The new digital archive at the Neanderthal Museum

Florian Quirin [1], Bärbel Auffermann [2], Gianpiero Di Maida [3]

1. Selbach Information Systems GmbH. Klutenscheuer 6, 40822 Mettmann, Germany.
Email: florian.quirin@selbach-gmbh.com
2. Neanderthal Museum. Talstraße 300, 40822 Mettmann, Germany. Email: auffermann@neanderthal.de
3. Nds. Landesamt für Denkmalpflege, Scharnhorststr. 1, 30175 Hannover, Germany.
Email: gianpiero.dimaida@ndl.niedersachsen.de

Abstract:

The management and publication of digital data on online repositories constitutes a task that is becoming more and more relevant for museums and cultural heritage institutions alike. To ensure the accessibility and usefulness of the ever-increasing amount of data generated each day, it is crucial to not only archive it, but also find innovative ways to share it with a diverse audience comprising researchers, students, personnel from public and private institutions, the general community and enthusiasts.

The Neanderthal Museum, which has been engaged in the development and establishment of diversified digital offerings for the past two decades, has recently launched a new online platform that serves as a centrepiece of its digital activities. This development was largely initiated by the completion of the DISAPALE project, which involved the creation of hundreds of 3D models that are now available on the platform. The history of the new digital archive, the issues encountered in the planning and realization phases (*e.g.*, requirements for efficiency, security and long-term maintainability, freedom from previous dependencies on proprietary software), as well as the design decisions and technical solutions are critically discussed, highlighting strengths, weaknesses and possible future improvements. An introduction to the client-server architecture of the platform is provided, detailing the development of the back-end using a lightweight Java server component with integrated PostgreSQL relational database, and the creation of the user-facing front-end using modern web technologies to build a tailor-made, flexible and clutter-free web application. Finally, a short look at the current status of scientific enterprises in digital archaeology is given, with a specific focus on the entanglements between museums, cultural heritage institutions and government entities.

Keywords: digital archaeology; archive; 3D scans; web application; data availability

1. Introduction and background

The Neanderthal Museum is located between the Rhineland and the Ruhr region at the eponymous site where the Neanderthal 1 skull specimen was first found in 1856. Its mission is to present and convey information about human evolution to a wide audience. Since its complete redesign in 1996, it offers a varied and research-based exhibition and educational program and attracts numerous visitors every year. As a site museum, the Neanderthal

Journal of Lithic Studies (2024) vol. 11, nr. 2, 11 p. DOI: https://doi.org/10.2218/jls.7321

Published by the School of History, Classics and Archaeology, University of Edinburgh
ISSN: 2055-0472. URL: http://journals.ed.ac.uk/lithicstudies/

Museum does not have a significant collection of its own. Therefore, the development and expansion of databases and 3D databases has always been an important pillar of our externally funded research.

Within various EU-funded projects, we have had the opportunity to exchange ideas with several European institutions and colleagues from the archaeology, evolutionary anthropology and IT areas, concerning the possibility to realize common databases such as Europe@ncestors (2000 to 2001), Object Rich Information Network (Orion, 2002 to 2003), The Neanderthal Tools (TNT, 2004 to 2006) and European Virtual Anthropology Network (EVAN, 2006 to 2010).

1.1. Lessons learned from previous digitalization projects

The experiences gathered with the previous projects laid the foundation for the development of the Neandertal Studies Professional Online Service (NESPOS). NESPOS was publicly available via the web address www.nespos.org from 2006 to 2021 and its goal was to collect digital scientific data on the ice age. CT and 3D surface scans of human fossils and artifacts, literature in digital form, illustrations and photographs were made available in a collaborative manner using a software similar to Wikipedia, which allowed users to integrate data into the platform themselves or to discuss the quality of the data through comments. The service was used as data storage, for research and presentation of results, as well as for educational purposes. Unfortunately, the performance of NESPOS degraded over time, updates and security patches could no longer be installed and thus the server became unstable and insecure and was hacked several times in spring 2019. Consequently, the service was taken offline for safety reasons. Outdated and no longer supported proprietary software components made it uneconomic and almost impossible to further maintain the project. Therefore, it was decided to transfer the data from NESPOS into a new digital archive of the Neanderthal Museum as completely as possible.

With insights gained from two decades of experience, the following requirements were formulated for the new digital archive:

- One-time new setup of the site (structure, layout).
- Maintenance, support for feature and security updates, support for hosting.
- Fast and user-friendly site, also for laymen.
- Content updates and changes must be easy to perform by current and future museum staff with only basic IT skills.
- Flexible and easy to make changes to the graphical parts of the site (banners, highlight area, main content, sidebar, bottom, *etc*.).
- Ability to execute advanced database searches.
- In the long term, the possibility to interface with large, international databases, for example www.europeana.eu (Europeana Foundation 2008).

