).

Figure 1. The mosaic icon of Pammakaristos (920x620mm)

1.1. Related approaches and studies

The field of conservation is rapidly adapting to contemporary digital technologies and exploits them in many applications and artwork studies of varied, materials, sizes and complexity. Numerous studies are presented in the literature, which focus on 3D digitization approaches, aiming at the reconstruction of digital models which can be utilized as documentation and diagnosis agents, while fostering the promotion heritage assets. Heritage items of multivariate structural and artistic characteristics often demand the employment of diverse 3D digitization methodologies. The synthesis of spatial, spectral and temporal information, in parallel with physical and chemical data have appeared in the beginning of 2000, aiming towards the integrated documentation and archiving of cultural artifacts (Tsirliganis et al.,2002; Tsirliganis et al., 2004) and the improvement of examination frameworks for religious icons (Stratis et al., 2014). The application of multi-image digital photogrammetric monitoring and surface measuring tools is reported in (Robson et al., 2004) on the treatment of the Westminster Retable, consisting of an oak support panel with several ornamental and structural elements of varied materials. In-situ CT scan analysis is also reported by (Morigi et al., 2010) in the study of two Japanese wooden statues of the XIII and the XVII century, where significant information on the construction technique and the past restoration efforts was obtained. The advantages of tomography in the investigation of artefacts of complex geometry and diverse materials is also attested in case of the Taiefmutmut's coffin lid (Re et al., 2016), where it revealed a large volume of information that could not be obtained by classical radiography. Manufacturing techniques, inner distribution of materials, state of preservation as well as previous interventions were detected and documented. An exemplary application of 3D scanning in the conservation of artworks has been demonstrated in the case of the Da Vinci's Da Vinci's renowned wood panel painting of 'Adorazione dei Magi', where the artwork's aging behavior was evaluated on the basis of the comparative study of 3D data obtained in two monitoring phases. The results of the monitoring enabled the readjustment of the interventions in order to prevent future deformation effects (Palma et al., 2019). Finally, the recent work of (Montaina et al., 2021) has demonstrated the further exploitation of Clinical Multislice Computed Tomography in the identification of the morphological and structural characteristics of the wooden panel of a 17thcentury painting, thought the study of radiodensity values. While the above studies employ the advantages of one documentation technique for each case, the current study attempts to employ all the aforementioned techniques in one case study and to compare and integrate their results.

Research aims

The aim of this paper is to report on the application of 3D modeling approaches that could benefit the study and conservation of portable mosaic icons. Moreover, a comprehensive assessment of the diverse quantitative and qualitative capabilities of the techniques employed in the study of the Pammakaristos icon is undertaken.

The goal of the Pammakaristos icon conservation program is twofold and involves both the study of materials and construction technology as well as the conservation, restoration and long-term preservation, ensuring its longevity. The research tasks are related to the diagnosis and documentation of its pathology, the evaluation of its condition and finally the decision making and planning of interventions. An additional objective within the framework of this project is to study the integration of 3D modeling techniques based on visible and non-visible spectrum documentation methodologies. A multidisciplinary team of conservation, computational and medical application experts joined their forces on a full 3D digital imaging survey targeting the comprehensive examination and documentation from structural, geometrical and topographical perspectives.

2. Documentation framework – protocol

A documentation framework was set up that would allow us to trace the alterations in the con's surface and inner structure through time and to investigate the factors or actions which ~~may~~ have contributed to its current state of preservation. To achieve these tasks, we undertake a three tiers acquisition protocol based on visible and non-visible spectrum acquisition methods. Due to the fine detail analysis, the icon's digitization and modelling requires a submillimeter overall precision. Three major techniques were adopted: a passive close-range documentation technique, namely photogrammetry, and two active range documentation techniques, namely optical 3D laser scanning and dual energy computed tomography. The latter has the advantage of unfolding the inner structure and condition of the artwork in its totality thus, supplementing the results of photogrammetry and optical laser 3D scanning, both focusing on the visible parts of it, providing detailed and accurate documentation of its surface. To our knowledge, this approach has not been reported in the relevant literature so far.

3. Visiblespectrumdocumentation techniques

The visible spectrum techniques are based on both active and passive documentation methods (Georgopoulos, Stathopoulou, 2017). The former involves the utilization of self-emitted radiation, typically in the form of laser light, to acquire spatial data points. While the later comprise capturing the light or radiation that comes from an external source (like the sun or artificial lights) and gets reflected off the object being studied. In particular, the former includes optical 3D laser scanning, while the latter comprises of structure-from-motion techniques based on close-range multi-imaging process. Both methods could provide detailed geometric and color information of the exterior surfaces and materials. In this project, photogrammetry and in particular Structure-from-Motion (SfM)and Multiple-View-Stereo (MVS),as well as laser scanning were implemented to accurate 3D document the icon's morphological characteristics and to detect surface anomalies caused by physical or anthropogenic factors before any conservation treatment is undertaken.

3.1. Multi-image close-range photogrammetry

The aim of SfM/MVS survey was to process a detailed 3D textured mesh, and a high resolution orthophoto. For the multi-view multi-image capture, a mirrorless Digital Single-Lens Reflex (DSLR) camera, Sony A6000 featuring an APS-C CMOS (23.5x15.6mm^2, with pixel size of 3,88μm) sensor of 24.3 megapixel, with a lens of 35 focal length, was used. The photographic process follows three different paths all around the icon, the front panel, the back panel and the sides. The main challenge was the safety of the icon during image acquisition.

Figure 2. Camera positions and frames on the front and the sides of the icon (processed in Agisoft Metashape Pro)

An extensive fracture was revealed by the initial radiographic investigation, which seemed to run along the entire length and depth of the wooden panel (Chlouveraki 2020). In order to eliminate the risk, the icon had to remain stationary while the camera was moved around it. A series of 936 images were acquired with an overlap of 70-80%.Uniform and stable lighting conditions were set to avoid shading. This was achieved with the use of two illumination sources, symmetrically positioned to the vertical axis of the icon. For each path the shooting scenarios results an average of 350 images, captured from a distance of approximately 45cm from the surface of the icon (Fig. 2). The acquired images were converted to .jpg format, white balance and color calibration were corrected, and the background was removed. In order to accurately scale the object photogrammetry targets were placed around the icon on the supporting board, while supplementary measurements were taken on the icon itself. The reconstruction of the 3D model took place in the Agisoft's Metashpe Pro®, (1.5.4.8885) a commercial image-based 3D modelling software for creating 3D content from still images, with both fully-automated and customized choices. Color information was stored as a high detailed texture map associated to an .obj file (Fig. 3). The resulting 3D digital replica of the icon, processed in medium quality, has a verified average resolution of 0.45mm, and consists of approximately 6.6 million vertices, 13.3 million faces, in an 1.5 GB, .obj format file, (Fig. 3).

Figure 3. S-f-M – 3D mesh of the icon

3.1.1. Orthophoto procedure

Orthophotos of high resolution were produced for the front and back side of the icon in Agisoft Metashape Pro, 1.5.4.8885, on a MacOS 64 platform. For each side, an orthomosaic was built in geographic projection, using local coordinates and average blending mode. The extracted orthophotos were in TIFF format of 2 GB (Fig. 4).

3.2. Self-positioned hand-held optical laser scanner

High level metrology acquisition systems have been introduced in cultural heritage studies with great success. Although they were initially produced for industrial design and production processes, they can be applied to a wide range of cultural heritage assets including movable and immovable archaeological finds as well as artworks of various forms.

3.2.1. Pre-acquisition stages

The characteristics of the particular artwork, consisting of minute glass tesserae and its aesthetic qualities demand a high-resolution capture setup. Therefore, the icon has to be divided into separate scans according to the desirable resolution, which was obtained through the appropriate adjustment of the scanner's acquisition volume. Considering the icon's magnitude (920mmx620mm), scanning acquisition was subdivided into twenty zones of about 200mm^3. The subdivision procedure was carefully adjusted to satisfy an overlapping area between each scan. Each zone overlapped with the adjacent ones by a percentage of about 3% to 6% in all relative directions with regard to the required accuracy level and precision of the individual acquisitions' registration. A number of 120 reflective targets were carefully placed on the surface of the icon and the perimeter of the supporting table, ensuring that the final alignment of the acquisition surfaces is gained through the common reference system that is defined by the targets and at the same time that delicate details of the tesserae artistic and aesthetic continuity are not obscured.

Figure 4. S-f-M – Orthophoto of front and back side of the icon

3.2.2. Survey operation

The ‘Pammakaristos’ icon’s surface presents different texture and colors intensities due to the different constituent (glass, gold and stone tesserae) and intervention materials (various types of fillers and overpaints). The glass tesserae by their nature reflect light up to a certain degree, as a result, during data acquisition a sort of noise was created. To tackle such phenomena first, we define our approach with longer scanning capture time. Second, due to the fragility of the icons’ materials, the conservator has been stabilized the condition of the icons’ surfaces with the application of a . As a result, the acquisition process did not experience any levels of refraction and / or reflection. Therefore, complex characteristics of the portable mosaic icons demand diverse acquisition approaches. The intricate materials can cause minor reflections which result in the occurrence of artificial holes within some irregular cavities of the surfaces. If necessary, the scanner’s data processing software can restore the holes based on neighborhood curvature detection embedded algorithms. In a similar manner the software can also restore the areas that are covered by the targets, according to neighboring surface values. The tesserae materials, especially the glass and gold ones, have particular characteristics, which challenge the optical laser scanner and make it impossible to follow one specific calibration, owing to the continuous variation in the surface’s light reflectance. The different kinds of materials, (wood, glass tesserae and stone), coexisting within small areas, necessitates the readjustment of the scanner’s calibration (laser beam power and the camera’s shutter count), resulting in rather time-consuming acquisitions. Each of the icon’s parts was registered onto the original coordinate system achieved by the targets’ initial acquisition.

3.2.3. Data processing

The resulting scanning data sets were initially edited within the scanner’s embedded software application. Functions, like noise reduction and filling holes, eliminate numerous minor geometric and topologic irregularities, mainly on the sides of the icon. Each scanned partition was recorded in a single scan file, with a 0.39mm verified resolution (Fig. 5).

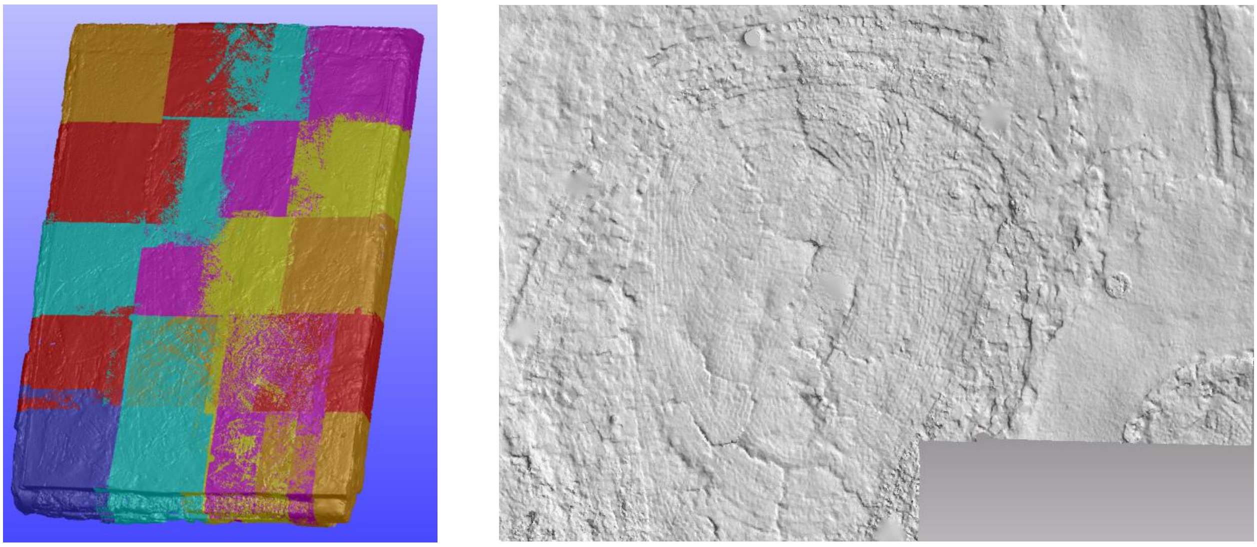


Figure 5. 3D scanning – Scanned partitions

The models were exported in .obj format with texture color information, and were processed and registered in Meshlab, a 3D mesh editing and processing software (Cignoni et al., 2008). In all partial 3D meshes were applied only relative filters that tackle topological inconsistencies without affecting the original acquisitions. The final alignment of the separate scans was accomplished in Meshlab based on the targets' common reference coordinate system and the partial meshes were merged into one continuous three-dimensional model. The resulting 3D digital replica of the icon has a verified average resolution of 0.39mm, and consists of approximately 7.3 million vertices, 15.4 million faces, in an 1.5 GB .obj format file (Fig. 6).

Figure 6. 3D scanning – front side of the icon

4. Non-visible spectrum acquisition

In recent years conservators have employed X-ray Computed Tomography (CT-scan) for non-invasive investigation of the morphology and inner structure of composite and delicate artworks. Tomography studies can provide valuable information on diverse documentation topics related to artworks, among which is the investigation of the structural characteristics of wood and wooden artifacts. Three-dimensional volume and surface renderings based on CT imaging reveals the diverse types of joints and assemblies utilized in wooden artifacts and structures, as well as the activity of wood boring (xylophagous) insects (Zhao et al., 2018). Moreover, the detection of wood density variations (HU values) on two-dimensional CT images and three-dimensional reconstructions enables the identification of internal log features such as the pith, growth rings, heartwood, sapwood, knots or other defects, thus allowing the identification of the part of the log utilized for the construction of an artefact and occasionally the type of wood (Montaina et al., 2021). Other materials of different radiodensities, such as inorganic materials (metallic elements, fillers, etc.) are highlighted and can be identified based on the HU values, while mechanical damage can be detected and accurately documented. In the present study the implementation of CT scanning is concerned with a twofold scope, on one hand, the documentation of the structural characteristics of the wood and of the icon as a composite artefact, and on the other hand, the diagnosis of its pathology and evaluation of its overall state of preservation. The dual-energy Computed Tomography proved to be a powerful tool in the non-invasive diagnostic approach. It has provided explicit evidence of the icon’s inner condition, while the 3D reconstruction by VRT revealed the different components of the icon’s internal structure and highlighted all the inorganic elements that were added to the icon, such as restoration mortars and metallic elements (nails) as well as their exact location in the composite structure.

4.1. From DICOM to 3D modelling

The examination of the icon was carried at the American Hospital of Istanbul on a Somatom Defitition Flash, SIEMENS MED CT Dual Energy system. Abdomen total body spiral acquisition was performed, using the following setting parameters: 504mA XRay tube current, 50 kW power to the x-ray generator, exposure of 210mAs, exposure time of 500msec, Spiral Pitch Factor of 1.2mm, matrix of 512x512, Pixel Spacing 0.9765625\0.9765625.The obtained data are in DICOM format, of 1634 slices, with a slice thickness of 0,6mm. Image reconstruction by a B30f Kernel for multi-planar (MPR) and 3D reconstruction by volume rendering technique (VRT) were also obtained. Post-processing image analysis was evaluated by Syngo CT2012B software and was further studied in Osirix Lite and Horos Project software. In our experiment, a Somatom Definiton MSCT (Siemens Healthcare) with a gantry opening of 80 cm was used. Abdomen total body spiral acquisition was performed setting the following parameters: 120 kV energy tube voltage, 35 mA current, slice thickness of 0.6 mm, pitch factor of 1 mm, field of view (FOV) of 445 mm, matrix of 512x512, spatial resolution of 0.87 mm, and scan time of 14.84s. Image reconstruction by a B60f kernel for multi-planar (MPR) and 3D reconstruction by volume rendering technique (VRT) were obtained. Post-processing image analysis was evaluated by SYNGO Siemens and Horos Project software. A limitation of the method is the width coverage of medical CT scan, which is smaller than the width of the icon. However, an effective width of 495mm out of the 620mm of its total width was captured.

5. Observations based on 2D and 3D reconstructions

Research in the literature provided some insights in the condition of the icon and the past interventions which shaped the initial objectives of the digital diagnostic and documentation techniques. Images of the icon published at the end of the 19th century, in 1915 and 1933 (Kondakov 1915: 200, Sotiriou 1933: 359, Millas 2013:81), provide clues about the condition of the icon before 1933. Sotiriou (1933) reported the critical condition of the icon and the conservation interventions undertaken by the artist K. Vasmatzidis in 1933, without elaborating the technical details (Fig. 7), (Millas2013).

The icon’s complexities are examined in three levels: the tessellated surface, the original wooden panel and the second panel, which was added later as a supplementary support. Comprehensive documentation, through digital reconstructions was driven by the fundamental objectives of the conservation program to study the materials, structure, technology and the pathology of the icon and finally to trace the history of past conservation interventions.The 2D reconstructions in axial, sagittal and coronal planes revealed information that could not be acquired by other means. Hidden elements and variations in the condition of the panels with respect to depth were disclosed. The structural condition and the extent of biodeterioration due to the activity of wood boring insects was evaluated in respect to depth (Fig. 8).

We were also able to understand and assess the severe mechanical damage that was encountered during the long and troubled history of the icon. A severe fracture runs along the entire length of the icon and through the entire depth of the original panel, which explains the application of the second two-component wooden support. The materials and techniques used for the assembly of these additions as well as discontinuities between its two components were identified. Further discontinuities between the original panel and the second support were also detected (Fig. 9 -10).

Figure 7. Images of the icon of Pammakaristos as illustrated in the publications of Kondakov, 1915, pp. 220 (left) and Sotiriou 1933, pp. 359 (middle, before conservation and on the right, after Vasmatzidis' interventions), (Chlouveraki 2020)

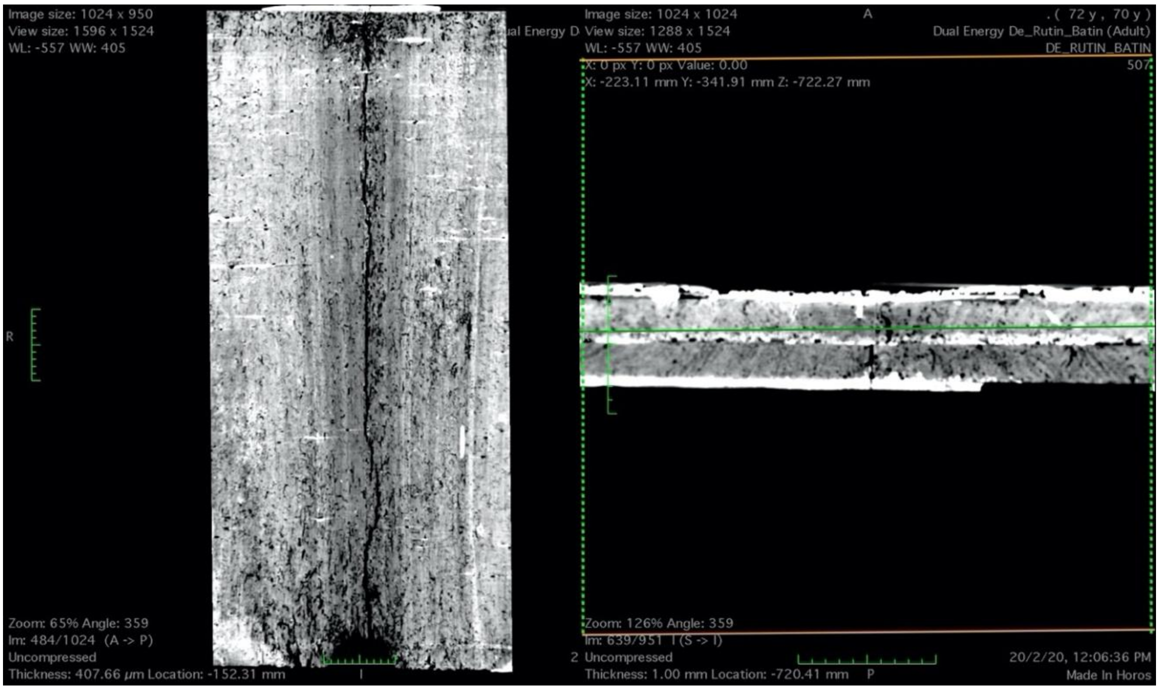


Figure 8. The structural condition and the extent of biodeterioration as documented in axial and coronal planes (made in the Horos Project)

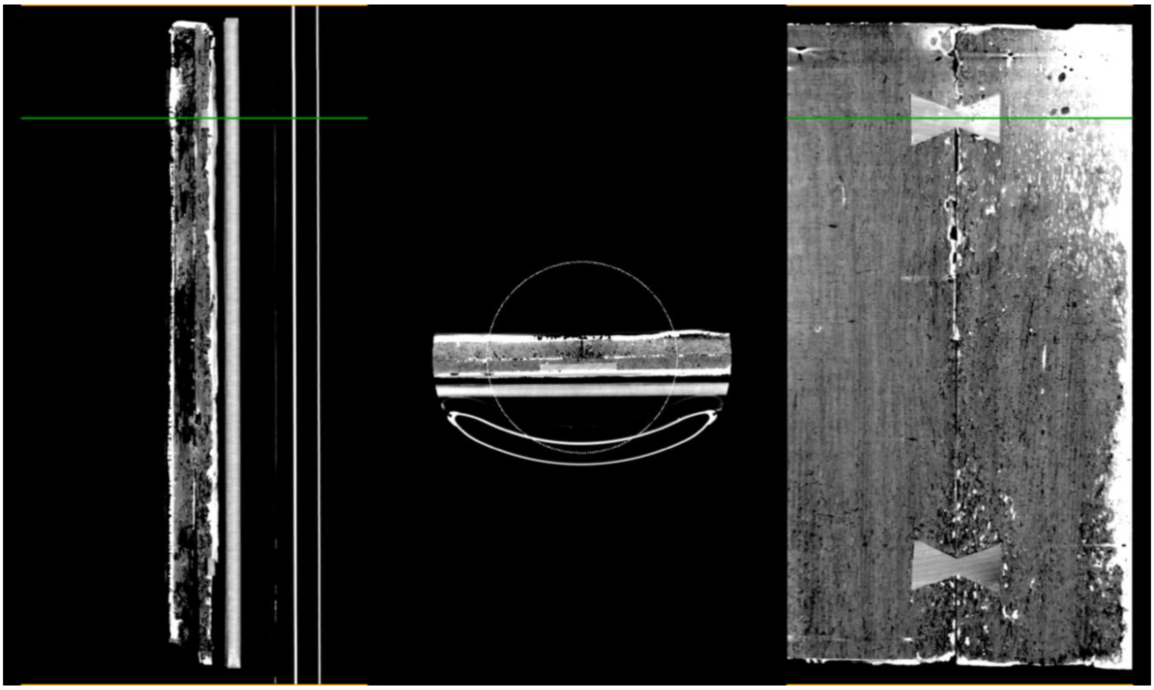


Figure 9. Structural details of the second supporting panel in sagittal, axial and coronal planes. Butterfly type of joins are used to connect the two boards of the second panel (extracted in the Horos Project)

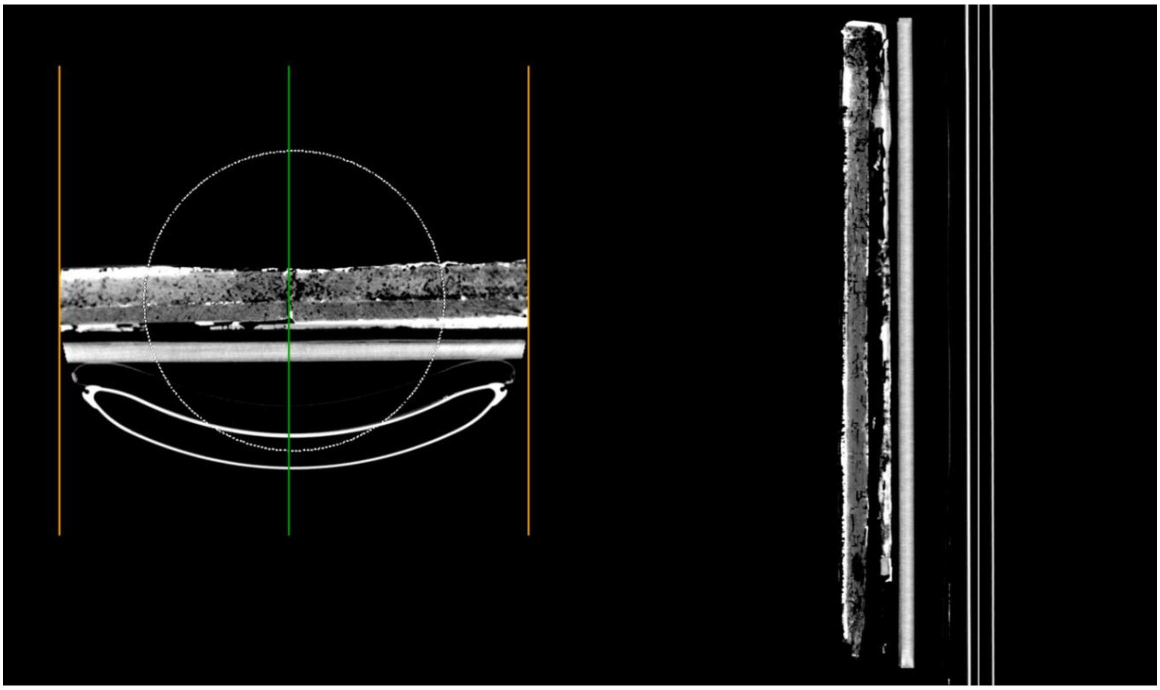


Figure 10. Axial slice of the composite structure of the icon (left) and discontinuities in the assembly of the parts of the second panel shown as black areas in a sagittal slice (right) (extracted in the Horos Project).

Figure 11. Image of the icon taken at the end of the 19th century, before the interventions of 1933 (Sebah et Joaillier, German Archeological Institute of Istanbul, cited in Millas 2013, pp. 81)

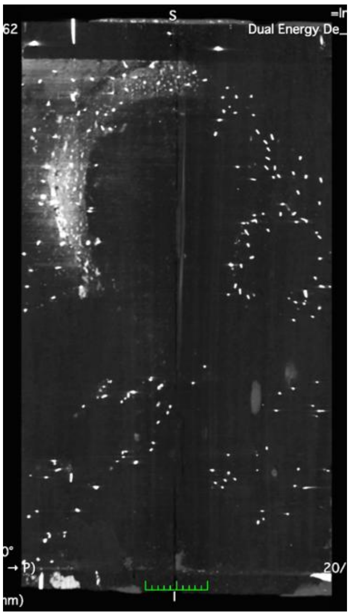


Figure 12. CT Scan slice (coronal plane) showing the distribution of metallic elements (nails) and their correlation with the votive metal sheets and the interventions (made in OsiriX Lite)

The distribution of the metallic elements (nails) as illustrated in axial slices provides insights into their function. It was determined that nails were used for the attachment of votive metal sheets which were placed around the halos and the hands of both figures and on body of the Christ, as well as for anchoring the restoration mortar in re-laid areas of the 'golden' background and the fillers used to reintegrate the severely damaged halos (Fig. 11-12) (Chlouveraki 2020).

The 3D reconstructions of the CT Scan, further provide an explicit view of the icon's inner structure and the geometric and morphological characteristics of all its constituents. The geometry of the dense sequence of metallic elements (Fig. 13), their location and3D spatial distribution is better understood in the 3D model of the inner structure of the icon. In some instances, their shape and bending degree can challenge the attempt to remove them (Fig. 14). Computed Tomography is expected to be instrumental in guiding their safe removal, an operation which resembles fine surgery.

Finally, the examination of the CT 3D model contributed to the verification and interpretation of the preliminary macroscopic observations on the structural characteristics and the current condition of the icon. A curved socket in the lower-middle part confirmed that the icon was indented for processions, which to some extent, justifies the loss of the tesselatum in this area as a result of increased mechanical stress.

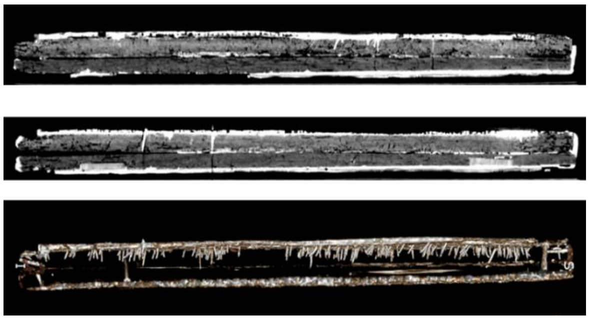


Figure 13. Nails as detected in 2D slices in sagittal plane (top and middle) and 3D reconstruction (bottom)

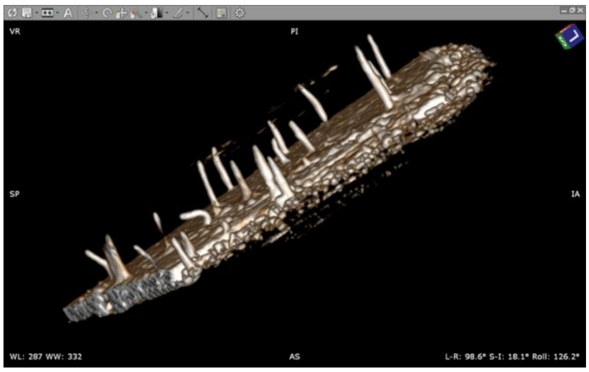


Figure 14. Nails of various shapes and bending degree in 3D reconstruction of the inner structure of the icon

5.1. Digital shading – highlight variations

The collection of the raw sampled data of SfM and laser scanning enabled the production of the icon's complete digital model. The post–processing phase included the individual models' (as resulted from the two methodologies) alignment based on the sampled data at the required accuracy and resolution. A preliminary analysis of the surface's topography (based on the high-fidelity 3D mesh), without the texture, illuminates the condition of the surface in its full extent, in an objective and instant visual approximation. Initially, we alter the virtual light's angles to intensify the surface geometry (Fig. 16).

Next, we deploy Radiance Scaling, a specific shader in Meshlab, that enables the detail adjustment of the reflected light intensities based on different material's characteristics and surface morphology. In particular, some parts display significant bulges (Fig. 17), which in turn result in extended anomalies in the continuity of the icon's surface.

The utilization of a specific variation of Radiance scaling, that of 'Lit Sphere Radiance Scale' highlights the surface concavities and convexities instantaneously. We can control the variations of the lighting source angles in precise steps along the three spatial axes, to achieve an improved visual inspection and perception of the detail of the surface under examination.

6. Discussion – conclusions

The on-going project to investigate, conserve and safely reutilize the Pammakaristos icon for pray and veneration, demanded the full exploitation of non-destructive approaches. A full three-dimensional documentation combined with medical diagnostic techniques was employed and evaluated in order to develop a protocol for the optimal methodology for the investigation of portable mosaic icons or other artworks with similar characteristics. Three specific digital imaging techniques, structure-from-motion, optical laser scan and computed tomography were combined in order to meet the requirements of diagnosis and to provide feedback in the conservation decision process.

The combination of the three methods contributed to a comprehensive and perspicuous documentation of the icon's exterior surface and inner structure. The several deliverables of the documentation process provided meaningful information for the interpretation of the icon and contributed towards the a) creation of high-resolution 3D models and orthophotos for the documentation of the icon's surface relief and mapping of surface deterioration phenomena, b) analysis of the constituent structural components of the icon including number and distribution of nails c) study of the inner layers and in-depth investigation of deterioration features of the wooden panels, d) documentation and assessment of its overall current condition status, e) estimation of deformations of the wooden support and of the tessellated surface, and finally f) clarification and documentation of the types and extent of past restoration and other human interventions. The presented operational workflow comprises an efficient methodology for the examination of mosaic icons and more generally of panel icons and other artworks of parallel characteristics and complexity, provided that they meet the size limits of the medical CT Scan gantry. It has been demonstrated that combined application and a final combined 3D model resulted fromthe three methods ensures a high level of geometric and color detail and accuracy and provides the full range of information that is required for the non-invasive examination of complex and delicate artworks in order to reach the optimal conservation decisions. The post-processing of the acquired data (for example within applications of finite elements methods), could provide a wide range of options that can be further utilized to guide the implementation of rather challenging operations/interventions which require high precision. Finally, monitoring of the interventions, as well as evaluation of their effectiveness, can be achieved through intermediate and post-treatment investigation. The combined application of Computed Tomography, Photogrammetry and Laser scanning has enhanced the conservator's knowledge and capabilities towards a better understanding of the artifacts' condition and pathology, thus enabling the optimization of conservation approaches and the advancement of the broader field of conservation. Last but not least, the various interpretive schemes facilitate the exchange of information among ICT experts, conservators and stakeholders and its dissemination to a larger audience.

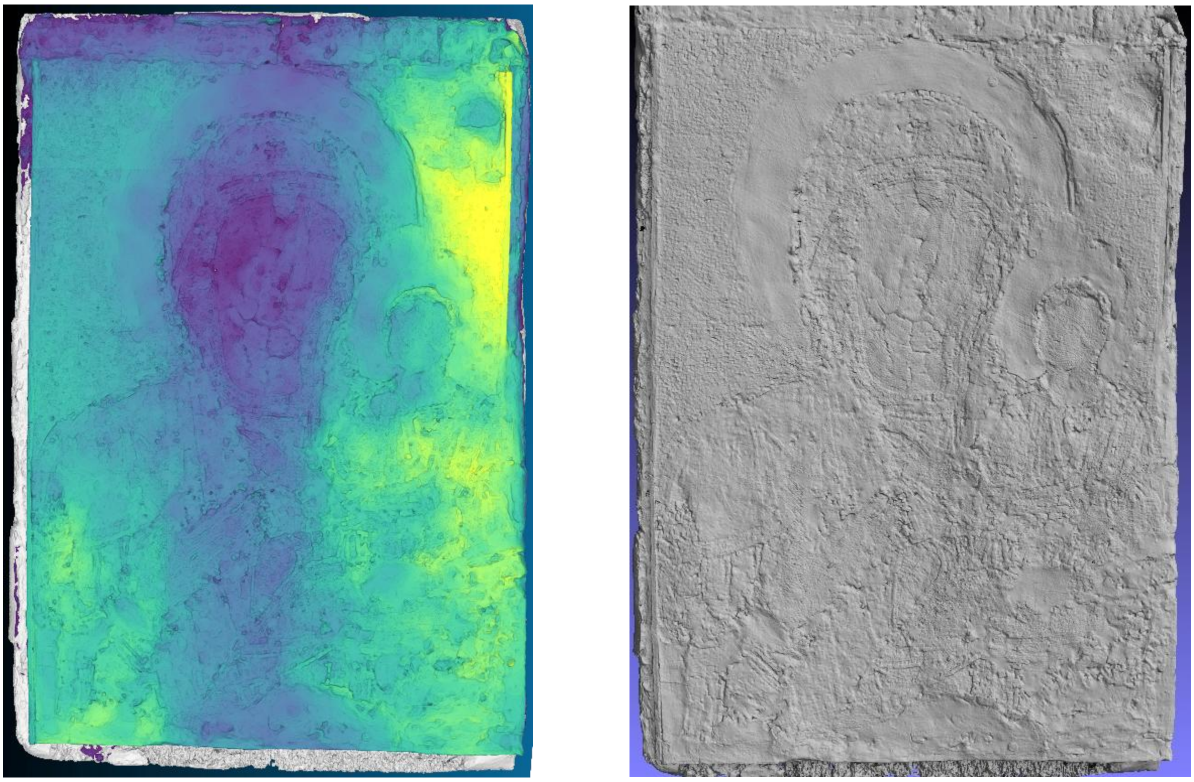


Figure 16. Icon's surface intensifiedby virtual light

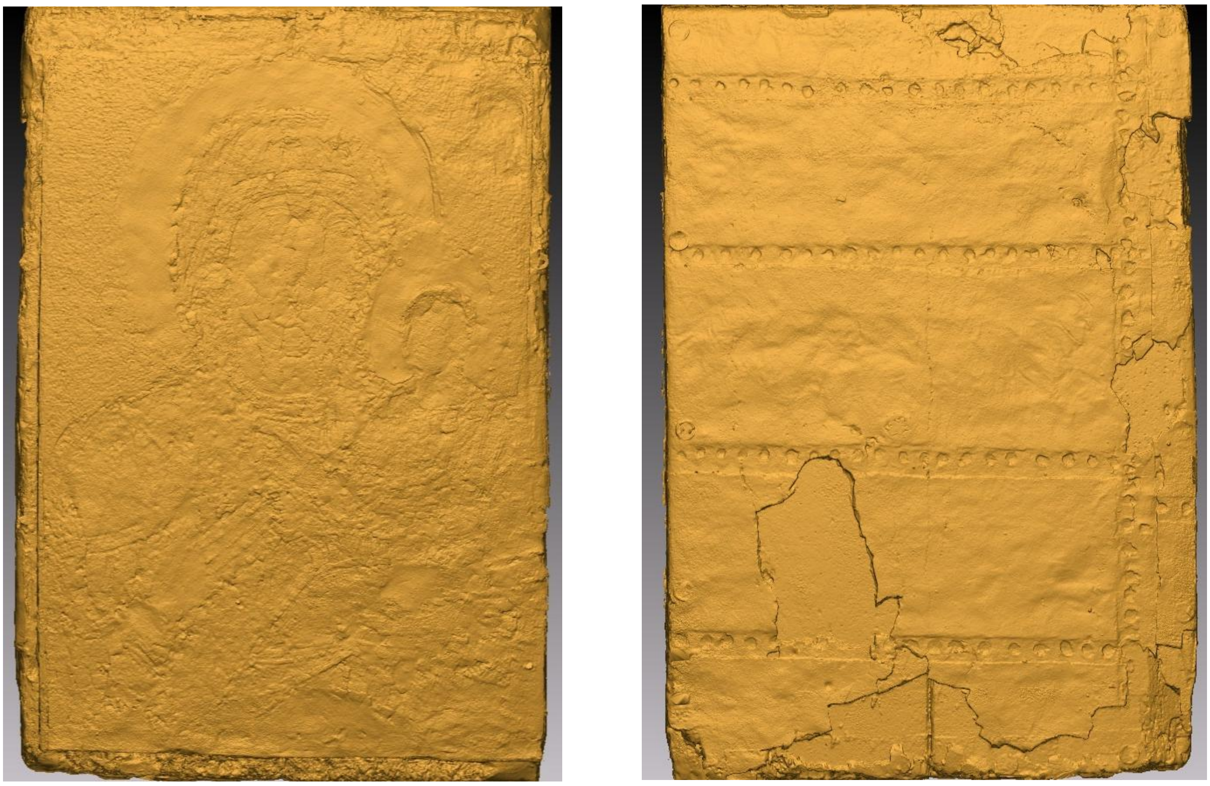


Figure 17. Icon's surface morphology with extended anomalies

Acknowledgments

The present study was undertaken in the framework of the conservation program of the Pammakaristos Icon, upon request and with the permission of His All Holiness the Ecumenical Patriarch Bartholomew. The program is coordinated by Dr. Chlouveraki and is funded by the Ekaterini Laskaridis Foundation with the technical and scientific support of the Dept of Conservation of Antiquities and Works of Art, of the University of West Attica (UniWA). Laser Scanning and post processing of data was undertaken by Dr. Makris with a scanner provided by the Dept of Interior Architecture of UniWA to whom we are thankful. The CT Scan data was obtained and processed by Dr. Akpek and his team at the American Hospital of Istanbul, who's help and advice was invaluable. We are grateful to the anonymous reviewer for their constructive comments and recommendations.

References

Aicardi, I., Chiabrando, F., Maria Lingua, A., Noardo, F., 2018, 'Recent trends in cultural heritage 3D survey: The photogrammetric computer vision approach', *Journal of Cultural Heritage*, 32, pp. 257–266.

Bornaz, L., Botteri, P., Porporato, C., Rinaudo, F., 2007, 'New application field for the handyscan 3D: The Res Gestae DiviAugusti of the Augustus Temple (Ankara), a restoration survey', in Georgopoulos, A. (ed.) *Anticipating the Future of the Cultural Past*, 21st CIPA symposium, 01–06 October, 2007, Athens, Greece, The ISPRS International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XXXVI–5/C53, pp. 182–187.

Chlouveraki, S., 2020, 'The Mosaic Icon of 'Pammakaristos Mother of God': Documentation and Condition Assessment', in Nardi R., &PugesiDorca M. (eds), *What comes to mind when you hear mosaic? Conserving mosaics from ancient to modern*, Proceedings of the 13th Conference of the International Committee for the Conservation of Mosaics, Barcelona 15-20 October 2017, EDIFIR-Edizioni Firenze, pp. 83-97.

Cignoni, P., Callieri, M., Corsini, M., Dellepiane, M., Ganovelli, F., Ranzuglia, G., 2008, 'Meshlab: an open-source mesh processing tool', in Eurographics Italian chapter conference, vol. 2008, pp. 129-136.

Dellepiane, M., Callieri, M., Corsini, M. & Scopigno, R., 2011 'Using digital 3D models for study and restoration of cultural heritage artifacts', in Gallo G., (ed.) *Digital Imaging for cultural heritage preservation: analysis, restoration and reconstruction of ancient works*. Florida: CRC Press. pp. 37–67.

Eros, A., Bornaz, L., 2017, '3D models in cultural heritage: Approaches for their creation and use', *International Journal of Computational Methods in Heritage Science*, Volume 1, Issue 11(1), pp. 1–9. https://doi.org/10.4018/ijcmhs.2017010101.

Georgopoulos, A., Stathopoulou, E.K., 2017, Data acquisition for 3D geometric recording: state of the art and recent innovations. In: Vincent, M.L., López-MencheroBendicho, V.M., Ioannides, M., Levy, T.E. (eds.) Heritage and Archaeology in the Digital Age, pp. 1–26. Springer, Cham. https://doi.org/10.1007/978-3-319-65370-9_1

Hess, M., Korenberg, C., Ward, C., Robson, S., Entwistle, C., 2015, 'Use of 3D laser scanning for monitoring the dimensional stability of a Byzantine ivory panel', *Studies in Conservation* 60:sup1, S126–S133. https ://doi.org/10.1179/0039363015 Z.00000 00002 17.

Ioannidis, C., Piniotis, G., Soile, S., Bourexis, F., Boutsi, A.-M., Chliverou, R., and Tsakiri, M., 2019, laser and multi-image reverse engineering systems for accurate 3d modelling of complex cultural artefacts, Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci., XLII-2/W11, 623–629, https://doi.org/10.5194/isprs-archives-XLII-2-W11-623-2019.

Kersten, T.P., Lindstaedt, M., Starosta, D., 2018, 'Comparative geometrical accuracy investigations of Hand-Held 3D scanning systems - an update', in International Archives of the Photogrammetry. Remote Sens Spatial Inf Sci - ISPRS Arch 42(2) pp. 487–494. https://doi.org/10.5194/isprs-archives-XLII-2-487-2018.

Kondakov, N. P., 1915, 'Iconography of the Mother of God' (volume 2), St. Petersburg (in Russian). Н. П. Кондаков. 1915. ИконографияБогоматери (Том 2). Санкт- Петербурге.

Koutsoudis, A., Vidmar, B., Ioannakis, G., Arnaoutoglou, F., Pavlidis, G., Chamzas, C., 2014, Multi-image 3D reconstruction data evaluation. J. Cult. Herit. 2014, 15, 73–79.
https://doi.org/10.1016/j.culher.2012.12.003

Makris, D., Skaltsas, I., Fotiou, S., Karampinis, L. and Vlachou, M. A., 2018, 'Digitization of Athens School of Fine Arts Artworks Based on Optical 3-D Scanning and Photogrammetry', 9th International Conference on

Information, Intelligence, Systems and Applications (IISA), 2018, pp. 1-7.
doi: 10.1109/IISA.2018.8633603.

