Journal of Open Humanities Data

Enhancing Archaeological Heritage Management: The PERAIA Project's Open Data Database and Web Application

COLLECTION:
REPRESENTING THE ANCIENT WORLD THROUGH DATA

DATA PAPER

DAVID LAGUNA-PALMA

ubiquity press

ABSTRACT

This paper underscores the critical need for open data practices in archaeo-historical research, particularly in the face of limited historical data availability, especially in underrepresented regions. Our study was conducted within the framework of the PERAIA project, which introduced an innovative approach by creating an open data database and web application to enhance data accessibility and reusability. This platform offers a comprehensive gazetteer that covers archaeological and heritage sites within the Eastern Mediterranean area (Crete and Marmarica) spanning Late Prehistory and Antiquity. To achieve this, we employed a digital research methodology that integrated legacy data with systematic aerial and satellite imagery analysis. This integration allowed us to identify previously known and unknown sites, enriched with associated (meta)data. Our commitment to open science is embodied in our open data practices. We also make the data accessible through Zenodo, ensuring that such data are available for potential reuse.

CORRESPONDING AUTHOR:
David Laguna-Palma
Department of Prehistory and Archaeology, University of Granada (Spain), ES
dlaguna@ugr.es

KEYWORDS:
archaeological heritage management; open data; LOUD+FAIR principles; digital humanities

TO CITE THIS ARTICLE:
Laguna-Palma, D. (2024). Enhancing Archaeological Heritage Management: The PERAIA Project's Open Data Database and Web Application. *Journal of Open Humanities Data*, 10: 2, pp. 1–7. DOI: https://doi.org/10.5334/johd.158

(1) OVERVIEW

CONTEXT

Archaeological research, by nature, demands extensive data. Thus, a thorough understanding and preservation of our shared heritage hinges on the careful collection, analysis, and sharing of myriad forms of data. Digital technologies and the emergence of Linked Open Usable Data (LOUD) and Findable, Accessible, Interoperable, and Reusable (FAIR) principles have revolutionized data development and management in this field, emphasizing openness, fair access to information, and enhancing data reusability (Marwick et al., 2017; Thiery, 2019; Wilkinson et al., 2016). Nonetheless, venues that prioritize the publication of historical data remain scarce, especially for underrepresented regions. This scenario presents a significant challenge in the field and delves into the strategies we have employed to ensure open practices – e.g., reuse potential for legacy data (Allison, 2008), for the publication and dissemination of such historical data.

This study was developed within the framework of the PERAIA project, which focuses on the Eastern Mediterranean area (Crete and Marmarica [northeastern Libya/northwestern Egypt]) (Figure 1) and introduces an innovative approach to managing and sharing historical data. By creating an open data database and accompanying web application, PERAIA delivers a comprehensive gazetteer of archaeological and heritage sites dating back to late prehistory and antiquity. Our data collection methods relied heavily on legacy data and a cross-verification process, enabling the accurate capture of the geographic locations of sites and generated (meta)data such as typology, chronology, or documented remains associated with each site. Furthermore, a distinguishing feature of our dataset was the inclusion of specific validation scales and the ecological context of the sites. The former helps reduce the uncertainty surrounding the data, and the latter allows us to consider the complex and variable environmental conditions of the Eastern Mediterranean area, particularly concerning their potential impact on the region's archaeological heritage.

