Applied Computing and Big Data: A Case Study of the Pancyprian Gymnasium Historical Archive

Avgoustinos Avgousti^{1,†}, Athanasios Koutoupas^{1,†} and Nikolas Bakirtzis ^{1,*,†}

Abstract

The Pancyprian Gymnasium, founded in 1812, is Cyprus' oldest continuously operating high school and a cornerstone of Cyprus's educational and historical heritage. Central to this legacy is its extensive Historical Archive, which houses many documents and artifacts spanning centuries. Despite the richness of its contents, access to these archives is limited, creating challenges for students, researchers, and the public alike. This article uses the Pancyprian Gymnasium as a case study to illustrate how digital technology can transform the preservation and democratization of cultural heritage. The CyprusArk Content Management System, designed for small museums and similar-sized cultural heritage organizations with a focus on Cyprus, is the backbone of this initiative. The article documents the first real-world implementation of the system, highlighting the challenges encountered, giving recommendations, and detailing the evolution of CyprusArk from a research prototype to a production-ready system. This project sets a precedent for similar initiatives worldwide, emphasizing the pivotal role of digital solutions in safeguarding cultural heritage for future generations.

Keywords

Applied Computing, Computing, Big Data, Digital Cultural Heritage, Content, and Collection Management Systems

1. Introduction

The historical archive of Pancyprian Gymnasium in Nicosia, Cyrpus, is a repository of invaluable historical documents and artifacts that faces significant challenges due to its restricted access and sensitive materials. Located within the school premises, the archive is accessible only to students and faculty staff during school hours from 7:45 to 1:35, Monday to Friday. This restricted time frame is problematic for both researchers who have to physically travel to the archive, adding logistical and financial challenges, and local students who often have to deal with classes.

Furthermore, a lot of archived objects are sensitive, which means that handling, environment, and inappropriate storage might destroy them. Careful management is required to maintain the historical narrative contained in these resources. In addition, the resources are dispersed and kept in many places and forms, frequently in inaccessible rooms and boxes. This disarray makes it even more difficult for researchers and students to use the archival resources efficiently.

A digital transformation of the Pancyprian Gymnasium Historical Archive was essential to address these challenges. The implementation of the CyprusArk CMS [1]. Aimed to overcome physical limitations by providing online access to the archive. This would grant students, researchers, and the public access to the archive's resources anytime, anywhere. The digital platform would centralize and organize the archive's contents, improving searchability and user experience while lowering the risk of damage to fragile items. This digital implementation is crucial for protecting and disseminating the Pancyprian Gymnasium heritage and fostering a deeper understanding of Cypriot history among students and the public.

The purpose of this article is to document and share the experience of implementing CyprusArk CMS at the Pancyprian Gymnasium Historical Archive. It highlights the challenges encountered

© 2024Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR Ceur-ws.org
Workshop ISSN 1613-0073
Proceedings

¹ The Cyprus Institute (CyI), Konstantinou Kavafi 20, 2121 Nicosia, Cyprus

MBS2024: 3rd International Conference On Museum Big Data, November 18-19, 2024, Athens, Greece

^{*}Corresponding author.

[†] These authors contributed equally.

a.avgousti@cyi.ac.cy (A. Avgousti); a.koutoupas@cyi.ac.cy (A. Koutoupas); n.bakirtzis@cyi.ac.cy (N. Bakirtzis)

^{© 0000-0002-9116-8217 (}A. Avgousti); 0000-0003-4913-2752 (A. Koutoupas); 0000-0002-4081-9331 (N. Bakirtzis)

during the transition from a prototype to a production system and outlines the steps taken to overcome these challenges. Further, it illustrates the benefits and improvements achieved through this implementation. Additionally, the article seeks to provide insights and best practices for other institutions considering similar digital archiving projects. Or institutions that like to use CyprusArk for their projects.