Makris, D., Sakellariou, C., Karampinis L., 2021, 'Emerging materiality through dynamic digital conservation', *Digital Applications in Archaeology and Cultural Heritage*, (23).
https://doi.org/10.1016/j.daach.2021.e00198.

Metashape Pro, 2022, *Agisoft*. Available at https://www.agisoft.com/ (Accessed 17 February 2022).

Millas, A., 2013, *Constantine Polis, Orthodoxy within the walls*, (in Greek), Μήλλας, Α. (2013) *ΚωνσταντίνουΠόλις, Ηεντός των τειχών ορθοδοξία*, Εκδόσεις Μίλητος, ISBN:9608033330

Montaina, L., Longo, S., Galotta, G., Tranquilli, G., Saccuman, R., Capuani, S., 2021, 'Assessment of the Panel Support of a Seventeenth-Century Dutch Painting by Clinical Multislice Computed Tomography', *Studies in Conservation*, 66(3), pp. 174-181, 10.1080/00393630.2020.1757881

Morigi, M.P., Casali, F., Bettuzzi, M., Brancaccio, R., D'Errico, V., 2010, 'Application of X-ray Computed Tomography to Cultural Heritage diagnostics', *Applied Physics* A 100, pp. 653–661. 10.1007/s00339-010-5648-6

Palma, G., Pingi, P., Siotto, E., Bellucci, R., Guidi, G., Scopigno, R., 2019, 'Deformation analysis of Leonardo da Vinci's "Adorazione dei Magi" through temporal unrelated 3D digitization', *Journal of Cultural Heritage*, 38, pp. 174–185.

Re, A., Lo Giudice, A., Nervo, M., Buscaglia, P., Luciani, P., Borla, M., Greco, C., 2016, 'The importance of tomography studying wooden artefacts: A comparison with radiography in the case of a coffin lid from ancient Egypt', *International Journal of Conservation Science*, 7, pp. 935–944.

Remondino, F., Spera, M.G., Nocerino, E., Menna, F., Nex, F., 2014, 'State of the art in high density image matching', *The Photogrammetric Record*, 29(146), pp. 144–166.

Robson, S., Bucklow, S., Woodhouse, N., Papadaki, H., 2004, 'Periodic photogrammetric monitoring and surface reconstruction of a historical wood panel painting for restoration purposes', *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 35, pp. 395–400.

Sotiriou, Y., 1933, 'The Icon of Pammakaristos' (in Greek). Σωτηρίου Γ. 1933. 'Ἡ εἰκών τῆς Παμμακαρίστου'. Πρακτικά 97 της Ακαδημίας Αθηνών, (ΠΑΑ) 8, Αθήνα, Ακαδημία Αθηνών. 359-368.

Sotiriou, Y., 1936, 'The Icon of John the Baptist', (in Greek), Σωτηρίου Γ. (1936) 'Ψηφιδωτή είκών του Προδρόμου του πατριαρχικού ναού'. ΠρακτικάτηςΑρχαιολογικήςΕταιρίας (ΠΑΕ) 11. 70-76.

Stratis, J. A., Makarona, C., Lazidou, D., Sánchez, E. G., Koutsoudis, A., Pamplona, M., Pauswein, R., Pavlidis, G., Simon, S., &Tsirliganis, N., 2014, Enhancing the examination workflow for byzantine icons: Implementation of information technology tools in a traditional context. Journal of Cultural Heritage, 15(1), 85–91.

Tsirliganis, N., Pavlidis, G., Koutsoudis, A., Papadopoulou, D., Tsompanopoulos, A., Stavroglou, K., Loukou, Z., &Chamzas, C., 2002, Archiving 3D cultural objects with surface point-wise database information. In Proceedings of first international symposium on 3D data processing visualization and transmission (pp. 766–769). IEEE.

Tsirliganis, N., Pavlidis, G., Koutsoudis, A., Papadopoulou, D., Tsompanopoulos, A., Stavroglou, K., Loukou, Z., &Chamzas, C., 2004, Archiving cultural objects in the 21st century. Journal of Cultural Heritage, 5(4), 379–384.

Zhao, G., Qiu, Z., Shen, J., Deng, Z., Gong, J., and Liu, D., 2018, 'Internal structural imaging of cultural wooden relics based on three-dimensional computed tomography', *BioResources* 13(1), pp. 1548-1562.

RURAL ECONOMY AND SOCIETY IN EARLY MODERN CYPRUS (RURAL-CY): AN INTRODUCTION TO PROJECT OBJECTIVES AND TECHNOLOGICAL INFRASTRUCTURES

CH. PARASKEVA, P. HADJITTOFI, E. RIZOPOULOU-EGOUMENIDOU AND A. VIONIS

Archaeological Research Unit, University of Cyprus,

nethphp@hotmail.com, phadji13@ucy.ac.cy, aregoume@ucy.ac.cy, vionis@ucy.ac.cy

Abstract

Following the Ottoman conquest of 1571, the Christian Orthodox Church of Cyprus became a major political, economic, and social actor on the island that sought to consolidate and expand its power by undertaking various agriculture-oriented income generating activities, which in turn embedded it in wide economic networks spanning the island and beyond. Those activities are documented in several textual sources, including Church property lists, which are the focal point of the RURAL-CY project, an internal research project of the Archaeological Research Unit of the University of Cyprus. Specifically, the project aims to explore the dynamics, strategies, and patterns of rural-based economic activities by fully digitising, systematically recording, and statistically-spatially analysing the Grand Manorial Codex and other auxiliary texts, which record in detail the movable and immovable property of ecclesiastical institutions located mainly in rural areas across Cyprus. This paper aims to introduce the reader to the project's objectives and outline the novel technological infrastructures developed as part of it, which include an archaeological information management system and an end-to-end workflow that allows the transfer of data between the data repository and third-party software packages for spatial-statistical analyses.

Keywords*: Rural economy, Cyprus, textual sources, MySQL database, information management system*

1. Introduction

In-depth documentation and examination of material culture and, consequently, the study of social, economic, and other relations, practices and perceptions traced through artefacts, have emerged in recent decades as imperative issues and major *desiderata* of historical and archaeological research focusing on the Eastern Mediterranean during the Ottoman period. As scholarship on post-Medieval material studies of this particular region has already acknowledged (see Silberman 1989; Baram and Carroll 2002, 3-10; Vionis 2016; Bintliff 2007, 221; Vorderstrasse 2014, 292-293; Yenişehirlioğlu 2015, 361), the negative perception and unpopular legacy of the Muslim Ottoman Empire, the rise of nationalistic political movements and the formation of national identities in the 19th and early 20th centuries, the consideration of Ottoman period remains as too modern or otherwise unworthy of being recorded, and the ideologically motivated direction of research towards certain past eras (e.g., the Biblical past or Iron Age kingdoms or even the Ottoman classical period [16th-17th centuries] versus the modernisation period), hindered the development of Ottoman and post-Ottoman material culture studies at least up to the 1980s. Correspondingly, in Cyprus, owing in large part to the current political situation, the historiographical and anthropological discourse on the Early Modern and Modern eras mainly revolved around the thematic axis of conflicts, either between the larger communities of the island (i.e., Greek- and Turkish-Cypriots; see Papadakis *et al.* 2006; Given and Hadjianastasis 2010, 38), or among the Greek-Cypriot community and the Ottomans or the British. Thus, scholarship has only recently started to explore the potential of macro-scale approaches to post-Medieval material culture, economic history, and landscape (see for example: Hadjikyriacou 2016; Hadjikyriacou *et al.* 2021).

Given the need for comprehensive, wide-reaching studies based on combined methodological approaches to the Ottoman material culture of Cyprus, the project *Rural Life In Cyprus (RURAL-CY): Digitising and Reconstructing Material Culture through Monastic Property Lists (18th-19th centuries)* was conceived, developed and implemented as an interdisciplinary project, exploring written records analysis to shed light on issues of daily life, society, and economy in Ottoman Cyprus. The project entailed the digitisation, processing, and interpretive analysis of material culture data contained in Cypriot church records spanning the late 18th and 19th centuries. The analysis built on the digitisation and interpretation of written accounts with the aid of palaeography and historical approaches, material culture and folk-life studies, landscape archaeology, and information and communication technology.

Ecclesiastical written accounts constitute a large body of extant evidence, crucially important for the study of Ottoman Cyprus. After the 1571 Ottoman conquest, the Christian Orthodox Church of Cyprus developed into one of the main economic, political, and social actors on the island. With a view to ensure its autonomy, bolster its economic position, and expand its power, the Church undertook a variety of income-generating activities, thus becoming involved in wide economic networks (on the status, the role, and the economic activities of the Cypriot Church, the latter examined through the case of Kykkos Monastery, see Μιχαήλ 2005a; Roudoumetof and Michael 2010; Michael 2021, 245-251). As agriculture remained the backbone of the Cypriot economy throughout the Ottoman period, land acquisition and the production of marketable agricultural products (wine, silk, cotton, olive oil, and wheat) was the focal point in the economic endeavours of ecclesiastical institutions. Moreover, extant written sources show that churches, monasteries, and *metochia* (dependencies) were hubs of extensive, multi-faceted economic activities, which among others also included animal husbandry, sericulture, beekeeping, the operation of owned olive presses, winepresses, and flourmills, as well as the practicing of various crafts. Ecclesiastical institutions functioned as agents of development in their areas of economic operations, interacted with the local populations in a closely-knit network of economic, social, and religious relations, and even influenced the character and use of the Cypriot countryside landscape. Through their activities, material culture, and general way of life, the church institutions of the Ottoman period became inextricably linked to the pre-industrial rural communities of Cyprus and their particular lifestyle.

The publication of ecclesiastical property lists and other written accounts in previous works focusing primarily on Church History (see mainly Μιχαήλ 2001; Σταυρίδης 2001; Μιχαήλ 2003; Θεοχαρίδης 2004; Μιχαήλ 2005; Παπαδόπουλλος 2008), had already made evident their potential for the documentation of daily-life material culture and rural economy; yet no such records were systematically explored towards this end. The RURAL-CY project undertook the detailed study of Cypriot ecclesiastical written records to reconstruct aspects of rural life in Cyprus, document a wide range of activities that comprise a microcosm and reflection of rural life across the island; trace the productive potentiality and use of agricultural lands belonging to the Church; combinedly examine the terminology, form and use of different kinds of domestic utensils, agricultural implements and craft tools most of which remaining in use up to the 20th century; record locally-produced craft objects versus imported goods; and explore dietary elements and food preferences.

Analysis conducted in the framework of the project was based on the Grand Manorial Codex, a largely unpublished document of 1,188 numbered pages that dates principally to the era of Archbishop Chrysanthos (1767-1810),[1] also known as the 'golden age' in the history of the Church of Cyprus (see Stavrides 2013). The Codex records in detail the property of 212 ecclesiastical institutions (churches and monasteries with *metochia*) predominantly located in rural areas across Cyprus and falling under the authority of the Holy Archbishopric. A significant part of the registered property lists is dedicated to landholdings (including types of cultivations and topography of estates), productive animals (mainly sheep and goats) and animals for draught, transport, and work in the fields (donkeys, horses, camels, and oxen), wine- and olive-presses, watermills, and other built elements (houses, stockyards, water tanks etc.). The 'household material culture' (*τα εν τοις οσπητίοις*, comprising agricultural and craft tools, furniture, vessels, installations, husbandry equipment, agricultural products etc.) forms a separate sub-category of the listed properties that is also closely related to rural economy.

The study of the Grand Manorial Codex, as a primary source of data, entailed on the one hand the digitisation, selection, segmentation, transcription, interpretation, and analysis of data related to numerous categories of material culture, and the post-collection processing of digital data via quantitative and qualitative statistical and spatial analyses, as well as the visualisation of mappable data resulting from text processing, such as the location of monasteries and estates, as well as the distribution of agricultural land and pre-industrial facilities. The project also fused

[1] Later additions, as well as supplementary notes to original records, date up to the mid- or, more rarely, the late 19th century.

supplementary data located in relevant bibliography, as well as in unpublished Codices from the Paphos and Larnaca Bishoprics Archives. Finally, especially for the study of dietary elements, the project also analysed day-to-day food purchases logs of the Holy Archbishopric during the years 1867-1869 and 1870-1872, contained in Codices 130 and 131 of the Holy Archbishopric Archive, respectively.

2. Technological Infrastructures

Prior to the commencement of data collection, it was considered necessary to establish an end-to-end project workflow, and to determine the functional requirements for the technological infrastructures that would manage the digital data input, editing, management, and analysis. To begin with, the adoption of any technical solution required that the latter would *de minimis* be able to also work in offline situations, allow data transfer between the database and third-party software designed to work in Win32/Win64 environment subsystems, conform fully or partially to an established ontology (e.g., Doerr *et al.* 2017), have the capacity to split the entirety of the digital textual content of any written source in segments determined by the user; perform transcription, transliteration, translation, and multivocal analysis of each text segment; tolerate expansion of the core schema to support new fields of enquiry as deemed necessary, and offer an accessible graphical user interface that would require minimal maintenance from the researcher performing data input.

Bearing in mind the project requirements, a wide literature review was undertaken with a view to locate suitable technical solutions that could be adopted, if necessary adapted, and implemented as part of the RURAL-CY project. Specifically, the technical solutions examined included at their core a data storage facility in the form of a graph, relational, cloud, object-oriented, network or NoSQL database coupled with an online or offline composite user interface or an online virtual research environment. Added to the core components, many applications and services were accompanied by extra modules for performing automated transcription via OCR and occasionally machine learning assisted OCR, data mining, named entity recognition usually by utilising natural language processing algorithms, data segmentation (sentence splitting, part-of-speech tagging, tokenisation), data annotation and augmentation (e.g., place, entity, event linking), morphological (stemming, lemmatisation) and syntax analysis (chunking, parsing), metadata creation, curation, management, cross-referencing, and enrichment; geolocation and georeferencing of data, and cross-source data querying and linking. Most solutions incorporated more than one of the above components, and were developed as part of numerous research projects, including Recogito/Pelagios (Simon *et al.* 2017), CLARIN-D (CMDI Explorer: Arnold *et al.* 2021; WebLicht: Hinrichs *et al.* 2010; Dima *et al.* 2012; WebAnno: de Castilho *et al.* 2015;de Castilho *et al.* 2016), DARIAH-DE/TextGrid (TextGrid Consortium 2006-2014), EpiDoc (Bodard 2010; Elliott *et al.* 2006-2021), Manuscriptorium/ENRICH (Uhlířand Knoll 2009; Tîrziman 2013), STAR/STELLAR (Richards *et al.* 2015), Interedition/CollateX (Dekker *et al.* 2015), LDAB-Trismegistos (Depauw and Gheldof 2014), LaQuAT (Blanke *et al.* 2010), CAMENA (Schibel 2001; Niehl 2003), eAQUA (Schubert 2010; Bünte 2011; Schubert 2015), VRE-SDM (Bowman *et al.* 2010; see also: Tarte *et al.* 2009), NEUMES (Barton 2002), DBBE (De Groot 2020), ByzAD (Bender *et al.* 2007-2022), MTA Software(Valovič *et al.* [2022]), ArcheoBERTje (Brandsen 2022), Fragmentarium (Duba 2019),AGNES (Brandsen 2022), as well as by individual researchers (Murrieta-Flores and Gregory 2015; Kinable2018; Schulz 2018;Chen *et al.* 2019; Liceras-Garrido *et al.* 2019; Schulzand Ketschik 2019).

A key conclusion of the literature review was that the analysis of written records has been the object of numerous projects in recent years, which have rendered a vast corpus of data available to a much wider audience, whilst also making significant technological contributions to the fields of Digital Humanities, Digital Classics, Text-Aided Archaeology, and Historical Archaeology (for more in-depth reviews see: Babeu 2011;Rutz and Kersel 2014; Sula and Hill 2019;van Lit 2019; Liu*et al.* 2021; Moudgil*et al.* 2022). However, with regard to the minimum functional requirements outlined for the RURAL-CY project, it was not possible to settle on one or more solutions that would satisfy them, as certain promising solutions lacked support due to termination of the project developing them (on this issue see also: Dombrowski 2014), while in other cases the main obstacle to adoption was the lack of interoperability between different solutions that also required different and, in several cases, costly hardware for use in offline situations. Added to the above, many of the software-as-a-service solutions developed by large-scale supra-national projects (e.g., CLARIN-D, Recogito, TextGrid) require constant online access and seem to aim principally at high volume aggregation and/or digitisation of written sources, rather than the in-depth complex documentation and interpretation/analysis of source contents. Finally, the annotation tools available in many utilities do allow a researcher to tag and comment on the texts, and in certain cases to also link the text to a named entity, but they do not produce or demand extra investment of time to create the necessary associations between annotations that would render them useful for statistical and spatial analyses. Given the above, it was decided to produce a custom solution for the RURAL-CY project, which involved the development of an archaeological information management system and a workflow that ensured transfer of data between the data repository and the software used for analysis.

2.1 Archaeological Information Management System

Commencing the design of a custom system dedicated to the in-depth analysis of written sources and bearing in mind the intricacies of language, as well as the immense variability of the data included in the texts under analysis, the principal issue requiring resolution concerned the logical schema of the underlying database. The schema needed on the one hand to satisfy the prerequisite for a formal structure required by domain ontologies (Doerr*et al.* 2017;Barzaghi*et al.* 2020), text and material culture analysis guidelines and best practices(Baca *et al.* 2006; TEI Consortium 2021), relational database design principles (Lemahieu *et al.* 2018; Vanier *et al.* 2019), and quantitative analysis prerequisites; and on the other hand to afford the researcher flexible mechanisms to capture and record information that escapes the narrow definitions of structured data and/or is necessary to manage researcher biases and preserve/represent the various forms of ambiguity that are inexorably associated with the complexities of the written word (e.g., lexical, syntactic, semantic, ontological, spatiotemporal; for a review of structured data issues see: Hacıgüzeller*et al.* 2021).As rigidity was inevitable for ensuring referential integrity of the database tables and by extension of the relations between data in the system, it was opted to infuse the latter with flexibility by employing a set of mechanisms hereafter discussed:

a. Allowing the researcher to define the least common denominator of analysis, identified in the system as a data unit. This is understood as the minimum ethically interpreted as complete set of information that captures a specific intelligible meaning, as discerned and decided by the human user of the system. In practice, a data unit can be equated with a variable text string ranging from one word to any number of words (e.g., string of words, phrase(s), sentence(s), etc.) satisfying the meaning completeness criterium outlined above. A data unit may float anywhere within a written source, which borrowing a term from Formal Concept Analysis is viewed as a conceptual landscape (Wille 1999; Fig. 1). In practice, the data unit forms the core concept of the system and is the hub between each written source and all the data associated with it.

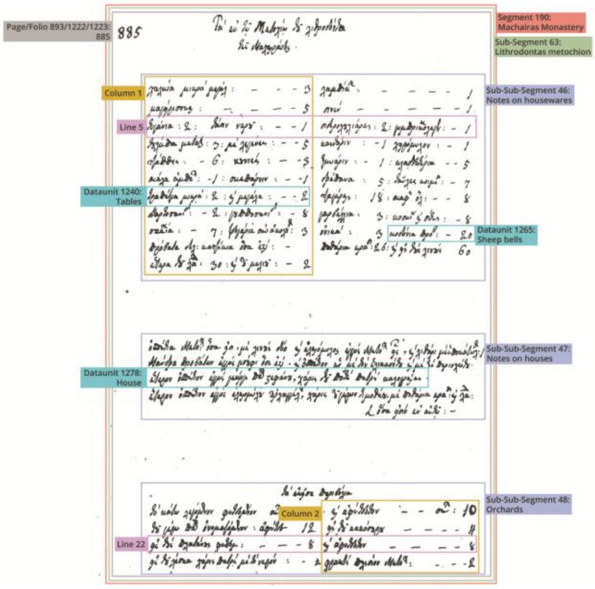


Figure 1. Example of a page from the Grand Manorial Codex with marked text segments. Conceptual/ imaginary spaces and nested subspaces are depicted on the right-hand side, while physical/measurable spaces are on the left. Examples of data units indicated by the aqua label. Photograph: ©Petroula Hadjittofi; Illustration: Charalambos Paraskeva.

b. Disentangling physical from conceptual spaces in the definition of text segments within each written source. If a written source is perceived as a conceptual landscape per the afore definition, it follows that any such source may be subdivided topologically into logically cognisable spaces that are in turn multiply related to each determined data unit. Such topological divisions include both conceptual/imaginary spaces (e.g., foreword, chapters, epilogue, sections, other book parts, etc.), and physical/measurable spaces (e.g., folia or pages, columns, lines, margins, etc.). Per the definitions above, the system was designed to allow both the segmentation of the written source in physical and conceptual spaces and nested spaces, as well as the recording of multiple relations between segments and data units (Fig. 1). At this point, it is recognised that any attempt at segmenting a written source and discerning data units for analysis effectively produces structured data that function as representations ossifying and iterating the researcher's predilections, biases, and perceptions (Hacıgüzeller *et al.* 2021: 1722-1723). However, in this case, the system does not lock the researcher into a single path of structuring data, as the existence of multiple segmentations and data units is established by design, which in turn promotes interpretative nuance, multivocality, and the diversification of our collective understanding of the written record.

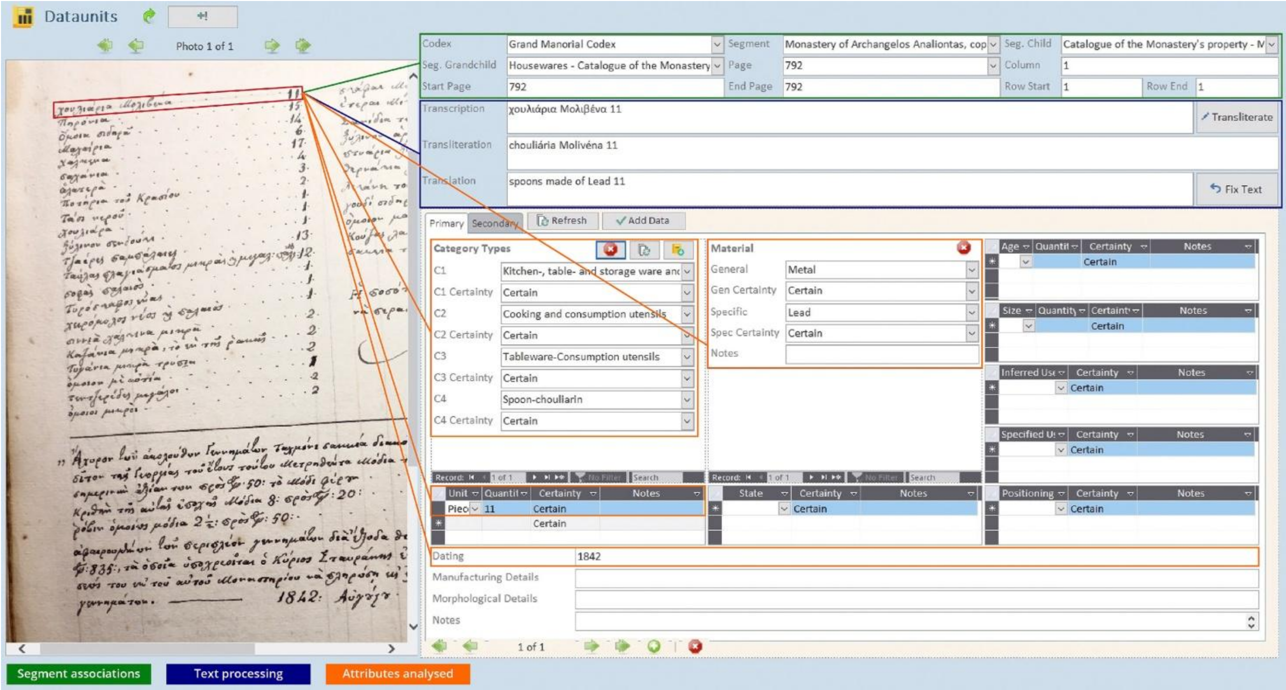


Figure 2: Screenshot of a data unit in the RURAL-CY archaeological information management system. Each data unit is multiply associated with text segments, a text processing facility, and numerous attributes. Illustration: Charalambos Paraskeva.

c. Incorporating devices for knowledge extension and assessment at each data entry deemed the result of interpretation. Given that the system was designed primarily for text data interpretation and analysis, the group of researchers and advisers - based on prior knowledge of the material culture categories and written sources under analysis- provided from the beginning a set of attributes representing analytical concepts that were expected to allow extraction of quantifiable data for the material culture entities described or mentioned in the written sources. Specifically, twenty attributes were modelled including category/type with four levels of nested subcategories, general and specific material, colour/shade/hue, relative/qualitative size, quantity by unit of measurement, reported and inferred use, specific entity provenance as described in the source, category/type provenance, relative/qualitative age, absolute or relative dating, condition/state, manufacturing method, morphological details, positioning in relation to another entity or the wider built environment, location at specific toponyms, location at specific villages, mode of acquisition by its current owner, and type of owner with owner name and associated rights if available. Nearly all the above attributes were implemented as drop-down lists carrying vocabulary terms, where each term was enriched with additional data from bibliographic sources (Fig. 2). It is worth noting that the system is not restrictive in terms of the attributes, as its structure can accommodate infinite additions and modifications, as deemed necessary, while all attributes are considered to be *de facto* related/complementary due to their association with the same data unit.[2] To return to the beginning of the discussion, though, recognising that each of these attributes was to be populated with data by a researcher who would essentially be offering their interpretation/view of the text, required the addition of mechanisms to reduce bias and increase transparency and nuance. Regarding bias, a self-assessment/reflexive mechanism was incorporated that required the researcher to state their

[2] It would bepossible to add a further mechanism indicating-clarifying the specific relationships between attributes, but for the sake of expediency this was not implemented in the current system.

level of certainty/confidence for any new data inputted to any of these attributes. Though qualitative, this evaluation can later be fruitfully incorporated in statistical analysis either by using weighs or fuzzy statistics to reach more insightful inferences. Turning to transparency and nuance, it was *ab initio* realised that a drop-down list of terms would be restrictive and reductionist, whereas the text could potentially bear more contextual information and carry subtle details that a structured list would effectively mask. To avoid leaving behind knowledge, an extra free-text field was associated to each attribute list, where the researcher would be able to add any comments, observations, or extra information they felt necessary to accompany the specific data entry. Beyond enriching the interpretation record, each of these commentaries can also render more nuanced any type of analysis by adding explanatory notes regarding any exclusion of data (e.g., deviation/exception, non-commensurability, incompleteness, etc.), and any deemed necessary homogenisation of the data pool.

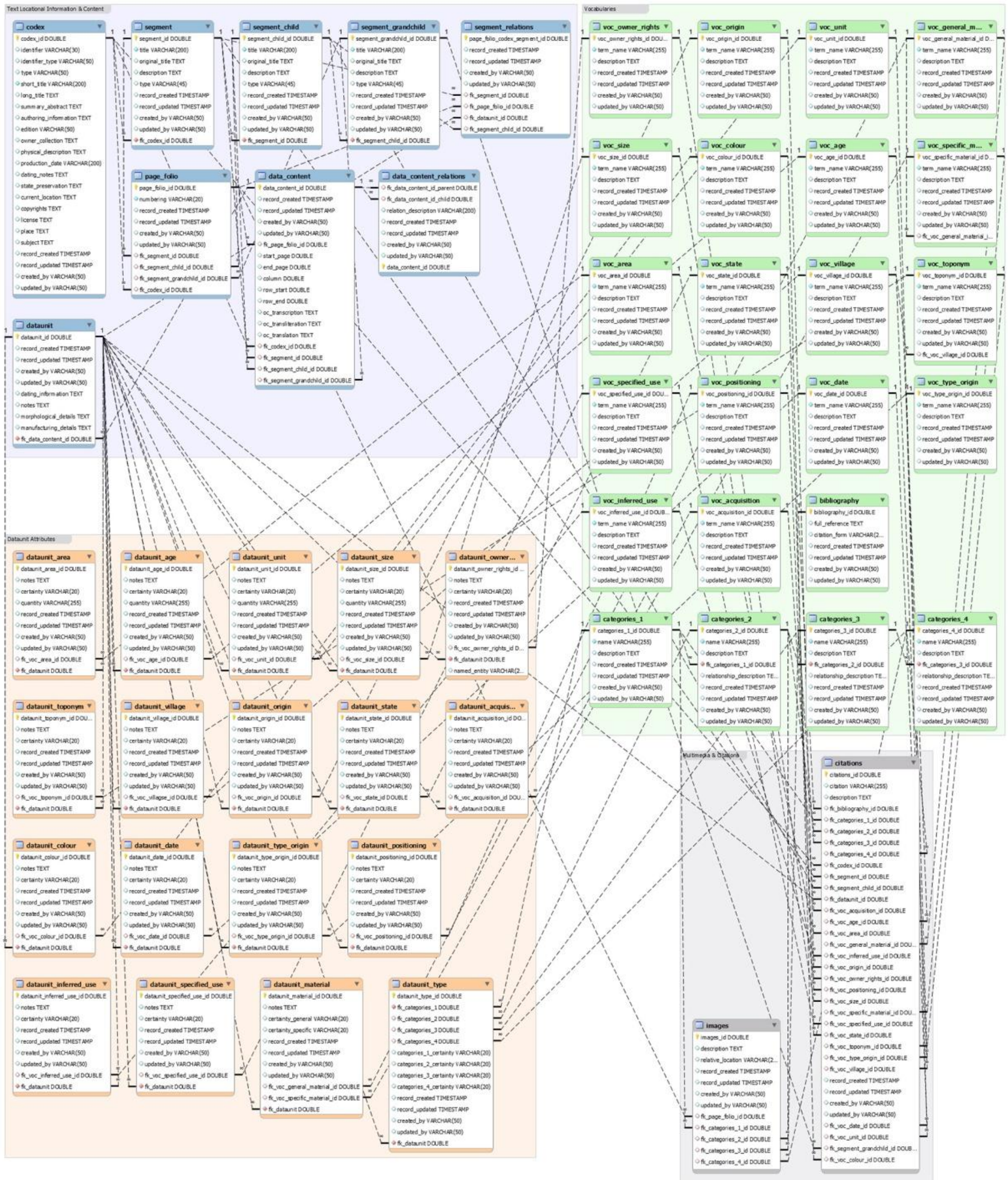


Figure 3: Schema for the RURAL-CY database. Vocabulary tables in green, text segmentation tables in blue, attributes analysis in orange, auxiliary tables for citations and multimedia attachment in gray. Illustration: Charalambos Paraskeva.

d. Ensuring that each data unit and attribute analysed support multiple entries to accommodate multivocality and ambiguities. Analysis of any written source is an extraordinarily complex process fraught with issues that inadvertently render plausible multiple interpretations of the same text. To avoid limiting the system's end-user to one interpretation, it was elected to structure anything that is a matter of interpretation with a one-to-many or many-to-many relationship to the original text (Fig. 3).

Beyond the schema of the system, another major issue during the design phase concerned metadata both for the text under analysis, and any multimedia associated with it. After consideration of select metadata standards (Cowles *et al.* 2007; Catepano *et al.* 2019; DCMI Usage Board 2020), and bibliography on metadata management (Sicilia 2014; Haynes 2018), it was elected to model metadata for the written sources table after the Dublin Core standard (DCMI Usage Board 2020), provide a stored procedure mechanism that timestamps and records the active researcher's name for every data input or update in all tables, and finally introduce a class module for the automatic appending of all metadata associated with a linked multimedia file to its description. The latter mechanism creates native code x86COM objects in memory similar to a shell property handler DLL, while VBA code then extracts metadata, selects those conforming to the MPEG7-Part 5 ISO/IEC 15938-5:2003 standard (International Organization for Standardization 2003; Martínez *et al.* 2002), structures them in a human-readable form, and appends them in the linked object's description. Incorporation of these mechanisms in the system ensures that the written sources are properly documented, any part of the analysis is attributable to the researcher performing the analysis, and any metadata associated with a linked multimedia file are also contained in its database record.

Having elucidated design issues and the data schema, development of the archaeological information management system was initiated utilising MySQL 8 for the backend and Microsoft Access 365 for the frontend. To begin with, the use of MySQL for the backend was a deliberate choice, as beyond its advantages as a relational database (e.g., high performance, security, scalability, comprehensive transaction support, open source),it will also enable reuse of the project's database as the underlying resource for a prospective website. As regards the frontend, Microsoft Access 365 was preferred, as it offers direct connectivity to the MySQL backend via the Open Database Connectivity open standard API, as well as a host of software design tools that permit easy development of the graphical user interface and enhancement of its functions (e.g., automations, error trapping, data validation, multi-user support) via macros and custom code compiled in the underlying Visual Basic for Applications.

Turning to the backend, tables were assembled using the InnoDB storage engine with a view to ensure data security and sustainability. A total of fifty-two fully normalised in accord to the Fourth Normal Form rules and multiply inter-related tables constitute the backend. Roughly based on their functions, the backend tables may be divided in four general groups (Fig. 3), namely text segmentation and text content storage (9 tables), vocabularies (23 tables), attributes analysis (18 tables), and citations and multimedia storage (2 tables). All tables use the utf8mb4 Character Set to ensure that any potential special characters outside the Basic Multilingual Plane can be introduced into the database, while all foreign keys are properly indexed to ensure faster performance of the system. Also, all data in the backend tables are encrypted using the AES256 algorithm when in storage, and with the SSL/TLS cryptographic protocol when in transit to and from the graphical user interface.

The latter was compiled in Microsoft Access 365 32-bit and beyond the 52 linked via ODBC backend tables, it consists of 131 queries and 74 forms, while several of its functions are supported by hundreds of macros and about 7600 lines of VBA code. Beyond the expected CRUD (Create, Read, Update, Delete) actions, the GUI features several functions to assist the user in introducing and managing data. For instance, it allows navigation to all forms via a dashboard that launches after user login, supports data discovery using a variety of filters, lists and custom thematic search forms (Fig. 4); includes several auxiliary forms functioning as vocabulary editors for the terms used for attribute analysis, facilitates linking of multimedia and citations with notes to any vocabulary term or data unit (Fig. 5); and displays a preview of the text source being processed with zooming ability natively (Fig. 2). Regarding user assistance and error management, the interface offers in-form tooltip explainers on how to complete each field, records the user's name and timestamps new entries and the latest update for any data unit or vocabulary term, while automated data validation and error trapping mechanisms silently operate in the background. One final notable feature of the GUI concerns automation, as a module was compiled that performs automatic transliteration of the transcribed Greek text into Latin characters based on the ELOT 743 standard (Ελληνικός Οργανισμός Τυποποίησης 2001; for an example see Fig. 2). The latter has proven a very efficient tool that drastically reduces transcription time.

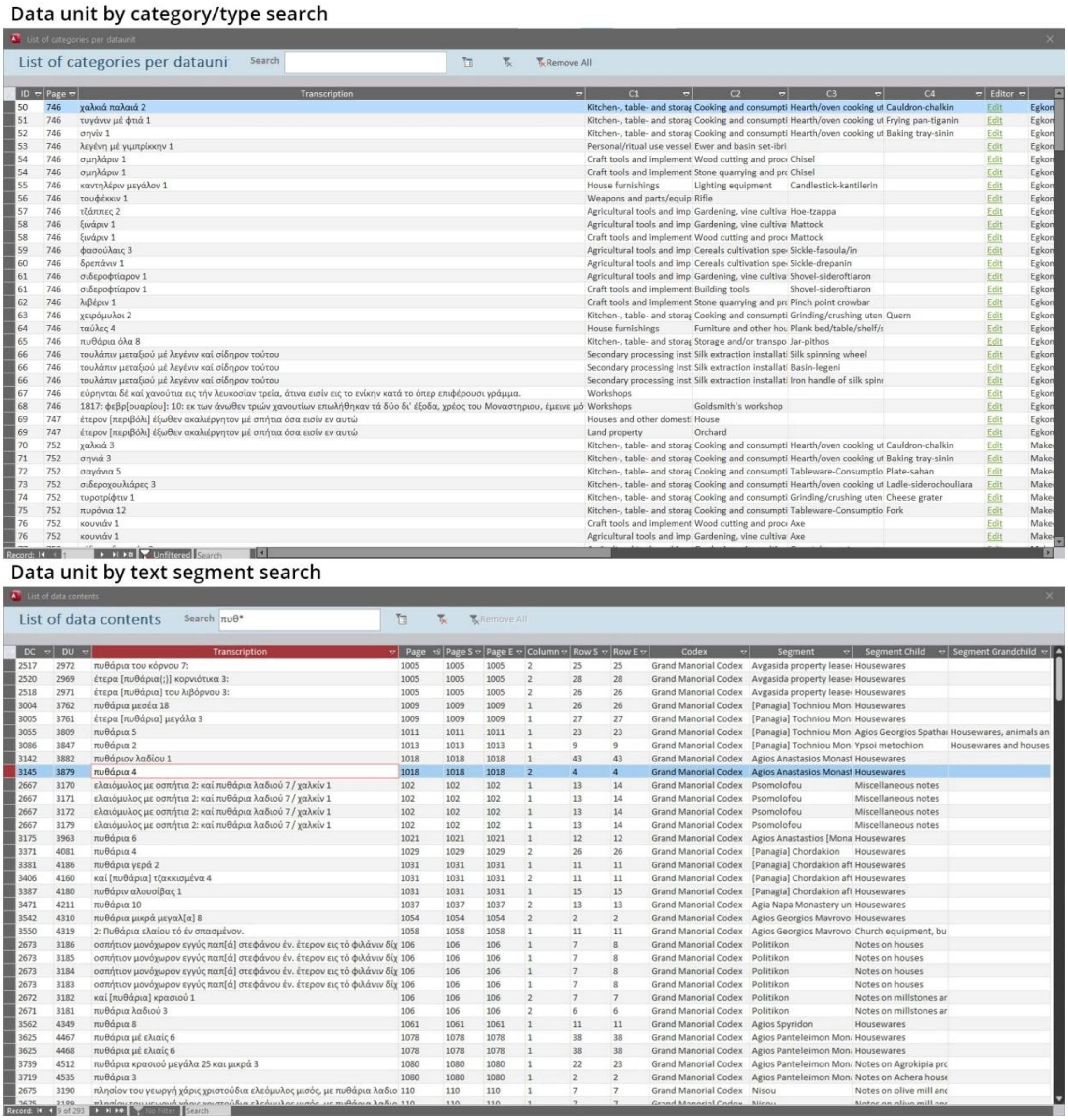


Figure 4. Two examples of custom thematic search forms developed as part of the RURAL-CY graphical user interface. The upper form is used to search for specific data units based on their material culture typology, while the lower form reveals the text segment associations for all the data units. Each column in the forms can be sorted based on content and queried using filtering functionalities of the software (e.g., the lower form is searching for the term "πυθ*" for pithoi in all columns). Illustration: Charalambos Paraskeva.

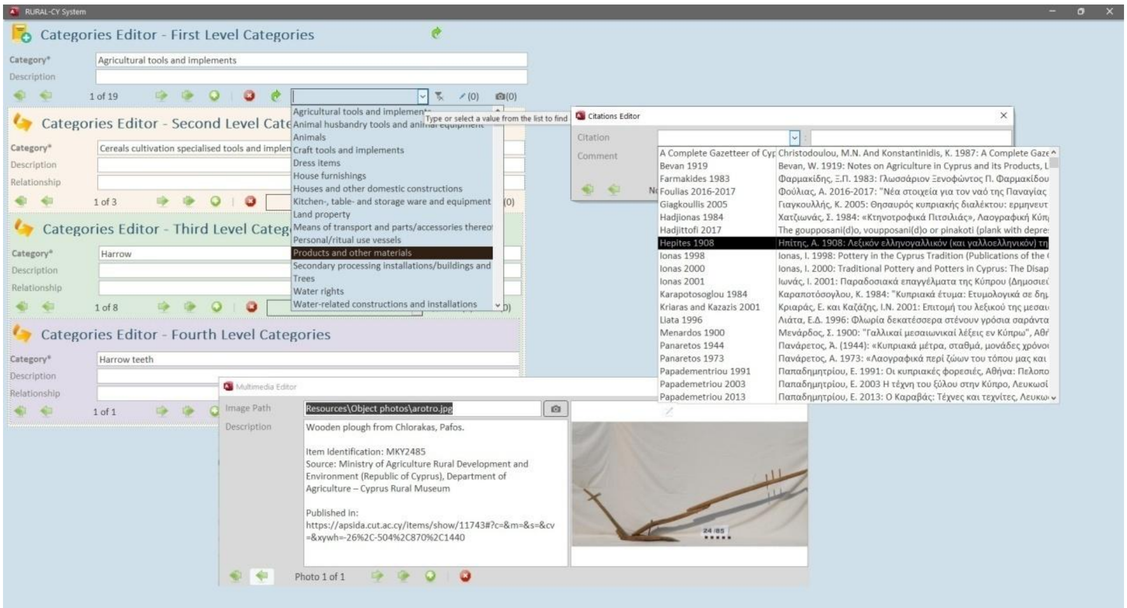


Figure 5. Auxiliary form/vocabulary editor for the category/type attribute. The form supports input of terms with descriptions for up to four nested levels, description of the relationship between the term and its parent term/category, form filtering by term name, introduction of multimedia files with descriptions carrying both user text and appended metadata, and addition of citations with commentary. Illustration: Charalambos Paraskeva.

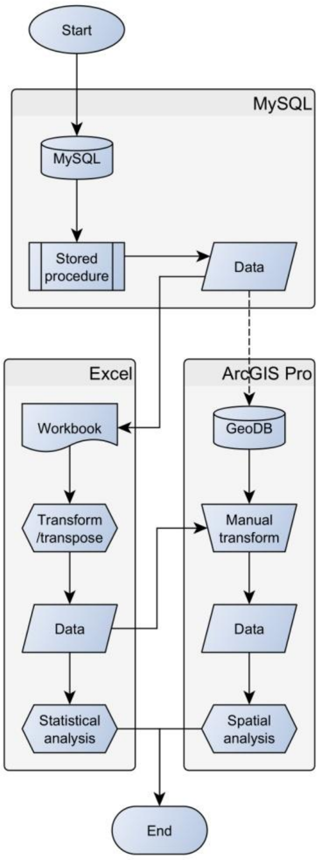


Figure 6: Flowchart of the end-to-end workflow adopted by the RURAL-CY project. Groups indicate software packages used for analysis, while arrows indicate the movement of data. Illustration: Charalambos Paraskeva.

2.2 Geodatabase and Data Processing Workflow

Moving past the archaeological information management system, a second technical infrastructure developed concerns the end-to-end workflow developed for the data transfer between the MySQL backend and statistical-spatial analysis software. In summary, the workflow facilitates the compilation of data in MySQL, data transfer to Microsoft Access 365, processing, structuring, and analysis of data in Microsoft Excel 365, and, linking of data in ArcGIS Pro (Fig. 6).

Specifically, the workflow commences with a complex query that is saved in the MySQL backend database as a stored procedure. When executed, the query links all the data units with their associated text segments and the content of their attribute analysis. Since the aim is to produce a master table with all the data, all the one-to-many and many-to-many relationships between the data unit table and other tables are represented as multiple rows of data. Whenever necessary, this stored procedure can be called in Microsoft Access 365 via a button on the Dashboard of the graphical user interface, which executes VBA code and populates a pass-through query in the background, effectively transferring the totality of data from MySQL to Access. Following this, a System DSN for the Microsoft Access 365 frontend was created utilising ODBC Data Sources to let third-party software packages access the data stored in the abovementioned query. Using this mechanism, the query populates a table in Microsoft Excel 365. Several intermediary tables, pivot tables, and dashboards referring to or mining data from the linked master table, transpose, transform, and order the assembled data into suitable structures either for statistical analysis within Excel and other statistical analysis software packages supporting spreadsheets, or to prepare them for analysis in GIS mapping software.