1.2. Release of the new archive

Version one of the new Neanderthal Museum Digital Archive has been completed in February 2022 and made publicly accessible via the web address archiv.neanderthal.de (Neanderthal Museum 2022) (*Das neue digitale Archiv im Neanderthal*). It contains reorganized collections from NESPOS, as well as completely new ones obtained in various research projects at the Neanderthal Museum and has already attracted hundreds of registered users since its release. The current list of public collections includes for example:

- Fossils: A collection of Hominin finds from all over the world.
- Apes & Monkeys: A collection of referential anatomic scans and pictures.

• DISAPALE: A research project that provides accessible 3D representations of known type forms of prehistoric stone artifacts (Di Maida & Hageneuer 2022).
• The Wendel Collection: A comprehensive archive of Ice Age cave art from France and Spain.

2. The Neanderthal Museum Digital Archive

The design plan for the new digital archive comprised three main goals, which resulted from the previously defined requirements of the Neanderthal Museum. The first one was to create an efficient platform that is self-contained using open-source software, is free of third-party proprietary plug-ins and will be easy to maintain and update over a large timespan, keeping the archive working well and secure. The second goal was to build a dynamic and extendable system that can adapt to new projects and new types of data, while building a database that would remain readable by humans without requiring extra tools. Finally, the third goal was to create a clearly structured, interactive and easy to use user interface that is always accessible no matter where the users are and what devices they use.

In this section we will give a technical overview of the new digital archive and describe in detail the different components we had to build and tools we integrated to achieve our goals.

To meet our requirements for efficiency and long-term maintainability and to remove all previous dependencies on proprietary software, we decided to build the digital archive from the ground up and divide it into several components that can be upgraded individually and fine-tuned to our specific needs. The core components are an open-source (non-proprietary) database for structured project and user data, a modular web application as interface for the users and a lightweight server application that serves files, manages user accounts, runs maintenance tasks and most importantly acts as a central HUB that connects all other components in a secure way (see Figure 1). In the following paragraph we will have a closer look at each of these components.

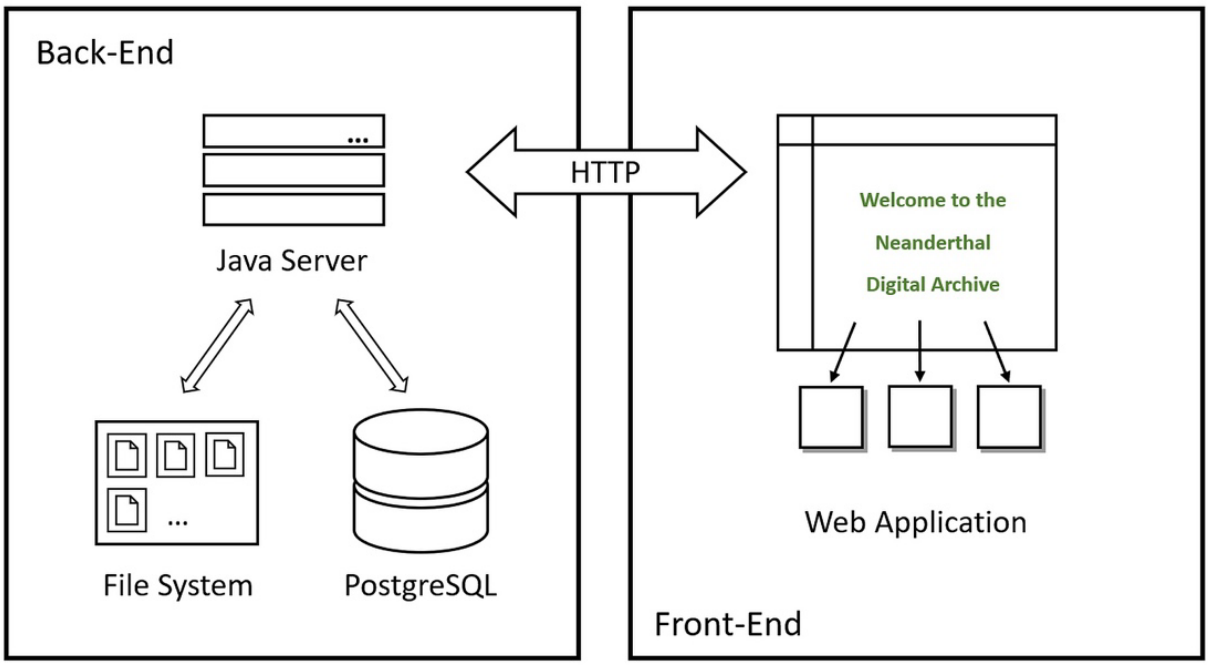

Figure 1. Sketch of the Digital Archive architecture.