2. Overview of the Platform

CyprusArk is a multi-instance open-source Web Content Management System, a software platform designed to create, manage, and modify content on websites without requiring specialized technical knowledge (fig. 1). It is designed for small museums, archives and similar-sized cultural heritage institutions with digitized resources that wish to disseminate their resources online for the benefit of sharing human knowledge [1]. It was developed and evaluated based on qualitative data collected from small museums in Cyprus using semi-structured interviews [2]. It has been created using an open-source stack consisting of the leading-edge Django, a high-level Python Web framework, Postgres database, Bootstrap, and running on virtualized Docker containers. Further, it employs Schema.org, a semantic vocabulary or ontology that enables the embedding of structured semantic markup (metadata) into web pages for direct consumption by major search engines, web crawlers, cultural heritage metadata aggregators, and other machines and can be considered a component of the Semantic Web [3].

The goal of the Semantic Web is to make data on the internet machine consumable. It is an extension of the Web, where vocabularies, ontologies, interoperability, and linked open data are essential concepts. Together with established data standards, they allow machines to understand the semantics (i.e. the meaning of data). Moreover, Schema.org is suitable for describing cultural heritage objects and is widely used in the cultural heritage domain [4].

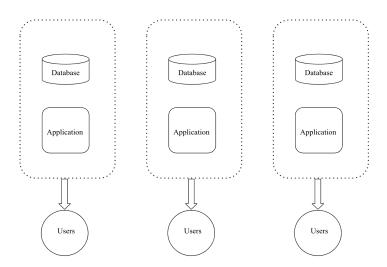


Figure 1: Multi-Instance Architecture. Graphic by the authors.

3. Similar systems

Cultural Heritage institutions are increasingly using Content Management Systems (CMS) as a means to make their collections accessible and available online. A CMS is a software tool that empowers users to manage, modify content, and disseminate digital content, including but not limited to images and text without any technical knowledge [5] In this section, we address similar CMSs to CyprusArk.

Omeka: an open-source Content Management System tailored explicitly for online collections and exhibitions, is a popular CMS utilized by many cultural heritage institutions. Omeka offers an intuitive user interface, customizable templates, and a diverse range of modules to enhance functionality. Moreover, Omeka incorporates pre-built metadata standards and supports various file formats, rendering it an excellent option for cultural heritage institutions seeking to give access to their collections online [6].

<u>WordPress</u>: an open-source CMS commonly used by cultural heritage. While WordPress is not designed for online collections, it is a popular choice for small cultural heritage institutions looking to create a digital online presence. Additionally, WordPress offers several modules that allow for immersive virtual tours of museum collections [7]. A similar approach to the use of generic open-source CMSs related to the study of the authors observed that small cultural heritage institutions often encounter difficulties in making their collections available online, owing to the expenses and intricate nature of commercial products [8]. To address this issue, they proposed an innovative approach centered on the creation of a plugin, designed to enhance the capabilities of the Drupal open-source CMS explicitly for museums.

<u>Mukurtu</u>: is a Content Management System (CMS), developed for the needs of Indigenous communities and their cultural heritage. It offers a suite of tools for managing, sharing, and displaying digital cultural heritage, such as photos, audio, and video files. Moreover, Mukurtu incorporates cultural protocols and permissions that enable Indigenous communities to control how their cultural materials are accessed and used. As such, Mukurtu serves as an excellent example of a CMS that meets the unique needs of specific communities and supports their cultural sovereignty [9].

<u>ViMuseo</u>: is a Content Management System that emerged from a collaboration between the Department of Arts and Cultural Studies at the University of Jyväskylä in Finland and the university museum. The system was custom-tailored to cater to the requirements of small Finnish cultural heritage institutions seeking a more comprehensive digital presence. ViMuseo offers a multimedia presentation platform designed for online virtual exhibitions and projects. The development of ViMuseo was an integral component of Laine-Zamojska's doctoral research, which investigated diverse strategies for improving the accessibility of cultural heritage materials via digital media [10].