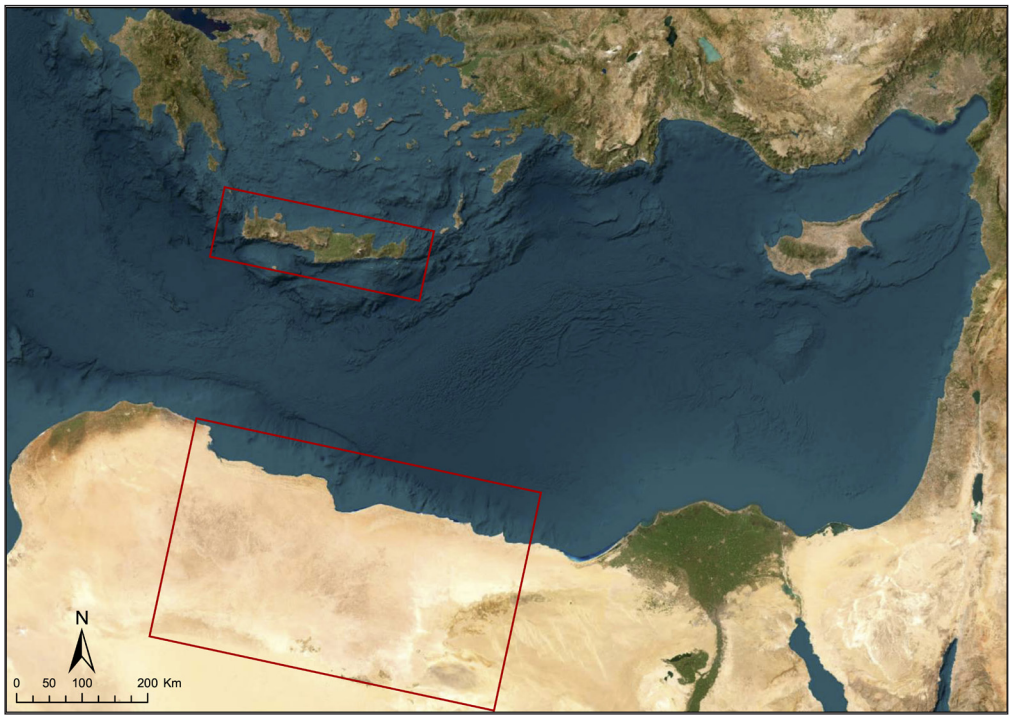

Figure 1 Map displaying the regions under study (Crete and Marmarica [northeastern Libya/northwestern Egypt]) (Basemap: © ESRI).

(2) METHODS

The methodology employed has enriched our data collection process, which was a multifaceted endeavor aimed at constructing a comprehensive database based on, (i) legacy data: we conducted an exhaustive review of existing archaeological records, published field surveys, and historical maps; and (ii) a remote aerial survey: aerial and satellite imagery were used to identify and record the geographical location of sites, ensuring a more accurate position within the study area (Figure 2).