2.1. Database

Developers find themselves in a very comfortable position today, as they can choose from a wide array of efficient open-source databases that work virtually “out-of-the-box”. The challenge is to find the right product for a specific use-case. We looked at several old and new projects of the Neanderthal Museum that we wanted to import into the digital archive and realized that the structured project data was often stored in tables (CSV, Excel, spreadsheet files) or could easily be extracted into the same format. Rows usually represented items of a collection and columns contained the item properties. Between different datasets we sometimes saw references to common properties like dig sites, countries of origin or time periods *etc*. Combined with the idea of offering extensive search and filter functions to the users we decided that a classic, SQL compliant, relational database management system (RDBMS) would be the best option for our archive. Ultimately, we chose to use PostgreSQL, one of today's most popular, reliable and modern high-performance, open-source RDBMS (Solid IT 2022) (DB-Engines). It supports a broad spectrum of different data types that fit well to the metadata in our projects, like short and long texts, integer and floating-point numbers, times and dates and even GPS coordinates that can for example be used to filter items by distance. An additional and somewhat unique feature for SQL databases is that PostgreSQL can natively store and query JSON data, a special format for structured data that is often used in web-services and that comes in handy if one needs to store data that changes its format dynamically during the lifetime of a project.

Besides saving structured data, there is also the task of saving files. Often project items have images, 3D models, documents or other types of files attached and in theory you can store these files as so-called blobs inside the PostgreSQL as well, but this has a few disadvantages. There are certain size limits that apply and it's harder to optimize for streaming data transfer. In addition, the plan was to keep the files accessible even without the RDBMS and make sure they were well organized with readable names and a self-explaining folder structure. For these reasons we decided to only keep file names as references inside the PostgreSQL and introduce a separate file storage component that uses the hard drives of the server directly.

2.2. Modular web application

For the user interface (front-end) we defined a list of features that we considered important to create an application that makes the digital archive a productive tool for curators of the Neanderthal Museum and an interesting place for researchers and visitors at the same time. First of all, users should be able to browse different collections and search the database using individual sets of filters to narrow down results. Those search results should be presented in an interactive and appealing way taking the different data types (such as texts, locations, dates) and the attached files (like images, 3D models or documents) into account instead of simply showing plain tables. In addition, we wanted to enable users to view different search results simultaneously, open items and files next to each other and compare everything on one screen. Finally, we needed to support user accounts and groups to handle varying levels of access, ranging from basic visitors who can view public projects, to registered users with read access to selected projects, to staff, editors and curators who can add and modify data, upload files or create entire collections using additional tools.

To realize all this, we decided to build a modular, single-page web application that runs on modern desktop and mobile browsers (Firefox, Chrome, Safari, Edge, *etc*.) and uses an extendable set of HTML components that can handle certain tasks independently inside the app while remaining tightly coupled to our custom back-end system (see Figure 2, search-widget example). As foundation for this application, we chose the Selbach web framework

that was previously developed in-house at Selbach Information Systems GmbH (Selbach 2022). The framework itself is designed to load multiple, individual HTML pages and widgets in separate panels that can be opened and closed as needed, moved around freely, resized and minimized similar to a classic desktop environment but inside a single browser tab. The widgets can be sand-boxed if required to show user and third-party content in a secure environment while remaining able to communicate with our back-end and with each other via a common interface and data channels using a broadcast or subscribe feature. Furthermore, the chosen framework can manage multiple users with different access rights out-of-the-box and offers several standard tools for things like pop-up messages, notifications, input forms and more. Inside the framework we developed a central control app for the digital archive with separate pages for introduction, search, management of collections and administration of user accounts. Each of these pages can launch a number of specialized widgets, designed for tasks like displaying search results, uploading/downloading files or importing/exporting whole datasets (for example from CSV tables). Besides that, we optimized some of the existing file-widgets of the framework to make sure users get the best possible experience when working with 3D models and images inside the browser. For example, we've created a new widget for 3D models using the popular open-source visualization tool 3DHOP (Visual Computing Lab of CNR-ISTI 2022) (3DHOP) that is able to load very large files in NXZ format by streaming them bit by bit from our server and render them right away with increasing levels of detail so they can be viewed quickly inside our application (non-compressed files can always be downloaded for offline use as well). The widget supports rotation and zooming of the model, modification of the lighting, measurement of distances and more (see Figure 3).