Other important systems, such as Access to Memory (AtoM), DSpace, Islandora Alfresco, and Omeka (including Omeka S), are strong contenders for archive and content management.

4. Addressing Technical & Operational Challenges

<u>System Architecture</u>: We used CyprusArk as our primary system. However, during the implementation process, we noticed that CyprusArk needs certain functionalities for this specific case study.

<u>Big Data</u>: The school archive contains thousands or even millions of images and scanned documents. With big data capabilities, this wealth of information will soon be imported into the system. This means that we had to change the way the system stores static files. To solve this problem we implemented an external Object Storage Server (S3) to store static files such as high-resolution images and other media that may be added in the future. This setup will give us the flexibility to store a large amount of static files without worrying about storage. The below image shows the system architecture.

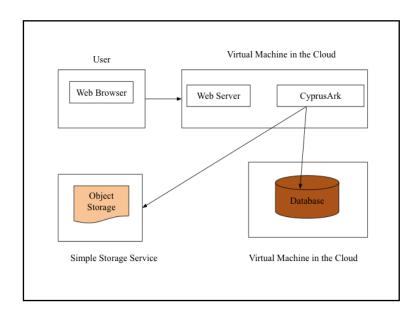


Figure 2: System Architecture Diagram. Graphic by the authors.

<u>Language</u>: another major issue was that CyprusArk was originally developed using the English language. However, the school language and the archival materials are in Greek. To solve this we manually changed all the HTML templates from English to Greek. Furthermore, we implemented a Google, translation form that can translate the online database to almost any language. This will help any non-Greek speaker understand the content of the archive.

<u>Extension of the Content Model</u>: The default CyprusArk content model for objects consists of basic fields such as title, artform, art medium, artwork surface, type, date, description and solo image, accession number, same, culture, period, rights reproduction, and citation. However, those fields were not sufficient for describing every object. For example, the school has a numismatic collection. In this case, coins needed additional fields such as axis, diameter, and mint. We added the extra fields necessary and we are planning to keep those fields in the CyprusArk core system for future similar cases.

<u>Images</u>: CyprusArk was initially able to handle only a single image per object. We extended this capability to six images per object based on the new requirements. Some of the objects such as coins have more than one image available. Additionally, we implemented a web viewer for these images to enhance accessibility and presentation.

<u>Permissions and Groups</u>: The uploading of the content and the descriptions of the content happen by different school teachers with different academic backgrounds. For example, the photography teachers are responsible for documenting and photographing the material and then upload to the online system. The history teachers are responsible for the historical descriptions of objects. For example, if the object is a painting, the photography teacher will take images of the painting and upload them to the system. On the other hand, the art history teachers will add all necessary content related to that object, such as the description, date, creator, etc. For this need, we created two main groups. Administrators, who have access to all data etc, and editors who have access only to the object's content type. We may have to create more groups because may be involved in the population of the database shortly.

<u>Training</u>: To ensure that Pancyprian Gymnasium teachers are comfortable with the CyprusArk platform, we provided them with the necessary training. We demonstrated how to navigate through the CyprusArk interface, upload and manage images, enter and edit metadata, and delete, edit, and upload content. After the initial training, new staff training is conducted internally by teachers who are comfortable using the system (fig. 3).

At this point, we found out that the terminology used in the system was very difficult to understand by the school staff. For example, CyprusArk used the term "bundle" to represent a

collection of items, and "creative works" for the objects within that bundle. This wording was confusing for teachers, so we renamed "bundle" to "collection" and "creative works" to "objects."



Figure 3: Photography teachers digitizing an old manuscript and importing to the database. Photograph by the authors.