DOI: 10.5334/johd.158

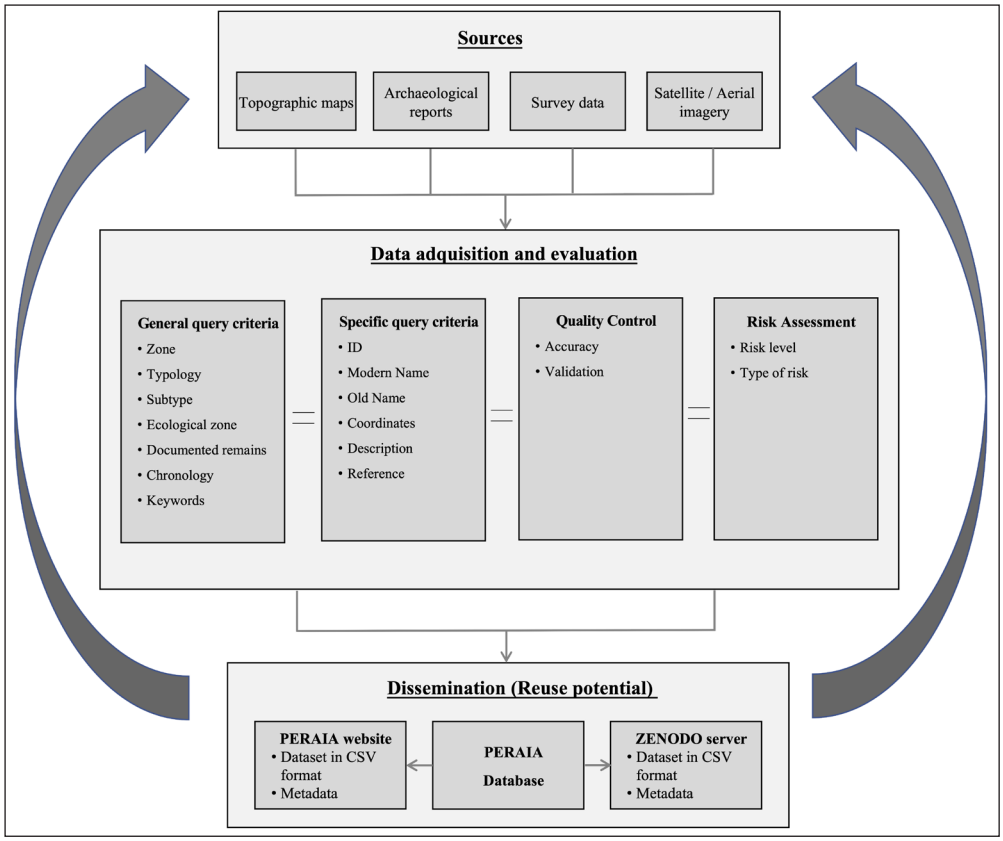

Figure 2 Streamlined workflow for data acquisition, processing, and dissemination.

All the sources employed are relevant because they provide detailed insights into both anthropogenic and natural factors, including topographical and hydrographical features, the exact locations of settlements, burial sites, production areas, and more. Additionally, the information contained in these sources allows us to capture a snapshot of landscapes before the advent of extensive agricultural, urban, and industrial development, thereby offering a glimpse of landscapes that have since been altered or entirely lost today.

In this ever-evolving landscape, new data management approaches are continually emerging. However, the fundamental question of how to enhance data management, storage, and dissemination remains open. Within the scope of the PERAIA project, we employed digital methods not only to streamline cross-verification processes but also to harmonize data derived from multiple sources. Open data approaches play a pivotal role in nurturing research transparency and scientific collaboration (Costa et al., 2014; Huggett, 2015). To achieve this, all generated data must adhere to the FAIR principles (Wilkinson et al., 2016). Furthermore, we aimed to establish best practices for data publication to ensure their widespread sharing and reutilization. For this purpose, we are working to ensure that all generated data also adhere to LOUD principles, in terms of creation, development, and dissemination (Thiery, 2019). This effort serves the purpose of facilitating access to and reuse of historical data by the scientific community and local administrations as well as any interested users, thereby fulfilling our commitment to open science principles (Marwick et al., 2017).

STEPS

Our primary step involved data modeling with the aim of creating a robust and modular structure to store all the information associated with each recorded site (Table 1). In open data initiatives, data modeling involves structuring information in a manner that ensures consistency, clarity, and ease of integration. To realize this goal, we considered semantic languages shared with other projects that also work in open data (e.g., Orbis; Viator-E; EAMENA; Desert Networks).[1] This semantic enrichment enables not only precise data representation but also the establishment of meaningful relationships between different datasets, fostering

1 Orbis: https://orbis.stanford.edu/; Viator-E: http://viatore.icac.cat/; EAMENA: https://eamena.org/; Desert Networks: https://desertnetworks.huma-num.fr/ (last accessed: 1 November 2023).