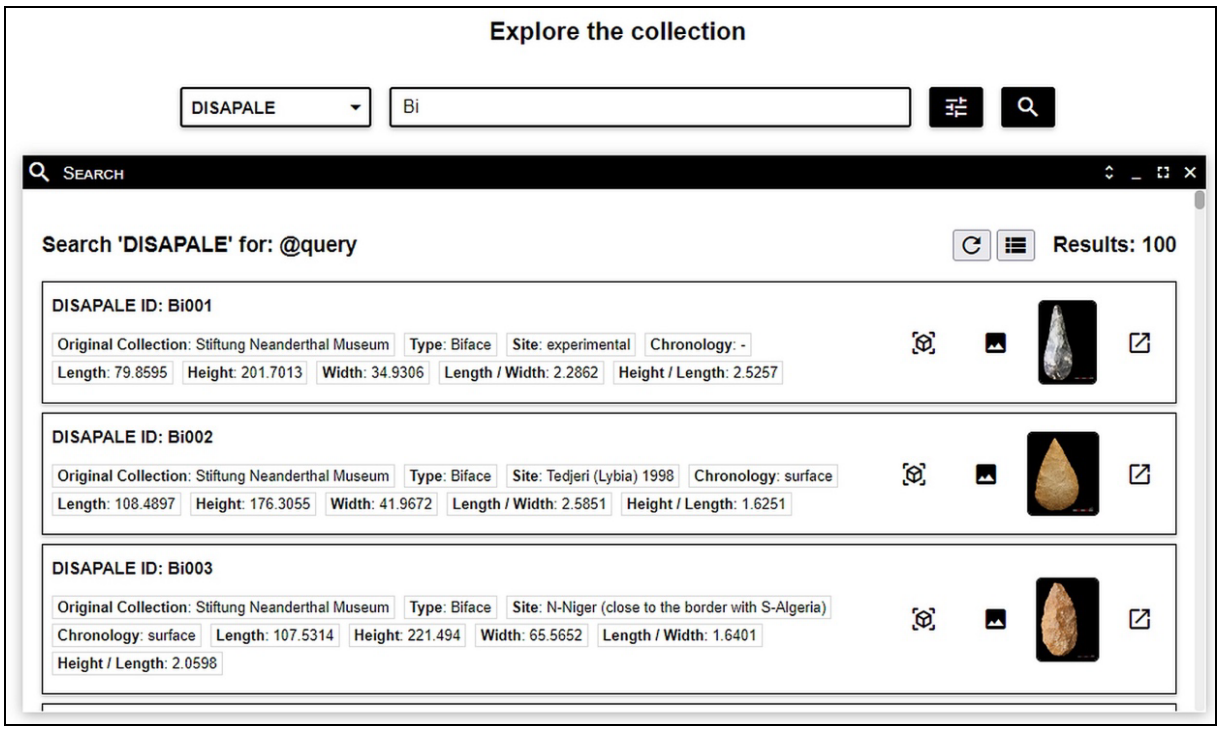

Figure 2. Search results are displayed inside a flexible panel.