5. Data Access Layer: Backend

In the current version of the system, the login user (administrator or editor) can upload data and related information in an easy and intuitive interface. The two main content items are bundles (collections) and creative works (Objects). Those two content types were originally introduced in the prototype of the system back in 2023. However, as mentioned above we had to change the naming convention and provide additional help, as many users couldn't understand what a "bundle" or "creative work" represented. A bundle is a collection. For example, in this case (fig. 4), a bundle is the (paintings collection) and a creative work (painting) is an object of this bundle (collection).

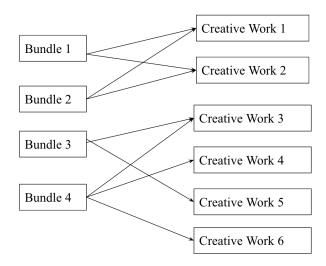


Figure 4: The relation between the content type bundle and creative work. Graphic by the authors.

The login user can create a bundle(s) (collection(s)) and relate any creative works (object) to that bundle. This way the user adds only one bundle and then simply links the creative works (objects).

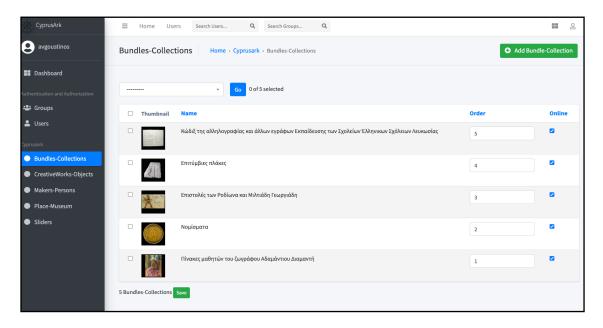


Figure 5: The Bundles (collections) dashboard shows that the system has five collections. Screenshot by the authors.

The platform currently hosts five collections (fig. 5) with objects (fig. 6), all of which are displayed online. This is due to the online checkbox being set to true. The order number allows the login user to display in order the collections. For example, the number 1 will display first in the front end. Further, the following image shows the individual objects. Each object is related to a collection or collections. Using the top menu the user can filter the database by collection, art form, art medium, date, maker, whether it is online or not, and the possibility of free text search.

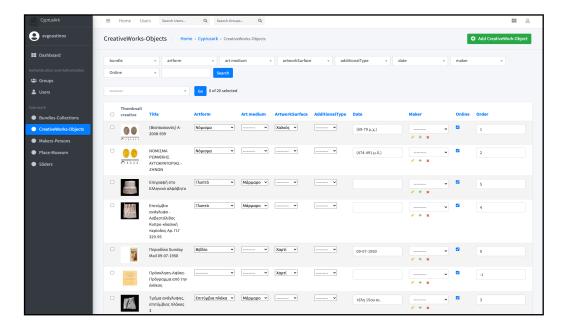


Figure 6: The Creative Works (objects) dashboard. Screenshot by the authors.

Additionally, each object can be related to a maker (creator), and a creator can be related to one or more creative works (objects) (fig. 7). At this point, we identify another problem related to the maker content type. A maker is solely seen as a creator in their role. However, there may be cases

where the maker assumes a different role, such as the inker or penciler of a specific object. We are working on improving this functionality.

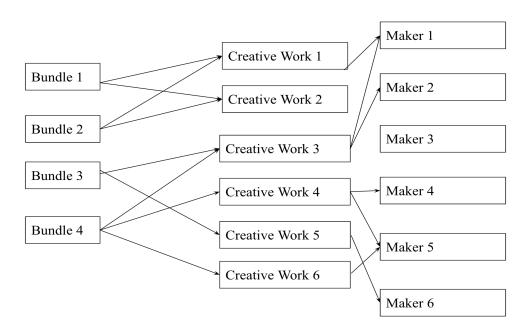


Figure 7: Relation between Bundles, Creative Works and Makers. Graphic by the authors.

Here we see the relation of a bundle (collections) to creative works (objects) and in the case of a maker (creator). So far we have demonstrated how the back-end works.

