



This PDF is a simplified version of the original article published in Internet Archaeology. All links also go to the online version.

Please cite this as: Botica, N. 2021 Digital Archaeology Archives in Portugal, Internet Archaeology 58. <https://doi.org/10.11141/ia.58.13>

Digital Archaeology Archives in Portugal

Natália Botica

Summary

Archaeology produces large quantities of data, much of it already in a born-digital format. The preservation of these data is vital, as they result from unrepeatable actions, and access to and sharing of these data are the engine of research activities and knowledge development. The creation of archaeological data repositories, in addition to safeguarding data, is a valuable step in supporting research, but it requires a series of work processes, namely data standardisation and use of common nomenclatures, so published data are findable and reusable.

The creation of an archaeological data repository for Portugal is a fundamental requirement. It would allow the safeguarding of data resulting from archaeological fieldwork, along with those resulting from research, making them findable, accessible and liable to be used in future research or even to be re-analysed using new technologies and tools. This article presents the evolution of the digitisation of archaeological information in Portugal, the responsibilities for its production and safeguarding, as well as the steps to be followed for its availability in a data repository.

1. Introduction

Archaeology has incorporated much born-digital information into its sites and monuments registers. Many countries have created repositories where reports of archaeological works in digital format are deposited and accessible. However, these are primarily repositories of technical and scientific documents, where access to data is not open and/or lacks necessary metadata, which makes it difficult or even impossible to locate, access and reuse (Richards and Wright [2019](#)).

The objective of this article is to evaluate the current state of archaeological data archives in Portugal, highlighting the national, regional and individual responsibilities of the actors involved in the production and management of archaeological data.



2. Portuguese Digital Archaeology Today

The structures for archiving archaeological data and the practical and ethical considerations that constrain their administration must be analysed from the perspective of the producers and users of the data and that of the administrators of the structures. To better understand these perspectives, we will trace the general state-of-play in Portugal with regard to digital archives for archaeological data. This included data used by archaeologists working in either the public or private sector, during the process of surveying, excavation and laboratory work. For a better understanding of who bears responsibility for creating and safeguarding the data, we will address the issue across three different levels: individual, regional and national archives.

2.1 Individual archives

Archaeological data are produced during the process of identification, registration, study, protection and enhancement of archaeological heritage. For this purpose, surveys, excavations, and other archaeological activities to record contexts, structures and the stratigraphy of archaeological sites are carried out. In Portugal, the licence to perform archaeological works requires prior authorisation from the DGPC (Directorate-General for Cultural Heritage) and must comply with certain legal requirements, namely the submission of preliminary and/or final reports. During archaeological work, a vast amount of data is produced, much of which is born-digital, using various technologies and producing files in different formats as text, images, GIS/CAD or 3D models. These elements will allow the preparation of archaeological reports, subject to specific rules and regulations c.f. Presidência do Conselho de Ministros [2014](#). This decree-law (164/2014, 4 November) establishes very clearly all the elements that must appear in these excavation reports, which must be delivered to the DGPC within one year of completion of the excavation. These elements include: the historical-archaeological framework; characterisation of the scope and dating of the works; location of the site, including military and orthophoto map excerpts showing the location; description of the methodology and work carried out; detailed description and interpretation of the nature, chronology and typology of the stratigraphic and structural contexts identified; inventory, description and preliminary study of the finds collected and graphic documentation.

The information included in the reports delivered to the DGPC feeds the Endovélico database, and is made available through the [DGPC Portal do Arqueólogo website](#) after a five-year embargo period. This corresponds with a period of 'scientific reserve' that gives the archaeological scientific director exclusivity to study the archaeological site and findings. These studies, when carried out, also give rise to datasets, but syntheses only are published in technical-scientific meetings, magazines, books, conference proceedings or university scientific repositories.

Both the creation and management of the archaeological data produced during the archaeological work and studies constitute individual archives that continue to be held on personal computers or the institutional servers of the authors. Most of these datasets are very dispersed and fragmented, and so will remain hidden. Only a few of these data



can be found in the Portal do Arqueólogo and in grey literature, which makes it difficult to locate and reuse in future research work.