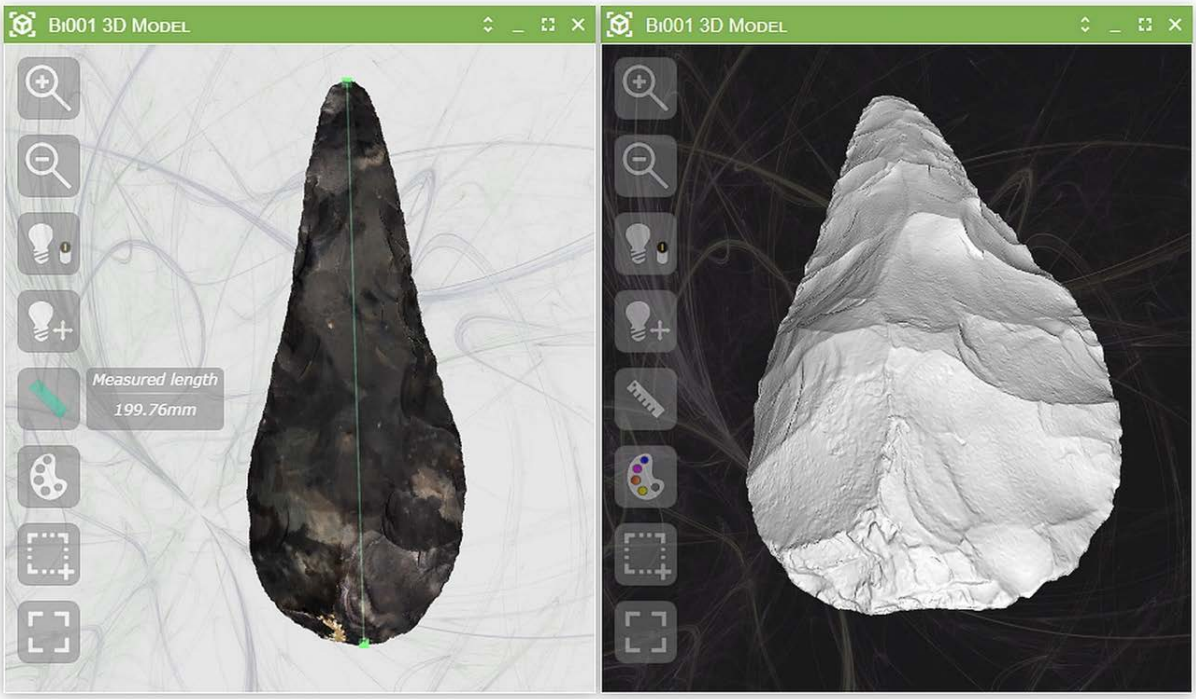

Figure 3. The panel system allows users to compare content next to each other, minimize widgets into the taskbar or for example play interactively with 3D models and images.

Finally, we wanted to make sure that the view of a collection (project) item accessed via the search results is interactive as well and not just a list of plain text fields, so we decided to generate the whole widget dynamically from the data types and properties defined inside a collection definition. This way you can open specialized widgets right next to the item itself for example to view long texts with proper text formatting, show GPS coordinates in OpenStreetMap or other map providers, show calendars for dates and times, view PDFs, images, videos, 3D models and more or even open references to other items.

A few notable other features of the web app are dynamic pages that depend on user access rights, data export to tables and graphs, user defined info sections and action buttons that can for example be used to create descriptions for collections and attach files like PDFs *etc*. (see Figure 4).

2.3. Java server

The third component we've developed is the heart of the digital archive. It is a server written in Java that acts similar to a content management system with the main task of connecting the web application to our database and project files in a secure way. It generates dynamic SQL statements to query the database, taking into account the access permissions of each user, based on their user account and groups configuration. For example, a visitor without account may only see public collections, while a registered user with the user-role “curator” may create and edit collections and a user who is in the user-group “staff” may have access to specific, internal collections and projects. In addition, the Java server offers a number of services like the conversion of files and data formats and supports automatic user sign-up via email. To ensure the platform's health and safety, it collects various statistics, such as failed access attempts, downloaded data and other suspicious activities. It can then issue warnings to administrators via email and completely block potential attackers if necessary.