It is important to mention that all the data imported into the system was done by the teachers of the Pancyprian Gymnasium in Nicosia, Cyprus.

6. Presentation Layer: Frontend

At this point, we showcase the front end of the Pancyprian Gymnasium online database, where anyone can access the materials 24/7 from the comfort of their own home anywhere around the globe. The following image (fug. 8) shows the landing page, with the first four collections available online: paintings collection, letters collection, numismatic collections, and inscriptions collection.

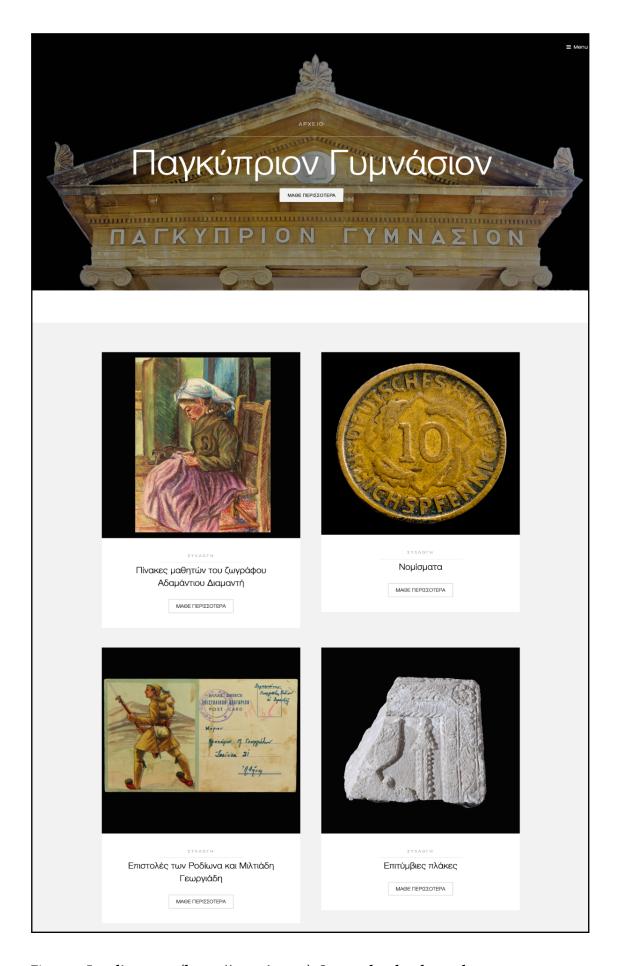


Figure 8: Landing page (https://pg.cyi.ac.cy). Screenshot by the authors.

When a user selects to read more from a specific collection for example the letters collection can be redirected to the detail page of each collection. Here they can find out the total number of items found in the collections and the user can filter down the collection (fig. 9). Additionally, the school can add a short description of the collection.

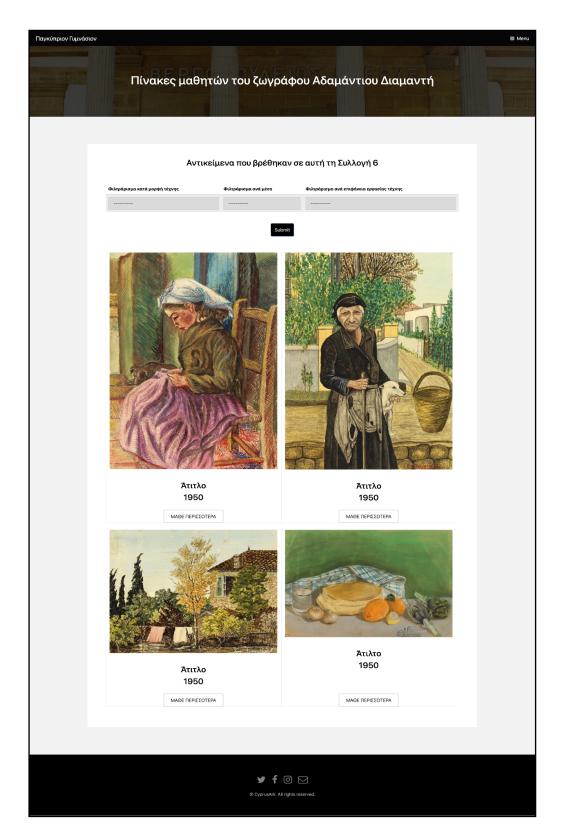


Figure 9: Items (objects) in a collection. Screenshot by the authors.