2.2 Regional archives

The Regional Directorates of Culture (DRC), along with the Municipal Councils and museums, deal with the data from various archaeological sites, and some of them have created digital archives at the regional level. These institutions have accumulated different responsibilities with regard to the creation, management and safeguarding of digital archives.

The DRCs are regional institutions dependent on DGPC and are responsible for organising and maintaining information systems for cultural heritage and making that information available to the public but which can only be used after a three- or five-year embargo period, dependent upon the type of archaeological work.

The Municipal Councils are responsible for the Municipal Master Plans (PDM) that define the heritage protection areas and the inspection and evaluation of archaeological works under their tutelage. Although the approximately 300 Municipal Councils have different approaches and different stages of digital archives, most of them already have Geographic Information Systems for internal management and public disclosure of the architectural and archaeological heritage. Some Municipal Councils also intend to publish data on the archaeological works carried out, with the aim of making these processes more transparent to their municipalities. However, Municipal Councils that have archaeology departments may also be responsible for carrying out archaeological work. In such cases, they also have the responsibility to create and manage their own digital data. These archives, whose data are produced within the scope of their archaeological works, are considered in this analysis as individual archives, sharing the same characteristics described previously, that is, the lack of publication and access to data.

Museums and some Municipal Councils are responsible for managing the deposit of the physical archive from archaeological work when it is considered to be national heritage. The archaeological collection consists of artefacts, ecofacts and samples, along with their respective inventories, and any documentation produced from the archaeological work that is fundamental for understanding the archaeological context (Presidência do Conselho de Ministros 2014).

In the 1980s, some museums began to create regional data archives to safeguard the data associated with archaeological artefacts. These databases resulted from individual initiatives, using very different data structures and tools. In the 1990s, the Institute of Museums and Conservation (IMC), subsequently integrated by DGPC, began the implementation of national policies with the aim of creating an integrated approach to conservation, research, publications and communication of collections, along with intangible cultural heritage within museums. The Matriz 2.0 application was implemented in all museums to centralise the registration of tangible and intangible assets in a national database. This application, which later evolved into Matriz 3.0, allows integrated management of archaeological artefacts and was designed to facilitate automated publication on the Web through MatrizNet. It was also designed to be interoperable with national platforms like PORBASE and international platforms like Europeana (Costa [2013](#)).



The use of a common application for museums, involving the migration of a variety of regional digital archives into a national archive integrating all the information produced in museums, was designed to streamline the process of standardisation and centralisation of information. However, museums faced some difficulties during and after this process, namely:

- Communication problems: the application installed in museums communicated with an IMC server, which offered a low capacity to respond to simultaneous use. Network connections were also very slow at the time, requiring long waiting times for data recording;
- Problems with mapping the data in the new database: some museums had already customised their databases to accommodate specific requirements inherent to their work, and the new centralised application did not include some of these specific data needs;
- The inherent difficulties in adjusting existing workflows into the registration framework offered by the new application added failures in communication and some deficiencies in training. This led to some museums keeping the registration on both platforms simultaneously: in the museum's databases and the registration in Matriz 3.0.

The Matriz 3.0 application allows a centralised and uniform registration of archaeological finds, with the possibility of online dissemination in collections and collective catalogues of museums. However, datasets are not open, and some data registered in museum databases will vanish when the staff members maintaining them move on.

2.3 National archives

The Direção-Geral do Património Cultural (DGPC) is the governmental organisation that oversees all archaeological activity in Portugal, and is responsible for the management and safeguarding of the documentation and inventory of immovable, movable and intangible heritage. The DGPC, which inherited functions from institutions that are now defunct, continued some work already started, namely the inventories of heritage and archaeological activity, implemented in databases of sectoral digital information: Ulysses for immovable cultural heritage, Endovélico for archaeological sites, BMCI, Matriz 3.0 and MatrizPix for movable heritage and MatrizPCI for intangible heritage (Costa [2013](#)).