Turning to the spatial analysis section of the workflow, a geodatabase was developed in ArcGIS Pro to support spatial analysis of the collected data from the written sources. This geodatabase stores layers mapping and to the extent possible geolocating any textual data referring to identifiable and extant spatial entities, such as villages, monasteries, churches, metochia, industrial installations, and toponyms. The latter geodata was compiled by visually comparing the locational data referred to in the text sources (i.e., toponym and village names) to georeferenceable data found in the 1:5000 Cadastral and Topographic Maps of Cyprus that date to the early to mid-20th century AD.[3] Though a long and challenging task, the first case study attempting to geolocate the spatial data referred to in the Grand Manorial Codex for the Machairas monastery and its dependencies proved highly successful, as it was possible to geolocate with confidence about 70% of all locations mentioned in the source text (Fig. 7). After compilation of the georeferenced data for the historical toponyms, the pre-processed Excel worksheets carrying suitably structured attribute analysis data were imported in ArcGIS Pro and using table joins enhance the geodata and facilitate both the visualisation and spatial analysis of the collected attribute data. Finally, it is worth noting that the georeferenced historical toponyms were further enriched using spatial joins to other geodata layers (e.g., hydrology, geology, soil maps, etc.), which can potentially lead to an inverse feedback loop and supplement the text analysis with new relevant geoinformation.

To sum up, an end-to-end workflow for linking data from the backend MySQL database to statistical and spatial analysis third-party software was developed as an integral part of the RURAL-CY project. The workflow has been evaluated and proven effective, as even though it demands dedication of time for the initial setup, it allows the non-expert users of the system to coalesce all the data and transfer them to other software packages for analysis.

3. Conclusions

This paper has presented the objectives, technological infrastructure and workflow adopted and maintained by the RURAL-CY project. The principal and novel infrastructure developed is the archaeological information management system that was designed with a view to assist the rapid and efficient input, segmentation, analysis, management, and output of data from a complex set of written sources. Given the challenging nature of the sources, the main concern has been to create a structure that is both rigid enough for data quantification and spatial-statistical analysis, and at the same time flexible and fluid to accommodate ambiguity, partial knowledge, and researcher bias. It is acknowledged that the system developed has only been tested using a very specific type of source that is the focus of the project, namely Church property lists, but it is contented that due to the openness and expandability of its structure, it can accommodate the addition of any new attribute/concept required for analysis of any type of text.

In terms of project progress and the future, RURAL-CY has digitised all the written sources of the project via digital photography; transcribed in Greek, transliterated and translated in English the text of the Grand Manorial Codex, extracted and analysed in the custom-designed system all the data concerning material culture found in the written sources, and enriched the vocabularies supporting these data with additional information from unpublished and published primary and secondary sources, such as codices, traveller texts, and other relevant bibliography. After

[3]Toponyms in the written sources are always listed under a village, metochion, estate or anotheradministrative unit, while the properties associated with toponyms almost exclusively concern land cultivation (e.g., orchards, vineyards, fields, etc.). Thus, to identify these historical toponyms on the Cadastral and Topographic maps, the larger administrative units, were considered as places of origin. Given their association with agricultural activities, an assumption for one round rip per outing and an estimate for a maximum 4-6 hours of donkey travel per day covering 24-30 km (Boas-Vedder 1985: 140; Adams 2007: 245 (n. 89); Raepsaet 2008: Tbl. 23.4; Mitchell 2018: 131 (Tbl. 5.2); see also similar estimates in: Feseha *et al.* 2004: 49 (Tbl. 3)) were used to draw a 15km buffer from each place or origin. All modern-day toponyms falling within the buffer area were cross-checked with the historical toponyms collected from the written sources for the specific place of origin. Any toponyms found to be associated with and situated at a relatively secure distance, namely did not obviously fall within the territory of another administrative entity, were considered potential matches for the historical toponym under examination and were geolocated as points.

completion of the data collection, the project has extracted the data from the system, conducted in-depth statistical analyses of the assembled data, mapped a subset of the data and to the extent possible geolocated churches, monasteries, metochia, agricultural lands, pre-industrial workshops, and other installations. Finally, it is worth mentioning that the project is currently in the final phase of carrying out spatial-statistical analyses that explore matters of interest, while it is also seeking to secure appropriate permissions from the intellectual property rights holders of the written sources under analysis to render all the data contained in the system open access based on FAIR principles.

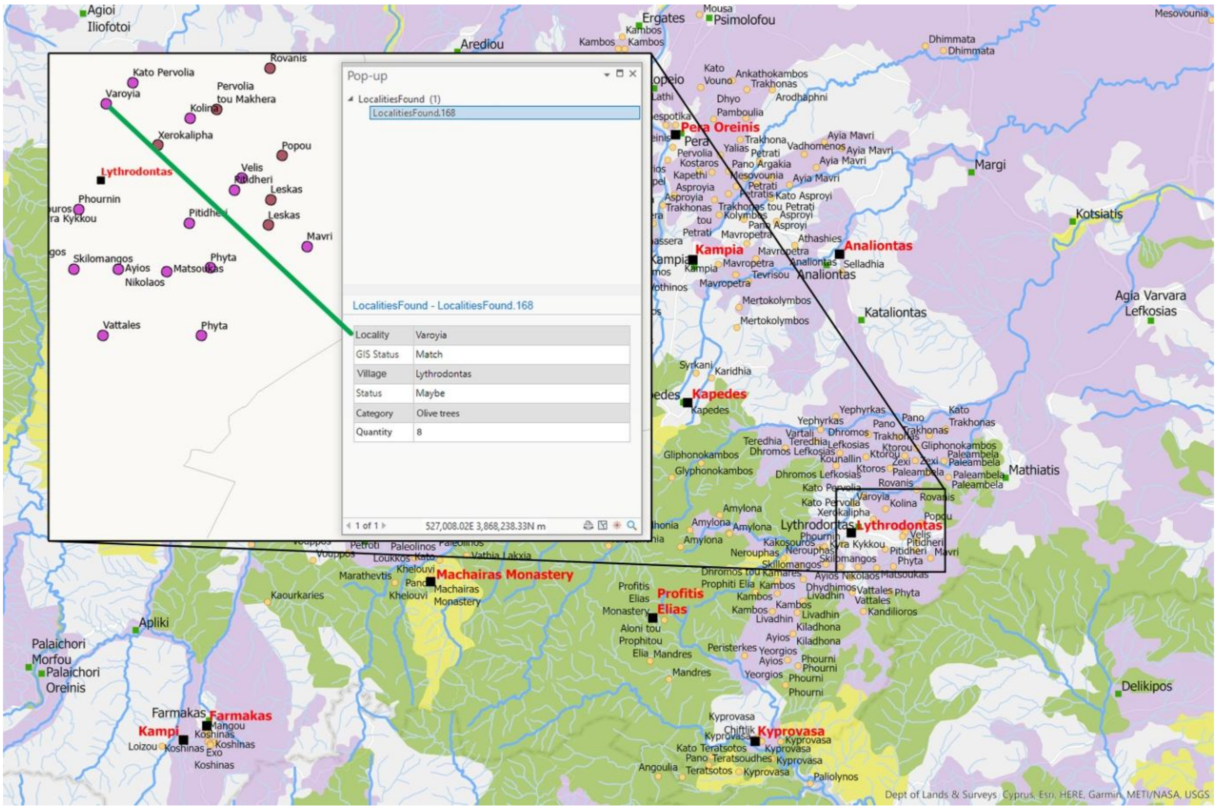


Figure 7. Map of locality-toponym geolocations with linked property data. All the illustrated localities are associated with properties owned by the Machairas Monastery according to the analysis of the Grand Manorial Codex. The inset zoom map depicts an example of property data associated with the locality-toponym "Varoyia". Illustration: Charalambos Paraskeva; River, administrative boundaries geodata: ©Department of Lands and Surveys Cyprus; CORINE2006 land cover geodata: ©Department of Lands and Surveys Cyprus, Copernicus Programme; Basemap: ©ESRI, HERE, Garmin, METI/NASA, USGS.

Acknowledgements

The RURAL-CY project was funded by internal funds of the University of Cyprus and received an additional research allowance grant within the framework of the Research Network Dedicated to the History of the Monastic Economy for expanding its technological infrastructures from the Centre for Advanced Study Sofia, Bulgaria. The project was implemented at the Archaeological Research Unit, Department of History and Archaeology, University of Cyprus (2019-2021).The research team consisted of Associate Professor Athanasios Vionis (Coordinator), Euphrosyne Rizopoulou-Egoumenidou (Folk-Life Studies, Advisor), Michalis N. Michael (Historical-Studies, Advisor), Petroula Hadjittofi (Special Scientist-Principal Researcher, 2019-2021) and Charalambos Paraskeva (External Collaborator, 2019-2021; Special Scientist, 2021).Project participants express their gratitude to the Holy Archbishopric of Cyprus, the Archives of the Holy Archbishopric of Cyprus, and the Holy Bishoprics of Kition and Paphos for allowing access to unpublished written sources.

Contributions

Charalambos Paraskeva prepared the text for sections 2-3, prepared all illustrations, and developed the technological infrastructure and workflows for the project; Petroula Hadjittofi prepared the text for section 1, photographed the Grand Manorial Codex, and performed text segmentation and attribute analysis of all the written sources of the project; Euphrosyne Rizopoulou-Egoumenidou assisted with and contributed to the analysis of the Grand Manorial Codex and Codices 130 and 131 of the Holy Archbishopric Archive; and Athanasios Vionis coordinated the project and reviewed the current paper.

References

Adams, C., 2007, *Land Transport in Roman Egypt: A Study of Economics and Administration in a Roman Province, Oxford Classical Monographs*, Oxford: Oxford University Press.

Arnold, D., Campbell, B., Eckart, T., Fisseni, B., Trippel, T. and Zinn, C., 2021, "CMDIExplorer", in: C. Navarretta and M. Eskevich (eds.),*Selected papers from the CLARIN Annual Conference 2020, Linköping Electronic Conference Proceedings* 180: 8-15, https://doi.org/10.3384/ecp1802.

Babeu, A., 2011, *"Rome Wasn't Digitized in a Day": Building a Cyber infrastructure for Digital Classicists*, Washington DC: Council on Library and Information Resources.

Baca, M., Harpring, P., Lanzi, E., McRae, L. and Whiteside, A., 2006, *Cataloging Cultural Objects: A Guide to Describing Cultural Works and Their Image*, Chicago: American Library Association.

Baram, U. and Carroll, L., 2002, "The Future of the Ottoman Past", in: U. Baram and L. Carroll (eds.), *A Historical Archaeology of the Ottoman Empire: Breaking New Ground, Contributions to Global Historical Archaeology*, New York: Kluwer Academic Publishers, 3-32.

Barton, L.W.G., 2002, "The NEUMES Project: Digital Transcription of Medieval Chant Manuscripts", in: C. Busch, M. Arnold, P. Nesi and M. Schmucker (eds.), *Proceedings of the Second International Conference on Web Delivering of Music, 2002, WEDELMUSIC 2002*, [Piscataway, NJ]: Institute of Electrical and Electronics Engineers Incorporated, 211-218.
https://doi.org/10.1109/WDM.2002.1176213

Barzaghi, S., Palmirani, M. and Peroni, S., 2020, "Development of an Ontology for Modelling Medieval Manuscripts: The Case of Progetto IRNERIO", *Umanistica Digitale*, 5(9), 117-140, https://doi.org/10.6092/issn.2532-8816/11187

Bender, L., Parani, M., Pitarakis, B., Spieser, J.-M., and Vuilloud, A., 2007-2022, *Artefacts and Raw Materials in Byzantine Archival Documents/Objets et Matériaux dans les Documents d'Archives Byzantines*,Online material, http://typika.cfeb.org.

Bintliff, J.L., 2007, "Considerations for Creating an Ottoman Archaeology in Greece", in: S. Davies and J.L. Davis (eds.), *Between Venice and Istanbul: Colonial Landscapes in Early Modern Greece, Hesperia Supplement*40, Princeton, New Jersey: The American School of Classical Studies at Athens, 221-236.

Blanke, T., Bodard, G., Dunn, S., Hedges, M., Jackson, M.and Rajbhandari, S., 2010, "LaQuAT: Integrating and Querying Diverse Digital Resources in Classical Epigraphy", in: B. Frischer, J.W. Crawford and D. Koller (eds.), *Making History Interactive: Computer Applications and Quantitative Methods in Archaeology (CAA), Proceedings of the 37th International Conference, Williamsburg, Virginia, BAR International Series*2079, Oxford: Archaeopress, 29-35, http://proceedings.caaconference.org/files/2009/04_Blanke_et_al_CAA2009.pdf

Boas-Vedder, I. D., 1985, "Appendix II: Jerusalem: The Wall of the MBII Period", in: M. L. Steiner (ed.), *Excavations by Kathleen M. Kenyon in Jerusalem 1961-1967,Volume III, The Settlement in the Bronze and Iron Ages*, London-New York: Sheffield Academic Press, 137-153.

Bodard, G., 2010, "EpiDoc: Epigraphic Documents in XML for Publication and Interchange" , in: F. Feraudi-Gruénais (ed.), *Latin on Stone: Epigraphic Research and Electronic Archives*, Lanham, MD: Lexington Books, 101-118.

Bowman, A.K., Crowther, C.V., Kirkham, R. and Pybus. J., 2010, "A Virtual Research Environment for the Study of Documents and Manuscripts", in:G. Bodard and S. Mahony (eds.), *Digital Research in the Study of Classical Antiquity*, Burlington, VT: Ashgate Publishing, 87-103.

Brandsen, A., 2022, *Digging in Documents: Using Text Mining to Access the Hidden Knowledge in Dutch Archaeological Excavation Reports,* PhD thesis, Universiteit Leiden, Leiden, viewed 27 February 2022, https://hdl.handle.net/1887/3274287

Brandsen, A., Verberne, S., Lambers, K. and Wansleeben, M., 2022, "Can BERT Dig It? Named Entity Recognition for Information Retrieval in the Archaeology Domain", *ACM Journal on Computing and Cultural Heritage* Just Accepted, https://doi.org/10.1145/3497842

Bünte, A., 2011, Documentation for the Use of the eAQUA Function 'Explorative Search', in: C. Schubert (ed.), *Neue Methoden der geisteswissenschaftlichen Forschung. Eine Einführung in das Portal eAQUA II*, *Working Papers Contested Order* 3, 19-32.

Catepano, T., (ed.), Technical Subcommittee for Encoded Archival Description, Schema Development Team 2019: *Encoded Archival Description EAD3, Release 1.1.1*, Society of American Archivists, https://github.com/SAA-SDT/EAD3/blob/v1.1.1/ead3.rng.

Chen, C.-M., Chen, Y.-T. and Liu, C.-Y., 2019, "Development and Evaluation of an Automatic Text Annotation System for Supporting Digital Humanities Research", *Library Hi Tech* 37(3), 436-455, https://doi.org/10.1108/LHT-10-2017-0219.

Cowles, E., Batchelor, H., Beaubien, R., Benzing, M., Boughida, K., Clarke, S., Eklund, J., Elings, M., Hamilton, F., Kessler, B., Kreymer, O., Morgan, E., Proffitt, M., Rose, T., Stedfeld, E., Waibel, G., Wendler, R., Westbrook, B. and Whiteside, A., 2007, *VRA (Visual Resources Association) Core 4 XML Schema*, *version 0.42*, Visual Resources Association, https://www.loc.gov/standards/vracore/vra.xsd.

DCMI Usage Board, 2020, DCMI Metadata Terms, Dublin Core Metadata Initiative, https://www.dublincore.org/specifications/dublin-core/dcmi-terms/.

de Castilho, R.E., Biemann, C., Gurevych, I. and Yimam, S.M., 2015, "WebAnno: AFlexible, Web-based Annotation Tool for CLARIN", in: *Proceedings of the CLARIN Annual Conference (CAC)*, [n.p.]: CLARIN, https://www.clarin.eu/sites/default/files/cac2014_submission_6_0.pdf

de Castilho, R.E., Mújdricza-Maydt, E., Yimam, S.M., Hartmann, S., Gurevych, I., Frank, A. and Biemann, C., 2016, "A Web-based Tool for the Integrated Annotation of Semantic and Syntactic Structures", in: E. Hinrichs, M. Hinrichs and T. Trippel (eds.)*Proceedings of the Workshop on Language Technology Resources and Tools for Digital Humanities (LT4DH)*, [Stroudsburg, PA]: [Association for Computational Linguistics], 76-84.

De Groot, S., 2020, "Byzantine Book-Epigrams: From Manuscript to Edition", in: M. Berghöfer, A. Beron, F. Etling, G. Hedderich, M. Stralla and A. Wilken (eds.),*(un)documented: Was bleibt vom Dokument in der Edition?*, Berlin-Boston: De Gruyter, 39-60.

Dekker, R.H., van Hulle, D., Middell, G., Neyt, V.,and van Zundert, J., 2015, "Computer-supported Collation of Modern Manuscripts: CollateX and the Beckett Digital Manuscript Project", *Digital Scholarship in the Humanities* 30(3): 452-470, https://doi.org/10.1093/llc/fqu007.

Depauw ,M. and Gheldof, T., 2014, "Trismegistos: An Interdisciplinary Platform for Ancient World Texts and Related Information",in: Ł.Bolikowski, V. Casarosa, P. Goodale, N. Houssos, P. Manghi and J. Schirrwagen (eds.),*Theory and Practice of Digital Libraries 2013, Selected Workshops*, *Communications in Computer and Information Science* 416, Switzerland: Springer, Cham, 40-52, https://doi.org/10.1007/978-3-319-08425-1_5.

Dima, E., Hinrichs, E., Hinrichs, M., Kislev, A., Trippel, T. and Zastrow, T., 2012, "Integration of Weblicht into the Clarin Infrastructure", in: *Proceedings, Service-Oriented Architectures (SOAs) for the Humanities: Solutions and Impacts, Joint CLARIN-D/DARIAH Workshop at Digital Humanities Conference 2012*, [n.p.]: CLARIN-D, DARIAH-DE, 17-23.

Doerr, M., Murano, F. and Felicetti, A., 2017, *Definition of the CRMtex*, *An Extension of CIDOC CRM to Model Ancient Textual Entities, Version 0.8*, Online material https://cidoc-crm.org/crmtex/sites/default/files/CRMtex_v0.8.pdf

Dombrowski, Q., 2014, "What Ever Happened to Project Bamboo?", *Literary and Linguistic Computing* 29(3), 326-339, https://doi.org/10.1093/llc/fqu026.

Duba, W., 2019, "Fragmentarium", *Das Mittelalter* 24: *Digitale Mediävistik und der deutschsprachige Raum*, 221-223, https://doi.org/10.1515/mial-2019-0015.

Elliott, T., Bodard, G., Cayless, H., DiGiulio, S., Filosa, M., Mylonas, E., Werning, D., Vagionakis, I., Farrugia, K., Stoyanova, S., Liuzzo, P.M., Tupman, C., Vanderbilt, S., *et al.*, 2006-2021, *EpiDoc: Epigraphic Documents in TEI XML*, Online material, http://epidoc.stoa.org/.

Feseha, G., Alemu, G.W., Friew, K., Abule, I. and Ketema, Y., 2004, “Donkey Utilisation and Management in Ethiopia”, in: P. Starkey and D. Fielding (eds.), *Donkeys, People and Development: A Resource Book of the Animal Traction Network for Eastern and Southern Africa (ATNESA),* Wageningen, The Netherlands: ACP-EU Technical Centre for Agricultural and Rural Cooperation (CTA), 46-52.

Hacıgüzeller, P., Taylor, J. and Perry, S., 2021,“On the Emerging Supremacy of Structured Digital Data in Archaeology: A Preliminary Assessment of Information, Knowledge and Wisdom Left Behind”,*Open Archaeology*7(1), 1709-1730, https://doi.org/10.1515/opar-2020-0220.

Hadjikyriacou, A., 2016, “The Ottomanization of Cyprus: Towards a Spatial Imagination beyond the Centre-Province Binary”, *Journal of Mediterranean Studies* 25(2), 81-96.

Hadjikyriacou, A., Papadias, E., Vradis, C. and Chalkias, C., 2021, “Combining Historical Maps and Censuses of Cyprus from the Sixteenth to the Twentieth Century: A Geospatial Approach”, *Proceedings of the International Cartographic Association* 3(7), https://doi.org/10.5194/ica-proc-3-7-2021.

Haynes, D., 2018, *Metadata for Information Management and Retrieval: Understanding Metadata and Its Use*, London: Facet Publishing.

Hinrichs, E.W., Hinrichs, M. and Zastrow, T., 2010, “WebLicht: Web-Based LRT Services for German” ,in: S. Kübler (ed.), *Proceedings of the ACL 2010 System Demonstrations*, Stroudsburg, PA: Association for Computational Linguistics, 25-29.

International Organization for Standardization, 2003, *ISO/IEC 15938-5:2003, Information Technology - Multimedia Content Description Interface - Part 5: Multimedia Description Schemes,* International Organization for Standardization.

Kinable, D.C.J., 2018, “At the Crossroads of Digital Humanities and Historical Lexicography: The Middle Dutch ‘Seemly Play (Abel Spel) of Winter and Summer’ as a Research Case”, *Digital Scholarship in the Humanities* 33(3), 557-574, https://doi.org/10.1093/llc/fqx051.

Lemahieu, W., Broucke, S.V. and Baesens, B., 2018, *Principles of Database Management: The Practical Guide to Storing, Managing and Analyzing Big and Small Data*, Cambridge: Cambridge University Press.

Liceras-Garrido, R., Favila-Vázquez, M., Bellamy, K., Murrieta-Flores, P., Jiménez-Badillo, D. and Martins, B., 2019, “Digital Approaches to Historical Archaeology: Exploring the Geographies of 16th Century New Spain”, *Open Access Journal of Archaeology and Anthropology* 2(1), https://doi.org/10.33552/OAJAA.2019.02.000526.

Liu R., McKay D. and Buchanan G., 2021, “Humanities Scholars and Digital Humanities Projects: Practice Barriers in Tools Usage”, in: G. Berget, M.M. Hall, D. Brenn and S. Kumpulainen (eds.), *Linking Theory and Practice of Digital Libraries. Proceedings of the 25th International Conference on Theory andPractice of Digital Libraries, TPDL 2021. Lecture Notes in Computer Science*12866. Switzerland: Springer, Cham,215-226.

Martínez, J.M., Koenen, R. and Pereira, F., 2002, “MPEG-7: the Generic Multimedia Content Description Standard”, *IEEE Multimedia*9 (2), 78-87, https://doi.org/10.1109/93.998074.

Michael, M. N., 2021, “The Church of Cyprus and the Transition from Ottoman Rule to British Modernity”, in: R. Avramov, A. Fotić, E. Kolovos and P. Kotzageorgis. (eds.), *Monastic Economy Across Time: Wealth Management, Patterns and Trends*, Sofia: Centre for Advanced Study Sofia, 244-260.

Mitchell, P., 2018, *The Donkey in Human History: An Archaeological Perspective*, Oxford: Oxford University Press.

Moudgil, A., Singh, S. and Gautam, V., 2022, "An Overview of Recent Trends in OCR Systems for Manuscripts", in: J.M.R.S. Tavares, P. Dutta, S. Dutta and D. Samanta (eds.), *Cyber Intelligence and Information Retrieval, Lecture Notes in Networks and Systems* 291, Singapore: Springer, 525-533, https://doi.org/10.1007/978-981-16-4284-5_46

Murrieta-Flores, P. and Gregory, I., 2015, “Further Frontiers in GIS: Extending Spatial Analysis to Textual Sources in Archaeology” *Open Archaeology*1(1), 166-175, https://doi.org/10.1515/opar-2015-0010.

Niehl, R., 2003,“CAMENA - Neulateinische Dichtung im Internet. On the Way from Text Corpus to Reading Environment”,*Neulateinisches Jahrbuch* 5, 302-314.

Raepsaet, G., 2008, “Land Transport, Part 2: Riding, Harnesses, and Vehicles”, in: J.P. Oleson (ed.), *Handbook of Engineering and Technology in the Classical World*, Oxford: Oxford University Press, 580-605.

Richards, J.D., Tudhope, D. and Vlachidis, A., 2015, “Text Mining in Archaeology: Extracting Information from Archaeological Reports”in: J. Barcelo and I. Bogdanovic (eds.), *Mathematics in Archaeology*, Florida: CRC Press, 240-254, https://doi.org/10.1201/b18530-15

Roudoumetof, V. and Michael, M.N., 2010, “Economic Functions of Monasticism in Cyprus: The Case of the Kykkos Monastery”, *Religions* 1, 54-77.

Rutz, M. and Kersel, M.M., 2014, *Archaeologies of Text: Archaeology, Technology and Ethics*. *Joukowsky Institute Publications*6, Oxford: Oxbow Books.

Schibel, W., 2001, “CAMENA - Neulateinische Dichtung im World Wide Web” *Neulateinisches Jahrbuch* 3, 211-219.

Schubert, C., 2010, “Zitationsprofile, Suchstrategien und Forschungsrichtungen”, in: C. Schubert and G. Heyer (eds.), *Das Portal eAQUA - Neue Methoden in der geisteswissenschaftlichen Forschung I*, *Working Papers Contested Order* 1, 42-55.

Schubert, C., 2015, “Digital Humanities: Laboratorium der Geisteswissenschaften oder Weg nach Atlantis?”, in: S. Keym and S. Wünsche (eds.), *Musikgeschichte zwischen Ost und West: von der musica sacra bis zur Kunstreligion, Festschrift für Helmut Loos zum 65Geburtstag*, Leipzig: Leipziger Universitätsverlag, 747-758.

Schulz, S., 2018, *The Taming of the Shrew - Non-standard Text Processing in the Digital Humanities*, PhD thesis, University of Stuttgart, Stuttgart, viewed 27 February 2022, http://dx.doi.org/10.18419/opus-9685.

Schulz, S. and Ketschik, N., 2019, “From 0 to 10 Million Annotated Words: Part-of-Speech Tagging for Middle High German”, *Language Resources and Evaluation* 53(4), 837-863. https://doi.org/10.1007/s10579-019-09462-8

Sicilia, M.-A., 2014, *Handbook of Metadata, Semantics and Ontologies*, Singapore: World Scientific.

Silberman, N. A., 1989, *Between Past and Present: Archaeology, Ideology, and Nationalism in the Modern Middle East*, New York: Henry Holt and Company.

Simon, R., Barker, E., Isaksen, L. and de Soto, C.P., 2017, “Linked Data Annotation without the Pointy Brackets: Introducing Recogito 2”, *Journal of Map and Geography Libraries* 13(1), 111-132.

Stavrides, T., 2013, “Chrysanthos, 1767-1810: Grappling with the Vicissitudes of Ottoman Power”, in: A. Varnava and M.N. Michael (eds.), *The Archbishops of Cyprus in the Modern Age: The Changing Role of the Archbishop-Ethnarch, their Identities and Politics*, Newcastle upon Tyne: Cambridge Scholars Publishing, 17-40.

Sula, C.A. and Hill, H. V., 2019, “The Early History of Digital Humanities: An Analysis of *Computers and the Humanities* (1966–2004) and *Literary and Linguistic Computing* (1986–2004)”, *Digital Scholarship in the Humanities* 34(Supplement1): i190-i206.

Tarte, S.M., Wallom, D., Hu, P., Tang, K. and Ma. T., 2009, “An Image Processing Portal and Web-Service for the Study of Ancient Documents,” in: *5th IEEE Conference on e-Science 2009*, 14-19, https://doi.org/10.1109/e-Science.2009.10.

TEI Consortium, 2021, *TEI P5: Guidelines for Electronic Text Encoding and InterchangeVersion 4.3.0. Last updated on 31st August 2021, revisionb4f72b1ff*, Text Encoding Initiative Consortium.

TextGrid Consortium, 2006-2014, *TextGrid: A Virtual Research Environment for the Humanities*, Göttingen: TextGrid Consortium, https://textgrid.de/.

Tîrziman, E., 2013, “An Experience of Preservation and Evaluation of Documentary Heritage through Digitization: Manuscriptorium ’Project”, *Infotheca* 14(1), 7-14.

Uhlíř, Z. and Knoll, A., 2009, “Manuscriptorium Digital Library and ENRICH Project: Means for Dealing with Digital Codicology and Palaeography” in: M. Rehbein, T. Schaßan and P. Sahle (eds.),*Kodikologie und Paläografie im digitalen Zeitalter - Codicology and Palaeography in the Digital Age,Schriften des Instituts für Dokumentologie und Editorik*2, Norderstedt: BoD, 67-78.

Valovič, R., Jakúbek, T., Přichystal, J. and Fuchs, J., 2022, “Multicriterial Text Analysis Software”, Working paper, viewed 27 February 2022, https://akela.mendelu.cz/~jprich/prichystal-mta.pdf

van Lit, L., 2019, *Among Digitized Manuscripts. Philology, Codicology, Paleography in a Digital World*, Leiden: Brill.

Vanie, E., Shah, B. and Malepati, T., 2019, *Advanced MySQL 8: Discover the Full Potential of MySQL and Ensure High Performance of Your Database*, Birmingham: Packt Publishing Ltd.

Vionis, A. K., 2016, "A Boom-Bust Cycle in Ottoman Greece and the Ceramic Legacy of Two Boeotian Villages",*Journal of Greek Archaeology* 1, 353-384.

Vorderstrasse, T., 2014, "The Archaeology of the Ottoman Empire and its Aftermath in the Middle East", *Middle Eastern Archaeology* 77, 292-298.

Wille R., 1999, "Conceptual Landscapes of Knowledge: A Pragmatic Paradigm for Knowledge Processing" in: W. Gaul and H. Locarek-Junge (eds.),*Classification in the Information Age, Studies in Classification, Data Analysis, and Knowledge Organization*,Berlin-Heidelberg: Springer,344-356, https://doi.org/10.1007/978-3-642-60187-3_36

Yenişehirlioğlu, F., 2015, "Whither Art History? How Global Is Ottoman Art and Architecture?", *Art Bulletin* 97(4), 359-363.

Ελληνικός Οργανισμός Τυποποίησης, 2001, *Ελληνικό Πρότυπο ΕΛΟΤ 743, 2η Έκδοση, Πληροφορίες και Τεκμηρίωση - Μετατροπή των Ελληνικών Χαρακτήρων με Χαρακτήρες Λατινικούς, Information and Documentation – Conversion of Greek Characters into Latin Characters*, Αθήνα: Ελληνικός Οργανισμός Τυποποίησης.

Θεοχαρίδης, Ι. Π., 2004, *Ο Κώδικας 53 της Ιεράς Μονής Κύκκου: Οι πολύπλευρες δραστηριότητες της μονής κατά την περίοδο 1843-1897*, Λευκωσία: Κέντρο Μελετών Ιεράς Μονής Κύκκου.

Μιχαήλ, Μ., 2001, *Ο Κώδικας 54 της Ιεράς Μονής Κύκκου και οι οικονομικές δραστηριότητες της, 1813-1819*, Λευκωσία: Κέντρο Μελετών Ιεράς Μονής Κύκκου.

Μιχαήλ, Μ., 2003, *Η Ιερά Μονή Κύκκου στο οθωμανικό οικονομικό πλαίσιο με βάση τον Κώδικα 56 (1844-1890)*, Λευκωσία: Κέντρο Μελετών Ιεράς Μονής Κύκκου.

Μιχαήλ, Μ., 2005a, *Η Εκκλησία της Κύπρου κατά την Οθωμανική περίοδο, 1571-1878: Η σταδιακή συγκρότησή της σε θεσμό πολιτικής εξουσίας*, Λευκωσία: Κέντρο Επιστημονικών Ερευνών.

Μιχαήλ, Μ., 2005b, *Το τσιφλίκι (ciftlik) της Οθωμανικής περιόδου: Μοναστηριακά τσιφλίκια στους Κώδικες 49 (1813-1841) και 51 (1818-1881) της Ιεράς Μονής Κύκκου*, Λευκωσία: Κέντρο Μελετών Ιεράς Μονής Κύκκου.

Παπαδόπουλλος, Θ., 2008, *Κώδιξ Μονής Κύκκου*, Λευκωσία: Κέντρο Μελετών Ιεράς Μονής Κύκκου.

Σταυρίδης, Θ., 2001, *Οικουμενικό Πατριαρχείο και Κύπρος: τα πατριαρχικά έγγραφα των ετών 1600- 1878*, Λευκωσία: Κέντρο Μελετών Ιεράς Μονής Κύκκου.

MEASURES OF POPULATION DIVERGENCE FOR BINARY DATA: IMPROVEMENTS VIA SIMULATIONS

E. NIKITA[1] AND P. NIKITAS[2]

[1]Science and Technology in Archaeology and Culture Research Center, The Cyprus Institute, 2121 Aglantzia, Nicosia, Cyprus

efi.nikita@gmail.com

[2]Department of Chemistry, Aristotle University of Thessaloniki, University Campus, 54124 Thessaloniki, Greece

nikitas@chem.auth.gr

Abstract

The performance of different distance measures for binary data as unbiased estimators of population divergence in bioarchaeological studies is examined using 10 datasets of nonmetric traits taken from the literature. The measures are based on the application of the probit, logit, and arcsine transformations to the sample proportions. The case of untransformed data is also examined. It is shown that the main source of biased estimations at low and high proportions is the data transformation adopted in these measures that results in variances that are asymptotically valid within a range of proportion values around 0.5. The estimation of variances via simulations improves their performance and the measures in most cases become nearly unbiased estimators. For further improvement, a new modification of these measures is proposed, which makes them strict unbiased estimators of population divergence throughout the proportion range. This modification does not alter the measure of divergence based on untransformed data, which is a strict unbiased estimator irrespective of the use of simulated or non-simulated variances. The computation of the proposed measures is implemented using custom functions in R.

Keywords*: measures of divergence; binary data; simulations*

1. Introduction

Bioarchaeology is the contextual study of human remains from archaeological settings, and it can provide key information regarding life in the past (diet, disease, mobility, physiological stress etc.) (Buikstra and Beck 2017). Among the key aspects bioarchaeologists examine is past gene flow, that is, kinship and mobility. Bioarchaeological studies of mobility have traditionally relied on so-called biodistance analysis (Pilloud and Hefner 2016). Biodistance analysis employs the phenotype (morphological and metric traits of the skeleton and the teeth) as proxies for the genotype under the premise that the former is largely controlled by the latter. In other words, individuals and groups who share the same phenotype (the same morphological skeletal and dental traits and/or cranial and dental dimensions) tend to be genetically related. This, in turn, elucidates patterns of kinship and larger scale gene flow (the mobility of people and their genes). Although palaeomobility studies are increasingly based on isotopic and ancient DNA analyses, biodistance studies are still very popular due to their non-destructive character and low-cost, and more advanced statistical methods for the processing of the phenotypic traits keep being developed.

Before proceeding to the presentation of the state of the art in the estimation of measures of divergence and the issue that the current paper addresses, some basic definitions are important. We hereby define a population as the entire group of individuals who ever comprised specific archaeological cultures. The sample is the number of individuals who are available for anthropological analysis; thus, the sample is a subset of the population. The term individuals is used to denote single skeletons from each sample, while the term dataset encompasses the individuals of different samples, depending on the comparisons performed.

The Mean Measure of Divergence (MMD) is the most commonly adopted measure of population divergence, that is, a mathematical distance used to estimate biodistances – phenotypic distances as a proxy to genetic ones(de Souza and Houghton 1977; Harris and Sjøvold2004; Irish 2010; Nikita 2015; Nikita 2017; Santos 2018; Sjøvold 1977; Zertuche and Meza-Peñaloza 2010). The use of the MMD began in the 1970s and continues until now. The MMD is based on the arcsine transformation of the traits proportion (relative frequency) in the samples and its estimation nowadays is carried out using the formulas presented by Sjøvold (1977) and Harris and Sjøvold (2004), whereas an interesting estimation of the MMD using the bootstrap method has been proposed by Zertuche and Meza-Peñaloza (2010).In a recent paper (Nikita and Nikitas2021), three new measures of divergence based on untransformed data (UMD) and the logit (LMD) and probit (PMD) transformations have been examined. It was found that the UMD outperforms all other measures in the sense that it is a strict unbiased estimator of population divergence and does not exhibit application problems at very low or very high trait proportions. The MMD is a satisfactory distance measure for binary data, although its application requires a careful test to avoid biased estimations of population divergence when there are small sample sizes and many traits with trait proportions lower than 0.1 or greater than 0.9. In what concerns the LMD and PMD, i.e. the measures of divergence based on the logit and probit transformations, these are more prone to biased estimations of population divergence than the MMD in datasets with small sample sizes and low/high trait proportions.

A limitation of the MMD and its currently available variants, with the exception of the UMD, is that these measures are not strict but asymptotic unbiased estimators of population divergence. This is due to the use of variances that are asymptotically valid in a limited range of proportions around 0.5.At this point we should clarify that in the arcsine transformations adopted in the MMD the variances are independent of the population trait proportion p, whereas in all other transformations discussed in Nikita and Nikitas (2021), the variances depend on p. Nonetheless, all these transformations are asymptotically valid for proportions around 0.5 and not particularly small sample sizes, n. In contrast, the variance of the untransformed data (i.e. when using UMD) is given by the exact expression p(1-p)/n, and it is valid throughout the p range, i.e., from p=0 to p=1. Noting that in biodistance studies the aim is to obtain reliable information on affinities between populations and not between samples, a straightforward approach to improve the validity of the measures of divergence based on transformed data is to replace the approximate variances by simulated ones. This is the purpose of the current study.

Based on these observations, in the present paper we use simulations to estimate variances and examine the performance of the MMD, LMD, and PMD when applied to the entire range of proportions, from 0 to 1. However, as we found that this procedure alone may not solve the problem of biased estimations, we also propose, apply, and discuss a new approach in defining these measures of divergence that makes them strict unbiased estimators of population divergence.

2. Mathematicalbackground

The general expression of a measure of divergence between two samples, 1 and 2, consisting of binary data may be given by (Sjøvold 1977; Nikita and Nikitas 2021, Nikitas and Nikita 2022):

$$MD = \sum_{i=1}^{r} \{(t_{1i} - t_{2i})^2 - Var(t_{1i}) - Var(t_{2i})\} \quad (1)$$

where r is the number of variables (traits) and t is any sample statistic, t = t(φ,n), with variance, Var(t), that tends to zero in the population. In general, Var(t) is a function of p and n, i.e., Var(t)=f(p,n). If n_{1i}, n_{2i} are the number of cases of trait i in samples 1 and 2, and k_{1i}, k_{2i} the number of individuals in the samples possessing the trait i, then t_{1i} and t_{2i} may be the proportions (relative frequencies)$\varphi_{1i}=k_{1i}/n_{1i}$and $\varphi_{2i}=k_{2i}/n_{2i}$or any proper transformation of φ_{1i} and φ_{2i}.The current study concerns cases where t = φ as well as where t is the arcsine, probit, and logit transformation of φ.

Up to now, except for the identity transformation, t = φ, only approximate expressions have been used for the variance Var(t). However, as already pointed out in the Introduction, the applicability of the MD can be extended over the entire range of φ values if we reject approximate expressions for the variance Var(t) in Eq. (1)and estimate values of Var(t) as accurately as possible via simulations. The details of the various simulation procedures adopted in the present study are described in section 3.2.

Taking into account that the variance of the sample statistic t tends to zero in the population, the application of Eq. (1) to calculate the divergence between two populations yields:

$$MDp = \sum_{i=1}^{r} \left(t_{1i}(p_{1i}) - t_{2i}(p_{2i})\right)^2 = \sum_{i=1}^{r} (T_{1i} - T_{2i})^2 \quad (2)$$

where p is the φ value in the population. In what concerns the expected value of MD, we have:

$$E[MD] = \sum_{i=1}^{r} \{E[(t_{1i} - t_{2i})^2] - E[Var(t_{1i})] - E[Var(t_{2i})]\}$$

Now, taking into account the definition of the variance of a random variable Z from $E[Z^2] = (E[Z])^2 + Var(Z)$, the above formula is transformed to:

$$E[MD] = \sum_{i=1}^{r} \{(E[(t_{1i} - t_{2i})])^2 + Var(t_{1i}) + Var(t_{2i}) - Var(t_{1i}) - Var(t_{2i})\}$$

because the variances $Var(t_{1i})$, $Var(t_{2i})$, as, in general, functions of p and n, are constants and therefore $E[Var(t_{ji})] = Var(t_{ji})$. Thus, we finally obtain:

$$E[MD] = \sum_{i=1}^{r} (E[(t_{1i} - t_{2i})])^2 = \sum_{i=1}^{r} (E[t_{1i}] - E[t_{2i}])^2 \qquad (3)$$

It is seen that the MD is an unbiased estimator of population divergence only if the population divergence is expressed by the sum of the squared differences between the expected t values and that this sum coincides or at least converges satisfactorily with the MDp defined from Eq. (2). From the transformations studied in this paper, this is strictly valid for the transformation that uses the identity function, i.e. t = φ, since, in this case, the equality E[t] = E[φ] = p = T holds for every φ and n value and, therefore, Eqs. (2) and (3) give always identical results. For the other transformations, E[t] = T is asymptotically valid in a range of φ values around 0.5, which is the range of φ where each transformation holds (see Sjøvold1977, equations 1.51). However, when Eq. (1) is applied to the entire range of φ values via simulations, it is improbable that the equality E[t] = T holds throughout the φ range, resulting in population biases, the magnitude of which is expected to depend on the percentage of φ values close to 0 or/and 1, as well as the presence of small sample sizes n. Additional problems may arise when a transformation depends upon n, t=t(φ,n), such as the Anscombe and Freeman-Tukey transformations (Harris and Sjøvold 2004). This relationship of t upon n may be weak, but it can cause issues when n is very small. In this case, the expected value E[t] depends upon n and, therefore, the population divergence defined from Eq. (2) depends also upon the sample size, i.e. a population property is defined from its sample size, which is problematic.

If the application of Eq. (1) to a certain dataset presents strong biases even using simulated variances, a simple approach to overcome them is via the following correction:

$$MDc = MD - E[MD] + MDp$$

where MD, MDp, and E[MD] are given by Eqs. (1) to (3). It is seen that for the computation of the MDc, we need first to simulate the populations from which the samples of the dataset originate and then draw randomly samples from these populations. The populations can be straightforwardly generated from uncorrelated multivariate binary variates with marginal probabilities the φvalues of the original samples, since φ is an unbiased estimator of the population p value and the correlations between variables affect neither the MDp nor the E[MD].From the simulated populations, the MDp can be calculated using Eq. (2), whereas the random samples drawn from these populations are used to estimate the $E[t_i]$values, which are further used to estimate E[MD] from Eq. (3).

From the definition of this new measure of divergence, we observe that MDc = MD when using untransformed data and that irrespective of the transformation adopted, the MDc is always an unbiased estimator of population divergence because $E[MDc] = E[MD] - E[MD] + MDp = MDp$.However, there is a critical point that we should consider when using Eq. (4) to compute the MDc. The validity of the calculated MDc depends upon the validity of the estimated E[MD]. Note that this quantity is estimated via iterations, where the number of iterations, N_{iter}, is the number of samples that are randomly drawn from the populations. Therefore, a preliminary test using different iteration numbers is necessary to determine the optimum N_{iter}.