From here the user can visit any individual object. All the information that the teachers add to the system will be displayed here in a chosen language (fig. 10). Here we can see an object (letter) that has four images related and all the metadata describing the object.

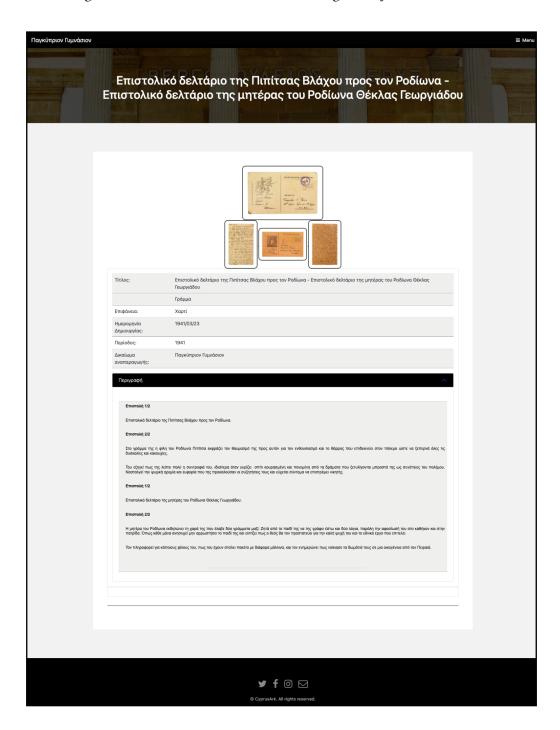


Figure 10: Detailed page of the object in Greek Language. Screenshot by the authors.

At this point, the user can enlarge the images to view any details more clearly. For example, in this case, we can see the four images on the right and we can zoom in to the main image to view more details.



Figure 11: Zoom in the previous screen (Figure 10). Screenshot by the authors.

7. Geographical Analytics

Over the past months, we have followed the traffic of the online database and we can see that many people are visiting the portal. For demonstration, we show the last 200 visitors in the database and their locations (fig. 12). As we can see the online platform can be a place for sharing knowledge.



Figure 12: Online visitors Map. Photograph by the authors.

This shows that many people, students, researchers, and the public worldwide are interested in studying or viewing the materials.

8. Teachers Feedback

Teachers expressed enthusiasm and provided very positive feedback about the CyprusArk platform and its implementation at the Pancyprian Gymnasium Historical Archive. Many of them found it to be user-friendly and easy to use. They also mentioned that the platform is easy to learn and use without the need for long training. They are enthusiastic about populating the system with new materials. The main plan is to make available online a large part of the archive over the next few years. Another goal is to engage students in the process and to provide them with access to populate the online database.

9. Recommendations

Based on the successful implementation of the CyprusArk CMS at the Pancyprian Gymnasium Historical Archive, several recommendations can be made for other institutions considering similar digital projects or using the CyprusArk CMS.

<u>Assessment</u>: Conduct a comprehensive assessment of the specific needs and challenges of the archive/museum before starting the project. Understand the types of materials, the volume of content, and the requirements for metadata and accessibility.

<u>Customization</u>: Ensure the system chosen is flexible and can be customized to meet the unique needs of the cultural heritage institution. The ability to add new fields, support multiple languages, and handle a large volume of static files was crucial for the success of CyprusArk.