VARIABLE	TYPE	DESCRIPTION
ID	string	This is the unique site number defined by the project (e.g., PE_00002). The two-letter prefix indicates the name of the project, the following numbers identify each site.
Zone	numeric	This number identifies the geographical location of a site (i.e., 1 = Crete 2 = Eastern Marmarica / 3 = Western Marmarica)
Modern_Name	string	Current name of the site or place.
Old_Name	string	Ancient name of the site as documented in the sources.
Coord_x	numeric	WGS 84 / UTM zone 35N [projected] [EPSG:32635]
Coord_y	numeric	WGS 84 / UTM zone 35N [projected] [EPSG:32635]
Accuracy	string	Scale that defines the accuracy of a settlement's location.
Validation	numeric	Scale that the defines the level of confidence provided by the sources.
Type	string	General typology of the sites (e.g., settlement, burial place, etc.).
Subtype	string	Specific typology of the sites (e.g., harbour, tomb, etc.).
Eco_Zone	string	Ecological zone where the site is situated.
Risk_Level	string	Scale that defines the likelihood of destruction related to the potential risks.
Cod_Risk	string	Number that identifies the type of threats.
Remains	dichotomous	If documented remains exist (Yes/No).
Chronology	string	Phases of human occupation of the sites.
Keywords	string	Any relevant term or information used to identify the site.
Description	string	The site description indicates the nature of the archaeological material observed, or any information related with the site.
Biblio_ref	string	Sources used to create this dataset.

Table 1 Format of gazetteer entries.

data interoperability. GIS (Geographic Information System) software was instrumental in the organization, analysis, and data visualization during this initial stage.

The next step involved site mapping of previously known and unknown archaeological and heritage sites within the regions under study. This process required the application of a remote aerial survey through photointerpretation. Therefore, historical sources, including aerial photographs, topographic maps, and open-source satellite imagery, along with other pertinent field data, were overlaid and analyzed within a GIS environment (ArcGIS 10.5). This framework proved essential for identifying subtle differences in the surface, which may indicate the presence of archaeological structures.

The last step was the creation of a project website to disseminate the historical data generated, which was a pivotal move to ensure accessibility and research transparency. This process began with the identification of a suitable web platform (WordPress) considering factors such as ease of use, security, and scalability. The data were then organized into structured datasets following standardized formats to enhance interoperability. To promote openness, the data were stored in a nonproprietary and accessible format (CSV file). Moreover, we integrated an interactive map developed using the Carto technology to visualize the sites within their geographic contexts (Figure 3).

SAMPLING STRATEGY

Historical sources and specific relevant field data were georeferenced and overlaid in ArcGIS 10.5. Next, we subdivided the study area into 10 × 10 km squares and conducted a comprehensive remote aerial survey by crosschecking the sources and imagery. Our approach to this research was twofold: (i) cross-referencing existing published data encompassing data from previous surveys and archaeological reports, and (ii) engaging in a systematic analysis encompassing topographic maps, historical aerial photographs, and open-source satellite imagery. This

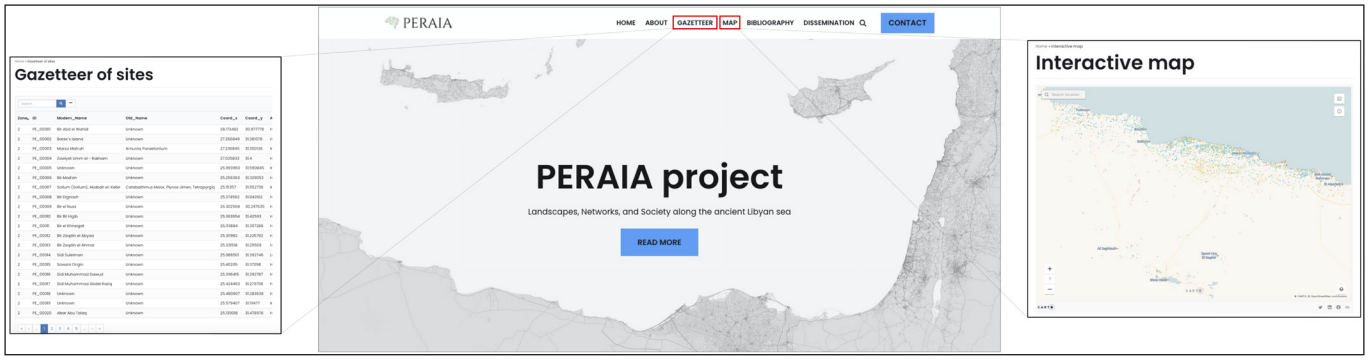

Figure 3 PERAIA web service: enabling gazetteer access and interactive map visualization.

methodology offers a unique perspective on landscapes, enabling access to areas that are not otherwise accessible and intricate patterns that may remain concealed from ground-level observations.