Finally, we should comment on the following two points: 1) A basic assumption in the theory of the MMD is that the size of the populations should tend to infinity (Sjøvold1977,page 16). It is under this assumption that the variances $Var(t_{1i}) = 1/n_{1i}, Var(t_{2i}) = 1/n_{2i}$ in Eq. (1) for the MMD tend to zero in the population and make the MMD an unbiased estimator provided that φtakes values around 0.5. In practice, this assumption means that the size of the populations should be as large as required for the variances to contribute insignificantly to the MDp. This is a reasonable assumption for biological populations since the population size includes the total number of members of the population for many generations. If, despite this general trend, there are populations with relatively small sizes, the expected value of the MMD, E[MMD], differs from the population MMD, MMDp, by the sum of the population variances. Therefore, in this case the MMD is no longer a strict unbiased estimator of population divergence. These observations apply to all measures of divergence considered in the present study. 2) In all distance measures defined from Eq. (1) negative and, therefore, biologically meaningless contributions to MD appear when $t_{1i} \approx t_{2i}$. These contributions must not be eliminated, even when they lead to negative MD values because in this case the MD ceases being an unbiased estimator. This issue is discussed in detail in Nikita and Nikitas (2021).

3. Materials and Methods

3.1. Materials

We selected initially five datasets to test the performance of the measures of divergence under study. These were obtained from the literature and in particular, two datasets from the Ossenberg database of cranial nonmetric traits (Ossenberg 2013), freely available online at: https://borealisdata.ca/dataset.xhtml?persistentId=hdl:10864/TTVHX, one dataset from Konigsberg's website (http://faculty.las.illinois.edu/lylek/)(Konigsberg 1990), and the other two from the supplementary materials at Zertuche and Meza-Peñaloza (2020) and Nikita and Nikitas (2021). From the Ossenberg database we created one dataset of eight African assemblages with 36 traits (AF-36) and another dataset of eight Eurasian assemblages also with 36 traits (EU-36). The other three datasets consisted of a) seven ancient Greek assemblages of 28 traits (GR-28), b) seven assemblages of 13 traits from the Basin of Mexico (MEX-13), and c) 14 assemblages of eight traits of the Lowilva dataset (LO-8), respectively. Note that the Lowilva dataset includes samples from the Lower Illinois River Valley and some sites from the Mississippi River Valley (Konigsberg 1990).
A common feature of all these datasets is the great percentage of traits with small and very small φ values, whereas the MEX-13 dataset includes traits with φ=1. For this reason, we proceeded to data editing by removing traits with φ=0 or/and φ=1. This procedure led to the creation of another five datasets characterized by a drastic reduction in the number of traits, since in most cases more than half of the traits were deleted. Table 1 presents the percentage of low and high trait frequencies in all datasets studied. The complete files of these data are available from the authors upon request. Note that in these datasets there are a lot of missing data. The missing data do not play any role in the treatment because we always calculate and use relative frequencies, φ. However, the logit and probit functions are not defined at φ = 0 and φ = 1. In this case, Bartlett's correction has been used (Harris and Sjøvold 2004), which entails the replacement of φ = 0 with φ = 1/4n and φ = 1 with φ = 1-1/4n.

Dataset	Traits	φ≤0.1	φ≤0.05	φ=0	φ≥0.9	φ≥0.95	φ=1
AF-36	36	44.1%	31.3%	27.1%	0	0	0
AF-14	14	9.8%	0.89%	0	0	0	0
EU-36	36	47.9%	34.0%	18.4%	0	0	0
EU-16	16	17.2%	7.0%	0	0	0	0
GR-28	28	51.0%	35.7%	20.9%	0	0	0
GR-12	12	8.3%	3.6%	0	0	0	0
LO-8	8	35.7%	26.8%	9.8%	6.3%	0	0
LO-5	5	7.1%	1.4%	0	10%	0	0
MEX-13	13	27.5%	19.8%	15.4%	5.5%	5.5%	4.4%
MEX-6	6	16.7%	4.8%	0	0	0	0

Table 1. Percentage of low and high trait frequencies in datasets adopted in the present study

3.2. Simulations and software

3.2.1. Simulation of variances

The variance of the binary variable of a certain trait and sample, V = Var(t), appearing in the definition of a measure of divergence, Eq. (1), as well as in all relevant expressions, is, in the general case, a function of the number n of the cases in the trait of question and its proportion p in the population, which is approximated by the corresponding sample proportionφ, i.e. V=V(n,φ).At a given (n,φ) pair the estimation of the variance via simulations is straightforward. In the R programming language we may use the function *rbinom(n, 1,* φ*)* to create a sample from n binary data that come from a population with relative frequency equal to φ. This step is repeated many times, in our study 10000 times, and at each iteration, the transformed value of the average value of the sample values is calculated. In this way a random value Z is computed, which is used to estimate the population variance V from $V=E[Z^2]-E[Z]^2$, where E denotes the mean value.

Note that this procedure usually requires long computational time. To reduce this time, we worked as follows. Consider the number of cases of trait i in sample s, n_{si}. At each n_{si} the value of V is estimated via the above simulation on a preset lattice of φ values. Then the V value at an intermediate φ value is estimated by means of cubic spline interpolation using the *splinefun()* function of R.

3.2.2.Generation of simulated data

To compute the MDc from Eq. (4), as well as to test whether a distance measure is an unbiased estimator of population divergence or not based on a dataset of samples with binary data, we need first to simulate the populations from which the samples of the dataset come. For this purpose, we followed the approach described in Nikita and Nikitas(2021). Thus, taking into account that the intercorrelations among the traits affect neither the biased/unbiased estimation of the population divergence nor the expected values $E[t_i]$ used in Eq. (3) and that a correlation matrix does not always conform with the marginal probabilities, we used the *rmvbin()* function of the *bindata* library of R to generate multivariate binary variates using as marginal probabilities the φvalues of the original samples. In the present study, populations of size 100000 were created from which a preset number of samples, i.e. number of iterations N_{iter}, usually ranging from 1000 to 5000, was randomly drawn. Each MD was calculated between the samples of the original dataset (MDs), between the corresponding populations (MDp) using Eq. (2) as well as between the samples drawn from the populations. The latter distances were averaged to obtain the averaged MD (avMDs). In addition, the random samples drawn from the populations were used to estimate the $E[t_i]$ values, which were further used to estimate E[MD] from Eq. (3).Note that the avMDs is in fact the long-run average value of MD and, therefore, it is an alternative estimate of the expected value of MD. Thus, E[MD] and avMDs are expected to converge.

When a MD is an unbiased estimator, the distances MDp, avMDs, and E[MD] coincide. Therefore, a scatterplot displaying these distances for all pairs of populations/samples can be used to visualize biases and detect their magnitude. An additional useful property is that these distances vary from the MDs estimated between the samples i and j by the quantity $\Delta V = V_i + V_j$ (Nikita and Nikitas 2021). This means that if the calculated distances MDp, avMDs, and E[MD] on the simulated data are corrected by subtracting the quantity ΔV, then these distances will coincide with the MDs provided that MD is an unbiased estimator of population divergence.

For a quantitative measure of the performance of a measure of divergence as an unbiased estimator, we may use the quantity:

$$SSD = \frac{100}{q(MDs_{max} - MDs_{min})^2} \sum_{i=1}^{q} (X_i - Y_i)^2 \qquad (7)$$

where q is the number of pairwise MDs in the original dataset and X, Y are the compared distances (MDp, avMDs, E[MD], MDs).This is the sum of squared differences between the compared distances normalized by the range of the MDs values (MDs_{max}-MDs_{min}). Therefore, for a MD to be an unbiased estimator of population divergence, the SSD values between MDp and E[MD] or between MDp and avMDs should be as small as possible, usually well below 0.1.

In what concerns the MDc, by its definition, this is always an unbiased estimator of population divergence. This means that if we compute the MDc between the samples drawn randomly from the populations and average them to obtain the averaged MDc (avMDc), the averaged MDc coincides always with the population distance MDpc. However, this is strictly valid if the variances have been computed via simulations. Therefore, there is need to verify that MDc is indeed an unbiased estimator. For this reason, we may use the scatterplot displaying the MDpc and avMDc for all pairs of populations/samples or the corresponding SSD values, i.e. the SSD values between MDpc and avMDc. Moreover, as for the MDs, we easily find out that the MDpc and avMDc vary from the MDc by the quantity $\Delta D = \Delta V + E[MD] - MDp$. Therefore, if MDpc and avMDc are corrected by subtracting ΔD, they will coincide with the MDc.However, we should note thatto use this correction for the MMD under Anscombe and Freeman-Tukey transformations, the population MDp should be computed using these transformations with the sample size and not the population size. We denote this population distance by MDp.adj.

3.2.3. Software

For each distance measure presented in this study two homemade functions have been written in R, except for the UMD for which only one function is necessary. The first function estimates the MDc using several preset numbers of iterations N_{iter}. From the obtained sets of MDc we can easily select the optimum N_{iter}, which is the minimum N_{iter} value that gives converged MDc values with the highest used N_{iter}. Such a function is not necessary for the UMD, since for this measure MDc = MD. The second function calculates among others: a) the MDs between all pairs of samples of the original binary dataset; b) the population measures of divergence MDp and E[MD]; c)the averaged distances avMDs between the samples randomly drawn from the populations; d) the distances MDc and avMDc, e) the standard deviations of MDs and avMDs; and f) SSD values between several pairs of distances. In addition, the functions provide p-values based on three standard approaches making use of the chi-squared distribution, the normal distribution, and the first four distance moments of the distance distribution. Note that the p-values are strictly valid under the assumption that the traits are independent. All functions, except for that of the UMD, have the option to subtract from the simulated distances the quantities ΔV and ΔD. For the UMD there is no need to subtract ΔD.Note that in all above computations the variances are estimated via simulations but there is also the option to use the approximate relationships discussed in Nikita and Nikitas (2021). All software material, along with detailed instructions, is available from the authors upon request and at the Zenodo repository (DOI: 10.5281/zenodo.5798100).

4. Results and Discussion

4.1. Number of iterations

First, we searched for the optimum number of iterations, N_{iter}, i.e. the number of samples that are randomly drawn from the populations, used in the computation of the MDc. For this purpose, we can use plots of MDc vs. distance pairs at various values of N_{iter} as well as plots of MDc vs. N_{iter} at various distance pairs. Such a plot is shown in Figure 1 for the MMDc under the Anscombe transformation. We observe that the convergence to a constant MDc value is achieved using relatively low N_{iter} values. In general, from the tests carried out in the 10 datasets adopted in the present study we found that a good value for N_{iter} is 1000, although for the LO dataset there was a small number of distances requiring slightly higher N_{iter} values and for this reason the value N_{iter} = 2000 was used for all pairwise MDc of this dataset.

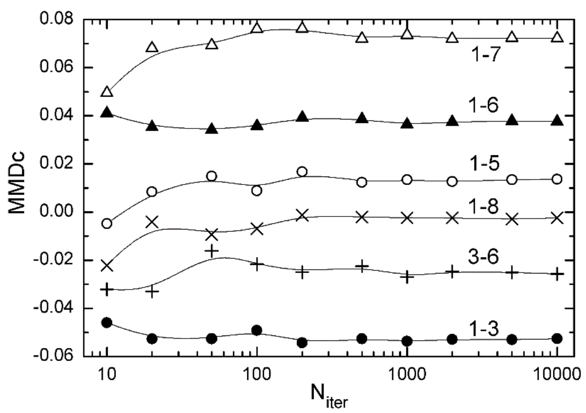


Figure 1. Plot of MMDc computed using Anscombe's transformation at various pairs of samples of the EU-16 dataset versus N_{iter}. Variances are estimated via simulations.

4.2. Distance biases

As pointed out in section 3.2.2, the performance of a distance measure as an unbiased estimator of population divergence can be estimated from scatterplots and the SSD values. For the MD, the scatterplots should display the pairwise measures MDp and avMDs, whereas for the MDc they should display the measures MDpc and avMDc. These plots may also include the measures MDs, E[MD], and MDc, respectively.

Selected scatterplots and SSD values are presented in Figures 2 to 5 and in Tables2 and 3. From the plots and the SSD values, we observe the following. The UMD is the only distance measure that is a strict unbiased estimator of population divergence under all conditions, from φ = 0 to φ = 1, irrespective of using simulated or not variances. This property makes its application simple and fast as it can employ non-simulated variances. The new proposed corrected measure MDc, i.e. the PMDc, LMDc, and MMDc using various arcsine transformations, is also a strict unbiased estimator, however, only when the variances are computed as accurately as possible via simulations. The conventional measures of divergence, i.e. those defined from Eq. (1), exhibit the following characteristics. The use of approximate formulas to estimate variances should exclude relative frequencies equal to 0 or 1 and frequencies close to these extreme values, which usually demands the deletion of a great number of traits. However, even in the range of φ values which excludes very low/high frequencies, relatively high SSD values may be detected (see for example the Smith MMD when applied to the AF-14 and EU-16 datasets in Table 3), which entails biased distance estimations. For this reason, a test that examines whether a distance measure is an unbiased or nearly unbiased estimator of population divergence is necessary. The replacement of approximate variances by simulated ones yields a considerable improvement of the results in the sense that all the measures of divergence discussed here become nearly unbiased estimators.

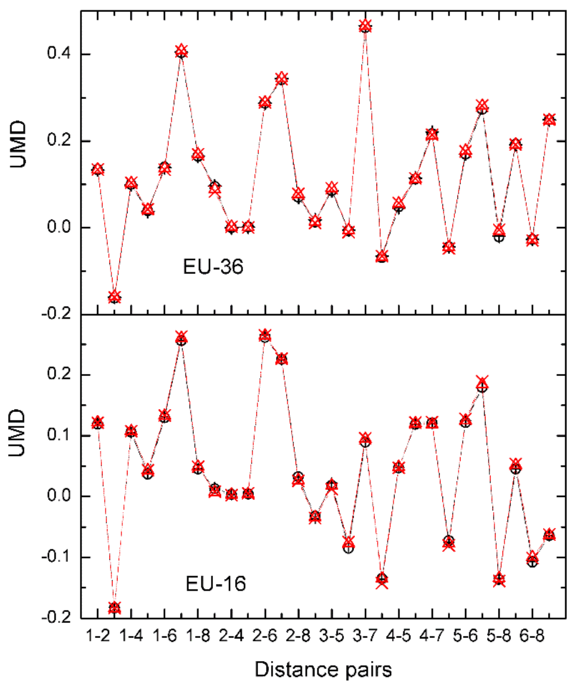


Figure 2. Plots of UMDs (o), UMDp (+), E[UMD] (Δ), and avUMDs (×)estimated on the EU-36 and EU-16 datasets.

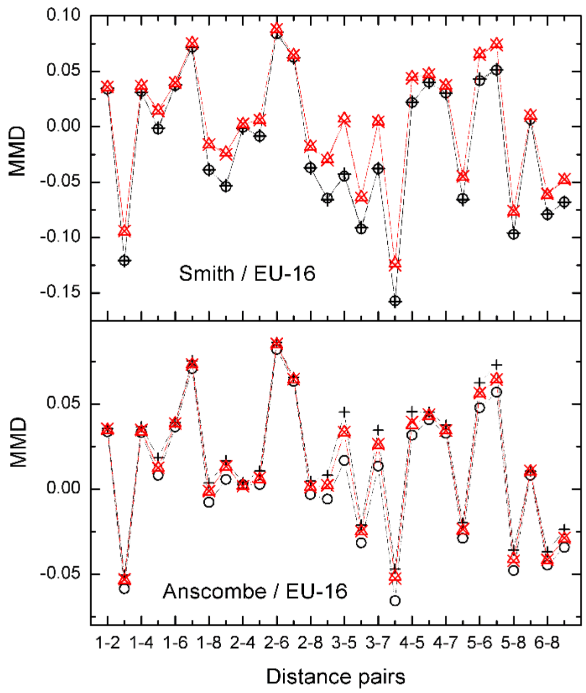


Figure 3. Plots of MMDs (o), MMDp (+), E[MMD] (Δ), and avMMDs (×)estimated on the EU-16 dataset using the Smith and Anscombe transformations.

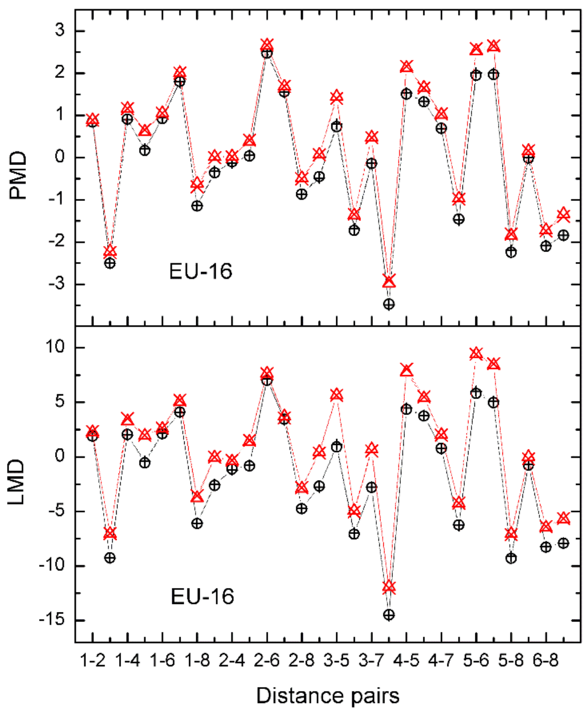


Figure 4. Plots of PMDs (o), PMDp (+), E[PMD] (Δ), avPMDs (×) (top), LMDs (o), LMDp (+), E[LMD] (Δ), avLMDs (×)(bottom) estimated on the EU-16 dataset.

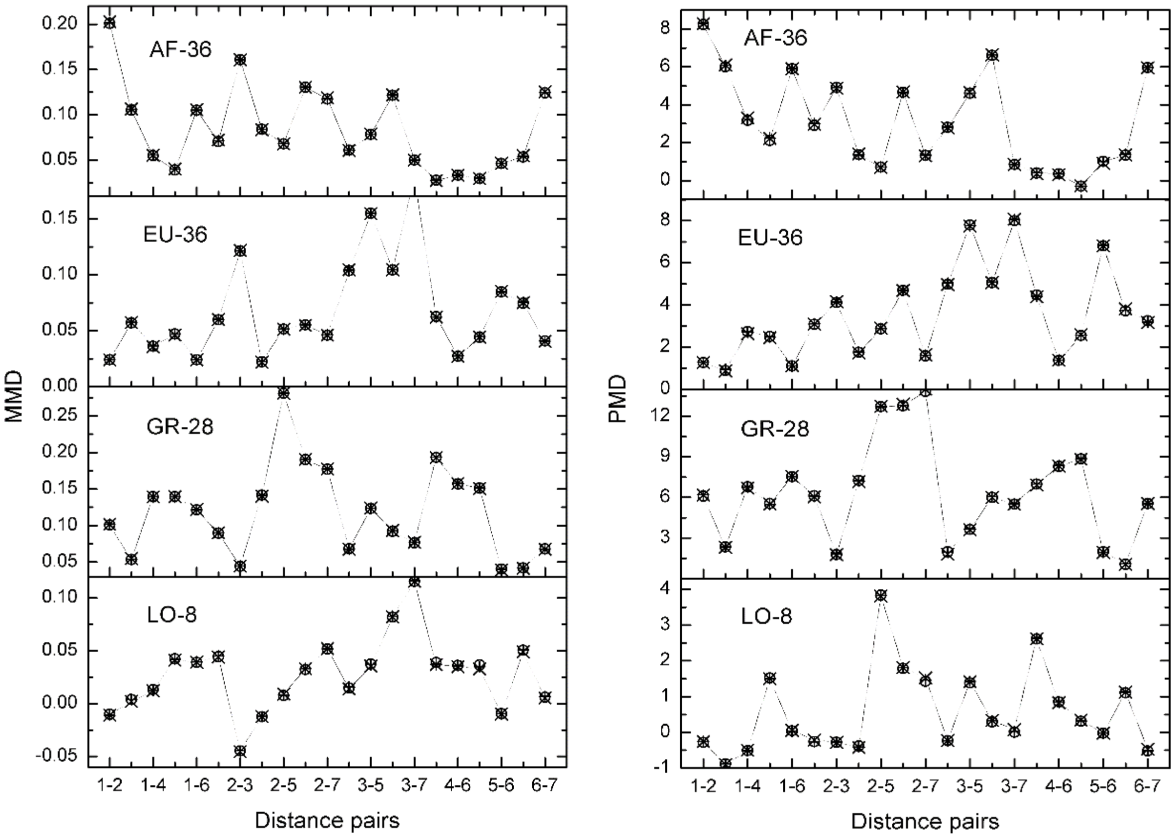


Figure 5. Plots of MMDc (o), MMDpc (+), avMMDc (×) (left) and PMDc (o), PMDpc (+), avPMDc (×) (right) estimated on the AF-36, EU-16, GR-28, and LO-8 datasets.

Datasets:	**AF-36**	**EU-36**	**GR-28**	**LO-8**	**MEX-13**
SSD between MDp (or MDp.adj for the Anscombe and Freeman-Tukey (FT) transformations) and avMDs					
UMD	0.0464	0.0077	0.0026	0.0041	0.0032
PMD	0.1796	0.3370	0.0885	0.0614	0.1162
LMD	0.3274	0.5659	0.1594	0.1072	0.1964
MMD(Smith)	0.4311	0.2083	0.0841	0.2036	0.2517
MMD(Anscombe)	0.0563	0.0669	0.0192	0.0277	0.0438
MMD(FT)	0.2108	0.1659	0.0303	0.0705	0.1273
SSD between MDpc and avMDc					
PMD	0.0168	0.0059	0.0011	0.0010	0.0008
LMD	0.0042	0.0040	0.0007	0.0007	0.0006
MMD(Smith)	0.0155	0.0024	0.0018	0.0007	0.0002
MMD(Anscombe)	0.0057	0.0050	0.0005	0.0018	0.0003
MMD(FT)	0.0065	0.0049	0.0008	0.0011	0.0002

Table 2. Selected SSD values between MDp and avMDs, and MDpc and avMDc when variances are estimated via simulations

Datasets:	AF-36	AF-14	EU-36	EU-16	GR-28	GR-12
SSD between MDp (or MDp.adj for the Anscombe and Freeman-Tukey (FT) transformations) and avMDs						
UMD	0.0706	0.0166	0.0179	0.0078	0.0060	0.0083
PMD	68.6410	1.3250	37.6392	0.1053	33.8548	0.8400
LMD	78.5198	3.6824	40.2800	0.6629	35.2913	2.0660
MMD(Smith)	1.5390	24.8204	0.5183	13.1246	0.2648	1.9970
MMD(Anscombe)	16.0384	0.2744	8.9356	0.1442	8.9356	0.1442
MMD(FT)	17.3181	0.2091	9.6644	0.0722	9.6644	0.0722
SSD between MDpc and avMDc						
PMD	68.4278	0.1746	37.1003	0.2880	33.1999	0.0423
LMD	78.2787	0.5055	39.6588	0.1960	34.0877	0.1354
MMD(Smith)	0.6955	11.2768	0.1440	5.1541	0.1255	0.3920
MMD(Anscombe)	16.4435	0.9087	8.3260	0.3995	2.8874	0.0361
MMD(FT)	15.9401	0.1924	8.4956	0.0827	2.1626	0.0273

Table 3. Selected SSD values between MDp and avMDs, and MDpc and avMDc when variances are estimated via approximate formulas

Thus, from the above results we conclude that the proposed corrections PMDc, LMDc, and MMDc using various arcsine transformations, along with the UMD, are strict unbiased estimators throughout the φ range, whereas the conventional MDs defined from Eq. (1) with simulated variances are nearly unbiased estimators and, therefore, they can be used to assess population divergence but cautiously and only after checking the extent of biases.

5. Conclusions

The proposed approach to estimate binary measures of divergence via Eq. (4) results in strict unbiased estimators of population divergence throughout the φ range, i.e. from $\varphi = 0$ to $\varphi = 1$, provided that the variances are computed via simulations. The conventional measures of divergence defined from Eq. (1), except for the UMD, which is always a strict unbiased estimator, are in most cases nearly unbiased estimators provided again that the variances are estimated via simulations. From this point of view, the use of simulated variances, although computationally more intensive, improves the results since it extends the range of applicability of Eq. (1) from $\varphi = 0$ to $\varphi = 1$ and reduces the biases. However, we should always check whether this reduction of biases is satisfactory or not. The UMD has the advantage that it is a strict unbiased estimator without the need to estimate the variances by simulations, which simplifies considerably its calculation.

Acknowledgements

Efthymia Nikita's contribution to this research was supported by the MetaMobility project. The project is implemented under the programme of social cohesion "THALIA 2021-2027" co-funded by the European Union, through the Research and Innovation Foundation [EXCELLENCE/0421/0376].

References

Buikstra, J.E. and Beck, L.A. (Eds.), 2017, *Bioarchaeology: the contextual analysis of human remains*. London: Routledge.

de Souza, P. and Houghton, P., 1977, The mean measure of divergence and the use of non-metric data in the estimation of biological distances. *Journal of Archaeological Science*, 4, 163-169. https://doi.org/10.1016/0305-4403(77)90063-2

Harris, E.F. andSjøvold, T., 2004, Calculation of Smith's mean measure of divergence for inter-group comparisons using nonmetric data. *Dental Anthropology*, 17, 83–93.

Irish, J.D., 2010, The mean measure of divergence: Its utility in model-free and model-bound analyses relative to the Mahalanobis D2 distance for nonmetric traits. *American Journal of Human Biology*, 22, 378–395. https://doi.org/10.1002/ajhb.21010

Konigsberg L.W., 1990, Analysis of prehistoric biological variation under a model of isolation by geographic and temporal distance. *Human Biology*, 62, 49-70.

Nikita, E., 2015, A critical review of the Mean Measure of Divergence and Mahalanobis Distances using artificial data and new approaches to estimate biodistances from non-metric traits. *American Journal of Physical Anthropology*, 157, 284-294.

Nikita, E., 2017, O*steoarchaeology: A guide to the macroscopic study of human skeletal remains*. San Diego: Academic Press.

Nikita, E. and Nikitas, P., 2021, Measures of divergence for binary data used in biodistance studies. *Archaeological and Anthropological Sciences*, 13, 1-14. https://doi.org/10.1007/s12520-021-01292-6

Nikita, E. and Nikitas, P., 2022, Simulation methods for squared Euclidean and Mahalanobis type distances for multivariate data and their application in assessing the uncertainty in hierarchical clustering. *Journal of Statistical Computation and Simulation*92, 2403-2424. https://doi.org/10.1080/00949655.2022.2030338

Ossenberg, N.S., 2013, Cranial nonmetric trait database on the internet. *American Journal of Physical Anthropology*, 152, 551-553. https://doi.org/10.1002/ajpa.22377.

Pilloud, M.A. and Hefner, J.T. (Eds.), 2016, *Biological distance analysis: forensic and bioarchaeological perspectives*. San Diego: Academic Press.

Santos, F., 2018, AnthropMMD: An R package with a graphical user interface for the mean measure of divergence. *American Journal of Physical Anthropology*, 165, 200-205. https://doi.org/10.1002/ ajpa.23336

Sjøvold, T., 1977, Non–metrical divergence between skeletal populations. The theoretical foundation and biological importance of C.A.B. Smith's mean measure of divergence. *OSSA*, 4 (suppl.), 1–133.

Zertuche, F. and Meza-Peñaloza, A., 2020, A parametric bootstrap for the Mean Measure of Divergence. *The International Journal of Biostatistics*, 16, 20190117. https://doi.org/10.1515/ijb-2019-0117

CONSIDERING SOCIAL NETWORK ANALYSIS: A BLACK SEA CASE STUDY INVESTIGATING TRADE DYNAMICS IN THE ANCIENT GREEK WORLD

H. BARTLET BALICKI[1] AND J. REMPEL[2]

[1]Institute of Nanoscience and Nanotechnology, N.C.S.R. "Demokritos", 15310 Aghia Paraskevi, Athens, Greece *h.balicki@inn.demokritos.gr,* [2] Faculty of Classics, University of Cambridge

Abstract

This paper presents the results of research conducted to investigate the contribution of network approaches when exploring past interaction and connectivity within trade, using Greek transport amphorae data. Using social network analysis (SNA), a tool used to investigate social structure, static amphorae distribution data are brought together in a relational and dynamic way. A specific dataset generated from amphorae data collected from nineteen sites around the Black Sea, between the seventh and first centuries BCE, was used in the SNA. The Black Sea was an integral part of the protoglobal ancient world, however, for historical reasons, this region has been secondary to the interests of western scholars. Using the Black Sea as a case study not only revealed patterns in trade dynamics in this region, but also served to bring Black Sea studies to the foreground when studying the ancient Mediterranean. SNA was conducted over two case studies, applying bipartite and co-occurrence network approaches. This approach transformed small and disparate datasets into a cohesive body of evidence, helping to reveal dynamic patterns within the dataset through network graphs and metrics. This research has provided a proof of concept for SNA as a tool that can be used to investigate patterns related to trade dynamics through transport amphorae distributions. The results help to understand nuances between methodological approaches to material culture networks investigating ancient trade and demonstrates how patterns revealed using SNA can encourage researchers to ask new and valuable questions about archaeological data.

Keywords: *social network analysis, Greek transport amphorae, Black Sea, Archaic period, Classical period, Hellenistic period, bipartite network, co-occurrence network*

1. Introduction

Understanding trade dynamics exclusively through material culture can be problematic, but analysis of its embedded role in networks offers crucial new perspectives to further knowledge of ancient interaction. Greek transport amphorae are considered the most representative finds for understanding changes in mechanisms for ancient trade and exchange (Lund 2007; Monakhov and Kuznetsova 2017), and their distribution data is often used to explain patterns in trade. Amphorae are highly typologised and often comprise high proportions of ceramic assemblages, making them obvious candidates for investigating trade patterns; however, issues surrounding ancient reuse and modern recovery and study biases mean that it is difficult to interpret amphora distributions as direct reflections of ancient long-distance trade (Lund and Gabrielsen 2005; Panagou 2015; Lawall and Graham 2018). Social network analysis (SNA) offers a new perspective to the interpretation of amphora data, as it emphasises the interaction and connectivity between actors rather than spatial distribution. By removing spatial constraints often imposed through amphora distribution analysis (e.g. Tzochev 2016) it is possible to identify patterns within a dataset that shed light on connectivity and trade dynamics.

This research builds upon a pilot project (Bartlet Balicki and Rempel in press) to investigate nuances in methodological application of SNA, primarily between bipartite and co-occurrence networks, to push the study of trade dynamics in the ancient Greek world into a novel way reflective of its dynamism. The results are a qualitative rather than quantitative interpretation of the amphora distributions (Lawall and Graham 2018), including preliminary conclusions about amphora producers and consumers in the network and a reflection on the analytical possibilities and new insights that can be gained when this approach is applied to larger datasets.

1.1 Trade Dynamics

Identification of trends in trade during a particular period and establishing the dynamics of import and export at a site are two key goals when studying trade dynamics (Monakhov and Kuznetsova 2017). Amphora distribution data sheds light on past connectivity, but interpretation of such data to measure strength of connectivity between the exporter and consumer (e.g. Bechtold 2008) is problematic when taken to represent direct shipments (Lawall and Graham 2018, 164). We know, however, that a variety of trading systems were in use in the ancient Black Sea, from direct shipments to down-the-line trade, as well as the predominance of cabotage along coastal routes (Davis et al. 2018). These trade dynamics can be investigated through shipwreck evidence, where cargoes can shed light on trends of produce being exported (Carlson 2012, 379); production centres, to interpret the scale and organisation of production and exportation (Whitbread 1995, 9); ancient sources, offering indications for the types of commodities being transported (see Table 1) and as evidence for scale of shipments (e.g. Demosthenes, *Against Lacritus,* 35.19-20). The often large quantities of imported ceramic material from coastal and inland sites have traditionally been the main tool for charting the development of ancient Greek commerce (Whitbread 1995, 24) but spatial and typological analyses have tended to rely on direct trading models. This paper uses SNA to investigate degrees of connectivity and centrality in the amphora assemblages from different Black Sea sites (two-mode) and within the assemblages themselves (one-mode).

1.2 Social Network Analysis

The primary aim of SNA is to recognise and interpret patterns and the implications of these relations (Wasserman and Faust 1994; Malkin et al. 2009; Brughmans 2010). Network approaches are relational as they focus on the relationships between social entities, rather than individual actors themselves (Mills 2017). Social entities, or actors, are represented as nodes, and the connections and relations between actors are referred to as edges (Brughmans 2010; 2013; Knappett 2016). Each connected pair of nodes and their relations is known as a dyad.SNA aims to reveal social structure within networks by presenting different forms of centrality, cliques, or structural holes (Burt 1992; Knappett 2016).

Early applications for studying the ancient Greek world in the first millennium BCE were primarily metaphorical (e.g.Horden and Purcell 2000; Malkin 2003, 2011; Broodbank 2013), but can span to formal, mathematical predictions (e.g. Verhagen et al. 2013; Rivers et al. 2013; Brughmans et al. 2015). Formal network analysis acts as a tool to highlight patterns within datasets which are identified without human bias, thus helping archaeologists to interpret complex datasets and thus provoke interesting research questions (Östborn and Gerding 2014); narrow research questions (e.g. Lawall and Graham 2018) and formulate new ones (Leidwanger *et al.* 2014); it can also offer new insights to old data and enable archaeological hypotheses to be tested as a network (Brughmans 2010).

Table 1 – Trade commodities from the Black Sea region attested in ancient sources.

Commodity	Ancient source
Cattle	Polybius (4.38.4)
Fish	Polybius (31.24, 4.38.4); Strabo (7.4.6, 7.6.2)
Goatskins	Demosthenes (*Against Lacritus* 35.34)
Grain	Demosthenes (*Against Leptines* 20.31); Isocrates (*Trapeziticus* 17.3-4); Herodotus (7.147); Polybius (31.25; 4.38.4); Strabo (7.4.6); Xenophon (*Anabasis* 6.1.16)
Honey	Pliny (21.45); Polybius (4.38.4); Strabo (11.2.17)
Linen	Strabo (11.2.17)
Oil	Polybius (4.38.4); Strabo (12.3.12)
Salt-fish	Demosthenes (*Against Lacritus* 35.34); Polybius (4.38.4)
Slaves	Herodotus (3.9.7, 5.6.1); Polybius 4.38.4); Xenophon (*Anabasis* 7.4.2)
Wax	Pliny (21.45); Polybius (4.38.4); Strabo (11.2.17)
Wine	Demosthenes (*Against Lacritus* 35.34); Polybius (4.38.4; 4.56); Xenophon (*Anabasis* 6.1.15)
Wood	Polybius (5.88); Strabo (11.2.17; 12.3.12); Theophrastus *(Enquiry into Plants* 4.5.5)
Wool	Demosthenes (*Against Lacritus* 35.34)

Formal SNA has only recently been attempted in relation to historical periods of the ancient Greek world (e.g. Donnellan 2016a; Greene 2018; Lawall and Graham 2018;Cline 2020), and only in a limited way in relation to the Black Sea region (notable exception, Donnellan 2021). This is an unfortunate omission, as alongside the well-documented evidence for trade between the Black Sea and the Aegean, increasingly the local and regional trade within the Black Sea is being studied, emphasising the likely ad hoc and indirect nature of this trade (e.g. Saprykin 2017; Davis et al. 2018; Bekhter et al. 2019; Parmenter 2020; Lewis 2022).

1.3 The Black Sea case study

The Black Sea was an integral part of the protoglobal ancient Mediterranean world and ancient sources attest to the trade of various commodities (Table 1) but for historical reasons, this region has been secondary to the interests of western scholars (Donnellan 2016b). Amphora data examined in this research spans from the end of the seventh to the mid first centuries BCE (Table 2). The earlier period corresponds to a time of Greek 'colonisation' when, in the last third of the seventh and early sixth centuries BCE, numerous Greek settlements were founded around the coast of the Black Sea (Tsetskhladze 2002; Greaves 2007; Ivantchick 2017). By the fifth century BCE, Greek settlements were functioning as poleis (city-states) and a robust Black Sea network had developed; this intensified in the fourth and third centuries BCE, a particularly prosperous time for the region (Rempel and Doonan 2020). The end date of this study, the mid first century BCE, corresponds with the end of the Hellenistic period, before the changes in trade and economy brought in with incorporation of region in Roman empire. Using the Black Sea as a case study not only enabled patterns in trade dynamics in this region, but also served to bring Black Sea studies to the foreground when studying the ancient Mediterranean.

Table 2 – Period names and dates used in the dataset.

Period names	Dates
Archaic	600-480 BCE
Early Classical	479-401 BCE
Late Classical	400-324 BCE
Early Hellenistic	323-251 BCE
Middle Hellenistic	250-151 BCE
Late Hellenistic	150-31 BCE

2. Methodology

A dataset was generated for the Black Sea case study to establish the possibilities made available through differing SNA approaches. The amphora data selected for the research was gathered from nineteen sites in the Black Sea region through published archaeological site reports and summaries of amphora distributions (Figure 1; Table 3). The selected data was gathered from datasets of assemblages of amphorae at individual sites (e.g. Olbia; Orgame, Pichvnari, Vani) and amphora producers (e.g. Sinopean amphora stamps), and the Ereğli E shipwreck from the Turkish coa st of the Black Sea . Ereğli E was included to test the applicability of integration of multiple source types into a large complex dataset. Using multiple source and site types, representing different processes within trade, is a way of bridging the multi-scalar nature of human society (Knappett 2011). The validity of the integration of multiple site types within one large dataset is tested against ideas of SNA being a useful tool to conduct multi-scalar research.

Table 3 – Black Sea sites included in the dataset and reference to publication containing amphorae data.

Map ID	Black Sea site	Coastal region	Source
1	Pistiros	West	Tušlová et al. (2010)
2	Vani	East	Akhvlédiani (2010, 141)
3	Berezan	North	Monakhov (1999); Chistov (2018, 93-96)
4	Olbia	North	Lawall et al. (2010, 356-405)
5	Panskoe 1	North	Kac et al. (2002, 107); Stolba (2012, 235-241)
6	Gorgippia	North	Monakhov and Kuznetsova (2017, 97-8)
7	Elizavetovskoe	North	Monakhov and Kuznetsova (2017, 97-8)

8	Phanagoreia	North	Monakhov and Kuznetsova (2017, 87-99)
9	Volna-4	North	Monakhov and Kuznetsova (2017, 87-91)
10	Beregovoĭ-4	North	Monakhov and Kuznetsova (2017, 87-91)
11	Chubovo	North	Monakhov and Kuznetsova (2017, 87-89)
12	Hermonassa	North	Monakhov and Kuznetsova (2017, 87-89)
13	Labrys	North	Monakhov and Kuznetsova (2017, 97-8)
14	Krasnoarmeĭskoe-1	North	Monakhov and Kuznetsova (2017, 89-91)
15	EreğliE shipwreck	South	Davis et al. (2018, 63)
16	Sinope	South	de Boer (2007, 8; 2013, 110); Garlan (2007, 145)
17	Apollonia Pontica	West	de Boer (2008, 119)
18	Orgame	West	Lungu (1992, 71)
19	Pichvnari	East	Akhvlédiani (2010 139-141)

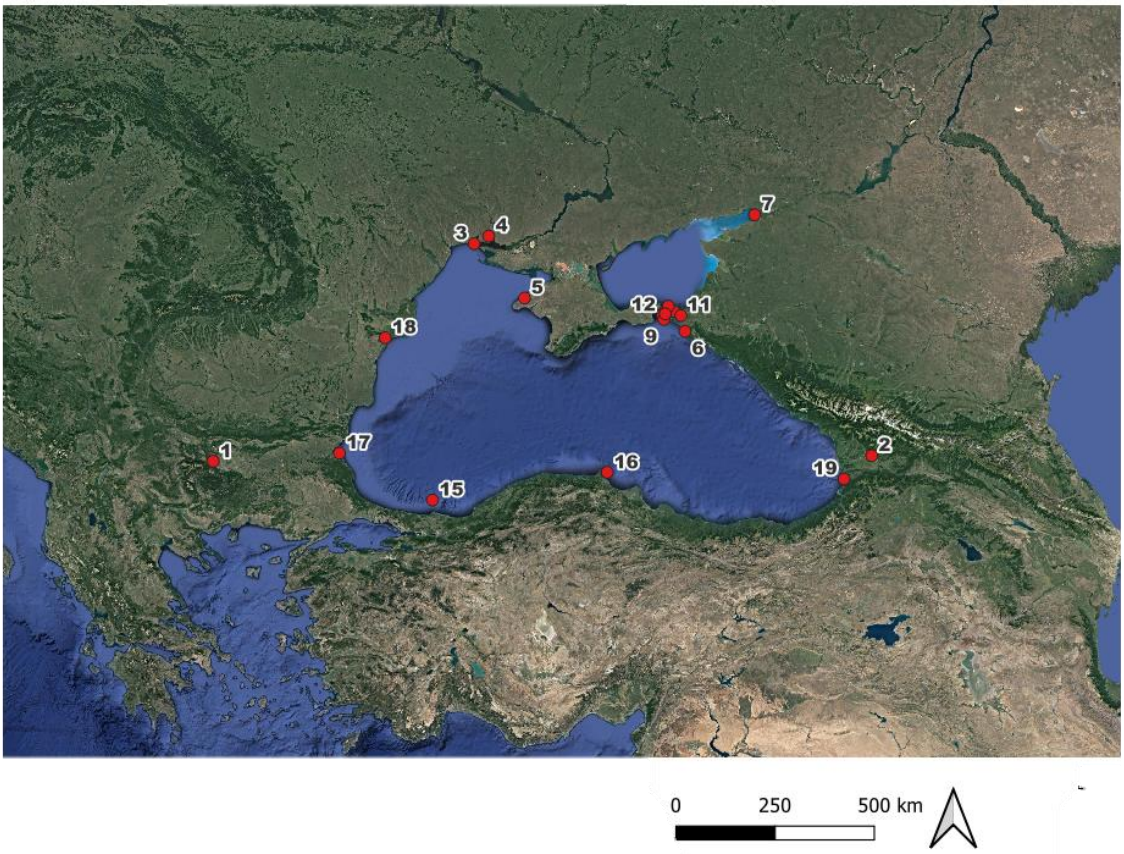


Figure 1 – Map of sites in the Black Sea region included in the dataset.(1) Pistiros; (2) Vani; (3) Berezan; (4) Olbia; (5) Panskoe 1; (6) Gorgippia; (7) Elizavetovskoe; (8) Phanagoreia; (9) Volna-4; (10) Beregovoĭ-4; (11) Chubovo; (12) Hermonassa; (13) Labrys; (14) Krasnoarmeĭskoe-1; (15) Ereğli E shipwreck; (16) Sinope; (17) Apollonia Pontica; (18) Orgame; (19) Pichvnari.

The final dataset included 12176 amphorae (see Bartlet Balicki and Rempel 2024), with a dominance of data from the Late Classical and Hellenistic periods. Data was coded by including appropriate numerical weighting, chronological slicing and geographic labels. The dataset was compiled in a spreadsheet and included origin, findspot, chronology and quantities of amphora type. Careful consideration was given when assigning data as actors (nodes) and when deciding which relations between actors (edges) were important and visible in the amphora data, as cautioned by Knappett (2013).

The SNA software package used for this research was UCINET (Borgatti et al. 2002), which offers suitable analytical capabilities and clarity when presenting the results of bipartite network analyses and has the capacity to convert bipartite networks into co-occurrence networks by creating co-occurrence matrices automatically for the researcher from two-mode edge lists. UCINET also has the capacity to analyse bipartite networks by node type (separated by context/affiliation). This is useful when studying different node types within a network and research on packages of material culture could benefit through the use of this software (e.g. Arthur et al. 2018).