<u>Involvement</u>: Engage all relevant stakeholders early in the project. Involve teachers, researchers, students, and technical staff in the planning and implementation phases to ensure the system meets their needs and they are invested in its success.

<u>Training and Support</u>: Provide thorough training and ongoing support to users of the system. This helps in the smooth adoption and effective utilization of the CMS. Peer-to-peer training, as seen with the Pancyprian Gymnasium teachers, can be particularly effective.

<u>Terminology</u>: Use clear and intuitive terminology to avoid confusion.

<u>User-Friendly</u>: Ensure the CMS has a user-friendly interface for both back-end users and frontend visitors. Intuitive navigation and search functionality enhance usability and accessibility for a wider audience.

<u>Ongoing Evaluation and Improvement</u>: Continuously evaluate the system and be open to improvements based on user feedback and needs. Regular updates and enhancements help maintain the system's effectiveness and relevance.

<u>Engagement</u>: Encourage community engagement by making the digital archive accessible to the public. This not only democratizes access to historical records but also promotes cultural heritage awareness and education.

<u>Preservation</u>: Prioritize data security and long-term preservation of digital assets. Implement robust backup solutions and ensure the system adheres to best practices in digital preservation.

10. Future Steps: Clowder Framework Integration

We are currently in the process of integrating the CyprusArk CMS with the Clowder Framework developed by The National Center for Supercomputing Applications (NCSA) at the University of

Illinois Urbana-Champaign [11]. This integration will allow for enhanced data management, enabling us to handle larger datasets and complex metadata more efficiently. By leveraging Clowder's capabilities, we aim to streamline the uploading and processing of digital assets while enabling advanced metadata extraction and visualization. This integration will also improve CyprusArk's ability to support big data needs, such as storing high-resolution images and extensive document archives, using Clowder's distributed storage infrastructure.

The goal is to combine CyprusArk's intuitive content management with Clowder's powerful data processing to create a robust digital archiving solution for cultural heritage institutions. This solution will be especially beneficial for museums, libraries, and other cultural institutions that require flexible, scalable storage and the ability to manage and process a wide variety of digital objects and metadata.

11.Conclusion

The implementation of CyprusArk at the Pancyprian Gymnasium Historical Archive marks a significant milestone in the preservation and accessibility of Cyprus' cultural and educational heritage. This initiative has successfully addressed the longstanding challenge of limited access to the extensive Historical Archive, ensuring that its documents and artifacts are now available to a global audience through digitization and online accessibility.

By leveraging the CyprusArk Content Management System, we have not only preserved the archives but also enhanced their educational and research value. This digital transformation facilitates a deeper understanding of Cyprus' rich history and cultural heritage, facilitating more extensive scholarly research and public participation.

Moreover, the adaptation of CyprusArk for this specific case study sets a precedent for similar initiatives in Cyprus and worldwide. The challenges encountered during the transition from a research-based prototype to a fully functional production system have been meticulously documented, offering insights and best practices for other cultural heritage organizations embarking on similar digital projects.

In the end, this project emphasizes how important digital solutions are to preserving and giving everyone access to historical records. As we look to the future, the lessons learned from this implementation will continue to shape how cultural heritage is preserved and shared globally, ensuring its enduring relevance and accessibility for generations to come.

Acknowledgements

The authors wish to thank the Pancyprian Gymnasium community for their warm support for our work. We want to acknowledge the wholehearted support of the school's directors, teachers, and students. Their knowledge and passion have been instrumental in the project's success. Specifically, we want to thank Morfii Maria principal of the school, Archimandrite Dr. Cristodoulou Georgios assistant principal A of the school, Zevlaris Haris, Elena Paisiou Photography teachers, and archive digitization managers. Further, Sorin Hermon Valentina Vassallo, and Andriana Nikolaidou at APAC Laboratories (https://apaclabs.cyi.ac.cy) of The Cyprus Institute STARC.