The Endovélico archaeological information system constitutes the main instrument for managing archaeological activity and storing archaeological information at the national level. This general inventory of archaeological sites in Portugal started in the 1980s with an archaeological map - the Carta Arqueológica, which was a set of paper files with technical, scientific, legal and administrative information about archaeological sites. At the end of the 1980s, the data contained in these forms were converted into a digital archive, first using Dbase database software, and later transferred to Filemaker. The Endovélico information system appeared in late 1995, supported by a relational database and designed to incorporate the information contained in the previous databases. New data were added and, most importantly, new features such as site georeferencing, and the inclusion of images and bibliographic references (Divisão de Inventário do IPA [2002](#)).

The contents of Endovélico are available online at the Portal do Arqueólogo. This portal allows different levels of access according to the user's profile: general public, heritage



professionals and professionals qualified to direct archaeological works (Gomes *et al.* [2012](#)), and searches can be made by site name or by location via a GeoPortal (DGPC [n.d.](#) b).

To implement the provision of technical and scientific information for archaeological sites online, DGPC has developed some fundamental features to streamline the process of incorporating data in the repository, as well as creating standards for metadata, and normalising the structure of the archaeological reports delivered on paper and/or on CD-ROM, through the DRC (Direção Regional da Cultura), namely:

- i. *Norms for naming and structuring files in a report*
By legal requirement (DGPC [2011](#)) a set of rules was created for technical-scientific reports delivered to the DGPC, namely defining the structure and nomenclature of the files. The root folder is identified with a code and the name of the archaeological site, and contains three subfolders: SiteFile, Report Text and Documentation, the last one being composed of three subfolders with Attachments, Appendices and Illustrations;
- ii. *Predefined file formats*
The DGPC also defined the file formats to be delivered with reports, namely PDF/A for Text, TIFF (preferred) or JPEG files for drawing and image files and DWG for vector files (DGPC [2011](#)).
- iii. *Metadata normalisation*
When a report of archaeological works is delivered to DGPC, an online form is also filled in with metadata relating to the characterisation and location of the site, including its geographic coordinates. These metadata are essential, allowing these documents, when integrated in the Endovélico repository, to be found by using a keyword or map-based search. For the standardisation of some descriptors, DGPC published guidelines to be adopted by scientific communities, establishing open and well-defined vocabularies for the characterisation of archaeological and architectural heritage sites, namely for the type of site, type of threat to the site, level of conservation, protection, surveillance and land use necessary for the site (DGPC [n.d. a](#)). DGPC also produced two documents establishing general rules for archaeological inventories: Normas Gerais and Cerâmica Utilitária (for ceramics inventories). However, for most attributes, a vocabulary list accepted within the community is still missing, including important descriptors such as chronological periods.

3. The Urgent Need for a Portuguese Archaeology Data Repository

The DGPC Endovélico information system is an important tool for the inventory of Portuguese archaeological heritage, allowing the management of information related to national archaeological activity. This system is complemented by another management system for the archaeological artefact inventory, the Matriz 3.0.

However, it is not a data repository, but rather an inventory of archaeological sites and artefacts with associated scientific and technical documentation. It was not designed to facilitate open access to datasets, making it difficult to find and access the data produced and delivered in these documents. It is, therefore, urgently necessary to create



an archaeological data repository that safeguards and makes accessible all datasets produced, either by development-led archaeological projects or research projects. This national data repository should be designed ensuring fundamental assumptions, namely:

- Interoperable with the Endovélico information system, where data and metadata files related to archaeological works would be deposited;
- Interoperable with the Matriz 3.0 information system, where data and metadata files related to the study and characterisation of artefacts and ecofacts, namely ceramics, coins and other metals, glass, lithic material, bones, etc. would be deposited;
- the repository of sets of (meta)data resulting from all research activity in archaeology, namely research projects, masters' dissertations, doctoral theses, etc;
- Interoperable with European digital archaeology infrastructures, such as the ARIADNE portal.

This repository of archaeological data could recover data that exists in dispersed files, reducing their probability of disappearing in the near future, while also making the data findable and accessible, contributing to the reuse of this data, and giving visibility to its authors and projects.