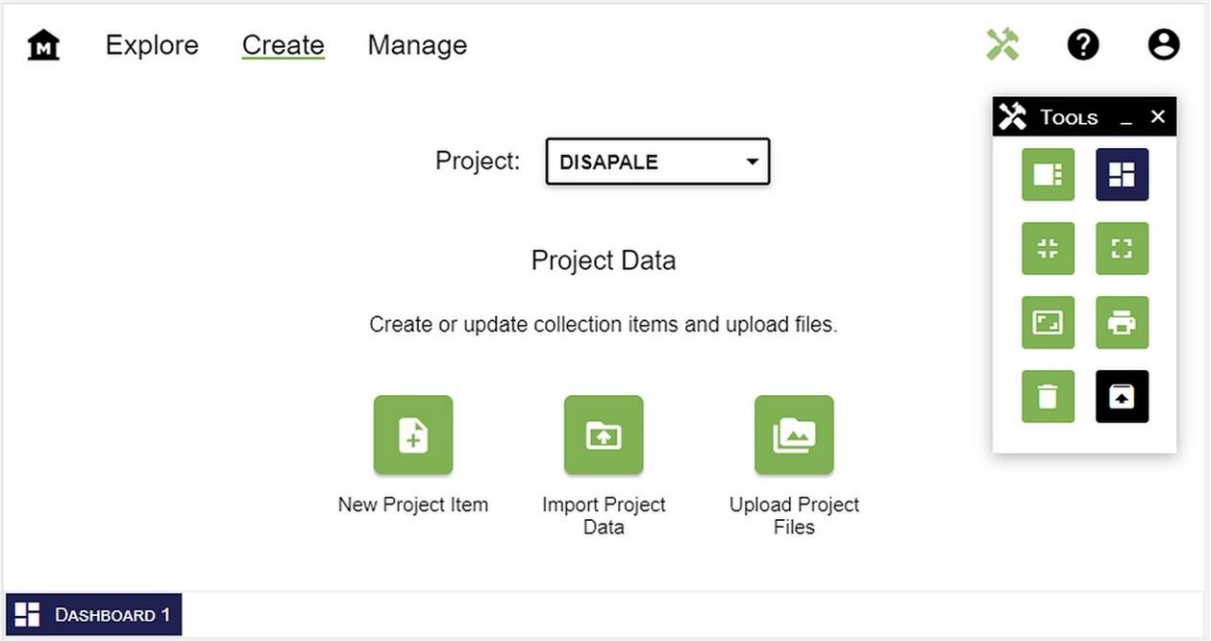

Figure 4. Curators have several tools available to create, import or edit content.

We've decided to use Java as the programming language because it has proven to be a rock-solid foundation for business-critical software for many decades. Our server is built using a modern, lightweight Java framework called Javalin (javalin.io) which removes a lot of overhead, keeps the code base small, and makes everything easy to maintain. The connection to PostgreSQL is handled via the JDBI library (jdbi.org) which provides a convenient, idiomatic access to relational data in Java based on JDBC, the standard database API.

The most challenging part during the development of the server was handling the different collections of the digital archive. Compared to static objects like user accounts where you can usually hard-code the object-relational mapping (ORM) because you know in advance what you need to store, we noticed early on that every collection and dataset has unique features and if we wanted to be prepared for new and unknown projects that will come in the future, we needed a system that can write this mapping in a flexible way. Furthermore, we wanted to allow administrators and curators to be able to create collections on the fly using the web application. To solve this problem, we decided to break down each collection into its core data types like long texts, short labels, integer numbers, decimal numbers, GPS coordinates, dates and times, file lists and more. Then we mapped each of these types to a specific data type of the SQL database and created a table to store what we call the "dynamic project definitions". Our digital archive server is then able to create the PostgreSQL table dynamically from the definition object and can write human readable database entries for collections that look very similar to the original table data (taken from Excel/CSV files *etc*.) just with all the advantages of a SQL compatible database.

All communication between web app and server is done via HTTP calls to the REST-like interface of the server and secured via access-tokens acquired during the login process. For example, the web app can send search requests as HTTP POST calls with a request body describing search term and filters in JSON format, the server will then evaluate user permissions, translate the request into a SQL query, get the data from the PostgreSQL and return it to the web app in the desired format. The returned data usually is a custom JSON object that can be visualized inside our web app using the corresponding widget, but we support a few common other formats as well like XML LIDO (Deutsche-Digitale-Bibliothek 2022. DDBInfo für Daten) to create a compatibility layer for institutions such as the Deutsche

Digitale Bibliothek (DDB). Having a server with public endpoints and a standardized, easy to understand API even makes it possible for third parties (researchers, institutes, *etc*.) to access the archive data via custom integrations and we hope this will further simplify the exchange of knowledge within the community.

Project and item files can be uploaded to the digital archive by using the dedicated files-endpoint. The server will store files on its hard drives and organize them in a clear folder structure using the unique collection ID, item ID and the original filename. After the upload it will write a reference into the database to make files searchable using the same SQL queries as for the datasets. In addition, we support manual reindexing as well where the server can read all folders and refresh the database entries. This can come in handy if you want to copy a large number of files directly to the server instead of uploading them via the user interface.

We've decided to use this approach because of bad experiences with the previous NESPOS system where recovering data became extremely challenging once the user interface was broken, as all the information within the database and on the hard drive utilized abstract and unreadable identifiers. To access files, we support what is called "seekable streams" which is typically used inside clients to load files in smaller chunks of data with an arbitrary start position. This is often used for music or video streaming but works very well for large, progressive images or 3D models as mentioned previously.

2.4. Flexible deployment

The components of the digital archive are platform agnostic and can be deployed on basically every cloud machine with at least 16 GB RAM. We are currently running the whole archive on a dedicated Hetzner server with Intel Xeon CPU, 64 GB RAM and Ubuntu Linux 20.04 LTS operating system. To host the web application, we use an Apache web server located on the same machine. Every component is available as a Docker container to make deployment even easier and further increase portability in case we ever want to move to a different cloud provider or deploy on-premises.