Furthermore, we would like to sincerely thank the Museology Lab at Ionian University, the Department of Archives, Library Science, and Museology, especially Dr. G. Papaioannou, the director; and the Andreas Pittas Art Characterization Laboratories (APAC) at the Science & Technology in Archaeology and Culture Research Center (STARC) of The Cyprus Institute (CyI), especially Dr. N. Bakirtzis, the director of the APAC Labs. Furthermore, I would like to thank Professor Feliz Gouveia from the Faculty of Science and Technology, Fernando Pessoa University. Many thanks to the Clowder Framework team developed by The National Center for Supercomputing Applications (NCSA) at the University of Illinois Urbana-Champaign, especially to Luigi Marini and McHenry Kenton.

Declaration on Generative Al

During the preparation of this work, the authors used X-GPT-4 in order to: Grammar and spelling check. After using this tool/service, the authors reviewed and edited the content as needed and take full responsibility for the publication's content.

References

- [1] A. Avgousti, G. Papaioannou, and F. R. Gouveia, "Building CyprusArk a Web Content Management System for Small Museums Collections Online," 2022, doi: https://journal.code4lib.org/articles/16722.
- [2] A. Avgousti and G. Papaioannou, "The Current State and Challenges in Democratizing Small Museums' Collections Online," Information Technology and Libraries, vol. 42, no. 1, Art. no. 1, Mar. 2023, doi: 10.6017/ital.v42i1.14099.
- [3] R. Wallis, A. Isaac, V. Charles, and H. Manguinhas, "Recommendations for the application of Schema.org to aggregated Cultural Heritage metadata to increase relevance and visibility to search engines: the case of Europeana," The Code4Lib Journal, no. 36, 2017, Accessed: Feb. 27, 2022. [Online]. Available: https://journal.code4lib.org/articles/12330.
- [4] N. Freire, R. Voorburg, R. Cornelissen, S. de Valk, E. Meijers, and A. Isaac, "Aggregation of Linked Data in the Cultural Heritage Domain: A Case Study in the Europeana Network," Information, vol. 10, no. 8, Art. no. 8, 2019, doi: 10.3390/info10080252.
- [5] "What Is a Content Management System (CMS)?," Kinsta®. Accessed: Apr. 12, 2023. [Online]. Available: https://kinsta.com/knowledgebase/content-management-system/
- [6] J. L. Hardesty, "Exhibiting library collections online: Omeka in context," New Library World, vol. 115, no. 3/4, pp. 75–86, Jan. 2014, doi: 10.1108/NLW-01-2014-0013.
- [7] I. Daradimos, C. Vassilakis, and A. Katifori, "A Drupal CMS Module for Managing Museum Collections," in Conference Paper, March, 2015.
- [8] K. Valeria, "How to Create a Museum Website With WordPress," Gloria Themes. Accessed: May 03, 2023. [Online]. Available: https://gloriathemes.com/create-museum-website/
- [9] C. Geiler, "A Digital Information Specialist at the National Institutes of Health," A Digital Information Specialist at the National Institutes of Health. 2017. Accessed: Sep. 08, 2020. [Online]. Available: /authors/christen-geiler/
- [10] M. Laine-Zamojska, "Virtual Museum and Small Museums: ViMuseo.fi Project | museumsandtheweb.com," Museum and the Web, 2011, Accessed: Apr. 29, 2023. [Online]. Available:
 - https://www.museumsandtheweb.com/mw2011/papers/virtual_museum_and_small_museums_vimuseofi_pro.
- [11] L. Marini et al., "Clowder: Open Source Data Management for Long Tail Data," in Proceedings of the Practice and Experience on Advanced Research Computing: Seamless Creativity, in PEARC '18. New York, NY, USA: Association for Computing Machinery, Jul. 2018, pp. 1–8. doi: 10.1145/3219104.3219159.