Two case studies are presented in the paper, which explored the use of bipartite and co-occurrence network methods in relation to archaeological research focusing on trade dynamics. Two-mode (bipartite) networks are useful for looking at patterns relating to the dynamics of import at a site and enabling comparisons of import patterns at particular sites within a dataset. One-mode (co-occurrence) networks focusing on amphora as actors can be used to reveal packages of amphora types around the Black Sea, useful for studying trends in trade during a particular period.

Visual inspection of network graphs is the first step in understanding SNA results in a qualitative way. As a result of the principles underlying SNA, the distance between nodes plotted on network graphs is indicative of the strength or number of ties between actors (Brughmans 2010). Related nodes are drawn closer in the network structure (de Nooy et al. 2005). This helps identify nodes with similar characteristics and those with a strong or weak position.

Exploration of statistical results helps to test hypotheses generated from visual inspection (Brughmans 2010). Centrality measures (Table 4) are the most popular in archaeological SNA (Brughmans 2013; Mills 2017) determining ‘how well connected a node is within its local environment’ (Malkin et al. 2009, 5). We argue that the best use of SNA in archaeological research is as a tool to explore patterns in a data set which allows the researcher to investigate different aspects of the project.

Table 4 – Summary of centrality measures.

Measure	Description
Local centrality	Considers the number or proportion of nodes a specific node is connected to
Global centrality	Determined by the distances between nodes in a graph
Relative centrality	Based on how much a node is ‘between’ other nodes in the same graph
Betweenness centrality	A node’s centrality is based on how frequently it is part of the shortest distance between other nodes
Closeness centrality	A node is central if it can interact quickly with all others
Eigenvector centrality	Node centrality is based on how important or influential their position is within the network compared to other nodes

3. Case Study 1: Bipartite Networks

This case study presents a series of relational bipartite networks, representing amphorae distribution patterns from the full dataset, in 6 time-sliced graphs (Figures2-7).Time-slicing networks is one approach to tackling a multi-scalar approach (Östborn and Gerding 2014). Node shape and colour distinguishes whether the node represents a site (blue square) or amphora (red disc).

3.1 Visual inspection

For the Archaic period (Figure2) there are two clusters within the network. The amphora cluster represents amphora types circulating in the Black Sea, present at every site in the Archaic dataset: Chian, Klazomenian, Lesbian, Samos-Milsian, and Unattributed. In the Early Classical network (Figure3) Thasos, Chios, Thasian Circle and Heraclea Pontica are amphora production centres involved in the central cluster.

In the Late Classical network (Figure4), amphorae from Sinope, Heraclea Pontica and Chios become central as they connect Olbia to the rest of the network. Sinope is connected to all sites except Pistiros and Pichvnari, and many of the sites are only connected to Sinope. Pistiros and Pichvnari form a secondary cluster. Pistiros, located to the west of the Black Sea, and Pichvnari, located to the east, are unconnected themselves and geographically peripheral to the central cluster of sites are located on the north coast of the Black Sea. To consider the similarities between small, trade-oriented sites like Pistiros and Pichvnari opens possibilities for new questions about the mobility created through trade at 'outposts' in the Black Sea world.

Sinope is connected to every site in the Early Hellenistic network (Figure5). The dominance of Sinope in the dataset is in part due to its recognisable amphora stamps which have been catalogued, but also reflects a marked dominance of coarse wares from Sinope at Black Sea sites during this period (e.g. Inaishvili and Khalvashi 2010). Chersonesos, Thasos and Rhodes also connect two clusters of Black Sea sites.

In the Middle Hellenistic network (Figure6) visual inspection shows that the network is limited to Olbia, Sinope, Rhodes and Apollonia Pontica. This is a result of the initial dataset not being broad enough to capture potential networks. Figure 7 shows no meaningful network for investigating trade dynamics in the Late Hellenistic dataset; however, it can be understood as an ego-network (see Greene 2018), suggesting an alternative way that data could be further examined. For example, the Late Hellenistic ego-network could be expanded to include additional sites which share the same amphora types. This method is useful as a process for expanding datasets and directing research.

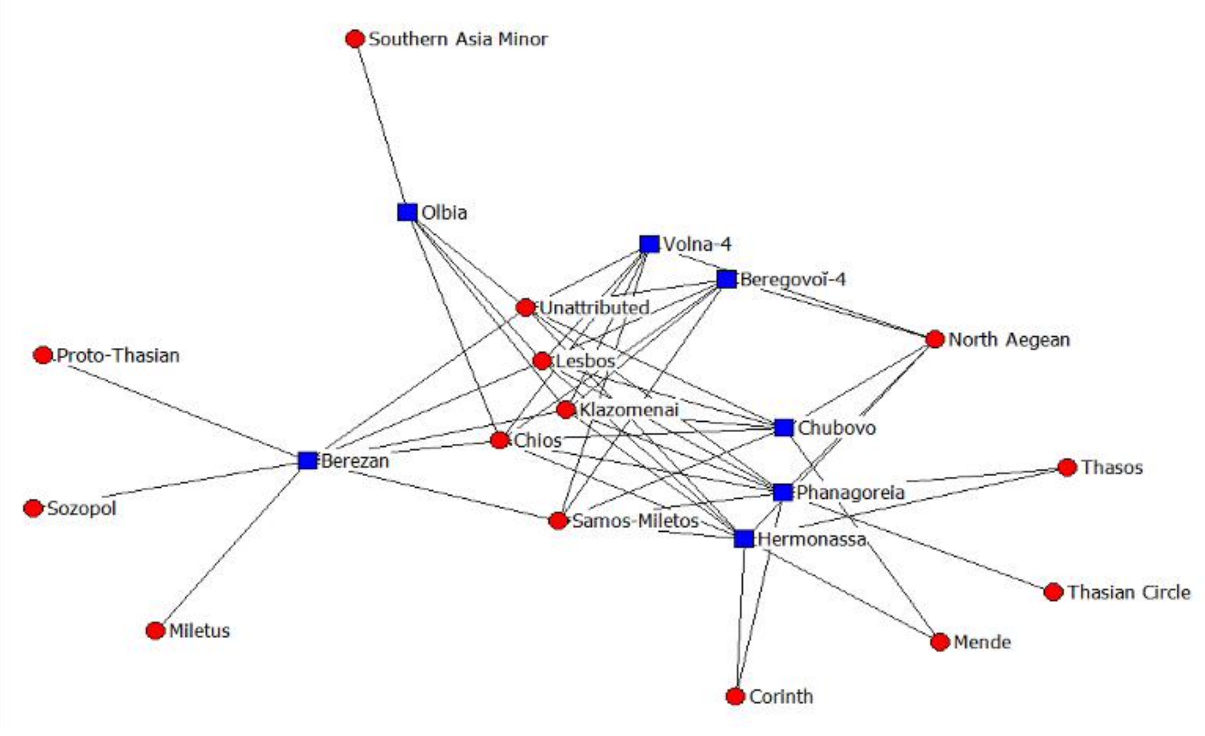


Figure 2– Archaic bipartite network graph.

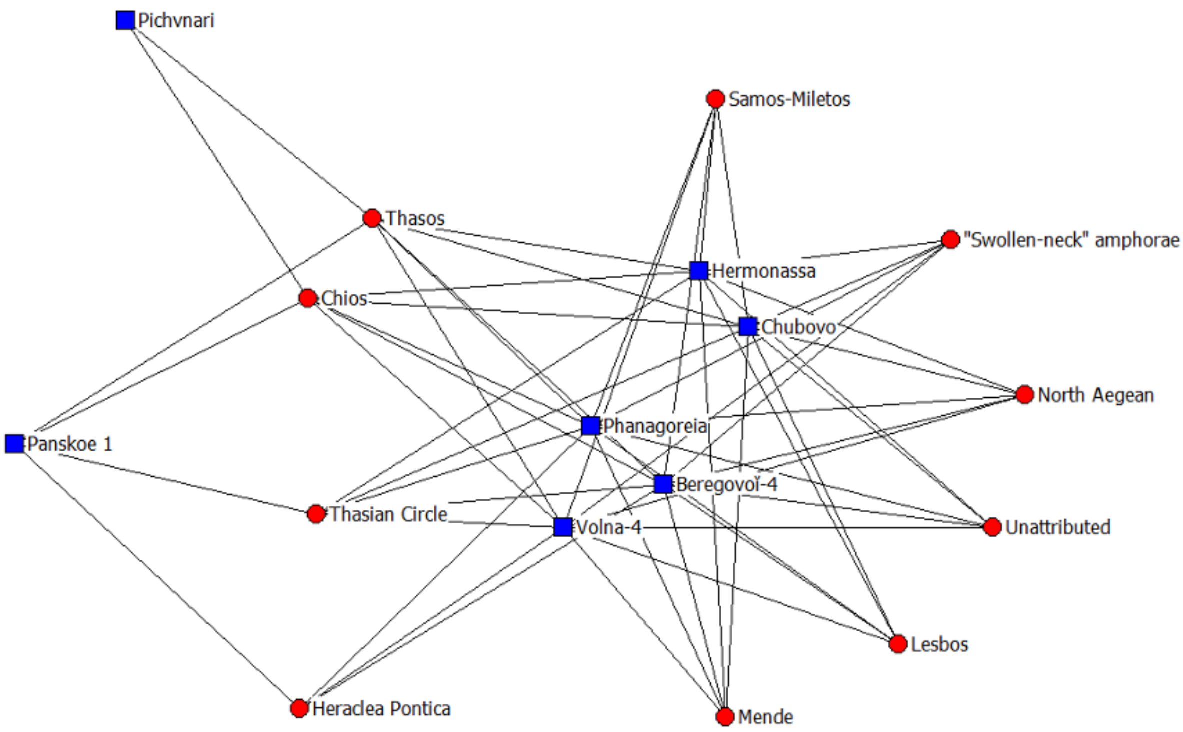


Figure 3– Early Classical bipartite network graph.

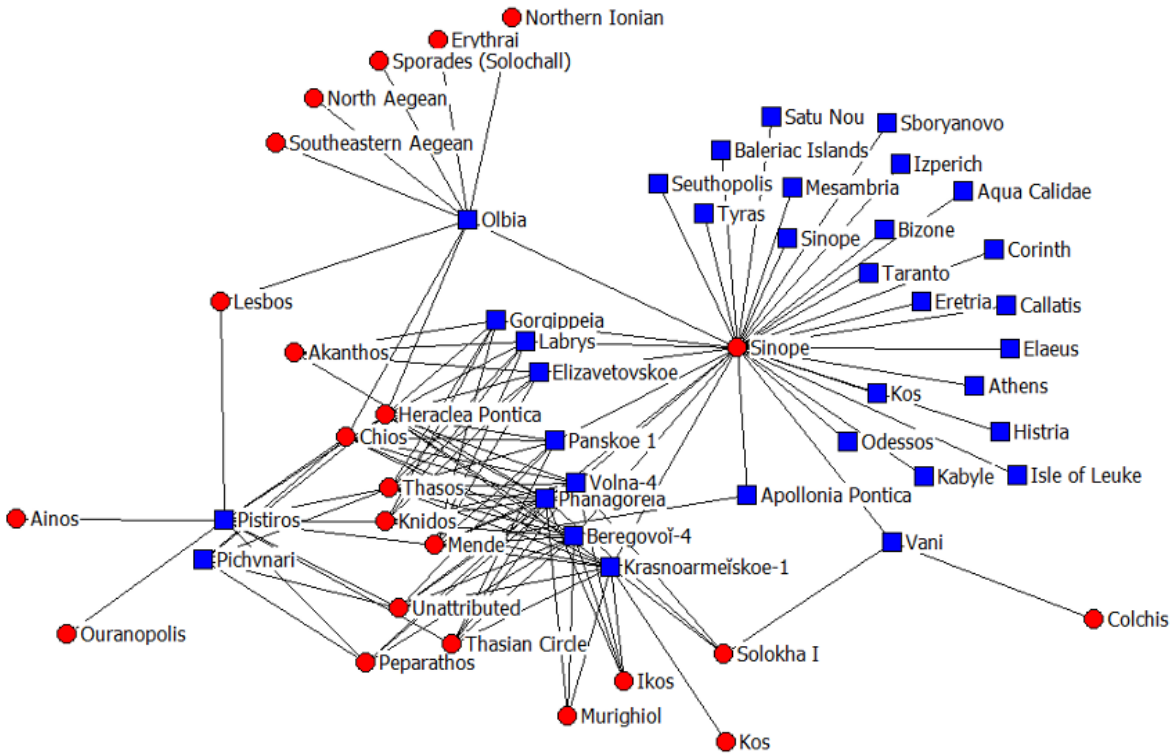


Figure 4 – Late Classical bipartite network graph.

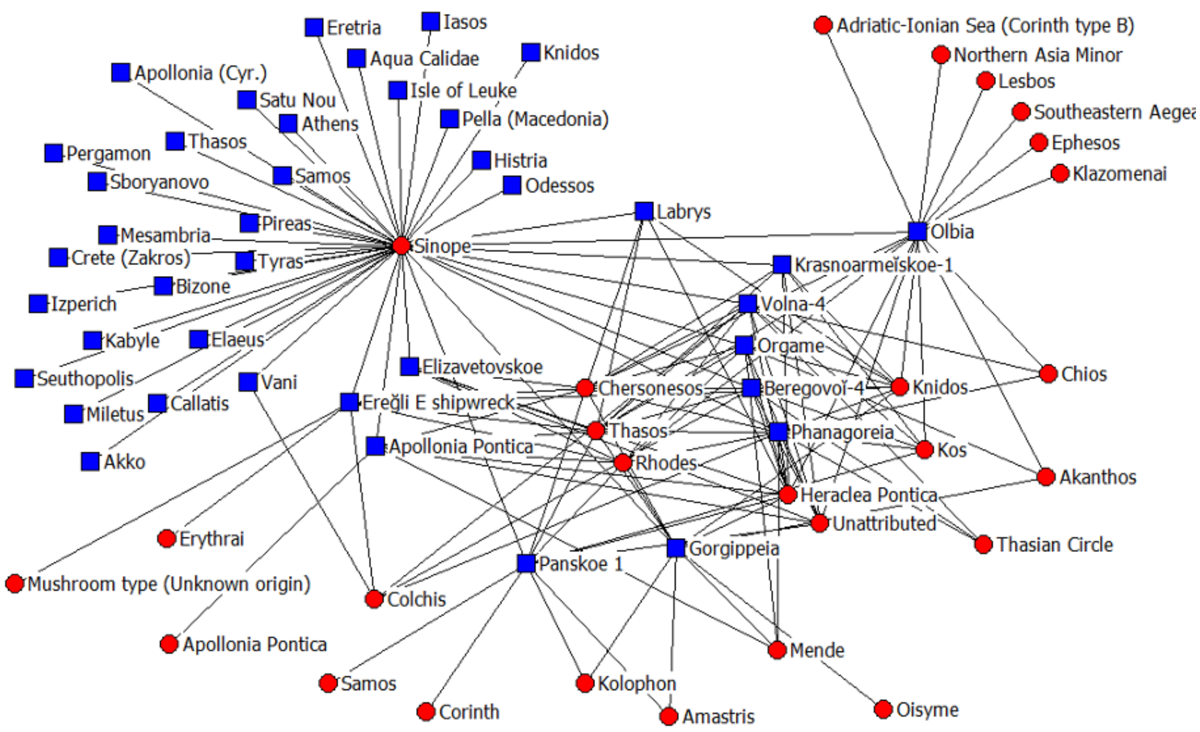


Figure 5 – Early Hellenistic bipartite network graph.

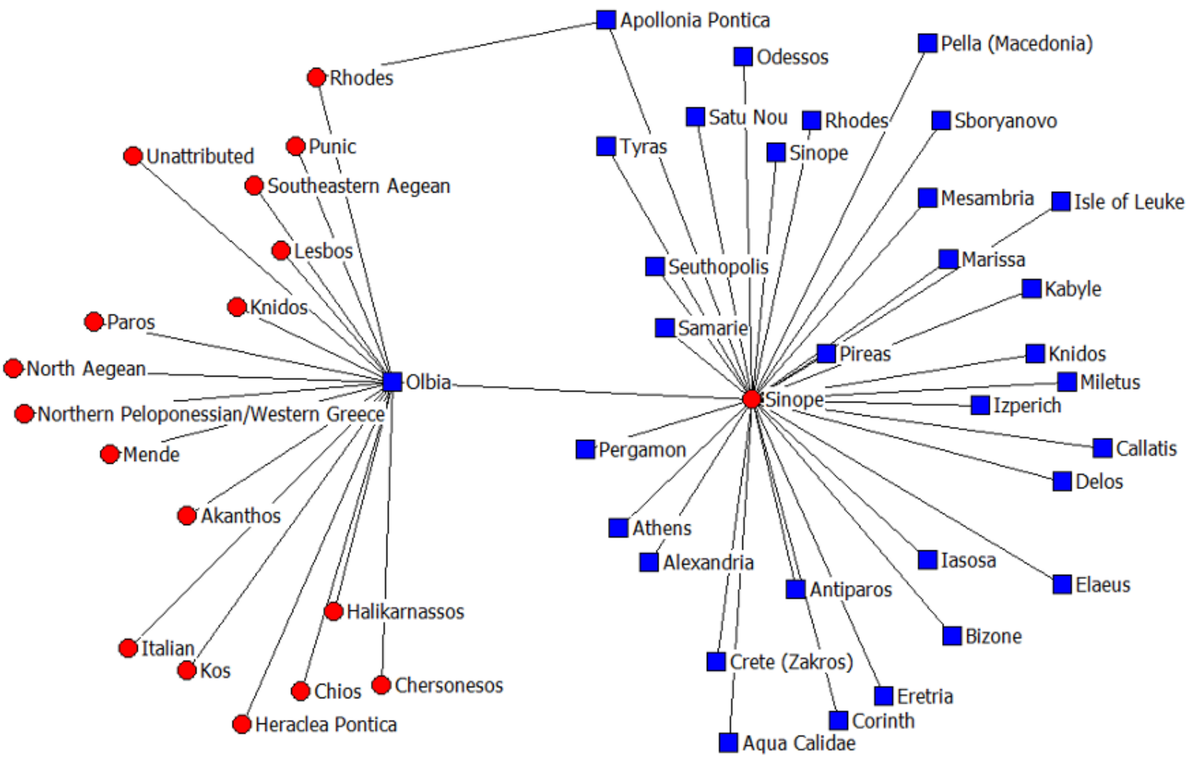


Figure 6– Middle Hellenistic bipartite network graph.

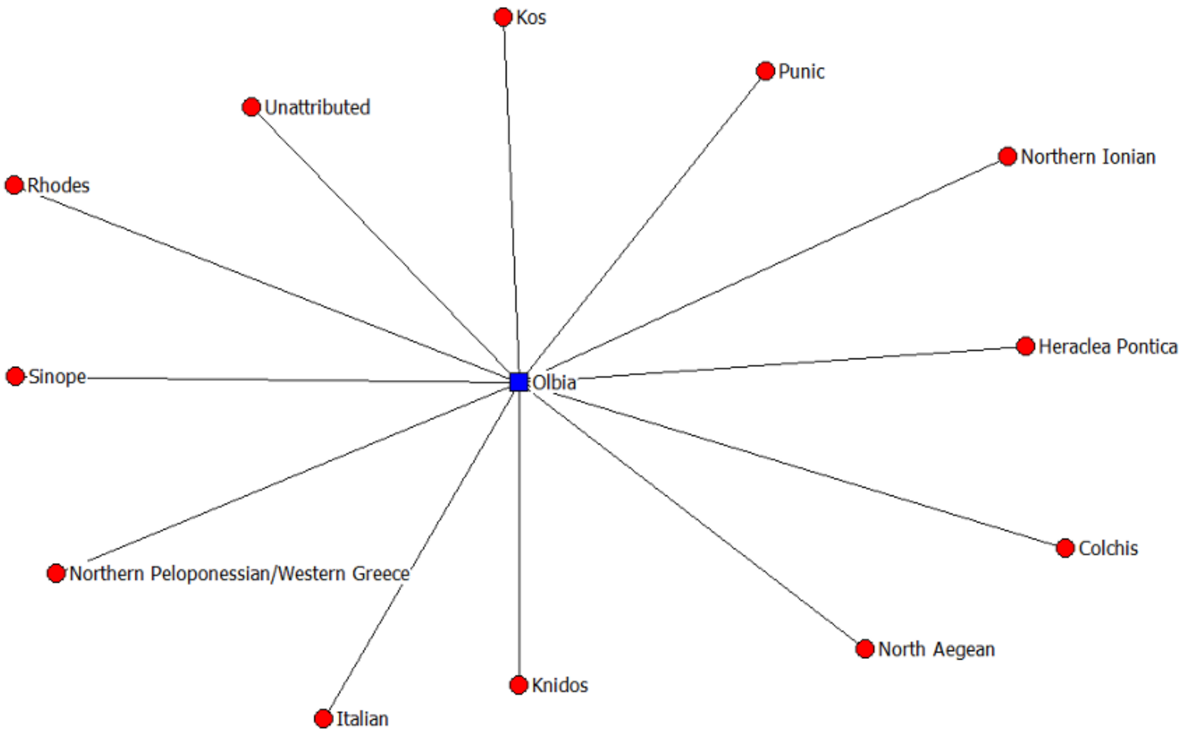


Figure 7 – Late Hellenistic bipartite network graph.

4. Case Study 2: Co-occurrence Networks

The second case study interprets the data through a co-occurrence network approach. This approach was taken to focus visual inspection of patterns within the network to focus primarily on the amphora types, to help consider questions of trends in trade dynamics based on the objects involved in trade. We hoped the use of co-occurrence networks of amphora types would lead to patterns of packages of amphorae to emerge. The co-occurrence network representing amphorae as nodes was analysed further to understand more about the attributes of their presence at different sites. The Late Classical and Early Hellenistic datasets are presented here due to their size.

4.1 Late Classical

There are two main clusters in the Late Classical co-occurrence network (Figure 8): Sporades, Northern Ionian, North Aegean, Erythrai, and South-eastern Aegean in one, and Knidos and the Thasian Circle in the other. The two main clusters are connected by Heraclea Pontica, Lesbos, Sinope and Chios. The betweenness centrality (Table 5) provides numerical values for this observation. Sinope, then Chios and Heraclea Pontica have the highest betweenness values indicating they connect the two main clusters the most. Lesbos has the next highest betweenness value, as it connects the two main clusters, but to a lesser extent. Solokha I is the next highest node for betweenness centrality; this is because these amphorae connect the third cluster (one node), Colchis, to the rest of the network.

Using betweenness centrality to highlight ‘packages’ of amphorae moving around the Black Sea indicates that amphorae from Sinope, Thasos, Chios and Heraclea Pontica were the most present amphora types across assemblages throughout the Late Classical network. Sinopean, Thasian, Chian and Heraclean amphorae may have been the most important amphorae in facilitating connectivity in Black Sea networks, as they are the amphora types connected with the most other types of amphorae.

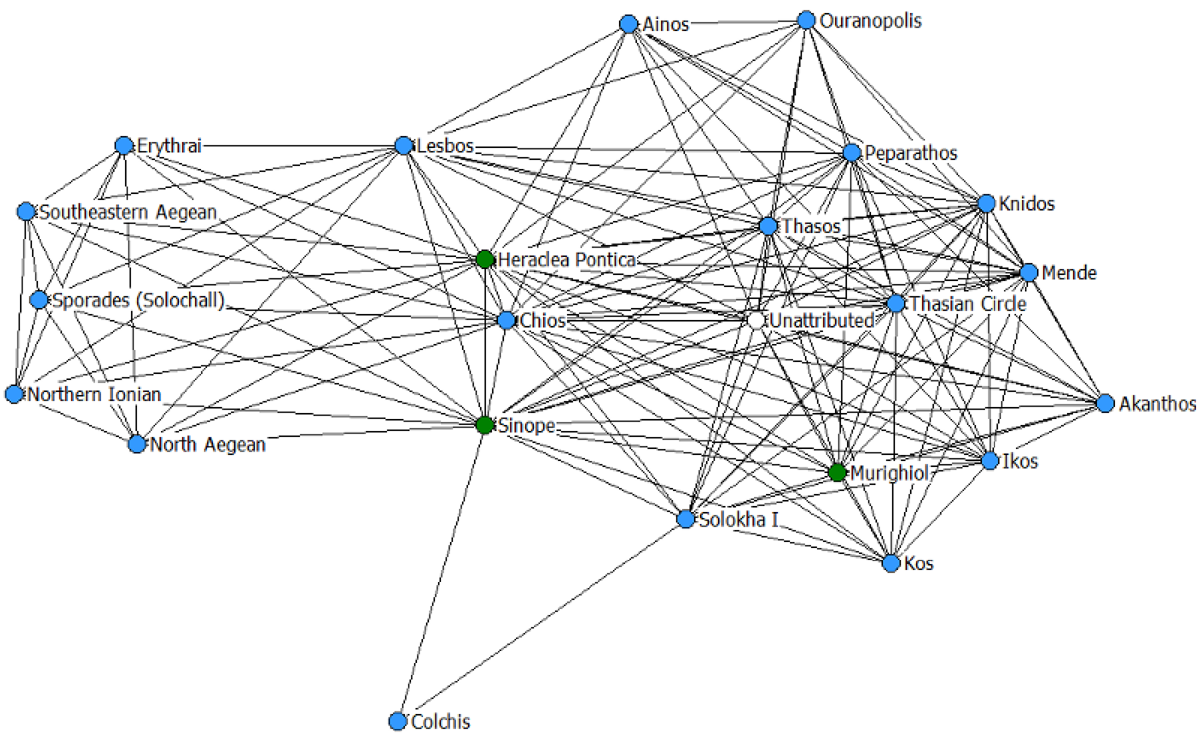


Figure 8 – Late Classical one-mode co-occurrence network graph (blue disc: amphora from Aegean centre; green disc: amphora from Black Sea centre; white disc: unknown origin).

Table 5 – Betweenness and Eigenvector centrality for Late Classical co-occurrence network.

Betweenness Centrality		**Eigenvector Centrality**	
Sinope	29.643	Thasos	1.000
Chios	21.661	Heraclea Pontica	0.520
Heraclea Pontica	21.661	Unattributed	0.397
Lesbos	11.173	Knidos	0.302
Solokha I	7.123	Chios	0.290
Knidos	2.494	Mende	0.259
Mende	2.494	Sinope	0.233
Thasian Circle	2.494	Peparathos	0.153
Thasos	2.494	Lesbos	0.119
Unattributed	2.494	Thasian Circle	0.085
Peparathos	1.902	Ikos	0.043
Ikos	0.182	Akanthos	0.021
Murighiol	0.182	Solokha I	0.016
Ainos	0	Ainos	0.013
Akanthos	0	Kos	0.009
Colchis	0	Murighiol	0.006
Erythrai	0	Ouranopolis	0.003
Kos	0	Colchis	0
North Aegean	0	Erythrai	0
Northern Ionian	0	North Aegean	0
Ouranopolis	0	Northern Ionian	0
Southeastern Aegean	0	Southeastern Aegean	0
Sporades (Solochall)	0	Sporades (Solochall)	0

4.2 Early Hellenistic

Visual inspection of the Early Hellenistic graph indicates that four amphora types from Black Sea centres are more prominent in the centre of the network, alongside five Aegean types (Figure 9).

In the Early Hellenistic period, amphorae from Rhodes, Sinope, Thasos and Chersonesos have the highest betweenness and eigenvector centralities (Table 6), showing they are the most active in connecting the network, and are the most influential amphora types. When looking at the Ereğli E shipwreck, we have direct evidence that Chersonesian, Sinopean, Chian, Thasian, Rhodian, Knidian and Mushroom type amphorae were being transported together (Davis et al. 2018). While the inclusion of this shipwreck does add to the centrality of these nodes, their co-occurrence at other sites within the Black Sea are the primary influencers of their centrality within the network. This method highlights the possibilities for patterns of centrality in co-occurrence networks to gain understanding as to which types of amphorae (and/or other commodities) may have been transported together. It is interesting to compare this with the Late Classical network and it begins to generate a sense of change over time. The Black Sea centres, Sinope and Heraclea Pontica are the only centres retaining their key position in terms of betweenness centrality, thus indicating the continued role of Black Sea centres in sustaining trade networks in the Black Sea region.

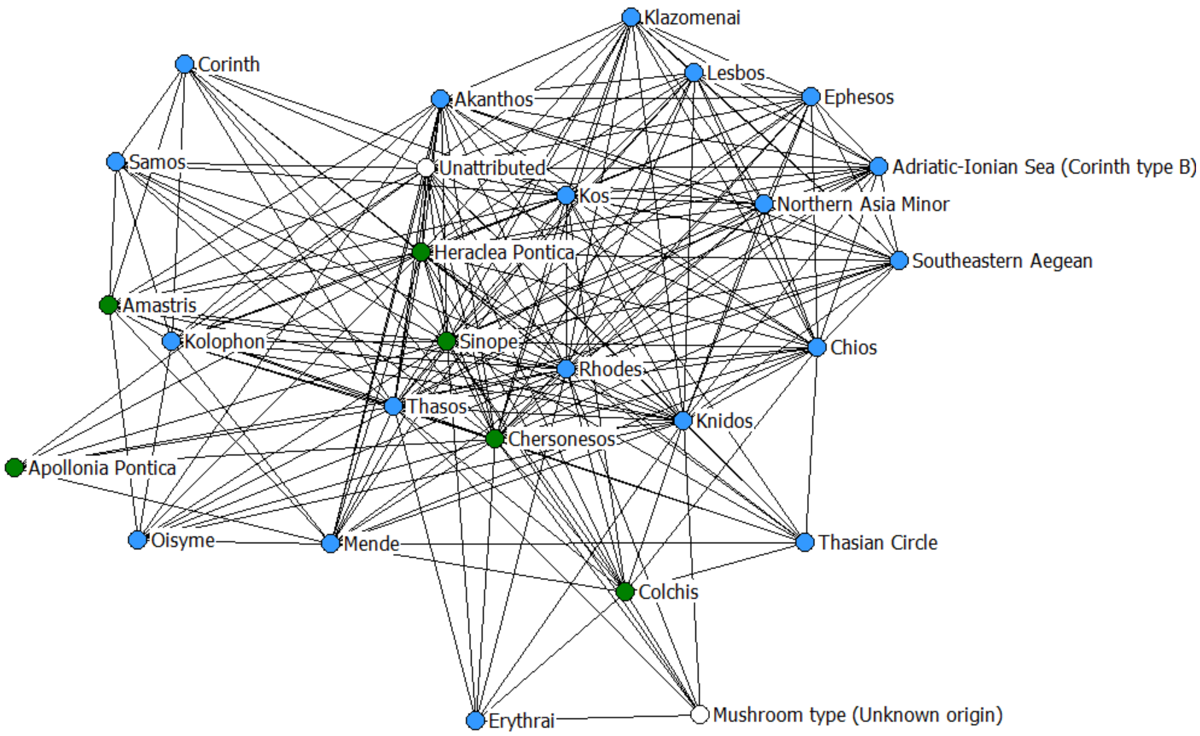


Figure 9 - Early Hellenistic one-mode co-occurrence network graph (blue disc: amphora from Aegean centre; green disc: amphora from Black Sea centre; white disc: unknown origin).

Table 6 – Betweenness and Eigenvector centrality for Early Hellenistic co-occurrence network.

Betweenness Centrality		Eigenvector Centrality	
Rhodes	19.246	Rhodes	1.000
Sinope	19.246	Sinope	1.000
Thasos	19.246	Thasos	1.000
Chersonesos	19.246	Chersonesos	1.000
Heraclea Pontica	11.746	Heraclea Pontica	0.967
Knidos	11.579	Unattributed	0.942
Unattributed	9.825	Knidos	0.932
Kos	7.349	Kos	0.932
Akanthos	3.315	Akanthos	0.840
Mende	2.765	Chios	0.809
Chios	2.133	Adriatic-Ionian Sea (Corinth type B)	0.714
Colchis	2.00	Klazomenai	0.714
Kolophon	1.152	Lesbos	0.714
Amastris	1.152	Northern Asia Minor	0.714
Apollonia Pontica	0	Southeastern Aegean	0.714
Erythrai	0	Ephesos	0.714
Mushroom type (unknown origin)	0	Mende	0.694
Oisyme	0	Kolophon	0.624
Samos	0	Amastris	0.624
Corinth	0	Colchis	0.575
Thasian Circle	0	Thasian Circle	0.541
Adriatic-Ionian Sea (Corinth type B)	0	Oisyme	0.478
Klazomenai	0	Samos	0.470
Lesbos	0	Corinth	0.470
Northern Asia Minor	0	Apollonia Pontica	0.363
Southeastern Aegean	0	Erythrai	0.321
Ephesos	0	Mushroom type (unknown origin)	0.321

5. Discussion

In the first case study, for time periods with sufficient data, SNA enabled recognition of key actors and relations in Black Sea trade networks. It provided a method of bringing together numerous sources of amphorae distributions in one cohesive analysis. Visualisation of the dataset allows influential nodes to be identified in terms of presence at sites and those in a position of controlling the flow of movement (of goods or information) in a network. Greek transport amphorae are acting as a proxy for certain types of connectivity and help to shed light on the dynamics of import at the particular sites in the dataset. SNA has the potential to illustrate new trends and patterns which have not been recognised before, offering a method for conducting original research on old data, thus enabling new research questions to be asked.

Overall, the visualisations confirm the general understanding of amphora consumption in the Black Sea, with eastern and northern Aegean vessels dominant until the Late Classical period, when Black Sea producers – Heraclea Pontica, Sinope, Chersonesos - become prominent (Monakhov and Kuznetsova 2017). Rhodes is also well-documented as a transhipment point in Hellenistic period (Garlan2007). But in addition, they demonstrate the existence of a dominant Aegean 'package' of amphora types in the Archaic period, which drove connectivity in the Black Sea network as well as the powerful shift in network connectivity with the introduction of Black Sea producers in the late Classical and early Hellenistic periods.

Due to the small size of the datasets for the Archaic, Early Classical, Middle Hellenistic and Late Hellenistic periods the results of the SNA and network graphs for these periods are only relevant to understanding trends within the dataset, as the datasets are not large enough to provide statistically meaningful quantitative interpretation of trade dynamics. However, more detailed analysis could be conducted for the Late Classical and Early Hellenistic periods, where more amphorae data are represented in the dataset. The import of Aegean amphorae is seen in the Archaic and Early Classical network graphs (Figures 2 and 3) and complements the trend reported by Monakhov and Kuznetsova (2017) who highlighted for the sixth and fifth centuries BCE for sites in the northern Black Sea region, imports to the Black Sea came mainly from the Aegean region. However, this observation cannot be extrapolated to suggest trends for the wider Black Sea region until a larger distribution of amphorae in these periods are analysed through SNA. This emphasis is in part a function of the data sets utilized, which include sites primarily occupied in these periods, but it also reflects the fact that the Late Classical and Early Hellenistic periods were a time of prosperity, increased agricultural production and robust trade in the Black Sea region (Rempel and Doonan 2020).

Formatting data into a co-occurrence network is the most appropriate method to focus on the archaeological material and was the approach used in the second case study. Using the amphora types as nodes, with edges representing the presence of nodes together at a site, can shed light on which amphora types are most central in the network. Here, the more central a node, the more likely it is to be circulating with other amphora types. SNA prompts us to ask further questions, such as, why were these nodes most central. For example, in the Late Classical dataset, Thasos has an eigenvector centrality of 1 despite having a low betweenness centrality of 2.494 (Table 5). This indicates its importance despite not being part of the shortest path for a lot of nodes, as it is connected to all the influential node in the network. Through visual inspection alone this is not visible, highlighting the importance of analysing metrics alongside graphs. The data around Thasian wine is regarded as particularly rich (Tzochev 2016) and bringing in ancient sources and inscriptions can shed light on the production and trade of Thasian wine (Salviat 1986; Tzochev 2016).

While it is true that Sinope has the highest betweenness in the bipartite network, as it is connecting sites with no other recorded amphorae to other sites, this observation is not significant when contextualising the SNA results in a broader understanding of ancient Black Sea trade dynamics, as the results derive from the formation of the dataset. Analysing a fuller dataset through co-occurrence networks of amphora types offers good potential for understanding trade dynamics.

The presence of Heraclea Pontica as second most influential in the Late Classical network, and having the second highest betweenness, is important to note. In the Early Classical co-occurrence graph metrics, Heraclea Pontica has the lowest eigenvector value (Table 7). The increase from the previous period from last to second in terms of influence, and to first for betweenness centrality, within the dataset reflects the general increase in regional trade within the Black Sea during this period and nuances our understanding of Heraclea Pontica as a driver.

Table 7 – Eigenvector centrality for Early Classical co-occurrence network.

Eigenvector Centrality	
Chios	1
Thasos	0.304
Unattributed	0.194
Mende	0.093
North Aegean	0.083
"Swollen-neck" amphorae	0.039
Samos-Miletos	0.038
Lesbos	0.031
Thasian Circle	0.027
Heraclea Pontica	0.014

The high number of connections seen in the Early Hellenistic period complements other archaeological research highlighting the robust trade networks in the Black Sea, Aegean and Mediterranean between the fourth and second centuries BCE (e.g. Archibald 1998; Fless 2008; Krapivina 2010; Monakhov and Kuznetsova 2017). By looking at data in different time periods it allows archaeologists to visualise aspects of networks changing over time.

When attempting to use SNA results to contextualise the past a co-occurrence network is better suited as it helps to remove bias within the dataset. For example, the large number of sites connected to Sinope, but no other amphora types seen in the bipartite network graphs only indicate the incompleteness of the dataset. However, when data was transformed into a co-occurrence network, where dyads represent the presence of two amphora types at least one site, the patterns displayed in the network graph shed light on amphorae most often found with other types. This is interesting to consider as it could help build an understanding of likely packages of amphorae being transported together in certain periods.

SNA allows multiple source types to be combined and analysed in a comparable way . The Ereğli E shipwreck was analysed alongside distribution data from archaeological sites, in addition to information on the distribution of one amphora type. The use of shipwreck data alongside import data is most useful in co-occurrence networks, where the aim is to look for patterns relating to trends in trade dynamics. Shipwreck data can help reinforce interpretations of packages of amphora types being transported together by providing direct evidence for such instances. The integration of multiple site types into one dataset is beneficial as often archaeological data is small and disparate. Shipwrecks are the only direct evidence to be certain of such understanding; however, this case study highlights the potential for understanding trends in trade dynamics and shipping practices where shipwreck evidence is not available. As more wrecks are discovered and datasets grow, such datasets become more accurate. A centralised dataset for Greek transport amphorae would provide an important source to further explore patterns and trends in trade dynamics using SNA.

The lack of amphora data for the Middle and Late Hellenistic periods is a reflection of the sources selected for the dataset and should not be inferred to represent a decrease in trade in these periods. It is widely recognised that there was a change in the Black Sea market around the middle of the third century BCE (Lazarov 1980; Avram 1996; Garlan 1999; Monakhov 1999), but this phenomenon is rarely discussed (Tzochev 2016). Within the scope of this research, it was not possible to offer further insight into the nature of this change, as the sources chosen did not cover this period extensively. Having a consolidated Greek transport amphora database would contribute greatly to research on both the ancient Greek world and Black Sea studies by making the extent of amphora data more widely known and accessible.

In Roman research, there are a number of substantial transport amphora databases including: Roman Amphorae (University of Southampton 2014); ICRATES (Bes et al. 2019); Potsherd (Tyres 2021); Recueil de Timbres sur Amphores Romaines (Université de Provence 2004); and Amphorae ex Hispania (Millet 2016). These databases provide extensive catalogues for Roman amphorae, amphora stamps and table ware sherds. For Greek transport amphorae, however, databases are not yet comparable to the scale of available Roman amphorae data, making compiling substantial datasets time consuming and difficult. Within the Black Sea, the new Amphora Stamps of the Northern Black Sea database (Kovalchuk 2020) offers great potential for both for Black Sea studies and broader studies relating to Greek transport amphorae. The database is accessible for researchers to add their own material and edit and supplement the 286 stamps published so far. Additionally, the Greek amphorae from Northern Pontus Euxinus database (Saratov State University 2018) has published 1195 vessels from Russian museums in its online catalogue.

It remains important to develop such databases for Greek transport amphorae across their entire geographic distribution. Quantified data for transport amphorae at many centres are fragmentary and difficult to source (Lawall and Graham 2018, 163). In addition, the Greek amphora data is not uniformly distributed chronologically; significantly more data is available for Late Classical and Early Hellenistic periods, while finds from the Archaic and Early Classical periods are more limited. Finally, varying protocol and techniques for recording amphora data makes it difficult to compare site reports. Large amounts of amphora data are either not collected or published, or summarily published in favour of more detailed publications and quantifications of amphora stamps. For example, of the c.250,000 amphorae at Elizavetovskoe only 1000 stamps were found (Garlan 1985; Whitbread 1995).

6. Conclusions

SNA helps buffer risk when using amphorae data directly as proxy for trade because it offers a representation of interaction and connectivity between actors in a dataset in addition to facilitating the dynamic processes within trade to be captured. Thinking about connectivity through actors (sites and amphora types) enables the dynamic processes within trade to be captured and considered. The use of bipartite network graphs proved beneficial when thinking about trade dynamics in relation to import patterns at sites. When using bipartite networks, the relationship between groups of sites and groups of actors can be visualised, encouraging further questions to be asked about the socioeconomic relations between sites in order to investigate why these 'packages' of amphorae (clusters on network graphs) have similar levels of influence and centrality. Bipartite network graphs do not, however, remove problems of inherent interpretation of direct line trade, as both exporter and importer are mentioned.

Transforming the dataset to a co-occurrence network focusing solely on the amphora actors proved useful for studying trends in trade during a particular period. Co-occurrence networks can help to explore patterns in datasets relating to key amphora types role in facilitating trade. Co-occurrence network graphs remove problems encountered with the implicit interpretation of trade being direct, often encountered when using traditional static distribution maps when thinking about trade dynamics. Exploring the relationships between amphora types highlights future questions for research such as the relations between amphorae and the centres they were from. Thinking about the amphorae clusters in network graphs as likely packages within cabotage type trade around the Black Sea proved the most innovative. Interpretations were focused on which groups of amphora types were most likely transported together or were key in connecting smaller clusters. The role of central actors (amphora types) connecting smaller clusters to one another in network graphs may offer an indication for crucial amphora types in one period for facilitating wider trade in the Black Sea region.

Combining multiple source types was achieved with SNA application. When interpreting patterns in datasets found using SNA, shipwreck evidence can help support interpretations of trends in trade dynamics and packages of amphora which were likely transported together. Further research could integrate multiple types of material culture, such as coins and red- and black-figure pottery, into both bipartite and co-occurrence networks, following the approach used by Arthur et al. (2018) for Byzantine amphorae.

Moving the focal point of Black Sea amphora studies away from typology, chronology and stamps (Lund and Gabrielsen 2005, 163), to more metaphorical interpretations used to explore trade and interaction, building an understanding of regional Black Sea connectivity as well as how the Black Sea interacted with the Mediterranean is essential for the development of Black Sea studies. The continued development of Greek amphorae databases is crucial for progression in this field. Having a centralised database on Greek transport amphorae across the Mediterranean and Black Sea is an essential development for future research. Such datasets have already been established within Roman amphora research demonstrating the feasibility of this within Greek amphora studies. The funding needed and time invested to such a project would be a large but hugely important task for research on ancient Greece.

SNA network graphs offer a representation of a dataset and methodology selected by a researcher and will never provide a full reproduction a past network. Using various layout algorithms helps to reveal different patterns in the dataset, thus, it is crucial that the researcher spends time with network graphs visualising it in different ways, and even through different pieces of software.