After identifying and recording the sites' geographical locations, the following step involved the integration of all associated information for each site into the dataset, systematically considering the variables listed in Table 1.

QUALITY CONTROL

Specific validation scales were introduced to identify and map the archaeological and heritage sites (see Figure 2) (Laguna-Palma et al., 2023). These scales serve two important purposes. First, they play a pivotal role in establishing a rigorous systematization framework, enabling consistent and standardized data handling. Additionally, we aimed to enhance the overall accuracy of our sites' geographical locations and evaluate the level of confidence placed in the sources employed to create this dataset. Second, these validation scales are crucial for acknowledging and addressing the inherent uncertainties that often accompany data, ensuring research transparency and the reusability of our results.

(3) DATASET DESCRIPTION

OBJECT NAME

Gazetteer of sites

FORMAT NAMES AND VERSIONS

CSV

CREATION DATES

2021-01-01 – 2022-12-31

DATASET CREATORS

David Laguna-Palma, Department of Prehistory and Archaeology, University of Granada (Spain)

LANGUAGE

English

LICENSE

CC BY 4.0

REPOSITORY NAME

https://peraia.ugr.es/gazetteer/

https://zenodo.org/records/7678852

PUBLICATION DATE

2022-12-19

(4) REUSE POTENTIAL

In the contemporary landscape of archaeological research, the integration of web information systems has emerged as a pivotal tool for the effective management, accessibility, and dissemination of data. The PERAIA project, with its emphasis on open data principles and innovative approaches to historical heritage management, leveraged web information systems to amplify its impact.

The project web application[2] hosted by the University of Granada (Spain), serves as a digital gateway to the comprehensive gazetteer representative of current knowledge of archaeo-historical information about these key areas within the Eastern Mediterranean context. This interface also allows users to explore historical data using specific queries and visualize the sites' geographical contexts. Search and query functionalities make it easier for researchers to access relevant information within the vast dataset. Thus, users can filter data based on various criteria, such as typology, chronology, geographical location, ecological context, etc. (see Figure 3). Furthermore, we connected the generated data and (meta)data with another open repository, Zenodo. Consequently, the project advances by adhering to the LOUD+FAIR principles, fostering accessibility, transparency, and equitable access to information.

In conclusion, open data practices, as employed by the PERAIA project, ensure the accessibility of historical data and facilitate data integration and reusability. This initiative offers an indispensable resource to scholars across disciplines, including heritage management, comparative archaeology, ancient trade studies, ecological history, and digital humanities. In addition, it is imperative to recognize historical data as a public good, and they should be freely available to foster and strengthen the knowledge generation process. The PERAIA project's commitment to open data and innovative data management practices encourages disciplinary interaction and transparency, enriches research through its use and reuse, and promotes the examination and reanalysis of datasets. These practices also have social outcomes, fostering knowledge creation, networking, and communication between academia and society.

ACKNOWLEDGEMENTS

We extend our gratitude to Maurizio Toscano from the Spanish Foundation for Science and Technology (FECYT) and Carlos Rodríguez Rellán from the University of Granada for their contribution to this project.

FUNDING INFORMATION

This project was funded by the Spanish Ministry of Universities (FPU17/06503) and the Vice-Rectorate for Research and Knowledge Transfer of the University of Granada (Ref: PPJIB2020.18).