3. Issues and perspectives

Overall, we are very satisfied with the current state of the digital archive, but as always, there is still room for improvement. For example, we've noticed that it can be difficult for users to search content when they partially lack the required knowledge on the subject (*e.g.*, lithic typology and technology names for the DISAPALE collection). Search filters are given as drop-down menus and adapt to the given data types (text, number, data, *etc*.), but if the right search term is unknown, one can still end up with 0 search results. The current search function can match partial results but isn't yet able to find similar results with a different or wrong spelling. There are well established methods of improving the search using auto-complete, fuzzy matching or even machine learning, all with a certain impact on performance and we are planning to optimize user experience further in the future. In the meantime, we've started to implement a feature that can list unique database entries for selected object properties to help users find the right search term when options are limited.

Besides that, we are thinking about how to offer users more curated content. We can already add a customizable HTML introduction section and dynamic action buttons to each collection which can be configured to show pre-defined search results, single items or complementary material like PDFs, but the space for this is currently limited. Because of that we'd like to explore new ways for curators of the digital archive to create interactive pages right inside our web application where they can guide the user through a specific collection or even combine content of multiple ones to build different narratives.

Finally, we are still working on fine tuning the interfaces to enable other digital archives to access our data. Implementing the LIDO XML format for the DDB was a first step and a challenging one too, because we had to find a way to dynamically add the missing info to each collection and then combine it with our data to build the correct XML format on the fly. There are unfortunately dozens of different formats out there and it remains to be seen if this method will properly scale.

4. Conclusions

The importance of digital archives for cultural heritage institutions and museums is crucial, not only to provide a space to store the growing number of digitized items and information, but also to offer old and new users a centralized and user-friendly place to go to consult, research and use said data.

The new archive at the Neanderthal Museum has been realized exactly with such goals in mind and - although there are, as it is natural, many aspects still to be developed and improved - it constitutes one of the first platforms of this kind in the field of prehistoric archaeology and human evolution. Only time will tell how often and for what purposes the hosted data will be used and consequently determine the success of the platform itself as it was designed.

In any case, there's no doubt whatsoever that the relevance of digital methods and data for research and educational purposes is now steadily increasing in our field of study, as it has been highlighted on several occasions and for several contexts (*e.g.*, Huvila 2018). The importance of these aspects is such that it might even be argued that a more integrated and centralized governance could be beneficial to the different kinds of enterprises. The current situation, in which decisions on how to set up and arrange digital data and archives are mainly left to the single institutions, presents a double-sided picture: on one hand the independence has the advantage of stimulating and encouraging new and creative solutions, while on the other the proliferation of different, often non-compatible technical and theoretical set-ups might constitute a major problem (and consequently cause higher costs), already in the near future. The existing fragmentation in the realization of digital enterprises by the institutions involved, moreover, tends to create obstacles in the communication and exchange of information and experiences, thus leading to unnecessary and additional difficulties.

The solution to such a delicate and wide-ranging problem certainly cannot be achieved in such a confined space and will instead require a large-scale discussion, involving all the actors engaged in the field, including the highest-level public institutions at both the federal and European level, which so far have mainly preferred to keep a lower profile.

Acknowledgements

The DISAPALE project was funded by the Federal Ministry of Education and Research (grant ID no. 01UG1885X). The integration of the Wendel Collection into the Digital Archive was funded by the Federal Government Commissioner for Culture and the Media. The authors would like to thank the Neanderthal Society for further financial support.

Data accessibility statement

Data stored in the digital archive at the Neanderthal Museum is publicly accessible and generally available under a permissive license referenced in the collection or item info. The DISAPALE data mentioned in this article is available under the creative commons CC BY 4.0 license. The archive URL is https://archiv.neanderthal.de. To view certain collections an archive account is required and can be created without any costs. This has mainly been

implemented as a security feature to prevent abuse of the website for example by automatized data download scripts *etc*.

References

Deutsche-Digitale-Bibliothek 2022, *DDBInfo für Daten*. Stiftung Preußischer Kulturbesitz. Retrieved: 2022-07-12. (in German) ("German Digital Library. DDBInfo for Data") URL: https://wiki.deutsche-digitale-bibliothek.de/display/DFD/Anforderungen+an+die+Lieferdaten

Di Maida, G., & Hageneuer, S. 2022, The DISAPALE Project: A New Digital Repository of Lithic and Bone Artefacts. *Lithic Technology*, 47(4): 283–295. DOI: https://doi.org/10.1080/01977261.2022.2048511

Europeana Foundation 2008, *Europeana*. Europeana Foundation. Retrieved: 2024-08-01. URL: https://www.europeana.eu

Huvila, I. (Ed.) 2018, *Archaeology and Archaeological Information in the Digital Society*. Taylor & Francis Ltd., London & New York, 184 p.