SNA can provide a mechanism for Greek transport amphorae to be seen as proxy for trade, through the interpretation of patterns in a network. The potential for new insights should encourage archaeologists to use SNA to visualise data, as it can highlight new patterns within networks of both new and existing data. SNA provides an innovative way to look at old and new data, and to combine otherwise disparate datasets. Crucially, the use of SNA, or other network analysis, approaches are most useful to archaeologists attempting to reveal patterns in datasets, proving especially useful to neatly consolidate large and/or disparate datasets, especially when using multiple site types in the research.

References

Akhvlédiani, D., 2010, The ratio of imported and local amphorae in Colchis in the 5th-2nd centuries B.C. according to the grave findings, in D. K. Tezgör and N. Inaishvili (eds.) *Patabs I: Production and Trade of Amphorae in the Black Sea. Actes de la Table Ronde internationale de Batoumi et Trabzon, 27-29 Avril 2006. Istanbul: French Institute of Anatolian Studies-Georges Dumézil.Varia Anatolica* 21, pp. 139-142.

Archibald, Z. H., 1998, *The Odrysian Kingdom of Thrace: Orpheus Unmasked.* Oxford: Oxford University Press.

Arthur, P., Imperiale, M.L. and Muci, G., 2018, Amphoras, Networks, and Byzantine Maritime Trade, in C, Knappett and J. Leidwanger (eds.)*Maritime Networks in the Ancient Mediterranean World*. Cambridge: Cambridge University Press, pp. 219-237.https://doi.org/10.1017/9781108555685.010

Avram, A., 1996, *Les timbres amphoriques. 1.* Thasos. Histria VIII. Bucharest-Paris.

Bartlet Balicki, H. and Rempel, J. 2024. *Greek transport amphorae dataset for considering social network analysis: a Black Sea case study*. Zenodo. https://doi.org/10.5281/zenodo.10550473

Bartlet Balicki, H. and Rempel, J., In press, *Social Network Analysis (SNA) approaches to material culture networks in the ancient Black Sea and Mediterranean*, in H. Oniz et al. (eds.) *The International Trade Around Mediterranean*, SOMA 23. Oxford: Archaeopress.

Bechtold, B., 2008, *Observations on the amphora repertoire of Middle Punic Carthage*. Ghent University.

Bekhter, A. P., Dana M. and Butyagin A. M., 2019, Orthographic Lead Letter from Myrmekeion. *Zeitschrift für Papyrologie und Epigraphik* 212: 161-171.https://www.jstor.org/stable/48632565

Bes, P., Willet, R., Poblome, J., Brughmans, T., 2019, *Inventory of Crafts and Trade in the Roman East (ICRATES): database of tableware*. York: Archaeology Data Service https://doi.org/10.5284/1050900

Borgatti, S. P., Everett, M. G., and Freeman, L. C., 2002, *Ucinet 6 for Windows: Software for Social Network Analysis*. Harvard, MA: Analytic Technologies.

Broodbank, C. 2013. *The Making of the Middle Sea: A History of the Mediterranean from the Beginning to the Emergence of the Classical World*. London: Thames and Hudson.

Brughmans, T., 2010, Connecting the dots: towards archaeological network analysis. *Oxford Journal of Archaeology* 29(3): 277-303. https://doi.org/10.1111/j.1468-0092.2010.00349.x

Brughmans, T., 2013, Thinking through networks: a review of formal network methods in archaeology. *Journal of Archaeological Method and Theory* 20(4):623-662.https://doi.org/10.1007/s10816-012-9133-8

Brughmans, T., Keay, S. and Earl, G., 2015, Understanding inter-settlement visibility in Iron Age and Roman Southern Spain with exponential random graph models for visibility networks.*Journal of Archaeological Method and Theory* 22(1):58–143.https://doi.org/10.1007/s10816-014-9231-x

Burt, R. S., 1992, *Structural Holes: The Social Structure of Competition*. Cambridge, MA: Harvard University Press.

Carlson, D.N., 2012, The Seafarers and Shipwrecks of Ancient Greece and Rome,in A. Catsambis, B. Ford and D.L (eds.)*The Oxford handbook of maritime archaeology*. New York; Oxford: Oxford University Press.https://doi.org/10.1093/oxfordhb/9780199336005.013.0017

Chistov, D., 2018, Amphorae Assemblages of the Second Quarter - Mid-6th Century BC from the North-Eastern Part of The Berezan Island Site, in A. Avram, L. Buzoianu and V. Lungu (eds.)*KOINÈ ET MOBILITÉ ARTISANALE ENTRE LA MÉDITERRANÉE ET LA MER NOIRE DANS L'ANTIQUITÉ Hommage à PIERRE DUPONT à son 70 e anniversaire (PONTICA LI SUPPLEMENTUM V) Constanța,* pp. 85-96.

Cline, D.H., 2020, Athens as a Small World. *Journal of Historical Network Research* 4:36-56.https://doi.org/10.25517/jhnr.v4i0.84

Davis, D., Brennan, M. L., Opait, A. and Beatrice, J. S. 2018. The Ereğli E Shipwreck, Turkey: an early Hellenistic merchant ship in the Black Sea. *International Journal of Nautical Archaeology* 47(1): 57–80.https://doi.org/10.1111/1095-9270.12276

de Boer, J., 2007, Sinopean Amphorae, South-East and North-West, in. S. L. Solovyov (ed.) *Greeks and Natives in the Cimmerian Bosporus 7th – 1st Centuries BC.*Oxford: BAR International Series 1729:pp. 7-10.

de Boer, J., 2008, Preliminary Observations on the Greek Transport, in: R. Docter, K. Panayotova, J. de Boer, L. Donnellan, W. van de Put and B. Bechtold (eds.) *Apollonia Pontica 2007*. Ghent University: Department of Archaeology, pp. 99-122.

de Boer, J., 2013, Stamped amphorae from the Greek Black Sea colony of Sinope in the Mediterranean during the Hellenistic period, in G.R. Tsetskhladze, S. Atasoy, A. Avram, Ş. Dönmez, J. Hargrave (eds.) *The Bosporus: Gateway between the Ancient West and East (1st Millenium BC–5th Century AD).* Proceedings of the Fourth International Congress on Black Sea Antiquities. Istanbul, 14th–18th September 2009. Oxford: BAR International Series 2517, pp. 109-114.

de Nooy, W., Mrvar, A. and Batagelj, V., 2005, *Exploratory Social Network Analysis with Pajek*. Cambridge; New York: Cambridge University Press.

Donnellan, L., 2016a, A networked view on 'Euboean' colonisation, in L. Donnellan, V. Nizzo and G. J. Burgers (eds.) *Conceptualising early Colonisation*. Turnhout: Brepols, pp. 149-166.

Donnellan, L., 2016b, ‘Connected Seas: Mediterranean networks and Black Sea regionality’ review of *Sanctuaries and the power of consumption. Networking and the formation of elites in the Archaic western Mediterranean world. Proceedings of the International Conference in Innsbruck, 20th-23rd March*, by E. Kistler, B. Öhlinger, M. Hoernes, (eds.) and*The Danubian lands between the Black, Aegean, and Adriatic Seas (7th centuries BC – 10th century AD)*. Proceedings of the fifth international congress on Black Sea antiquities, Belgrade, 17-23 September 2013, by G. Tsetskhladze, A. Avram and J. Hargrave (eds.). *Journal of Greek Archaeology*1:440-446.https://doi.org/10.32028/jga.v1i

Donnellan, L., 2021, Understanding Greek-native interaction in early Greek Black Sea colonisation. An example from Istros/Histria, in G.R. Tsetskhladze, A. Avram and J. Hargrave (eds.) *The Greeks and Romans in the Black Sea and the Importance of the Pontic Region for the Graeco-Roman World.* Oxford: Archaeopress, pp. 55-68.

Fless, F., 2008, Taste at the periphery of the Greek world: The Iberian Peninsula and the Black Sea, in K. Lapatin (ed.) *Papers on Special Techniques in Athenian Vases.* Los Angeles: Getty Publications, pp. 225-234

Garlan, Y., 1999, Réflexions sur le commerce des amphores grecques en mer Noire, in. Y. Garlan (ed.)*Production et commerce des amphores anciennes en Mer Noir. Colloque international organisé à Istanbul, 25– 28 mai* 1994. Aix-en Provence.

Garlan, Y., 1985, De l'usage par les historiens du matériel amphorique grec. *Dialogues d'histoire ancienne* 11(1) : 238-255.

Garlan, Y., 2007, Échanges d’amphores timbrées entre Sinope et la Méditerranée aux époques classique et hellénistique, in V. Gabrielsen and J. Lund (eds.)*The Black Sea in Antiquity: Regional and Interregional Economic Exchanges.* Aarhus: Aarhus University Press, pp. 143-148.

Greaves, A., 2007, Milesians in the Black Sea: Trade, Settlement and Religion, in V. Gabrielsen and J. Lund, eds. *The Black Sea in Antiquity: Regional and Interregional Economic Exchanges*. Aarhus: Aarhus University Press, pp. 9-22.

Greene, E.S., 2018, Shipwrecks as indices of Archaic Mediterranean trade networks, in J. Leidwanger and C. Knappett, eds., *Maritime networks in the ancient Mediterranean world*. Cambridge: Cambridge University Press, pp. 132-162.https://doi.org/10.1017/9781108555685.007

Horden, P. and Purcell, N., 2000, *The Corrupting Sea. A Study of Mediterranean History*. Oxford: Wiley-Blackwell.

Inaishvili, N., Khalvashi, M., 2010, Sinopean imports on the Black Sea Littoral of South-West Georgia,*Ancient Civilizations from Scythia to Siberia* 16:487–559.
https://doi.org/10.1163/157005711X560471

Ivantchick, A., 2017, The Greeks and the Black Sea: the earliest ideas about the region and the beginning of colonisation, inV. Kozlovskaya (ed.) *The Northern Black Sea in Antiquity: Networks, Connectivity, and Cultural Interactions*.New York: Cambridge University Press, pp. 7-25.https://doi.org/10.1017/9781139094702.003

Kac, V.I., Monakhov, S.Y., Stolba, V.F. and Ščeglov, A.N., 2002, Tiles and ceramic containers, in L. Hannestad, V.F. Stolba &A.N. Ščeglov (eds.)*Panskoye I*. Aarhus, pp. 101-126.

Knappett, C., 2011, *An Archaeology of Interaction. Network Perspectives on Material Culture and Society.* Oxford; New York: Oxford University Press.

Knappett, C., 2013, Introduction: why networks? in C. Knappett (ed.) *Network Analysis in Archaeology. New Approaches to Regional Interaction.* Oxford: Oxford University Press, pp. 3-15.
https://doi.org/10.1093/acprof:oso/9780199697090.003.0001

Knappett, C., 2016, Networks in Archaeology: Between Scientific Method and Humanistic Metaphor, in T. Brughmans, A. Collar, and F. Coward (eds.) *The Connected Past: Challenges to Network Studies in Archaeology and History*. Oxford: Oxford University Press, pp. 21 – 33.

Kovalchuk, A.V., 2020, *Amphora Stamps of the Northern Black Sea*. Available at: http://stampsofbosphorus.ru/en/ (Accessed: 14 January 2024).

Krapivina, V.V., 2010, Ceramics from Sinope in Olbia Pontica. *Ancient Civilizations from Scythia to Siberia* 16(1-2): 467-477.
https://doi.org/10.1163/157005711X560453

Lawall, M., and Graham, S., 2018, Netlogo simulations and the use of transport amphoras in antiquity, in J. Leidwanger and C. Knappett, eds. *Maritime networks in the ancient Mediterranean world*. Cambridge: Cambridge University Press, pp. 163-183.https://doi.org/10.1017/9781108555685.008

Lawall, M.L., Lejpunskaja, N.A., Diatroptov, P.D. and Samojlova, T.L., 2010, Transport amphoras, inN.A. Lejpunskaja, P. GuldagerBilde, J. MunkHøjte, V.V. Krapivina and S.D. Kryžickij (eds.) *The Lower City of Olbia (Sector NGS) in the 6th Century BC to the 4th Century AD*.Aarhus: Aarhus University Press, pp. 355-405.

Lazarov, M., 1980, Le commerce de Thasos avec la côte thrace du Pont à l'époque préromaine. *Actes du IIe congrès international de Thracologie (Bucarest, 1976), II*, pp.171-187.

Leidwanger, J. Knappett, C., Arnaud, P., Arthur, P., Blake, E., Broodbank, C., Brughmans, T., Evans, T., Graham, S., Greene, E.S., Kowalzig, B., Mills, B., Rivers, R., Tartaron T.F., and Van de Noort, R., 2014, A manifesto for the study of ancient Mediterranean maritime networks. *Antiquity*342(88): 1-5. https://www.antiquity.ac.uk/projgall/leidwanger342

Lewis, D.M., 2022, A cargo of slaves? Demosthenes 34.10. *The Mariner's Mirror* 108(2):135-148. https://doi.org/10.1080/00253359.2022.2048524

Lund, J., 2007, The circulation of ceramic fine wares and transport amphorae from the Black Sea region in the Mediterranean, c. 400 BC-AD 200, in. V. Gabrielsen and J. Lund, eds. *The Black Sea in Antiquity: Regional and Interregional Economic Exchanges*. Aarhus: Aarhus University Press, pp. 183-194.

Lund, J. and Gabrielsen, V., 2005, A fishy business: transport amphorae of the Black Sea region as a source for the trade in fish and fish products in the Classical and Hellenistic periods, in T. Bekker-Nielsen (ed.)*Ancient fishing and fish processing in the Black Sea region*. Aarhus: Aarhus University Press, pp.161-169.

Lungu, V., 1992, Circulaţia amforelor ştampilate în zona Capul Dolojman. *Pontica* 25:69-97.

Malkin, I., 2003, Networks and the emergence of Greek identity. *Mediterranean Historical Review* 18(2):56-74. https://doi.org/10.1080/0951896032000230480

Malkin, I., 2011, *A small Greek world: networks in the Ancient Mediterranean*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199734818.001.0001

Malkin, I., Constantakopoulou, C. and Panagopoulou, K., 2009, Introduction, in I. Malkin, C. Constantakopoulou and K. Panagopoulou (eds.) *Greek and Roman networks in the Mediterranean*. Abingdon: Routledge, pp.1-11.

Millet, P.B., 2016, *Amphorae ex Hispania*. Available at: http://amphorae.icac.cat/ (Accessed 14 January 2024).

Mills, B.J., 2017, Social network analysis in archaeology. *Annual Review of Anthropology* 46:379-397.https://doi.org/10.1146/annurev-anthro-102116-041423

Monakhov, S.Y., 1999, Grecheskie amfory v Prichernomor'e. Kompleksy keramicheskoi tary VII—II vekov do ne (Greek Amphorae in the Pontic Region. Complexes of Ceramic Ware from the 7th—2nd Centuries BC). Saratov: Saratov State University.

Monakhov, S. J. and Kuznetsova E. V., 2017, Overseas Trade in the Black Sea Region, inV. Kozlovskaya (ed.)*The Northern Black Sea in Antiquity: Networks, Connectivity, and Cultural Interactions*.New York: Cambridge University Press, pp.59-99.

Östborn, P. and Gerding, H., 2014, Network analysis of archaeological data: a systematic approach. *Journal of Archaeological Science* 46:75-88. https://doi.org/10.1016/j.jas.2014.03.015

Panagou, T., 2015, Patterns of amphora stamp distribution: Tracking down export tendencies, in E.M. Harris et al. (eds.) *The Ancient Greek Economy: Markets, Households and City-States.* Cambridge: Cambridge University Press, pp. 207-229.

Parmenter, C.S., 2020, Journeys into Slavery along the Black Sea Coast, c. 550-450 BCE. *Classical Antiquity* 39 (1): 57–94. https://doi.org/10.1525/ca.2020.39.1.57

Rempel, J. and Doonan, O., 2020, Rural hinterlands in the Black Sea during the fourth century BCE: Expansion, intensification and new connections. *Anatolian Studies* 70: 127-151. https://doi.org/10.1017/S0066154619000152

Rivers, R., Knappett, C. and Evans, T., 2013, Network models and archaeological spaces, in A. Bevanand M. Lake (eds.)*Computational Approaches to Archaeological Spaces*. Walnut Creek, CA: Left Coast Press, pp. 99– 126.

Salviat, F., 1986, Le vin de Thasos, amphores, vin et sources écrites, in J.Y. Empereur and Y. Garlan (eds.) *Recherches sur les amphores grecques. Actes du colloque international organisé par le Centre National de la Recherche Scientifique, l'Université de Rennes II et l'École française d'Athènes, 10–12 Septembre1984* (BCH Suppl. XIII). Athens and Paris, pp. 145–196.

Saprykin, S., 2017, Ancient sea routes in the Black Sea, in C. Buchet, P. Arnaud, and P. De Souza (eds.) *The Sea in History - The Ancient World*. Boydell & Brewer, pp. 345-361. https://doi.org/10.1017/9781782049081.030

Saratov State University, 2018, *Greek Amphoraefrom Northern Pontus Euxinus (VII–II c. BC)*. Available at: https://ape.sgu.ru/Amphora/complexes?lang=en (Accessed: 14 January 2024).

Stolba, V., 2012, Catalogues of Objects, in V. Stolba and E. Rogov (eds.)*Panskoye 1, Volume 2: The Necropolis*.Aarhus: Aarhus University Press, pp.230-356.

Tsetskhladze, G. R., 2002, Ionians abroad, in G. R. Tsetskhladze and A. M. Snodgrass (eds.)*Greek Settlements in the Eastern Mediterranean and the Black Sea*. Oxford: Archaeopress, pp. 81.96.

Tušlová, P., Kučová, S. and Weissová, B., 2010, Greek and Black Sea Transport Amphorae in Emporion Pistiros: Quantified Analysis of Material Excavated until 2009, in J. Bouzek, L. Domaradzka, and ZH Archibald, eds. *Pistiros IV: Excavations and Studies: Report on Joint Project of Excavations and Studies by the Bulgarian Academy of Sciences, Charles University in Prague, University of Liverpool, and Archaeological Museum "Prof. M. Domaradzki," Septemvri*. Prague, pp. 205-220.

Tyres, P.A., 2021, *Potsherd: Atlas of Roman Pottery*. Available at: https://potsherd.net/atlas/index (Accessed: 14 January 2024).

Tzochev, C., 2016, Markets, amphorae trade and wine industry: The case of Thasos, in E.M. Harris, D.M. Lewis, and M. Woolmer (eds.) *The ancient Greek economy: Markets, households and city-states*. Cambridge: Cambridge University Press, pp. 250-253.

Université de Provence, 2004, *AMPHORES, Recueil de Timbres sur Amphores Romaines.*Available at: https://rtar.univ-amu.fr/rtar/index_valid_droit.html (Accessed: 14 January 2024).

University of Southampton, 2014, *Roman Amphorae: a digital resource*. York: Archaeology Data Service.https://doi.org/10.5284/1028192

Verhagen, P., Brughmans, T., Nuninger, L. and Bertoncello, F., 2013, The long and winding road: Combining least cost paths and network analysis techniques for settlement location analysis and predictive modelling, in *Proceedings of Computer Applications and Quantitative Techniques in Archaeology conference 2012, Southampton*. Amsterdam: Amsterdam University Press, pp. 357–366.

Wasserman, S. and Faust, K., 1994, *Social network analysis: Methods and applications* (Vol. 8). Cambridge University Press.

Whitbread, I. K., 1995, *Greek Transport Amphorae: A petrological and archaeological study*. Athens: British School at Athens.

Ancient Sources

Demosthenes.*Against Lacritus*. Trans. A. T. Murray (1939). Cambridge, MA: Harvard University Press; London, William Heinemann Ltd.

Demosthenes.*Against Leptines*. Trans. C. A. Vince, M. A. and J. H. Vince (1926) Cambridge, MA, Harvard University Press; London, William Heinemann Ltd.

Herodotus.*The Histories*. Trans. A. D. Godley. (1920). Cambridge, Harvard University Press.

Isocrates. *Isocrates speeches*. Trans. G. Norlin, (1980). Cambridge, MA, Harvard University Press; London, William Heinemann Ltd.

Pliny the Elder. *The Natural History*. Trans. J. Bostock, (1855). B.A. London. Taylor and Francis, Red Lion Court, Fleet Street.

Polybius.*Histories*. Trans. Evelyn S. Shuckburgh. (1962). London, New York: Macmillan.

Strabo. *The Geography of Strabo*. Trans. H. L. Jones. (1924). Cambridge, Mass.: Harvard University Press; London: William Heinemann, Ltd.

Theophrastus. *Enquiry into Plants, Volume I: Books 1-5*. Trans. A. F. Hort. (1916). Loeb Classical Library 70. Cambridge, MA: Harvard University Press.

Xenophon. *Anabasis.* Trans. C.L. Brownson, C.L., (1998). Cambridge, MA: Harvard University Press.

CATEGORIZATION OF ARCHAEOLOGICAL CERAMICS BASED ON THEIR ELEMENTAL COMPOSITION USING SELF ORGANIZING MAPS (SOM)

A. HEIN[1] AND V. KILIKOGLOU[1]

[1]Institute of Nanoscience and Nanotechnology, N.C.S.R. "Demokritos", 15310 Aghia Paraskevi, Athens, Greece
a.hein@inn.demokritos.gr

Abstract

Ceramic provenance studies by analyzing elemental compositions are based on the assumption that there exist compositional differences between ceramics manufactured at different production sites. These differences are essentially related to the use of geochemically diverse raw material sources and provide, thus, information about the place of manufacture of the ceramics and about potential dissemination and trading. In order to determine the elemental composition of archaeological ceramics commonly laboratory methods, such as neutron activation analysis (NAA) or wavelength-dispersive X-ray fluorescence spectrometry (WD-XRF) are applied providing high precision as well as high accuracy. During the recent years portable energy dispersive XRF (pXRF) has been introduced in the study of archaeological materials allowing for fast and non-invasive measurements of their elemental compositions. In the case of pottery analyses, though, several issues arise. (1) The method is extremely surface sensitive, which has to be considered for measurements of slipped, painted or weathered surfaces. (2) The determined element concentrations are commonly less precise and accurate than compositional data collected with laboratory measurements. (3) The suite of elements potentially measured is smaller and concentrations might be below the lower limit of determination. (4) The sample geometry introduces additional uncertainty as the systems are calibrated for measurements of plane surfaces in direct contact. Hence, the statistical evaluation of pXRF data following traditional approaches of multivariate statistics applied to data collected by laboratory analyses becomes inappropriate. For the evaluation of the considerably fuzzier pXRF data a rather flexible approach is expected to be more feasible taking into account the actual data structure rather than assumptions regarding geochemistry. For this, the application of artificial neural networks (ANN) is tested and the multivariate data are evaluated using self organizing maps (SOM).

Keywords: *ceramics, compositional variation, portable XRF, artificial neural networks*

1. Introduction

The categorization of archaeological ceramics based on their elemental composition constitutes a well-established analytical approach in ceramic provenance studies (Hein 2018). The compositional differences between ceramics manufactured at different production sites are essentially related to the utilization of raw materials from geochemically diverse sources providing, thus, distinct compositional patterns propagated in the ceramics manufactured from them (Hein and Kilikoglou 2020). The straightforward linkage of ceramic compositions to compositions of a specific clay deposit, however, is obscured by the processing of the raw materials including refinement, tempering or even mixing. For this, the ceramics of unknown origin can be localized by means of reference data of ceramics from specific production sites. In this way the dissemination and trading of ceramic wares and in the case of transport jars their content can be investigated revealing eventually information about social, political or economic relations among ancient societies. In order to determine the elemental composition of archaeological ceramics commonly laboratory methods with high precision and accuracy, such as neutron activation analysis (NAA) or wavelength-dispersive X-ray fluorescence spectrometry (WD-XRF), are applied in view of minimizing analytical uncertainties, which could obscure the natural variation of the determined compositional patterns (Hein and Kilikoglou 2017). During the last 20 years, however, handheld portable energy dispersive XRF (pXRF) has been introduced in the study of archaeological materials, providing fast and non-invasive measurements of elemental compositions (Shugar and Mass 2012, Frahm and Doonan 2013). In the case of pottery analyses, though, several drawbacks have to be considered (Goren et al. 2011, Hein et al. 2021a). Because the method is extremely surface sensitive measurements of slipped, painted or even weathered surfaces cannot provide information about the composition of the actual ceramic body. Furthermore, the determined element concentrations are commonly less precise and accurate than compositional data collected with laboratory measurements. In this context also the suite of elements, which can be potentially measured, is smaller in comparison to laboratory measurements. In the case of trace elements the concentrations are frequently below the lower limit of determination. Eventually, also the sample geometry affects the measurements as the systems are commonly calibrated for measurements of plane surfaces in direct contact. All these issues increase the analytical uncertainties and impede the statistical evaluation of pXRF data following traditional approaches of multivariate statistics, such as hierarchical cluster analysis or principal component analysis, which are applied commonly to data collected by laboratory analyses (Baxter and Freestone 2006, Baxter 2008). Because pXRF data are considerably fuzzier, their categorization requires a more flexible and robust approach based on the actual data structure rather than on assumptions regarding geochemistry. As potential option the application of artificial neural networks (ANN) is tested and the multivariate data are evaluated using self organizing maps (SOM) (Kohonen 1995). In the present case study unsupervised learning is applied on a data set of Hellenistic transport amphorae, which has been previously evaluated using conventional multivariate data evaluation (Hein et al. 2021a). SOMs were generated considering varying element suites in order to explore the formation of stable elemental patterns. The resulting categories are compared with the previous statistical results using conventional multivariate statistics. Scope of the present study is to investigate to what extent ANN can contribute to an enhanced evaluation of pXRF data and possibly can reveal information, which is not detectable by conventional data analysis.

2. Conventional Approaches towards Evaluating Multivariate data

The elemental composition of a single ceramic specimen can be considered as a vector with indexed variables corresponding to the element concentrations determined. The analysis of an entire ceramic assemblage, thus, provides a two-dimensional data array or matrix defined by the number of samples and the number of element concentrations considered:

$$\boldsymbol{X} = (x_{ij})$$

- i = 1 to n (number of samples)
- j = 1 to p (number of elements)

In order to categorize the analyzed ceramics according to their elemental composition similar samples have to be recognized while dissimilar samples have to be distinguished. For this, a distance criterion has to be defined, such as the Euclidian distance of the scaled elemental concentrations. Statistical approaches can be classified essentially either as exploratory data analysis or as model based data analysis assuming for example a multivariate normal distribution (Baxter 2008). The most commonly used exploratory methods are Principal Component analysis (PCA) and Hierarchical Cluster Analysis (CA). Through PCA linear combinations of the original variables are determined, so-called principal components, which separate the data set with maximum variance (Baxter and Freestone 2006). Clusters of similar data or differences among particular data can be explored by plotting the data according to these new coordinates. In CA, on the other hand, individual data are successively merged to clusters, which can be illustrated in a dendrogram. The clustering algorithm depends on the distance criterion defined and on the linkage method (e.g. average linkage or complete linkage) (Baxter 2001).

Variability and covariance matrix

Concerning model based analysis under assumption of a multivariate normal distribution the variance of a specific elemental pattern x can be evaluated by determining the covariance matrix $\boldsymbol{S}_x$ based on the compositions of the individual samples. Theoretically, the covariance matrix can be separated according to different sources of variability (Beier and Mommsen, 1994; Bieber et al., 1976):

$$\boldsymbol{S}_x^{\,2} = \boldsymbol{S}_{xN}^{\,2} + \boldsymbol{S}_{xS}^{\,2} + \boldsymbol{S}_{xA}^{\,2}$$

Here, $\boldsymbol{S}_{x\mathrm{N}}$ is the natural variability of the true composition of the ceramic group, $\boldsymbol{S}_{xS}$ is the variability introduced by the sample selection and $\boldsymbol{S}_{xA}$ is the analytical uncertainty (Hein and Kilikoglou 2017). In order to distinguish different compositional patterns using conventional statistics the differences among these patterns have to exceed the differences within an individual pattern. However, in the case of pXRF the contribution of the analytical uncertainties are comparably increased. After all, they potentially obscure the differences among individual compositional patterns and the categorization of particular element patterns might be ambivalent. For this, a more flexible method for pattern recognition appears to be helpful without presumptions concerning variance and distinctiveness.

3. Self Organizing Maps (SOM)

Self Organizing Maps (SOM) constitute a specific type of artificial neural networks (ANN) (Kohonen 1995). Complex data, such as in the present case multivariate compositional data, are projected on a two dimensional array of discrete codebook vectors or output neurons. The SOM can have either a rectangular topology with each vector having four neighbors or a hexagonal topology with each vector having six neighbors (Kruse et al. 2022). Topological proximity of the output neurons in the SOM indicates similarity of the original data. In the beginning the output neurons are initialized with random values. During the training process the input vectors of the original data are compared with the output neurons taking into account a distance criterion, such as the Euclidian or the squared Euclidian distance of the standardized data vectors. For each input vector the output neuron with the smallest distance is determined and the weights of this output neuron and its neighbors are updated towards the input vector. The presentation of the training data is repeated over an initially defined number of training steps. Step per step the radius of the neighborhood and the learning rate is decreased. For the SOM generation in the present case study the *kohonen* package (version 3.0.10) was used, which is available for R (Wehrens and Kruisselbrink 2018), providing apart from the training of the SOM various visualization methods.

4. Dataset of Hellenistic Amphorae from Paphos

In the present case study a pXRF dataset comprising elemental compositions of Hellenistic transport amphorae from Nea Paphos (Cyprus) has been reevaluated. The dataset includes 366 pXRF measurements of 288 amphora fragments, which had been discovered at the public Agora of Paphos and the residential site of Maloutena (Hein et al. 2021a). According to the finding context they represent trade as well as consumption of amphorae or rather their content presenting an expectedly high variation in terms of types and origins. For the pXRF measurements a NITON XL3t GOLDD+ hand-held system (Thermo Fisher Scientific) was used (Fig. 1). Each fragment was analyzed for 120 s with the pre-set ‘soil’ method providing the concentrations of up to 33 elements from S (Z=16) to U (Z=92) (Hein 2021). The original pXRF data are available in the Mendeley Data Repository (Hein et al. 2021b).

Figure 1 – On-site pXRF measurement of amphora fragments.

Scope of the initial study was to investigate the compositional variation of the ceramics and to verify the provenance of local or regional Cypriot amphorae as well as to identify amphorae imported from the Aegean or even farther production regions. For this, the dataset had been divided into two parts representing regional Cypriot production and imported wares according to macroscopic fabric categorization (Fig. 1). This was done taking into account the comparably large number of measurements and expected clusters. The two datasets were evaluated separately using hierarchical clustering of the compositions of 14 elements (Ca, Cr, Cu, Fe, K, Mn, Ni, Pb, Rb, Sr, Ti, V, Zn and Zr). The element concentrations were standardized by log-ratio transformation using the Fe concentrations as common divisor. The hierarchical clustering of the Cypriot amphorae indicated seven amphora clusters, while the hierarchical clustering of the imported amphorae indicated ten compositional clusters. Based on the pXRF results 97 fragments were selected for NAA, representing the main compositional clusters apparent in the pXRF data. In this way the provenance of specific compositional groups could be investigated taking into account existing reference data of production places (Hein and Kilikoglou 2017, Hein et al. 2021a). Apart from this, the NAA data allowed for subdividing ware groups from clusters determined initially in the pXRF data. Thus, one issue to be investigated was, whether it was possible to confirm these sub-clusters in the pXRF data by using SOM.

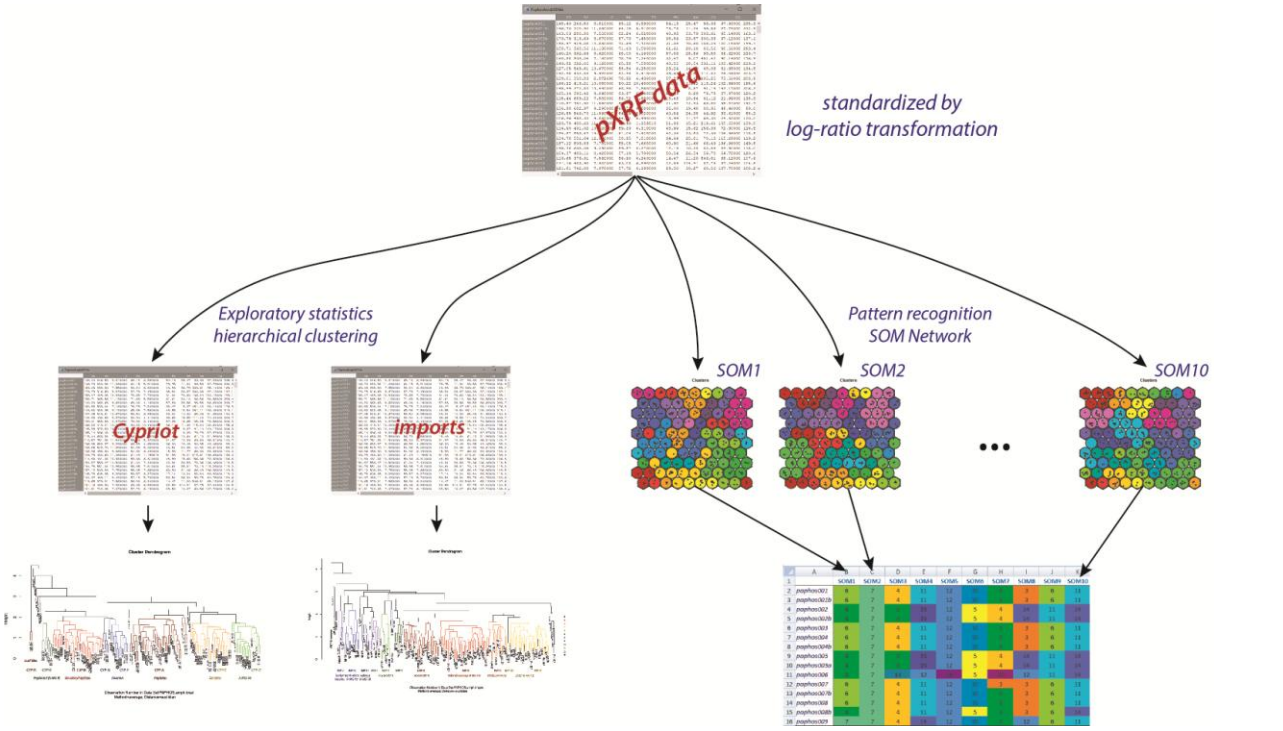


Figure 2 – Comparison of methodological approaches: Initial exploratory statistical evaluation of two separate datasets (left) Pattern recognition in the entire dataset in a series of SOM networks (right)

Results and Discussion

A toroidal 10×10 base map with hexagonal topology was defined comprising thus 100 codebook vectors. The same 14 element concentrations were used as for the initial statistical evaluation with the addition of the Th concentrations, for which missing values were replaced with a minimum concentration. According to the initial evaluation the data were standardized by log-ratio transformation with the Fe concentrations as common divisor. For the SOM training the squared Euclidian distance was defined as distance criterion. The entire dataset including Cypriot as well as imported wares was presented for a predefined number of training steps to the network, from 1.000 up to 10.000 times, with updating the codebook vectors after the presentation of each individual input vector (‘online’ algorithm). The resulting codebook vectors were evaluated with hierarchical clustering, so that similar fields could be combined into 20 clusters. Figure 3 presents one example SOM, in which codebook vectors with similar elemental compositions based on the hierarchical clustering are indicated with colors.

The clusters defined by the SOM network appear indeed to confirm the results of the above described evaluation using conventional exploratory statistics (Hein et al. 2021a) (Fig. 4). In addition, it appears to be possible to determine sub-clusters in the pXRF results, such in the case of the initial clusters CYP-A, CYP-B, IMP-A, IMP-B or IMP-D. Eventually, similarities among Cypriot and imported wares become apparent, which could not determined in the initial two pXRF datasets, which had been separated based on macroscopic fabric categorization.

It has to be considered, though, that the codebook vectors are initially loaded with random values. As a consequence of this the resulting SOM is not necessarily reproducible and measurements clustering together in one SOM might be separated in another SOM with new random initialization. In the case of specific measurements the SOM based clustering, thus, still appears to be ambivalent. In order to investigate the significance and uncertainty of clustering using a SOM network for the present pXRF data a series of ten SOM was generated, each with different random initialization, and the resulting clusters were examined (Fig. 2). Measurements, which were placed in the same cluster in at least eight out of the ten SOM networks, were combined in ware groups. Measurements, which were not clearly correlated to other samples, were defined either as chemical loner or as questionable. Table 1 present the resulting assignment. Some of the ware groups are rather clear and distinct, such as in the case of the Cypriot amphorae the red-brown wares CYP-RB and CYP-RB2, which were initially combined in a Cluster CYP-D and can be apparently sub-divided. Clear subdivisions are also indicated for the two largest clusters CYP-A and CYP-B initially defined in the Cypriot pXRF dataset. This coincides in fact with the NAA measurements of selected samples (Hein et al. 2021a).

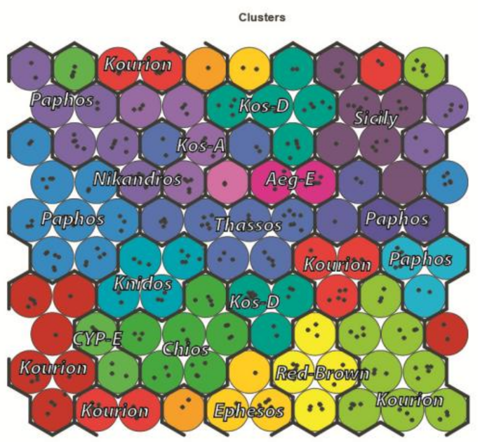


Figure 3 – Color coded SOM of the compositional data of 288 amphora fragments: The colors indicate codebook vectors with similar compositions. Indicated is the assumed origin of amphorae in specific areas of the SOM.

A similar picture can be observed in the second dataset of supposedly imported amphorae. Some of the initially defined larger clusters can be subdivided in ware groups, which coincide with compositional groups defined based on NAA. Apart from this it is possible to assign amphorae, which were originally categorized as probable imports based on their macroscopic fabric, to Cyrpiot ware groups, such as in the case of the initial Cluster IMP-F, which eventually appears to comprise amphorae from Kourion. Figure 4 presents a heat map comparing 20 clusters defined by SOM with 17 clusters initially defined with conventional statistics.

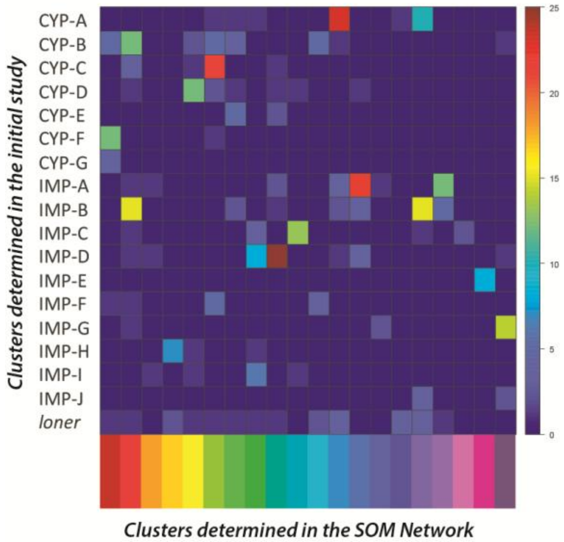


Figure 4 – Heat map comparing the clusters determined by SOM and conventional statistics: The clusters determined with the SOM network presented in Fig. 2 are indicated with the color coding on the horizontal axis, while the initial clusters defined by conventional exploratory statistics are indicated on the vertical axis (Hein et al. 2021a).

Conclusions

Pattern recognition using SOM networks provides a more detailed compositional categorization compared with conventional data evaluation of parts of the data using exploratory statistics. The details revealed by SOM networks can be indeed compared with the level achieved by the evaluation of NAA data. Thus, the application of SOM networks appears to offer more thorough pattern recognition than it can be achieved by common statistical methods. Beyond that the SOM approach is automated and non-biased. The present case study, though, indicates some uncertainties concerning the clustering, which could not be simply reproduced assumedly due to the random initial loading of the codebook vectors. For this, series of SOMs can be generated in order to investigate the variation of the resulting clustering. This variation can be reduced by selecting modeling parameters, such as the numbers of steps, the radius of the neighborhood or the learning rate. Another option could be the initializing of the codebook vectors with specific values.

	CYPRUS							IMPORTS												
	A	B	C	D	E	F	G	A	B	C	D	D2	E	F	G	G2	H	I	J	loner
PAPH-A	9+1(-)	1						1												1+2(-)
PAPH-B	7+1(-)								2											1
PAPH-C		3												2						1
KOUR-A		4				10	4							1						1
KOUR-B		14							1(-)											
KOUR-C		2	21+1(-)	1(-)		1								3						1
KOUR-E																				2
CYP-RB				11													1			
CYP-RB2				3											1(-)					
CYP-E					6				1											1
CYP-X3		2																		
CHIOS				1							3+1(-)							3		
EPHESOS																	7			
KNIDOS										6								2		1
KOS-A								8	5											1
KOS-D									1		25									
NIKA	1								13											
SICILY											1				12	2			2+1(-)	1
THASSOS								8					1			1				
BASKET																			3	1
AEG-D										2										1
AEG-E												3	4							
AEG-E2								2												
AEG-F									10+1(-)											
AEG-X1																		2		1
loner	6	1	1					2	5	2	3			1	1					3

Table 1 – Comparison of the ware groups defined by a series of SOM networks (rows) with the initially defined clusters (columns)

References

Baxter, M.J., 2001, *Statistical Modeling of Artefact Compositional Data*, Archaeometry 43 (1): 131-147.

Baxter, M.J., 2008, *Mathematics, Statistics and Archaeometry: The past 50 years or so*, Archaeometry 50, 6: 968-982. DOI: 10.1111/j.1475-4754.2008.00427.x

Baxter, M.J. and Freestone, I.C., 2006, *Log-ratio compositional data analysis in Arrchaeometry*, Archaeometry 48, 3: 511–531.

Beier, T.,Mommsen, H., 1994, *Modified Mahalanobis filters for grouping pottery by chemical composition*. Archaeometry 36, 287–306.

Bieber Jr., A.M., Brooks, D.W., Harbottle, G., Sayre, E.V., 1976, *Application of multivariate techniques to analytical data on Aegean ceramics*. Archaeometry 18, 59–74.

Frahm, E., Doonan, R.C.P., 2013. *The technological versus methodological revolution of portable XRF in archaeology*. J. Archaeol. Sci. 40, 1425–1434. https://doi.org/10.1016/j.jas.2012.10.013.

Goren, Y., Mommsen, H., Klinger, J., 2011, *Non-destructive provenance study of cuneiform tablets using portable X-ray fluorescence (pXRF)*, Journal of Archaeological Science 38: 684-696.

Hein, A., 2018, Elemental Analysis of Pottery, in: *The SAS Encyclopedia of Archaeological Sciences* (ed. S. L. López Varela), Wiley-Blackwell. DOI: 10.1002/9781119188230.saseas0211

Hein, A., 2021. Revisiting the groups: exploring the feasibility of portable EDXRF in provenance studies of transport amphorae in the Eastern Aegean. In: Schneider, G., Dazkiewicz, M., Hegewisch, M. (Eds.), *Application of Portable Energy-Dispersive X-ray Fluorescence to the Analysis of Archaeological Ceramics and Glass*. Topoi. Berlin Studies of the Ancient World - Vol. 75, Edition Topoi, Berlin, 43-61.

Hein, A. and Kilikoglou, V., 2017, *Compositional variability of archaeological ceramics in the Eastern Mediterranenan and implications for the design of provenance studies*, Journal of Archaeological Science: Reports 16: 564-572.