COMPETING INTERESTS

The author has no competing interests to declare.

AUTHOR AFFILIATIONS

David Laguna-Palma orcid.org/0000-0001-8425-9324
Department of Prehistory and Archaeology, University of Granada (Spain), ES

REFERENCES

Allison, P. (2008). Dealing with Legacy Data – An Introduction. *Internet Archaeology, 24*. DOI: https://doi.org/10.11141/ia.24.8

2 https://peraia.ugr.es/ (last accessed: 1 November 2023).

Costa, S., Beck, A., Bevan, A., & **Ogden, J.** (2014). Defining and advocating open data in Archaeology. In A. Chrysanthi, C. Papadopoulos, D. Wheatley, G. Earl, I. Romanowska & P. Murrieta-Flores (Eds.), *Archaeology in the Digital Era: Papers from the 40th Annual Conference of Computer Applications and Quantitative Methods in Archaeology (CAA)* (pp. 449–456). Southampton: Amsterdam University Press. DOI: https://doi.org/10.2307/j.ctt6wp7kg.50

Huggett, J. (2015). Digital Haystacks: Open data and the transformation of Archaeological knowledge. In A. T Wilson & B. Edwards (Eds.), *Open Source Archaeology: Ethics and Practice* (pp. 6–29). Berlin: De Gruyter Open. DOI: https://doi.org/10.1515/9783110440171-003

Laguna-Palma, D., Toscano, M., & **Rodríguez-Rellán, C.** (2023). Building an Integrated Database of North-Eastern African Archaeological and Heritage Sites for Mapping Complex Social Landscapes. *Journal of Open Archaeology Data, 11*(5), 1–10. DOI: https://doi.org/10.5334/joad.99

Marwick, B., Barton, C. M., Bates, L., Bollwerk, E., Bocinsky, R. K., Carter, A. K., Conrad, C., Costa, S., Crema, E., Davies, B., Drake, L., Dye, T., Giusti, D., Graham, S., Hawks, J., Huffer, D., Madsen, M., Neiman, F., Opitz, R., Riel-Salvatore, J., Riris, P., Romanowska, I., Ullah, I., & **Wren, C.** (2017). Open Science in Archaeology. *The SAA (Society of American Archaeology) Archaeological Record*, 8–14. DOI: https://doi.org/10.17605/OSF.IO/3D6XX; https://doi.org/10.31235/osf.io/72n8g

Thiery, F. (2019). Sphere 7 Data: LOUD and FAIR Data for the research community. *Linked Geodesy Working Papers*, 1–6. DOI: https://doi.org/10.5281/zenodo.2643470

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., Gonzalez-Beltran, A., Gray, A. J. G., Groth, P., Goble, C., Grethe, J. S., Heringa, J., 't Hoen, P. A. C., Hooft, R., Kuhn, T., Kok, R., Kok, J., Lusher, S. J., Martone, M. E., Mons, A., Packer, A. L., Persson, B., Rocca-Serra, P., Roos, M., van Schaik, R., Sansone, S.-A., Schultes, E., Sengstag, T., Slater, T., Strawn, G., Swertz, M. A., Thompson, M., van der Lei, J., van Mulligen, E., Velterop, J., Waagmeester, A., Wittenburg, P., Wolstencroft, K., Zhao, J., & **Mons, B.** (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data, 3*, 160018. DOI: https://doi.org/10.1038/sdata.2016.18

TO CITE THIS ARTICLE:
Laguna-Palma, D. (2024). Enhancing Archaeological Heritage Management: The PERAIA Project's Open Data Database and Web Application. *Journal of Open Humanities Data*, 10: 2, pp. 1–7. DOI: https://doi.org/10.5334/johd.158

Submitted: 05 September 2023
Accepted: 06 November 2023
Published: 02 January 2024

Journal of Open Humanities Data is a peer-reviewed open access journal published by Ubiquity Press.