Neanderthal Museum 2022, *Das neue digitale Archiv im Neanderthal*. Neanderthal Museum. Retrieved: 2022-03-01. URL: https://neanderthal-blog.de/2022/03/das-neue-digitale-archiv-im-neanderthal/

NESPOS (2021), *NESPOS Pleistocene People and Places*. NESPOS Society e.V., Mettmann. Retrieved: 2021-07-26. URL: https://www.nespos.org/ [Archived: https://web.archive.org/web/20210726134051/https://www.nespos.org/display/openspace/Home]

Selbach (2022), *Selbach Information Systems*. Selbach GmbH . Retrieved: 2022-07-12. URL: https://www.selbach-gmbh.de

Solid IT (2022), *DB-Engines*. PostgreSQL Global Development Group. Retrieved: 2022-07-12. URL: https://db-engines.com/en/system/PostgreSQL

Visual Computing Lab of CNR-ISTI (2022), 3DHOP. National Research Council of Italy, Italy. Retrieved: 2022-07-12. URL: https://3dhop.net

Das neue digitale Archiv des Neanderthal Museums

Florian Quirin [1], Bärbel Auffermann [2], Gianpiero Di Maida [3]

1. Selbach Information Systems GmbH. Klutenscheuer 6, 40822 Mettmann, Deutschland.
Email: florian.quirin@selbach-gmbh.com
2. Neanderthal Museum. Talstraße 300, 40822 Mettmann, Deutschland. Email: auffermann@neanderthal.de
3. Nds. Landesamt für Denkmalpflege, Scharnhorststr. 1, 30175 Hannover, Deutschland.
Email: gianpiero.dimaida@ndl.niedersachsen.de

Abstrakt:

Die Verwaltung und Veröffentlichung digitaler Daten in Online-Sammlungen ist eine Aufgabe, die für Museen und Einrichtungen des kulturellen Erbes gleichermaßen immer wichtiger wird. Um die Nutzbarkeit der täglich wachsenden Datenmenge und deren Zugänglichkeit zu gewährleisten, ist es von entscheidender Bedeutung, diese nicht nur zu archivieren, sondern auch innovative Wege zu finden, um sie mit einem vielfältigen Publikum zu teilen, wie zum Beispiel Forschern, Studenten, Mitarbeitern öffentlicher und privater Einrichtungen, der allgemeinen Öffentlichkeit und Enthusiasten.

Das Neanderthal Museum, das sich seit über zwei Jahrzehnten mit der Entwicklung und Etablierung eines breit gefächerten digitalen Angebots beschäftigt, hat kürzlich eine neue Online-Plattform ins Leben gerufen, die als Herzstück seiner digitalen Aktivitäten dient. Auslöser für diese Entwicklung war vor allem der Abschluss des DISAPALE-Projekts, in dessen Rahmen Hunderte von 3D-Modellen erstellt wurden, die nun auf der Plattform zur Verfügung stehen. Die Entstehungsgeschichte des neuen digitalen Archivs, die in der Planungs- und Realisierungsphase aufgetretenen Probleme (z.B. Anforderungen an Effizienz, Sicherheit und langfristige Wartbarkeit, Unabhängigkeit von bisheriger proprietärer Software) sowie die Designentscheidungen und technischen Lösungen werden kritisch diskutiert und Stärken, Schwächen und mögliche zukünftige Verbesserungen aufgezeigt. Es wird eine Einführung in die Client-Server-Architektur der Plattform gegeben, wobei die Entwicklung des Back-Ends unter Verwendung einer leichtgewichtigen Java-Server-Komponente mit integrierter relationaler PostgreSQL-Datenbank und die Erstellung des benutzerorientierten Front-Ends unter Verwendung moderner Web-Technologien zum Aufbau einer maßgeschneiderten, flexiblen und übersichtlichen Web-Anwendung beschrieben wird. Abschließend wird ein kurzer Blick auf den aktuellen Stand der wissenschaftlichen Unternehmungen im Bereich der digitalen Archäologie geworfen, mit besonderem Augenmerk auf die Verflechtungen zwischen Museen, Kulturerbe-Institutionen und staatlichen Stellen.

Keywords: Digitale Archäologie; Archivierung; 3D Scans; Web Anwendung; Datenverfügbarkeit