Hein, A. and Kilikoglou, V., 2020, *Ceramic raw materials. How to recognize them and locate the supply basins. I Chemistry*, Archaeological and Anthropological Science 12: 180. DOI: 10.1007/s12520-020-01129-8 Q1

Hein, A., Dobosz, A., Day, P.M. and Kilikoglou, V., 2021a, *Portable ED-XRF as tool for optimizing sampling strategy – Case study of a Hellenistic amphora assemblage from Paphos (Cyprus)*, Journal of Archaeological Science 133: 105436. DOI: 10.1016/j.jas.2021.105436

Hein, A., Dobosz, A., Kilikoglou, V., 2021b, Portable ED-XRF as a Tool for Optimizing Sampling Strategy:

the Case Study of a Hellenistic Amphora Assemblage from Paphos (Cyprus) - Supplementary Data. Mendeley Data. https://doi.org/10.17632/ydgs3z4g76.1 vol. 1.

Kohonen, T., 1995, *Self-Organizing Maps*. Springer-Verlag, Berlin. DOI:10.1007/978-3-642-97610-0.

Kruse, R., Mostaghim, S., Borgelt, C., Braune, C., Steinbrecher, M., 2022, *Computational Intelligence: A Methodological Introduction*, 3rd Edition, Cham: Springer Nature.

Shugar, A.N., Mass, J.L., 2012. Introduction. In: Shugar, A.N., Mass, J.L. (Eds.), *Handheld XRF for Art and Archaeology*. Studies in Archaeological Sciences 3. Leuven: University Press, Leuven, pp. 17–36.

Wehrens, R., Kruisselbrink, J., 2018, *Flexible Self-Organizing Maps in kohonen 3.0*, Journal of Statistical Software 87 (7): 1-18. DOI: 10.18637/jss.v087.i07

TRAINING ALEXNET TO CLASSIFY GROUND PENETRATING RADAR IMAGES FEATURING BURIED STRUCTURES

M. MANATAKI[1], A. VAFIDIS[2], A. SARRIS[3] & N. PAPADOPOULOS[4]

[1]Alma-SistemiSrl, 00012 Guidonia, Rome, Italy;
mma@alma-sistemi.com

[2]School of Mineral Recourses Engineering,Laboratory of Applied Geophysics, Technical University of Crete (TUC), Chania 73100,
vafidis@mred.tuc.gr

[3]Digital Humanities Geoinformatics Lab, Archaeological Research Unit (ARU), Department of History and Archaeology, University of Cyprus, Cyprus;
asarri01@ucy.ac.cy;

[4]Laboratory of Geophysical Satellite Remote Sensing and Archaeoenvironment (GeoSat ReSeArch),Institute for Mediterranean Studies-Foundation for Research and Technology Hellas (IMS-FORTH), Nik. Foka 130, Rethymno, 74100, Greece,
nikos@ims.forth.gr

Abstract

Ground Penetrating Radar (GPR) data derived from archaeological surveys are well known for their complexity, and thesubsequent interpretation is time-consuming, requiring experience, expert skills, and a deeper understanding of GPR attributes. Hence, a system capable of identifying patterns related to an archaeological context could significantly improve the interpretation process. Convolutional Neural Networks (CNNs) seem to be a promising approach towards an automated GPR data analysis, considering the exciting developments and improvements they brought in several computer vision tasks like image classification, segmentation, and object detection. In this study, the AlexNet CNN architecture was implemented in Python using Tensorflow and Keras libraries and trained to classify GPR C-scans collected from several archaeological sites. The training datasets were made from scratch, while the classification labels were set after patterns identified as buried structures, stripping noise, and other features of geological or anthropogenic origin. These three feature classes usually co-exist in GPR images and often exhibit similar patterns, making interpretation challenging and thusincreasing uncertainty. The Stochastic Gradient Descent with momentum was used in the training process, while Batch Normalization and Dropout techniques were also employed to improve the resulting classification. The learning curves of the loss function and classification accuracy were used for the training performance evaluation. The results showed that batch normalization is vital for the models to learn, while when combined with dropout and tunning of the batch size and the learning rate, very high classification accuracy was scored. The best-obtained model was used to make predictions on 100 new GPR images, reaching a classification accuracy of 92%.

***Keywords**: Ground Penetrating Radar (GPR), AlexNet, CNNs, Batch Normalization (BN), archaeological prospection*

1 Introduction

Ground Penetrating Radar (GPR) is a popular geophysical technique in archaeological exploration and is used mainly to investigate the near-surface and map ancient buried foundations (Conyers, 2004; Manataki et al., 2015; Sarris et al., 2017), road networks(Pakkanen et al., 2019), and tombs(Goodman, 2009). The acquired information can be invaluable to the archaeological communityas the compiled maps can guide future excavation projects more efficiently, saving valuable time and reducing costs by avoiding unnecessary excavation works.Over the last two decades, with the advent of more powerful hardware modules, GPR systems have reduced the associated acquisition time for data collection.As a result, larger amounts of data are being collected per survey site, which significantly increases the need for more efficient data interpretation schemes. In particular, the interpretation of GPR archaeological data is often tedious and time-consuming(Conyers, 2012; Manataki et al., 2021). GPR practitioners typically need to study and cross-correlate hundreds of images per site to exploit the collected information to its fullest extent. Further, data interpretation is prone to mistakes as the subsurface of archaeological sites is usually disturbed and can create similar and non-intuitive patterns that are difficult to discriminate. Hence, methods for the automatic analysis and recognition of patterns in archaeological GPR data are nowadays a necessity.

An early attempt to automatically detect buried walls from synthetic GPR tomographic images is presented by Pasolli et al. (2009). The proposed method combines Genetic Algorithms (GA) and Support Vector Machines (SVM) to recognize and classify the patterns of interest. However, despite the promising results, a hiatus was put intothe related research efforts. The main reasons are the lack of proper dataset required and the difficulty in designing suitable features for representing such complex data.However, with the latest developments and impressive performance of Deep Learning (DL) algorithms, especially Convolutional Neural Networks (CNNs)(LeCun et al., 2015), where features are learned automatically, the approach of automatic GPR data analysis seems genuinely feasible. The latter is shown in the results presented in the recent study by Küçükdemirci and Sarris (2020), where U-net architecture, developed to perform image segmentation in biomedical images (Ronneberger et al., 2015), was used to detect linear features from real GPR data. To further improve the results, the key point is to effectively train CNN architectures that can handle big datasets (i.e., hundreds of thousands or millionsof images) to use with GPR archaeological data comparatively of smaller size (i.e., thousands of images).The training process allows the DL algorithms to learn from the available data and interpret specific patterns automatically, guided by the human experience.

Motivated by the latter and aiming for a complete automated interpretational scheme,the present study focuses on training AlexNet architecture to classify patterns detected as ancient structures as well as noise and signals from the subsurface not related to the archaeological remains. First, proper datasets are made from scratch, using GPR data collected from several archaeological sites. Then, training follows using the Stochastic Gradient Descent (SGD) optimizer where Batch Normalization (BN)(Ioffe and Szegedy, 2015) and dropout (Srivastava et al., 2014) techniques are extensively explored to find the setup that yields the best performance. The results of this study contribute to a better understanding of CNNs with GPR data, show the potential, and set future work for further improvements.

2 Background

This section outlines some fundamental concepts and theories on which this research relies. These include the GPR data used for training, the chosen CNN architecture, the training optimizer, and two popular techniques, dropout, and batch normalization, which were introduced to improve training time and classification results.

2.1 GPR data

Ground Penetrating Radar (GPR) is an electromagnetic (EM) geophysical, real-time, and non-destructive technique (NDT) that is based on the propagation of EM waves into the subsurface and their reflection once they meet a boundary of different electrical properties (Annan, 2009). The GPR records are time series of the reflected waves' amplitudes, known as the A-scans, which are stacked when collected along the same survey line, producing the characteristic GPR tomographic images called the B-scans. The B-scans can be collected in survey grids, resulting in volumetric data, from which depth slices, called the C-scans, are obtained through specific processing. In archaeological prospection, C-scans are mainly used as the recorded structural patterns are easier to interpret thanB-scans' recorded patterns. In the latter, the structures appear as complex hyperbolic patterns that are difficult to discriminate from other buried objects unrelated to the archaeological context (i.e., geological layers, debris). This has led to the C-scans being the default choice by most users. However, aside from B-scans, many challenges also exist in interpreting C-scans. Because the near-surface is highly disturbed due to human actions or environmental factors, the extracted C-scans are often fuzzy and noisy, leading to mistakes or high uncertainty in the survey results even by more experienced users. The most common mistakes are to a) miss a pattern related to buried structures, b) interpret noise patterns as buried antiquities, and c) misinterpret the recorded signals as patterns from a different buried object which can look similar in the GPR data or due to personal bias(Manataki et al., 2021). Hence, a framework to assist GPR data analysis and interpretation is highly desirable. Considering recent breakthroughs in Deep Learning and their success in several Computer Vision tasks (i.e., image classification, image segmentation, and object detection/recognition), as well as the proliferation of available GPR data featuring ancient buried structures, a framework based on Deep Learning (DL) algorithms to automate the interpretation processseems a promising direction that remains unexplored.

2.2 AlexNet

AlexNet, proposed by Krizhevsky et al. (2012), is a popular deep CNN architecturethat performs image classification and employs GPUs to speed up learning. AlexNet is considered a tremendous milestone in the performance improvements and developments of CNNs. The input of AlexNet is a color image of 227x227x3 and has eight layers.The first five are convolutional (Conv) layers, with the Conv-1, Conv-2, and Conv-5 succeeded by overlapping max-pooling layers. This part of the architecture is responsible for automatic feature detection through the convolutional operation. At the same time, max-pooling is used for dimensionality reduction purposes while retaining the most important features. The remaining three layers are Fully Connected (FC), with the last (FC-3) being the output. The FC layers classify the detected features. An intermediate layer between the Conv and FC layers called the flattening also exists and vectorizes its input so that the FC layers can receive it as their input. Among the key characteristics of AlexNet is that it uses the Rectified Linear Unit(ReLU) activation functions (instead of $tanh$ that used to be the standard) after everyConv layer and after FC-1 and FC-2 layers. In addition, the softmax activation function is applied for the last FC layer, which produces a distribution over the total number of the class labels defined by the dataset used for training.

2.3 Stochastic Gradient Descent (SGD) with momentum

The optimizer is critical in a Feed-Forward Network (FFN) training process like CNNs,where the Backpropagation algorithm is used(Goodfellow et al., 2016). The optimizer is responsible for minimizing the training error,E_{train} , expressed by acost function that is parameterized by some weight values, $\boldsymbol{w}$, assigned to the network interconnection, also called synapses. For the case of CNNs, gradient-based optimizers are mainly used, where the weights are updated iteratively toward the direction that the gradient becomesminimum∇E_{train} . Similar to Krizhevsky et al. (2012), the optimizer used in this study is the SGD, a randomized version of Gradient Descent (GD) performed for a random training example or batches (Abu-Mostafa et al., 2012):

$$\boldsymbol{w}^{(t+1)} = \boldsymbol{w}^{(t)} - \eta \nabla E_{train}\left(\boldsymbol{w}^{(t)}\right) \qquad (1)$$

where (t) is the iteration number, $\eta$is a step size, also known as the learning rate,and expresses how quickly SGD will approach the minimum (local or global), while bold letters indicate vectors. When SGD is used to train deep CNNs architectures, the momentum technique(Sutskever et al., 2013) is employed to speed up learning. Momentum introduces velocity, $\boldsymbol{v}$, in the gradient descent updating process that helps in accelerating towards the direction of the local minima (Goodfellow et al., 2016):

$$\boldsymbol{v}^{(t+1)} = \mu \boldsymbol{v}^{t} - \eta \nabla C(\boldsymbol{w}^{(t)}) \qquad (2)$$

$$\boldsymbol{w}^{(t+1)} = \boldsymbol{w}^{(t)} - \boldsymbol{v}^{(t+1)} \qquad (3)$$

where $\mu \in [0,1]$ is the momentum coefficient (usually set to 0.9), and $\nabla C(\boldsymbol{w}^{(t)})$ is the gradient of the cost function at $\boldsymbol{w}^{(t)}$. The velocity vector in equation (2) accumulates the gradient elements of the previous steps that point to the same direction, which are then used to update the weights as indicated in equation (3).

2.4 Dropout technique

Dropout(Srivastava et al., 2014) is a regularization technique applied in FFNs to improve generalization by preventing overfitting. Because deep neural networks are capable of learning complicated relationships, they adapted very well to the training dataset, including the noise in the data. Hence, they will fail to correctly predict from any input other than the training dataset. Dropout is a computationally effective method that prevents the training network from adapting to noise in the data. This is achieved by randomly setting inactive or "drop out" neurons and their interconnections in the network's hidden and input layers while being trained. The number of neurons or units that are muted is expressed with a probability, p, of a neuron being trained. This probability is referred to as the dropout rate. The outcome is thinned versions of the original network while the output model used for prediction is obtained weights $\boldsymbol{w}$ multiplied by p.

2.5 Batch normalization

Batch normalization (BN) (Ioffe and Szegedy, 2015)is a proven beneficial technique in training deep neural networks as it accelerates training. Training times of deep FFNs, especially when using SGD, can be very long, either due to a slow convergence to the local minimum or skipping it. The latter can be caused even by tiny changes in the weight values of the first layers, resulting in more extensive changes in the weight values of the deeper layers due to their dependent interconnections. BN prevents this as it normalizes the input of each layer's activation function, allowing faster convergence for the optimizer and making the optimization problem more stable. As a result, fewer training epochs are required to minimize the training error. Aside from the gains in training time, BN has other benefits, like acting as a regularization technique and, in some cases, eliminating the need for dropout(Ioffe and Szegedy, 2015; Luo et al., 2018).

3 Methodology

This section briefly describes the methods employed and their implementation to classify GPR images through the training of AlexNet architecture.Firstly, annotated datasets were constructed due to the lack of available ones for this study's purposes, and then training of the CNNs followed, where the use of BN and dropout was tested extensively to improve the classification results and training performance. Then, the final model was evaluated by making predictions using new GPR examples.

3.1 Dataset construction

The data used for this research were collected from 52 different archaeological sites,most of them being located in Greece (**Figure 1a**), through integrated geophysical surveysmanaged by the Laboratory of Geophysical – Satellite Remote Sensing and Archaeo-Environment (GeoSa ReSeArch) of the Institute for Mediterranean Studies - Foundation and Research and Technology Hellas (IMS - FORTH).The selected sites exhibit traces of past civilizations from different historical periods, dating from the Neolithic to Ottoman years. The GPR system used was NOGGIN GPR, equipped with a 250MHz antenna. The data were collected using survey grids in parallel lines (i.e., B-scans)with a fixed 0.25 or 0.5m spacing between them, which is adequate for mapping foundation structures. The A-scans sampling was set to 0.05m or 0.025m along a scan line. The above-mentioned values were decided based on each terrain condition and the corresponding survey needs.

A total of 470 survey grids were selected based on their features, which were then imported and processed in MATLAB to extract C-scans. The processing workflow included the following filters and corrections: time zero correction, traces repositioning, dewow filter, inverse amplitude decay gain, average background removal, bandpass filtering, and instantaneous envelope calculation. The produced C-scans were studied, and examples were then manually selected to define three classes: a) structure, b) stripping noise, and c) more generic patterns that are not identified as structures and are described as geophysical anomalies (**Figure 2**). These three feature classes are fundamental in the interpretation of the archaeological data in order to understand the subsurface conditions.

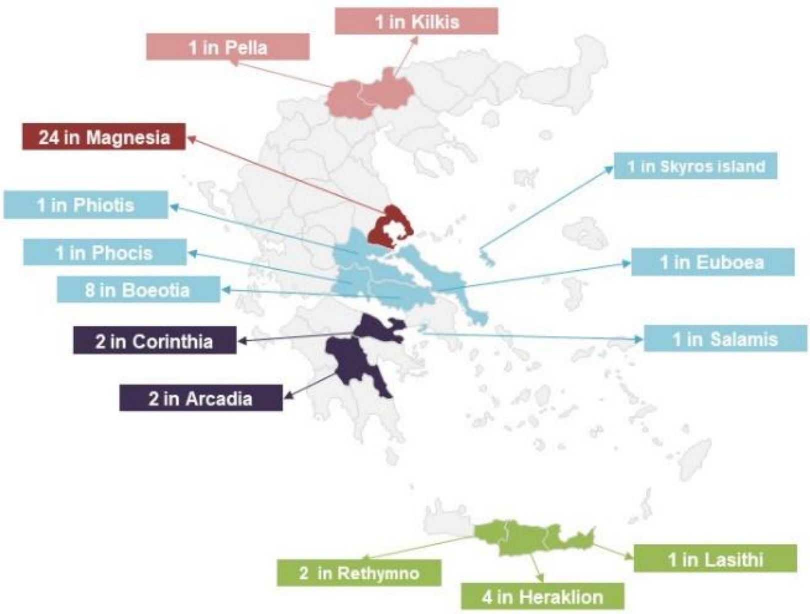


Figure 1: Greek prefectures where data collection was performed.The total number of GPR survey grids used for the dataset construction is also mentioned per area.

A challenge met during the dataset construction for training CNNs was locating enough images per class to avoid underfitting,indicating no learning from the data. The leading cause was the co-existence of the three feature classes in the selected C-scans. Hence, most images could not be assigned to asingle class. In order to overcome this issue,a sliding window was applied that crops square subregions of the input C-scan. The window size was set to 10x10m as it was found adequate to produce as many sub-images as possible per C-scan andsimultaneouslyresolve the patterns of interest satisfactorily. In addition, an overlapping step of 2.0m was used to increase the image number further. The effect of the sliding window crop is shown in **Figure 3**. All the images that were producedwere saved in .jpeg format using a pixel size of 256x256. This process was applied to all the selected C-scans, and a new manual selection of examples followed for each class, trying to avoid repetitive patterns and overly similar images. A total of 18,750 images were gathered,assigning 6,250 images per class. Next, the images were split into a training set and a test set following the 80%-20% rule.The result is 5,000 images per class dedicated to training models,while the remaining 1,250 are used to make predictions and test their generalization.

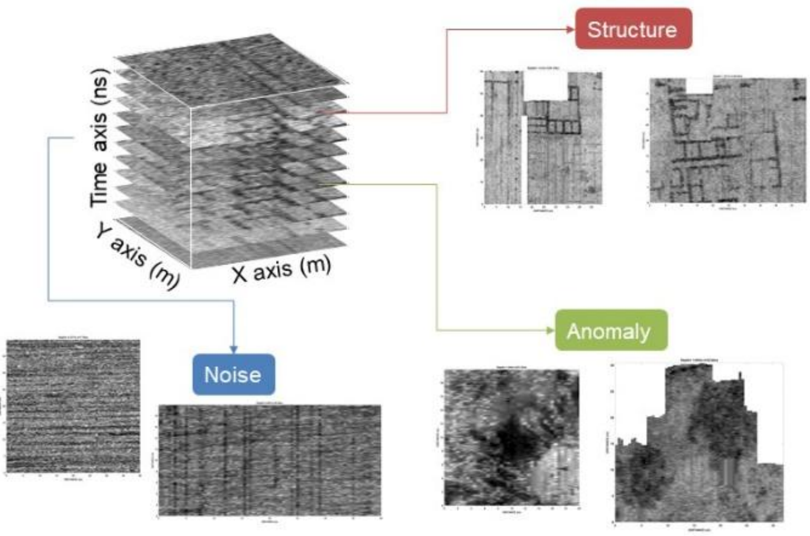


Figure 2: Process of the manual selection and annotation of the C-scans to construct the datasets with indicative examples for the structure, noise and geophysical anomaly class.

3.2 TrainingAlexNet

The annotated constructed dataset was used to train several models seeking performance improvements. The trials started with a baseline model that is a straightforward implementation of AlexNet as presented by Krizhevsky et al. (2012) and wastrained with SGD using a batch size of 128 examples, a learning rate of 0.01, and momentum of 0.9 for 100 epochs.The baseline model was used as a reference to monitor performance improvements and the gains of applying BN and dropout. For all the trials performed, BN was applied only after the convolutional layers and before the activation function, while dropout was used in the first two fully connected layers. Several different combinations were tested and are presented in **Figure 4**. For the combination of BN and dropout that returned the highest classification accuracy, tuning the SGD optimizer hyperparameters (i.e., batch size and learning rate)followed to improve performance further. Retraining was then held, using the best set of hyperparameters, to obtain the final model. AlexNet implementation and training models were performed in Python using the Tensorflow library and Keras API while employing GPU computations to speed up the process. As for tuning the optimizer's hyperparameters, therandom search algorithm ofthe Keras Tuner library was used. The latter trains models for a different combination of batch size and learning rate set by the user, using only the training set to find the pair that returns the highest classification results. The performance was evaluated using the Keras library's standard metrics, which are the classification accuracy and loss functions calculated for the training set and test set after an epoch is completed. These values are plotted with respect to the epochs producing the learning curves that can reveal helpful information regarding the training robustness and whether a model overfits. Last, the best-acquired model generalization was evaluated by making predictions on new GPR images.

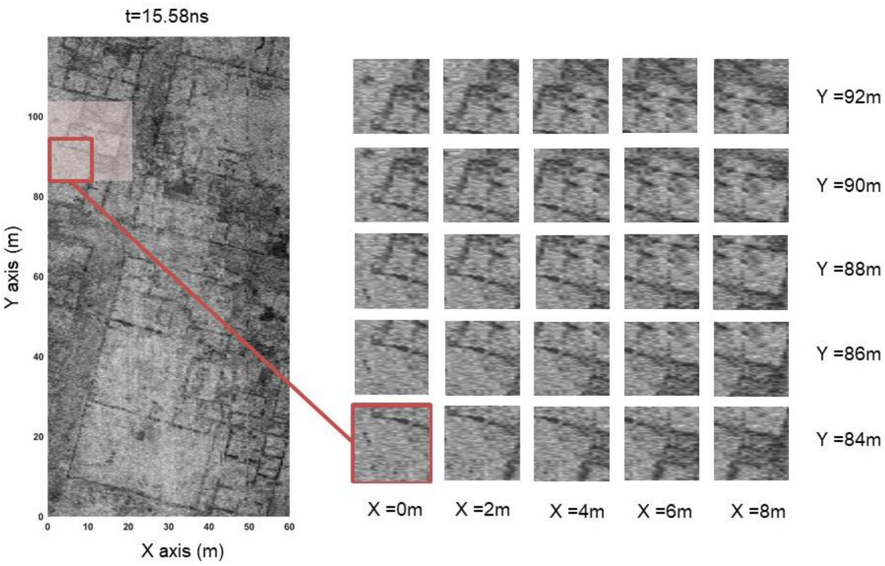


Figure 3: The effect of the sliding crop window of size 10mx10m with an overlapping step of 2.0m (both in the X and Y axis)applied in a region of a C-scan (fade red square) collected from Ancient Demetrias.If the starting point of the window is X=0m and Y=84m,it produces the sub-image framed with the red square. The rest of the sub-images shown here are produced by applying the overlapping step in both axis directions from the starting point.

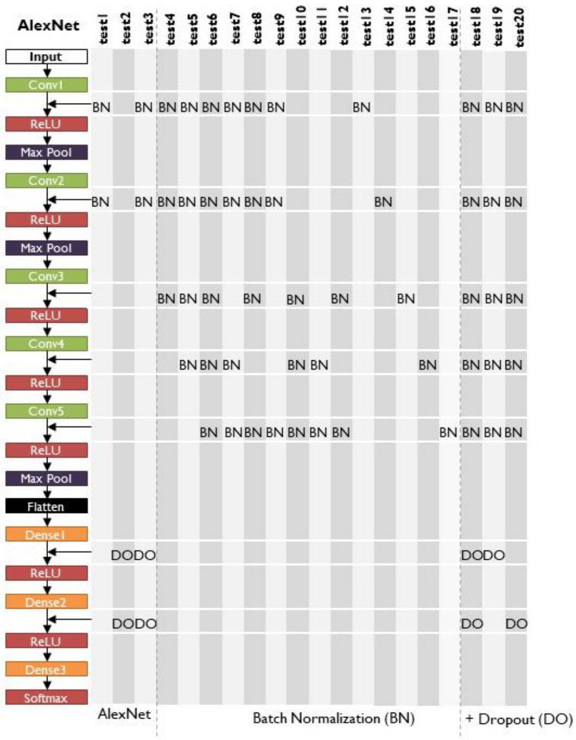


Figure 4: AlexNet architecture outline along with the different trials performed to test different combinations of BN and DP that yield better performance.

4 Results

The training results of each trial described in **Figure 4** are summarized in the corresponding learning curves of **Figure 5** for 50 epochs. On the left side are the loss and accuracy calculated on the training set, while on the right are the ones calculated on the test set.Learning was not feasible for most of the trials performed (gray color), including the baseline model. The accuracy is lowwhile the loss is high, and both do not change as the epochs progress. This behavior shows that no optimization is achieved. However, this changes when applying BN to the first four (test5) and first five (test6) convolutional layers, with the latter giving significantly better results. The validation accuracy increases while the validation loss is dropping near zero, indicating no signs of overfitting. In addition, the performance oftest6 is more robust than test5 which appears noisy.Further, the model produced by test6 reaches the best classification score faster. Hence, BN applied to all five convolutional layers is critical for learning. A dropout of 0.5 was used for FC-1 and FC-2, resulting in small performance gains and increasing the classification accuracy on the test set. Hence, the combination described by test18 (**Figure 4**) gave the best results and was used to tune the learning rate and batch size of the SGD optimizer to improve the results further.

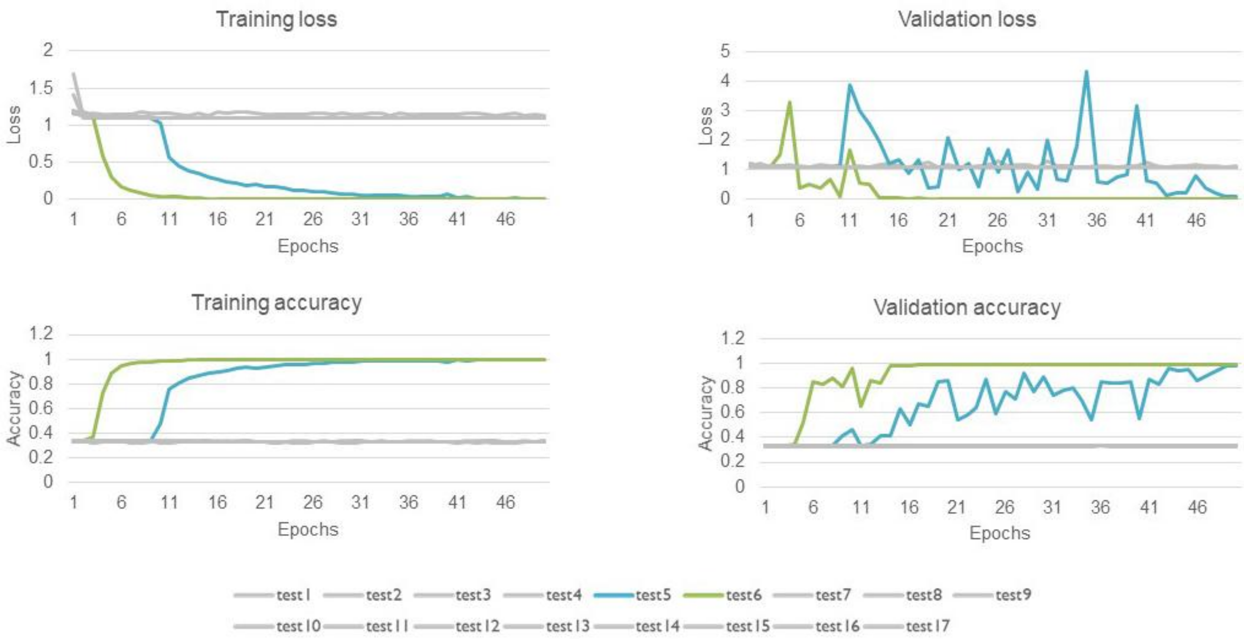


Figure 5: Learning curves showing gains in performance with Batch Normalization. Gray lines indicate poor performance with no learning. Learning was only feasible for the setup of test5 (blue) and test6 (green), with the latter being the best.

		Learning rates				
		0.01	**0.001**	**0.0001**	**0.00001**	**0.000001**
Batch sizes	**16**	0.966	0.978667	0.974333	0.955667	0.6673333
	32	0.333333	0.967667	0.966	0.918	0.6293333
	64	0.937667	0.969667	0.948333	0.865333	0.549
	128	0.957667	0.964333	0.939667	0.743333	0.5073333
	256	0.79	0.895	0.819333	0.655333	0.5023333

Table 1: Classification accuracy calculated for different pairs batch sizes and learning rates on a subset of the original training set.

Tunning the hyperparameters revealed several combinations of batch size and learning rate that return very high classification accuracies (>95%). These are shown in **Table 1** with the learning rate of 0.001 to concentrate the highestaccuracies which are of approximate values, making it difficult to understand which pair is better. Hence, new models were trained using the learning rate of 0.001 and the batch sizes of 16, 32, 64, and 128 to find the best pair. The obtained learning curves showed that the pair of batch size 64 and learning rate of 0.001 yields the highest prediction result, and at the same time, it does not overfit and has a more robust performance. The curves produced by the best-obtained model are shown in **Figure 6**.

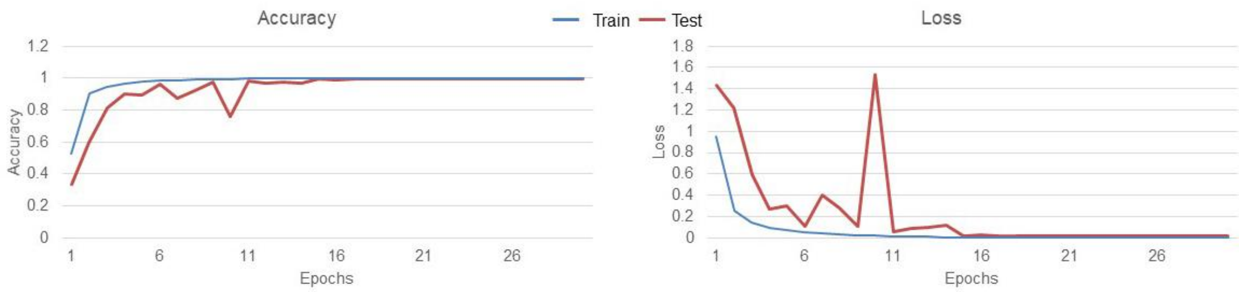


Figure 6: Learning curves of the best-obtained model reaching an accuracy of 99% on the test set. The model was trained with SGD and momentum 0.9, using the learning rate 0.001 and batch size 64. Batch normalization was applied to all five convolutional layers, and dropout was applied to FC1 and FC2.

This model was used to predict new GPR images to better evaluate its generalization. A total of 100 images containing examples from the three feature classes were used. These images were derived from the archaeological sites of Ancient Halos in Magnesia and Sissi in Heraklion, which were entirely excluded from the training process and the annotated dataset. The classification score for all predictions reached 92%, and the results are presented in **Figure 7,** where the GPR images used are shown along with details on the wrong predictions. For the anomaly class, 30 out of 32 examples were predicted correctly. Mistakes were made for examples #20 and #23, which were classified as structures. Even though the predictions are satisfactory, a lack of robustness is observed, as slightly different images of the same feature (i.e., #20, #23, #24) were not classified the same. The predictions on the noise class were also good, having classified 29 out of 32 examples correctly. Mistakes were made in examples #38, #39, and #40, which were classified as structures. Last, the predictions made for the structure class were also good, having correctly classified 33 out of 36 examples. The mistakes were made in examples #82, #84, and #87. These examples are patterns of poorly preserved structures classified as anomalies. Similar to noise class, predictions of anomaly class also lack robustness (i.e., #82, #83, and #84). Last, one encouraging observation concerns the examples #97, #98, and #99,which were classified correctly and are ground truth data through the excavations that have taken place in Sissi archaeological site.

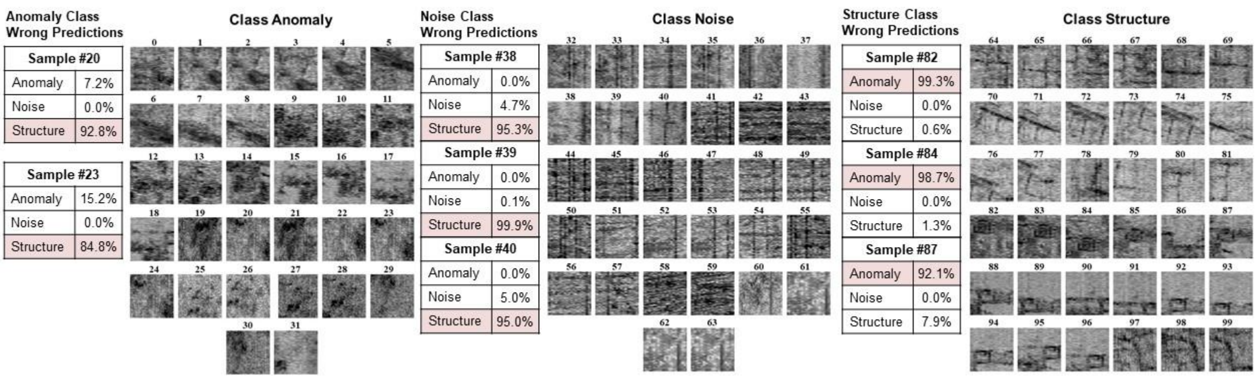


Figure 7: Predictions results on 100 new GPR images derived from Sissi and Halos archaeological sites. More details are given on the wrong predictions, while the rest of the examples were classified correctly, reaching a total accuracy of 92%.

5 Conclusions& Future Work

This study examined the automatic classification of archaeological GPR data using the AlexNet CNN architecture. This was performed through a series of steps, includingannotated dataset construction, training AlexNet, applying methods and techniques to improve performance, and evaluating learning using unseen data from the training process. Training a deep learning architecture like AlexNet with GPR data proved challenging. The initial data gatheredwere too few, and the first training attempts resulted in underfitting and no learning from the data. Eventually, the image number was increased by applying a sliding crop window with an overlapping step on the selected C-scans. This achieved an annotated dataset with 5,000 images as a training set and 1,245 as a test set. However, the training of AlexNet with SGD remained a challenge, as optimization was not achieved for most training trials, leading again to no learning. However, applying BN to all five convolutional layers of AlexNet significantly improved the overall performance, leading to very high validation accuracy.Further performance gains were also achieved when using dropout to FC1 and FC2 and tuning the learning rate and batch size of SGD. Finally, predictions were made using 100 new GPR examples to evaluate the generalization of the best-obtained model, reaching a classification accuracy of 92%.

Despite the training being challenging, the results are encouraging, showing that CNNs can work very well with GPR C-scans and are a promising approach that could make the automated data interpretation scheme achievable. As a result, this opens a new ground for further research and improvements. Among thoseis to improve the training dataset byincreasing the number and the variety of the examples. This can be achieved either through new data collection and data sharing that is strongly encouraged or through synthetic data. The increase in data examples can also lead to the definition of more complex feature classes, characterizing, for instance, the type and integrity of the buried structures, the type of noise, and the type of anomaly observed in the data. Regarding the training improvements, a couple of techniques recommended by many authors that were not tested in this study are ensemble learning and transfer learning. In the former, weights obtained from different models are averaged, while in the latter, the weights are initialized from a trained model. Both methods improve classification performance and the training process, even when the data are limited and small in numbers. Further, since the optimizer plays an important role, another suggestion is to examine optimizers other than SGD and use adaptive learning rates to reach the local minimum faster. The last is to explore more sophisticated CNNs architectures, including those performing image segmentation and object detection.

6 References

Abu-Mostafa, Y.S., Magdon-Ismail, M., Lin, H.-T., 2012, Learning from data. AMLBook New York, NY, USA:

Annan, A.P., 2009, Chapter 1 - Electromagnetic Principles of Ground Penetrating Radar, in: Harry M. Jol (Ed.), Ground Penetrating Radar Theory and Applications. Elsevier, Amsterdam, pp. 1–40.

Conyers, L.B., 2012, Interpreting Ground-penetrating Radar for Archaeology, 1st ed. Routledge, New York. https://doi.org/10.4324/9781315426334

Conyers, L.B., 2004, Ground-Penetrating Radar for Archaeology. AltaMira Press, Walnut Creek, CA.

Goodfellow, I., Bengio, Y., Courville, A., 2016, Deep Learning. The MIT Press.

Goodman, D., 2009, GPR methods for archaeology. Seeing Unseen Geophys. Landsc. Archaeol. 229–244.

Ioffe, S., Szegedy, C., 2015, Batch normalization: Accelerating deep network training by reducing internal covariate shift. ArXiv Prepr. ArXiv150203167.

Krizhevsky, A., Sutskever, I., Hinton, G.E., 2012, Imagenet classification with deep convolutional neural networks, in: Advances in Neural Information Processing Systems. pp. 1097–1105.

Küçükdemirci, M., Sarris, A., 2020, Deep learning based automated analysis of archaeo-geophysical images. Archaeol. Prospect. 27, 107–118. https://doi.org/10.1002/arp.1763

LeCun, Y., Bengio, Y., Hinton, G., 2015, Deep learning. Nature 521, 436–444. https://doi.org/10.1038/nature14539

Luo, P., Wang, X., Shao, W., Peng, Z., 2018, Towards understanding regularization in batch normalization. ArXiv Prepr. ArXiv180900846.

Manataki, M., Sarris, A., Donati, J.C., Cuenca Garcia, C., Kalayci, T., 2015, GPR: Theory and Practice in Archaeological Prospection, in: Best Practices of Geoinformatic Technologies for the Mapping of Archaeolandscapes. Archaeopress Archaeology, pp. 13–24.

Manataki, M., Vafidis, A., Sarris, A., 2021, GPR Data Interpretation Approaches in Archaeological Prospection. Appl. Sci. 11, 7531. https://doi.org/10.3390/app11167531

Pakkanen, J., Lentini, M.C., Sarris, A., Tikkala, E., Manataki, M., 2019, Recording and Reconstructing the Sacred Landscapes of Sicilian Naxos. Open Archaeol. 5, 416–433. https://doi.org/10.1515/opar-2019-0026

Ronneberger, O., Fischer, P., Brox, T., 2015, U-Net: Convolutional Networks for Biomedical Image Segmentation, in: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (Eds.), Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. Springer International Publishing, Cham, pp. 234–241. https://doi.org/10.1007/978-3-319-24574-4_28

Sarris, A., Kalayci, T., Simon, F.-X., Donati, J., Cuenca García, C., Manataki, M., Cantoro, G., Kalogiropoulou, E., Karampatsou, G., Argyriou, N., 2017, Opening a New Frontier in the Neolithic Settlement Patterns of Eastern Thessaly, Greece., in: Communities, Landscapes, and Interaction in Neolithic Greece, Archaeological Series. Presented at the International Conference: Communities, Landscapes, and Interactions in Neolithic Greece, International Monographs in Prehistory, Rethymno, pp. 27–48.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., Salakhutdinov, R., 2014, Dropout: a simple way to prevent neural networks from overfitting. J. Mach. Learn. Res. 15, 1929–1958.

Sutskever, I., Martens, J., Dahl, G., Hinton, G., 2013, On the importance of initialization and momentum in deep learning, in: International Conference on Machine Learning. pp. 1139–1147.

AUGMENTING EXISTING FOOD IMAGE DATASETS WITH GREEK DISHES

V. SEVETLIDIS[1], C. KIOURT[1], C. TZOUVARA[1,2], G. TASTZOGLOU[1,2], AND G. PAVLIDIS[1]

[1]Athena Research and Innovation Center, Kimmeria Campus, PO-BOX 159, GR-67100, Xanthi, Greece,
vasiseve@athenarc.gr, chairiq@athenarc.gr, gpavlid@athenarc.gr
[2] Department of Electrical and Computer Engineering, Democritus University of Thrace, Greece,
grigtast1@ee.duth.gr, chartzou2@ee.duth.gr

Abstract

Gastronomic tourism is a quest of individual experience. To assist tourists in their visits a food recognition service has been developed focused on dishes from the Greek cuisine. Since no dataset of Greek food images existed, a workflow is proposed to automatically construct such a dataset with samples acquired from the Web. The workflow addresses the utilization of a Greek food specific ontology to extract keywords from which the classes of the dataset are based on, discarding unwanted content i.e., the removal of duplicate and irrelevant images and the merging of underpopulated classes into broader categories. This dataset is employed as a training template for a deep learning architecture and its produced model is used as the core of a publicly available recognition service.

***Keywords**: gastronomic tourism, food recognition, work*

1. Introduction

Food, an essential product to sustain life, is also a fundamental pillar of family and social relationships as well as an expression of art and culture, according to the European Parliament's Committee on Culture and Education. Gastronomy's importance to EU regional development is shown by the promotion of events organized by the S3 Platform, which addressed food innovation as a driving force for growth. At the same time, mass tourism is reported to have reached a tipping point, and the quest of individual experiences is taking over. Such encounter offers the gastronomic tourism, which according to (Lin 2006) is *avisitto primary and secondary producers of food and beverages, gastronomy festivals, dining venues and specific locations, where tasting and experience of specialty local food features are a prime motivation for the visit*.

Food identification is a challenging task with broad range of applications in gastronomic tourism and health tracking that has met an increasing interest by the computer vision community, due to inherent complications present in all dishes alike: the lack of distinctive spatial layout in food images makes the recognition prone to inconsistencies. Ingredients, say of a salad, constitute mixtures that typically come in different shapes and sizes; furthermore, often the nature of a dish is defined by the different colours, shapes and textures of the ingredients (Kiourt, Pavlidis & Markantonatou 2020).The prerequisites of developing a recognition service include mainly the apprehension of a dataset that represents the topic of interest, a classification model trained on those data to perform the actual recognition and finally a mechanism for transforming the model into an application.

The GRE-Taste project targets culinary tourism and food tradition in northern Greece and has developed a suite of services and resources aiming at becoming a useful companion to those who travel in Greece for the purpose of experiencing local food. Among other services, the GRE-Taste offers food recognition, and this paper presents the approach which has been followed in order to develop an AI enabled service. The rest of the paper is organized in three sections: first a background review goes into recent developments regarding food datasets and recognition, second the approach of producing the necessary resources for having a classification model of Greek food dishes is presented and last a discussion of the results and the model's incorporation in the GRE-Taste list of services.

2. Background

Surveys on food recognition and food computing in general (Min et al. 2019, Allegra et al. 2020) address different tasks, their accompanying challenges, and proposed methods. In the earlier years, the computer vision community extracted various handcrafted features like SIFT, Visual Bag of Words, Colour Histograms, Gabor filters, Histogram of Oriented Gradients, Haar wavelets and many others alike from food images. These features have been used in textbook machine learning models such as Random Forests and Multiple Kernel Support Vector Machines to classify foods into categories in predefined classes in the context of Supervised Learning. Later, Convolutional Neural Networks have been exploited to produce automatically features that were encoded using the Fisher Vector technique. Recent advances in deep learning have gained significant attention due to its impressive performance. As a result, existing methods resort to deep learning for food recognition (Horiguchi et al. 2018, Martinel et al.). There are also works, which utilize additional context information, such as ingredients and location (Chen and Ngo, Min et al. 2019b) to improve the recognition performance designed a two-branch network to learn global and local features jointly to enable discriminative feature representation for food recognition(Min et al. 2020).

Other tasks, related to food recognition are the food quantity estimation. Food quantity estimation is very important in the context of health tracking applications since it provides assessments regarding food intakes (Chen et al. 2013, Ciocca, Napoletano & Schettini 2015). Finally, very few works specifically consider the problem of leftover estimation. Often the problem is theoretically treated as a special case of the problem of food recognition and quantity estimation (Zhu et al. 2010, Pouladzadeh, Yassine & Shirmohammadi 2015). Regardless of the objective, a dataset of food images is required to evaluate the performance of the different feature extraction and classification algorithms. To this end, the above research works either use existing datasets or introduce new ones.