The meritorious work that has meanwhile been developed by DGPC in the creation of metadata, allowing the reports of archaeological work to be found more easily, is an excellent starting point for defining guidelines to build a data repository. Thus, in continuity with the work already developed, it will be very important for the DGPC website to provide a set of rules and recommendations for (meta)data schemes for archaeological work, namely associated with the FAIR data principles (PARTHENOS [2019](#)).

This data repository, created under the guidance of the DGPC, could benefit from a legal requirement to deposit data, as the documentation must be delivered to the DGPC on CD-ROM. A simplified process for complementing the current delivery of the documentation to the DGPC and the deposit of the datasets in a digital repository would reduce the additional effort that this task implies, and would help to overcome resistance or inertia that may exist from depositors.

4. Conclusion

The creation of a Portuguese archaeological data repository is a fundamental requirement. These data are often the only record of archaeological sites, as excavation is a destructive activity. Thus, the systematic deposit of these data in a national data repository would be a guarantee for their usefulness in perpetuity, including for findability and reuse.

In addition, in most research where archaeological data are created, the time and effort spent collecting data is very high. We do not have an open data repository and most datasets are currently stored on archaeologists' personal computers or on institutional servers, which makes these data very difficult to find and access. Because the data used in research projects are not held by a data repository, they are not findable and accessible, and the task of collecting and validating the data is replicated in all research projects that could instead be reusing existing data.



For new research projects, namely those financed by the Fundação de Ciência e Tecnologia (FCT), which is a Portuguese governmental institution for scholarships and funding, it is already mandatory to deposit data in a data repository. This constitutes an advance in the process of making data open and accessible, but reinforces the need for a national repository for archaeological data, proposing schemes of (meta)data and procedures for FAIR data. This data repository could also boost the recovery of datasets contained in the individual databases held by archaeologists, research centres, museums, etc., which are in danger of being lost forever.

Costa, P.F. 2013 *Preservação Digital de Património: paradigmas, estratégias e instrumentos no âmbito da DGPC*, s.l., DGLAB.

DGPC 2011 *Património Cultural - Direção-Geral do Património Cultural*. http://www.patrimoniocultural.gov.pt/media/uploads/formularios/circulares_novas/CIRCULARDOCUMENTACAODIGITAL.pdf [Last accessed: 21 March 2020].

DGPC n.d. a *Certidões, Formulários, Regulamentos e Circulares*. <http://www.patrimoniocultural.gov.pt/pt/recursos/formularios-regulamentos-e-circulares/> [Last accessed: 1 March 2020].

DGPC n.d. b *Portal do Arqueólogo - Listas de campos para preenchimento da base de dados - Endovélico*. http://www.patrimoniocultural.gov.pt/static/data/patrimonio_arqueologico/listadecampos.pdf [Last accessed: 1 March 2020].

Divisão de Inventário do IPA 2002 'Endovélico - Sistema de Gestão e Informação Arqueológica', *Revista Portuguesa de Arqueologia* **5**, 277-83.

Gomes, A.S., Leite, S., Neto, F., Oliveira, C. and Bragança, F. 2012 'Inventariação e Gestão do Património Imóvel na Direção-Geral do Património Cultural', *Documentazione e conservazione del patrimonio architettonico ed urbano*, 3-8. <https://doi.org/10.6092/issn.1828-5961/3290>

PARTHENOS 2019 *PARTHENOS Guidelines to FAIRify data management and make data reusable*. <https://zenodo.org/record/3937183#.X2H022hKiUk> / <https://doi.org/10.5281/zenodo.3937182>

Presidência do Conselho de Ministros 2014 *Decreto-Lei Nº 164/2014 de 4 de novembro*.

Richards, J.D. and Wright, H. 2019 *COST 18128: SEADDA Management Committee*. https://archaeologydataservice.ac.uk/resources/images/presentations/2019/PDFa/SEADDA_introduction_4_March_2019.pdf [Last accessed: 21 March 2020].