Table ***1***summarizes the statistics of benchmark image datasets related to food. Non-automatic image acquisition procedures for food datasets are divided in in-situ acquisition (Chen et al. 2009) and browsing for them on the Web (Bossard, Guillaumin & Gool 2014, Joutou & Yanai 2009, Hoashi, Joutou & Yanai 2010). Automatic frameworks for food dataset construction are discussed in (Liu et al. 2018, Wang et al. 2015). Finding content in such uncontrolled environment might be faster than other approaches, but introduces additional challenges i.e., the sample per class distribution might not be the desired one and food co-occurrence class assignment is often disregarded. Food co-occurrence is discussed in (Matsuda, Hoashi &Yanai 2012), where various image descriptors were used to extract local and global features to recognize multiple-food photos considering co-occurrence statistics. Later, the same authors employed a manifold learning approach to improve their results (Matsuda & Yanai 2012). Finally, other frameworks (not explicitly related to food) for image acquisition from the Web are discussed in (Krizhevsky 2009, Krizhevsky, Sutskever & Hinton 2017).

Dataset	Number of images	Number of classes	Cuisine
PFID(Chenetal. 2009)	4,545	101	Japanese
UEC Food 100 (Matsuda, Hoashi&Yanai 2012)	14,361	100	Japanese
UEC Food 256(Kawano &Yanai 2015)	25,088	256	Japanese
Sushi-50(Qiu et al. 2019)	3,963	50	Japanese
Chinese FoodNet(Chen X. 2017)	192,000	206	Chinese
Vireo Food-172(Chen & Ngo 2016)	110,241	172	Chinese
ETH-Food 101(Bossard, Guillaumin& Gool 2014)	101,000	101	Western
UPMC Food 101(Wang et al. 2015)	90,840	101	Western
UNIMIB 2015(Ciocca, Napoletano&Schettini 2015)	2,000	15	Generic
UNIMIB 2016(Ciocca, Napoletano&Schettini 2017)	1,027	73	Generic
FoodX251(Kaur et al. 2019)	158,846	251	Generic
ISIA Food-200(Min et al. 2019b)	197,323	200	Generic
ISIA Food-500(Min et al. 2020)	399,726	500	Generic

Table 1. Cross-referencing food datasets and the cuisine of their focus

3. Towards developing a Greek Food Recognition Service

3.1. The dataset

Apparently, in GRE-Taste the problem of food recognition is even more challenging, as traditional and regional Greek food and dish images have not been used to develop models in the past. Thus, the first issue to be tackled was the selection of an appropriate dataset that would aid in the recognition of Greek dishes; however, such a dataset did not exist. The most popular datasets consist of general/universal food categories (Table ***1***), whereas some are only regional. To amend for this shortcoming a data acquisition framework had to be developed. To avoid the in-situ manual acquisition's expensive overheads (site visits planning, camera calibration, setting up the scene, removing obstacles and avoiding capturing humans and human parts to preserve their anonymity, human labor, etc.), we opted for creating a dataset of Greek food images automatically with samples acquired from the Web. To achieve this two mechanisms have to be devised, namely a crawling mechanism to obtain food images of Greek cuisine and a data sorting operation to discard unwanted content and distribute the samples into their respective classes.
The work most closely related to ours, is the one which introduced the ISIA Food-500 (Min et al. 2020). That dataset is publicly available and contains around 400K images not equally distributed in 500 categories of diverse cuisine origins. The authors followed a similar approach to ours to enrich their dataset i.e., they included images from the Web and removed duplicate and non-food images. Further data inspection was applied with crowdsourcing. However, the authors weren't specific about the details of the data acquisition procedure. It is not clear how strict the rule about removing duplicates was (i.e., allowing the inclusion of geometrically transformed images, as similar but not duplicates), and how they dealt with the assignment of labels to conflicting dish classes (food co-occurrence).

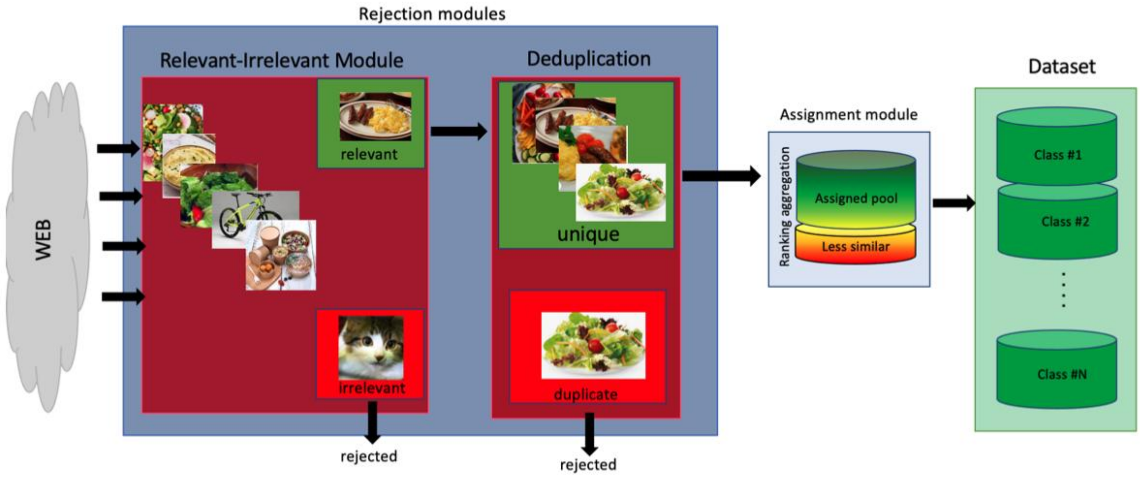


Figure 1. Overview of the proposed workflow

Regarding the dataset construction the proposed approach begins with enumerating the keywords for which the crawler seeks out on the Web. Since GRE-Taste's goal is to provide tourists with credible information about food, the list of terms must be complete, up-to-date and valid. To fulfill these conditions the terms representing Greek food dishes ideally should be extracted from a Greek food specific resource. Given that list of keywords, the crawler looks for each of the listings on the Web and captures the available sample images. These images are expected to contain the food represented by the keyword as well as depictions of topics loosely related to them. Of course, the less related images have to be sorted out as the more accurate the portrayal within the collection of the requested food the better the representation. In addition, the Web is an uncontrolled environment and incidentally its content is replicated. To avoid having the same image more than once or some geometrical deformation copy of another, the sorting mechanism has to be able to discard duplicate images as well. Finally, since food co-occurrence is a common phenomenon, the sample images have to match their respective classes decidedly. The overview of the aforementioned configuration is shown in Figure 1.

3.2. The classification model

As mentioned before, deep learning models demonstrated impressive performance at the task of food recognition. In the development of the food recognition service, the 'PureFoodNet' classification model was employed as a pivotal component of the proposed architecture. This model was developed by our research team, as detailed in our previous works

(Kiourt, Pavlidis & Markantonatou 2020 and Pavlidis et al. 2020), and has been instrumental in accurately categorizing Greek dishes by leveraging deep learning techniques. At this work (Pavlidis et al. 2020)several architectures were trained on the same food benchmark dataset and their performances were compared in terms of predictive accuracy and computational resources needed for their training. In a set of experiments PureFoodNet was tested against VGG16, InceptionV3, ResNet50, InceptionResNetV2, MobileNetV2, DenseNet121 and NASNetLarge. Apparently, most of the models performed comparably with respect to the accuracy, and the authors verdict about favoring one architecture over the other relied on the type of the application that the model would be used for as they demanded a light-weight solution. Considering this study, any of these architectures could provide similar indications about the accuracy to be achieved given a food dataset. So, we opted for the InceptionV2, which is known for its good performance in versatile recognition tasks.

3.3. The service

Modern day applications rarely operate as standalone computer programs, while most of the time they are cloud based and interactive. Again, GRE-Taste's goal is to be a companion to gastronomic tourists. In this spirit, the food recognition model should be accessible broadly (e.g., via the web) and readily available for devices that tourists carry and rely on mostly (e.g., smartphones). Thus, a REST-API is designed as a backbone infrastructure which operates in GRE-Taste's servers to provide with all necessary information food recognition requests. Moreover, to make the information presentable, the service is incorporated within the project's web portal as well as it has a smartphone app counterpart.

4. The proposed workflow architecture

4.1. Knowledgebase

Recently, a trilingual thesaurus of food served in restaurants in Eastern Macedonia and Thrace (North-Eastern Greece) has been developed, called "ΑΜΑΛΘΕΙΑ" (Amalthea) (Markantonatou et al. 2021). The authors designed and implemented a Web lexicographic environment, which accommodates information retrieved from restaurant menus. This infrastructure enabled the development of a thesaurus with information about dishes, their ingredients, recipes, concepts related to food, as well as dietary and cultural, which also considers:

- varieties in culinary language related with dialectic forms, local specialties, etc;
- dish variations, leading to the definition of specific types of dishes, e.g. "stifado" and "rabbit/beef/cuttlefish/stifado";
- synecdoche and ambiguity, i.e. when the same word denotes both the food and its source, e.g. "lettuce", which may be used to name the vegetable as a plant, the vegetable as an ingredient of a salad or the lettuce salad;
- categorization of dishes, i.e. some dishes may need to be classified under more than one general category (facet), e.g. "papoutsakia", literary small shoes, "eggplant with minced meat" may be classified both as meat dishes and vegetable dishes.

To incorporate this knowledge base (Amalthea) to the proposed approach a tree-like structure was implemented. A traversal mechanism granted the dish labels, their parent nodes and their corresponding ID codes. The dish labels were used as the queries to be provided to the Web search engines. We chose not to look for complex relations between dishes within the knowledge base even though label translations, recipes, related ingredients, and dish preparation procedures were, as stated, available.

4.2. Food/NonFood classification

To reject irrelevant content from the candidate pool of images, a Food/NonFood discriminator was trained via transfer learning. A binary class dataset with 200K colored images was constructed. The one class included 100K images from the ETH-Food101, while the other was a meticulous manual selection (with duplicate exclusion) of non-food images from benchmark datasets (Krizhevsky, Sutskever & Hinton 2017). The choice of ETH-Food101 was based on its content, which it was visually aligned to the criteria of the desired images for the GRE-Taste project. The chosen deep neural network model was the VGG-16 (Simonyan) pretrained on the ImageNet. The accuracy on the test set was 98.59%. This model was used as the Food/NonFood discrimination module.

4.3. Removing duplicate images.

To reject duplicate images, a deduplication procedure was developed to handle this task. Deduplication is used widely in many applications (forensics, closed-loop systems, etc.). Several approaches have been developed to deal with this task, including, but not limited to, relying on textbook image descriptors

(Sikora 2001) and (Ng 2003), as well as bespoke deep neural networks (Wang et al. 2014, Kansizoglou, Bampis & Gasteratos 2021).

Typically, in image processing a feature extractor combined with a distance measure, compare any two images to find out if they are similar or not. The application of a threshold θ defines the degree of similarity, which in this case reduces to being a duplicate or not. A reliable method is needed to (a) bring closer those images that are either exactly the same or geometrically transformed copies of one another, while at the same time to (b) push further away those, which have just a visual resemblance or are completely different. That said, image descriptors coupled with locality- sensitive hashing methods (Chi & Zhu 2017) offer the benefits of being fast feature extractors and reliable at detecting most affine transformations. At the same time, they keep the description of an image compact, which is computationally important for the comparison. In addition, Hamming distance fits nicely as a metric for these descriptors. To tackle the computational complexity a BK-tree structure (Burkhard & Keller 1973) is exploited, which indexes the binary hashes of the candidate images. Given a query hash and a threshold θ, this structure is used for the purpose of being faster in finding the hashes under that threshold. The complexity of this algorithm is approximately $O(logn)$.

The deduplication module in the proposed pipeline employs three image hashing methods, namely (a) the average image hash, (b) the perceptual hash and (c) the difference hash. An exclusive BK-tree is constructed for the three descriptors respectively. Each time an image is encountered its features are extracted and then compared with all candidate image hashes within the tree structures. A majority voting approach decides whether an image is duplicate or not; no tie braker is needed since there are three voters with equal voting weight. The query hash is stored within its respective BK-tree, if and only if there was no duplicate decision outcome (also the query image is added to the candidate pool).

4.4. Pruning candidate classes

The Web is an uncontrolled environment for the task of collecting representative images in respect to some label. It is expected for some classes to be overpopulated, whereas others be exactly the opposite. In the overpopulated class scenario, the reasonable thing to do is to keep the Kmost similar images based on some ranking. In contrast, underpopulated classes need a closer inspection to decide upon an action plan, since greedy approaches might result data losses. Usually, this issue is addressed with (a) seeking for more content, (b) discarding underqualified classes and (c) merging similar classes. Automating this procedure is not always trivial.

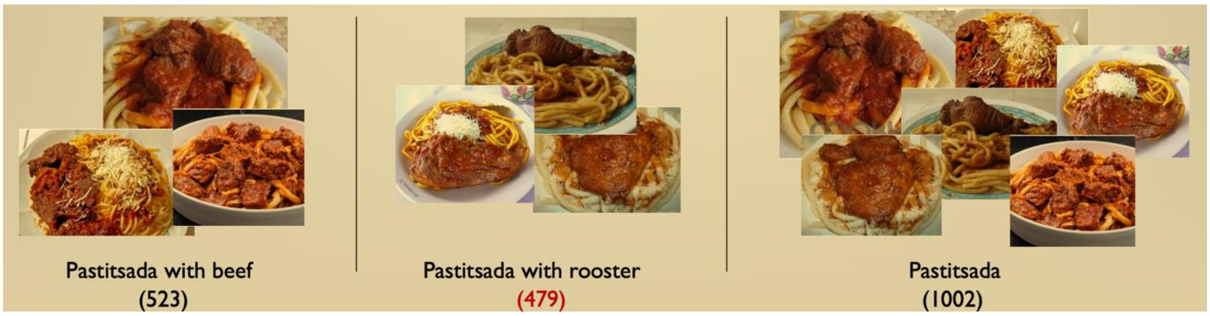


Figure 2. Two related classes merged into one super-class

To deal with this shortcoming the proposed workflow exploits the aforementioned knowledge base. When a class's candidate pool doesn't meet the completion criterion, then it is merged with a pertinent class within the same super-class group (children-nodes under the same parent). These relationships are extracted from Amalthea, the Greek food thesaurus. If no other class exists within a certain group, then the underpopulated one is merged with its parent. This pruning procedure is followed until all classes meet the completion criterion. Figure 2 shows a visual example of the *labels pastitsada with beef* and *pastitsada with rooster*. These dishes share the same base ingredients and recipe, but the protein is different. These labels can be merged into one super class, namely *pastisada.*

4.5. Class assignment based on aggregated rankings

Another shortcoming related to the uncontrolled environment of the Web, is that, given a query, no perfect retrieved pool of samples is expected to emerge. Actually, it is anticipated for some content to be misclassified. Duplicate and non-food retrieved images are discarded already by the respective modules within the workflow. Thus, the misclassified results contain only food samples fitted mainly in two categories: either they reside to a different class, or they belong to the desired class but only to some small degree.

To assign candidate images to a certain class an aggregation of two rankings is employed in the proposed pipeline. Under the assumption that search engines function as retrieval systems, which provide the results in a relevance order, the first ranking system emerges from their retrieved arrangement. The second ranking system arises from the visual similarity amongst the samples of a candidate pool. For this purpose, a rather simple yet effective image descriptor is devised, which is somehow close to the concept of the dominant color descriptor of the MPEG-7 standard (Sikora 2001), yet is simple and rather effective.

The descriptor by design considers color similarity between images, whilst is invariant to geometrical transformations. It analyzes colored images by matching the actual colors to a number of h hues and $v$intensity values in the HSV color space. This results in a quantized image with q = h + $v$channels. From the quantized image two further quantities are calculated per channel: (a) the coordinate of the center of mass and (b) the percentage coverage. Subsequently, these values are normalized within the range [0,1]. At the end, the normalized quantities are put together in a vector forming the final signature. We term this descriptor as Color Gravity Descriptor (CGD), due to its inherent characteristic to form color discs with specific mass and capture the attractions among them.

The procedure of ordering the candidate pool by its visual similarity is as follows: (a) extract the CGD signatures; (c) calculate a single cluster with Chebyshev distance; (d) calculate the mean $\mu$and standard deviation $\sigma$of that single cluster; and (e) use the $\mu$and $\sigma$to arrange the samples by their distance from the center of the cluster in a weighted manner.

As it happens, the two ranking systems have different arranging criteria. Thus, a method is needed to produce an optimal ranking that maximizes some sort of agreement. The approach followed in this pipeline is to calculate the pair wise disagreement τmeasured with Kendall Tau and then maximize the following loss function (van Blokland-Vogelesang 1989):

$$\text{argmax}\sum d(\bar{\tau},\tau)\leq\sum d(\tau,\tau)\ \forall\tau\in N(1)$$

This aggregation module can combine any ranking system, but a naive implementation can lead in trying to solve an NP-hard problem (Bartholdi, Tovey & Trick 1989). Approximations using weighted graphs have been proposed as well (Conitzer, Davenport & Kalagnanam n.d.).

5. Experimental verification

5.1. Setting up the parameters

The proposed workflow consists of several modules, which need certain parameters to work as expected i.e., the desired number $K$of samples per class, and the threshold θ per method which separates duplicates from the rest.

To find the optimal threshold for the image hashing algorithms, an experiment is conducted similar to the one proposed by (Ke, Sukthankar & Huston 2004). A random subset is taken comprised of 2000 unique food images from the ETH Food-101 dataset. For each image the following 40 transformations are applied: (a) colorization, (b) contrast adjustment, (c) cropping, (d) despeckling, (e) downsampling, (f) mirroring (flipping), (g) change of format, (h) framing, (i) rotation, (j) scaling, (k) saturation enhancement and (l) intensity adjustment. These transformations result in 82000images, of which the signatures are extracted respectively with each image hashing method. Distances and confusion matrices are calculated for all images. The desired threshold $\theta$is the one that minimizes the following function:

arg min $|FN - FP|$

To test the thresholds, another sample of 20000 images is taken from the same dataset with mutual exclusion in respect to the training one. Then, 500 of them are randomly selected and undergo the aforementioned transformations. This results in a test set of 40000 images. The achieved cross-validated accuracy for $\theta^{(\alpha)}$ = 3 (average image hash), $\theta^{(d))}$ = 14 (difference hash) and $\theta^{(p)}$ = 14 (perceptual hash) are 85.57%, 86.72% and 86.5% respectively.

5.2. Data acquisition

From the Amalthea ontology 361 labels of dishes served in Greek restaurants were extracted. The pipeline queried two search engines (Google and DuckDuckGo), in order to populate these classes with candidate samples. The Food/NonFood and deduplication modules rejected initially all the undesired content. A total of 401231 images were collected from the Web. In total, 128K were rejected as duplicate (32%) and 107K as non-food (26.8%). Note that, the deduplication module was not applied on the non-food images. Since the purpose of this pipeline is to produce a balanced dataset, the need for the pruning and the class assignment modules is evident in

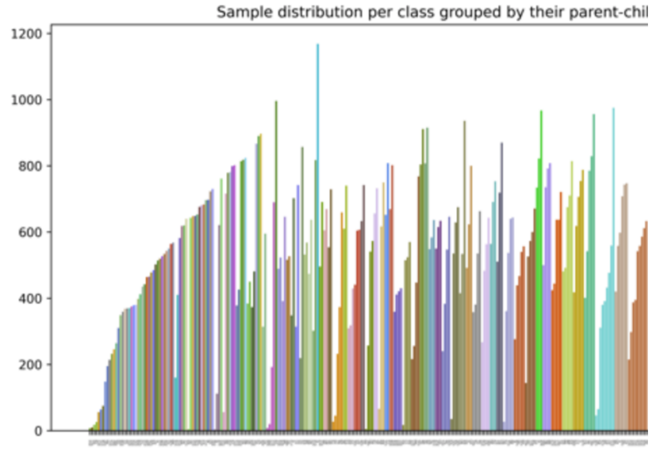


Figure ***3***, which shows many underpopulated classes within a group of semantically similar. After pruning the classes according to the procedure described previously, there were 169 classes left.

The pruning module by design merges underpopulated classes with other relevant classes in the same group or their parent nodes; this is repeated indefinitely until the completeness criterion is met. We chose to not let this happen in this work and we parameterized the module to stop the merging procedure when a class was merged with its parent. Thus, underdeveloped classes were removed completely (4 classes).

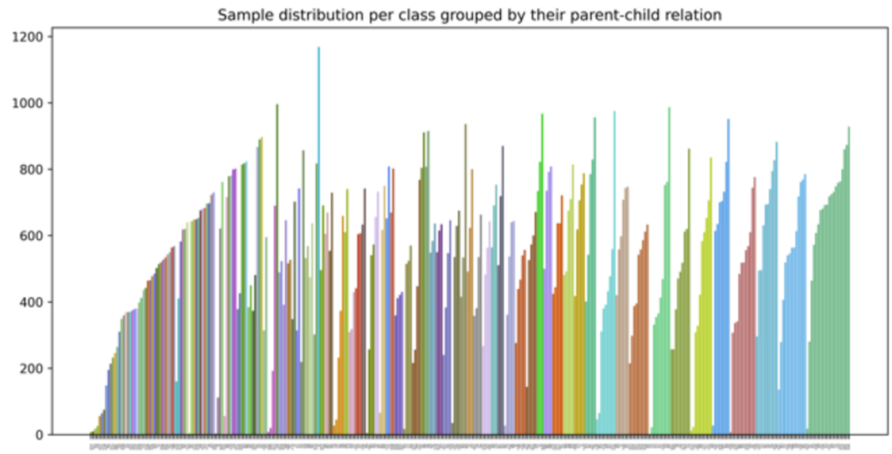


Figure 3. The sample distribution grouped by their relation to a common super-class

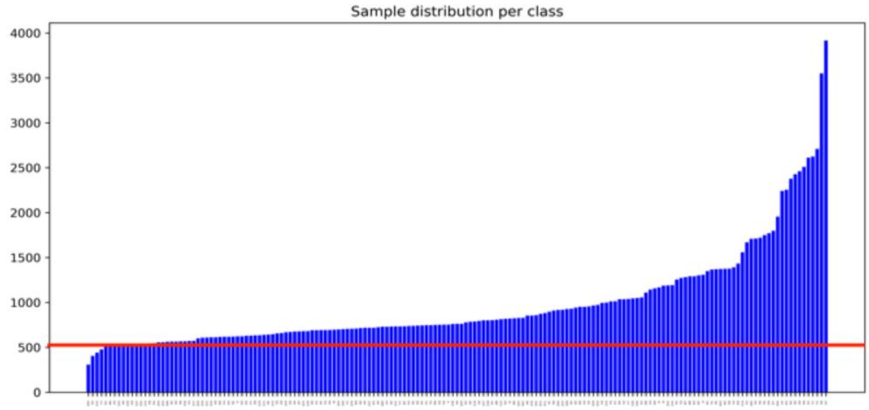


Figure 4. The final distribution of the dataset

The final module of this workflow is the one which assigns the candidate samples into their respective classes by aggregating a visual similarity and an indexing ranking system. Therefore, 165 classes were kept that met the dataset completion criterion (in this case 550 images per class), as shown in Figure 4.

5.3. Food recognition

A visual example of the created dataset, in which the samples were collected and curated by the proposed workflow, is shown in. To test its fitness for dish recognition, a classification experiment was conducted using the vanilla InceptionV2 architecture pretrained on the ImageNet. The dataset was split into training and testing instances with ratio 80 to 20 respectively. Affine transformations were applied on each training set. To avoid overfitting, regularization mechanisms were employed i.e., Dropout and batch normalization. The performance of classification on the test set is 76 (±4.81) %. The just above the average accuracy score indicates the classifier struggled to perform remarkably. Since we didn’t account for food co-occurrence and between classes’ similarity during the construction of the dataset, these are probable complications. However, classification tasks are attainable, in this dataset which has been constructed in a fully automatic manner.

6. GRE-Taste

The purpose of the GRE-Taste project is twofold: (a) creation of a platform for the promotion of gastronomic tourism as a holistic travel experience with the use of cutting-edge technologies and the cooperation of agri-food / catering / culture companies (b) scientific recording of the complex Greek cultural stock concerns food. The GRE-Taste project aims at an innovative rearrangement of the tourism applications landscape by adopting a thematic design of the tourism experience. In this context a theme/location-central planning system has been developed (Figure 5).

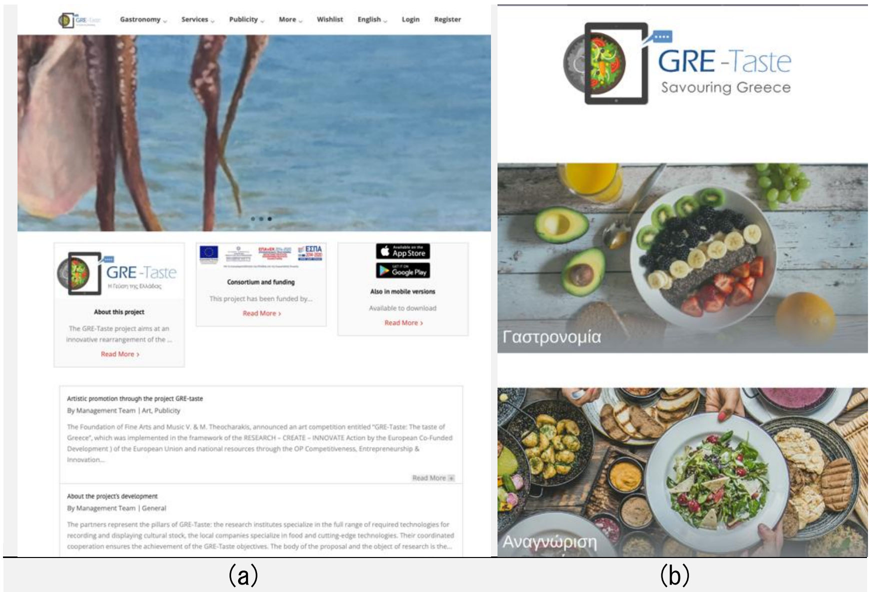


Figure 5. The GRE-Taste (a) web-portal's landing page and (b) the corresponding smartphone app.

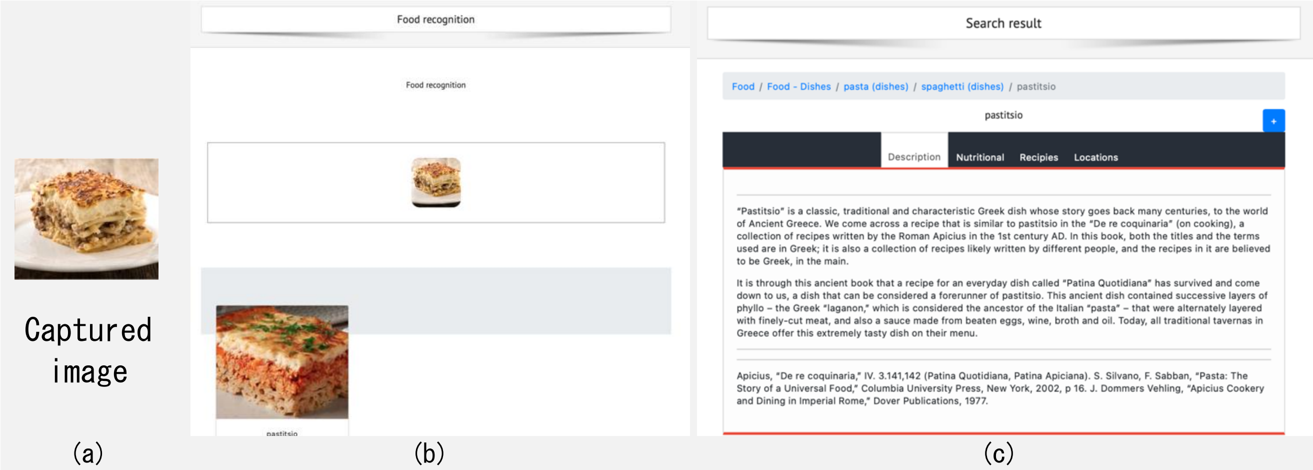


Figure 6. Using the recognition service offered at the GRE-Taste web portal. (a) the captured image is (b) uploaded at the service and the (c) result offers a multitude of information

Also, the project offers a trilingual (Greek, English, Russian) innovative tourist guide[1] with which the visitor first starts from what he or she wants to do, where and what he or she wants to eat and drink and based on these can arrange the places he or she can visit. All services are accessible via the project's web portal and its smartphone app. Particularly, the tourist can perform food recognition by capturing a photograph depicting a food and then upload it to the corresponding service. Then, internally the service performs the food recognition task using the model which was trained on the aforementioned dataset. A successful identification results in producing the identified label to the user prompting him to continue finding more about that plate (ingredients, similar foods, culture related notes, recipes, etc.). In case the identification turns ambiguous, the service replies with the TOP-5 label retrievals, so the user can choose from them and learn more. A visual example of the process is shown in

Figure ***6***.

7. Discussion

The development of a food recognition service for culinary tourism in northern Greece has been the primary focus of this work. However, the potential for the methodologies and workflows established to be applied in other domains, such as archaeology and history, has been identified. The adaptation of web crawling, data sorting, and machine learning model generation processes for the creation of image data repositories for archaeological artifacts, historical monuments, and cultural heritage sites is feasible. Nevertheless, the challenges associated with the initial collection of open access data in the humanities, particularly in archaeology and history, must be acknowledged. Rooted in concerns over preservation and copyright, hesitance towards open access data can hinder the aggregation of a comprehensive dataset. The adherence to FAIR data principles (Wilkinson 2016) — ensuring data is Findable, Accessible, Interoperable, and Reusable—is seen as crucial in overcoming these obstacles and fostering a culture of openness and collaboration in the humanities. The unlocking of new insights and fostering a deeper understanding of our cultural heritage can be facilitated by embracing these principles and leveraging the power of machine learning.

The integration of open access data principles in the humanities has been met with both challenges and opportunities. The digital era has brought about unprecedented access to information, yet a certain hesitance in fully embracing open access data has been exhibited by the humanities. This reluctance often originates from concerns over the preservation of sensitive cultural information, copyright issues, and the potential misuse of data (Novák 2021). Nevertheless, the enrichment of humanities research through the immense potential benefits of open access data has been recognized, offering possibilities for cross-

[1]A cross-platform online version of the system can be found in http://gre-taste.ceti.gr. The system can be used on a mobile device as a service and can be easily ported to a native mobile apphttps://gre-taste.athenarc.gr/?page_id=943.

disciplinary studies, wider dissemination of knowledge, and the democratization of information.

The FAIR data principles provide a robust framework for navigating these challenges. The ease of discovery and indexing in databases is ensured when data is made Findable. When data is made Accessible, it must be retrievable by both humans and machines, with clear and explicit conditions for its use. The ability of data to be integrated with other datasets, tools, and applications is referred to as Interoperability, which facilitates collaborative research. Reusability pertains to the capacity of data to be effectively reused in different contexts, thereby maximizing its value (Hiebel 2021).

The mitigation of challenges associated with open access data can be supported by adhering to FAIR data principles. A culture of transparency and collaboration is encouraged, fostering trust within the community. Furthermore, it ensures that collected and generated data is leveraged to its fullest potential, paving the way for innovative research and a deeper understanding of our cultural heritage. As the development of technologies and methodologies in fields like machine learning and data science continues, the alignment of these advancements with the FAIR principles to promote ethical, responsible, and impactful research in the humanities is deemed imperative.

8. Conclusions

In conclusion, an ontology is used to retrieve keywords related to Greek food so a crawler can query the web for data. The images collected from the web are processed via a module to filter out duplicate and irrelevant content. The valid ones form candidate pools in respect to their class. In case a candidate pool does not have the desired minimum number of samples, it is merged with the semantically closest class according to the information provided by the ontology. Additionally, not all of these images fit into their respective final class, since misclassified samples are expected to have been collected in the first place. To assign images into classes we assume intra-class visual coherence, thus a ranking based on visual similarity is used. Thus, to assign a candidate image to its respective class an aggregated ranking of its intra-class visual similarity and its position in the retrieved results is calculated. Finally, the top candidate images within the range of the minimum number of desired samples are harnessed and a Greek food image dataset is created. A deep learning architecture is trained on that dataset for the task of food recognition. The produced model is used at the core of a publicly accessible service via the GRE-Taste's web portal and its smartphone app counterpart.

Acknowledgments

Co-financed by the European Regional Development Fund of the European Union and Greek national funds through the Operational Program Competitiveness, Entrepreneurship and Innovation, under the call RESEARCH – CREATE - INNOVATE (project code:T1EDK-02015)

References

Allegra, D., Battiato, S., Ortis, A., Urso, S. &Polosa, R., 2020, 'A review on food recognition technology for health applications', Health Psychology Research, vol. 8, no. 3.

Bartholdi, J.J., Tovey, C.A. & Trick, M.A., 1989, 'The Computational Difficulty of Manipulating an Election', Social Choice and Welfare, vol. 6, no. 3, pp. 227–241, viewed 21 February 2022, https://www.jstor.org/stable/41105918

Bossard, L., Guillaumin, M., & Van Gool, L., 2014, Food-101–mining discriminative components with random forests. In *Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part VI 13* (pp. 446-461). Springer International Publishing.

Burkhard, W.A. & Keller, R.M., 1973, 'Some approaches to best-match file searching', Communications of the ACM, vol. 16, no. 4, pp. 230–236.

Chen, H.-C., Jia, W., Yue, Y., Li, Z., Sun, Y.-N., Fernstrom, J.D. & Sun, M., 2013, 'Model-based measurement of food portion size for image-based dietary assessment using 3D/2D registration', Measurement Science and Technology, vol. 24, no. 10, p. 105701.

Chen, J. & Ngo, C., 2016, 'Deep-based Ingredient Recognition for Cooking Recipe Retrieval', Proceedings of the 2016 ACM on Multimedia Conference - MM '16.

Chen, M., Dhingra, K., Wu, W., Yang, L., Sukthankar, R. & Yang, J., 2009, PFID: Pittsburgh fast-food image dataset, IEEE Xplore, pp. 289–292, viewed 4 August 2021, https://ieeexplore.ieee.org/document/5413511

Chi, L. & Zhu, X., 2017, 'Hashing Techniques', ACM Computing Surveys, vol. 50, no. 1, pp. 1–36.

Ciocca, G., Napoletano, P. & Schettini, R., 2015, 'Food Recognition and Leftover Estimation for Daily Diet

Monitoring’, New Trends in Image Analysis and Processing -- ICIAP 2015 Workshops, pp. 334–341.

Ciocca, G., Napoletano, P. & Schettini, R., 2017, ‘Food Recognition: A New Dataset, Experiments, and Results’, IEEE Journal of Biomedical and Health Informatics, vol. 21, no. 3, pp. 588–598.

Conitzer, V., Davenport, A. & Kalagnanam, J. n.d., Improved Bounds for Computing Kemeny Rankings *, viewed 21 February 2022, https://www.aaai.org/Papers/AAAI/2006/AAAI06-099.pdf

Hoashi, H., Joutou, T. & Yanai, K., 2010, ‘Image Recognition of 85 Food Categories by Feature Fusion’, 2010 IEEE International Symposium on Multimedia.

Horiguchi, S., Amano, S., Ogawa, M. & Aizawa, K., 2018, ‘Personalized Classifier for Food Image Recognition’, IEEE Transactions on Multimedia, vol. 20, no. 10, pp. 2836–2848, viewed 21 February 2022, https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=8316919

Joutou, T. &Yanai, K., 2009, A food image recognition system with Multiple Kernel Learning, IEEE Xplore, pp. 285–288, viewed 21 February 2022, https://ieeexplore.ieee.org/document/5413400

Kansizoglou, I., Bampis, L. & Gasteratos, A., 2021, ‘Deep Feature Space: A Geometrical Perspective’, IEEE Transactions on Pattern Analysis and Machine Intelligence, pp. 1–1, viewed 21 February 2022, https://ieeexplore.ieee.org/abstract/document/9477034

Kaur, P., Sikka, K., Wang, W., Belongie, S.J .& Divakaran, A., 2019, ‘FoodX-251: A Dataset for Fine-grained Food Classification’, ArXiv, viewed 21 February 2022, https://www.semanticscholar.org/paper/FoodX-251%3A-A-Dataset-for-Fine-grained-Food-Kaur-Sikka/fc01959b365de833fad2230464274ce4a34c0288

Kawano, Y. & Yanai, K., 2015, ‘Automatic Expansion of a Food Image Dataset Leveraging Existing Categories with Domain Adaptation’, Computer Vision - ECCV 2014 Workshops, pp. 3–17.

Ke, Y., Sukthankar, R., Huston, L., Ke, Y., & Sukthankar, R., 2004, Efficient near-duplicate detection and sub-image retrieval. In *ACM multimedia* (Vol. 4, No. 1, p. 5).

Kiourt, C., Pavlidis, G. & Markantonatou, S., 2020, ‘Deep Learning Approaches in Food Recognition’, Learning and Analytics in Intelligent Systems, pp. 83–108.

Krizhevsky, A., 2009, Learning multiple layers of features from tiny images, CiteSeer, viewed 21 February 2022 http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.222.9220

Krizhevsky, A., Sutskever, I. & Hinton, G.E., 2017, ‘ImageNet classification with deep convolutional neural networks’, Communications of the ACM, vol. 60, no. 6, pp. 84–90.

Lin, Y., 2006, ‘C.M. Hall, L. Sharples, R. Mitchell, N. Macionis and B. Cambourne, Editors, Food Tourism Around the World: Development, Management and Markets, Butterworth Heinemann, Oxford (2003) ISBN 0-7506-5503-8 (392pp., £24.99)’, International Journal of Hospitality Management.

Liu, C., Cao, Y., Luo, Y., Chen, G,. Vokkarane, V., Yunsheng, M., Chen, S. & Hou, P., 2018, ‘A New Deep Learning-Based Food Recognition System for Dietary Assessment on An Edge Computing Service Infrastructure’, IEEE Transactions on Services Computing, vol. 11, no. 2, pp. 249–261, viewed 19 March 2021, https://ieeexplore.ieee.org/document/7837725

Markantonatou, S., Toraki, K., Minos, P., Vacalopoulou, A., Stamou, V. &Pavlidis, G., 2021, ‘ΑΜΑΛΘΕΙΑ: A Dish-Driven Ontology in the Food Domain’, Data, vol. 6, no. 4, p. 41.

Martinel, N., Foresti, G.L. & Micheloni, C., 2018, Wide-Slice Residual Networks for Food Recognition, IEEE Xplore, pp. 567–576.

Matsuda, Y., Hoashi, H. & Yanai, K., 2012, Recognition of Multiple-Food Images by Detecting Candidate Regions, IEEE Xplore, pp. 25–30, viewed 11 May 2020, https://ieeexplore.ieee.org/abstract/document/6298369

Matsuda, Y. & Yanai, K., 2012, Multiple-food recognition considering co-occurrence employing manifold ranking, IEEE Xplore, pp. 2017–2020, viewed 21 February 2022, https://ieeexplore.ieee.org/document/6460555

Min, W., Jiang, S., Liu, L., Rui, Y. & Jain, R., 2019a, ‘A Survey on Food Computing’, ACM Computing Surveys, vol. 52, no. 5, pp. 1–36.

Min, W., Liu, L., Luo, Z. & Jiang, S., 2019b, ‘Ingredient-Guided Cascaded Multi-Attention Network for Food Recognition’, Proceedings of the 27th ACM International Conference on Multimedia.

Min, W., Liu, L., Wang, Z., Luo, Z., Wei, X., Wei, X. & Jiang, S., 2020, 'ISIA Food-500: A Dataset for Large-Scale Food Recognition via Stacked Global-Local Attention Network', Proceedings of the 28th ACM International Conference on Multimedia, viewed 21 February 2022, https://arxiv.org/pdf/2008.05655.pdf

Ng, P.C., 2003, 'SIFT: predicting amino acid changes that affect protein function', Nucleic Acids Research, vol. 31, no. 13, pp. 3812–3814.

Pavlidis, G., Markantonatou, S., Toraki, K., Vacalopoulou, A., Strouthopoulos, C., Varsamis, D., Tsimpiris, A., Mouroutsos, S., Kiourt, C., Sevetlidis, V. & Minos, P., 2020, AI in gastronomic tourism, Berlin, Germany.

Pouladzadeh, P., Yassine, A. & Shirmohammadi, S., 2015, 'FooDD: Food Detection Dataset for Calorie Measurement Using Food Images', New Trends in Image Analysis and Processing -- ICIAP 2015 Workshops, pp. 441–448.

Qiu, J., Frank, P.-W., Lo, Sun, Y., Wang, S. & Lo, B., n.d., Mining Discriminative Food Regions for Accurate Food Recognition, viewed 21 February 2022, https://bmvc2019.org/wp-content/uploads/papers/0839-paper.pdf

Sikora, T., 2001, 'The MPEG-7 visual standard for content description-an overview', IEEE Transactions on Circuits and Systems for Video Technology, vol. 11, no. 6, pp. 696–702.

Simonyan, n.d., Visual Geometry Group - University of Oxford, www.robots.ox.ac.uk, viewed 21 February 2022, https://www.robots.ox.ac.uk/~vgg/publications/2015/Simonyan15/

van Blokland-Vogelesang, R.A.W., 1989, Unfolding and Consensus Ranking: A Prestige Ladder for Technical Occupations, in G de Soete, H Feger& KC Klauer (eds), ScienceDirect, vol. 60, North-Holland, pp. 237–258, viewed 21 February 2022, https://www.sciencedirect.com/science/article/abs/pii/S0166411508602393

Wang, J., Song, Y., Leung, T., Rosenberg, C., Wang, J., Philbin, J., Chen, B. & Wu, Y., 2014, 'Learning Fine-grained Image Similarity with Deep Ranking', arXiv:1404.4661 [cs], viewed 21 February 2022, https://arxiv.org/abs/1404.4661

Wang, X., Kumar, D., Thome, N., Cord, M. & Precioso, F. 2015, Recipe recognition with large multimodal food dataset, IEEE Xplore, pp. 1–6, viewed 21 February 2022, https://ieeexplore.ieee.org/document/7169757

X, C., n.d., (PDF) ChineseFoodNet: A large-scale Image Dataset for Chinese Food Recognition, ResearchGate, viewed 21 February 2022 https://www.researchgate.net/publication/316779861_ChineseFoodNet_A_large-scale_Image_Dataset_for_Chinese_Food_Recognition

Zhu, F., Bosch, M., Woo, I., Kim, S., Boushey, C.J., Ebert, D.S. & Delp, E.J., 2010, 'The Use of Mobile Devices in Aiding Dietary Assessment and Evaluation', IEEE Journal of Selected Topics in Signal Processing, vol. 4, no. 4, pp. 756–766, viewed 19 March 2020, https://ieeexplore.ieee.org/abstract/document/5473089

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. and Bouwman, J., 2016, The FAIR Guiding Principles for scientific data management and stewardship. *Scientific data*, *3*(1), pp.1-9.

Novák, D., Kuna, M. and Lečbychová, O., 2021, Taming the Beast. Approaches to Digital Archiving in Czech Archaeology. *Internet Archaeology*, *58*.

Hiebel, G., Goldenberg, G., Grutsch, C., Hanke, K. and Staudt, M., 2021, FAIR data for prehistoric mining archaeology. *International Journal on Digital Libraries*, *22*, pp.267-277.

ISBN 000-000-00000-0-0