Please cite this as: Trognitz, M., Batlle Baró, S., Matskevich, S., Katalin Péter, R., Szabó, A.D., Gál, M. and Moitinho de Almeida, V. 2024 Preserving digital data without an archive: Illuminating the path towards digital preservation through knowledge of essential requirements and strategies, Internet Archaeology 67. https://doi.org/10.11141/ia.67.2

Preserving Digital Data without an Archive: Illuminating the path towards digital preservation through knowledge of essential requirements and strategies

Martina Trognitz, Sabina Batlle Baró, Sveta Matskevich, Réka Katalin Péter, Alexandra Diána Szabó, Márton Gál and Vera Moitinho de Almeida

This publication provides an understanding of the concepts and principles of digital (long-term) preservation (or archiving) of research data. On one hand, it shall serve as a guide for researchers who don't have access to a suitable trusted or certified digital archive and will provide information on the measures to take for preserving and sustainably disseminating research data. On the other hand, the publication provides an overview of the technical requirements for the implementation of a service for the long-term preservation of data, as well as non-technical necessities like workflows and policies. The need for archiving is contextualised by the presentation of the driving principles of good scientific practice.

1. Introduction

The idea of this publication is to provide guidelines for researchers who wish to deposit and preserve their digital research data but have no access to a suitable trusted digital archive in their home institution or even in their country. By "digital archive" in the context of this publication, we mean a dedicated institution that ensures the long-term preservation of the data entrusted to it. Long-term preservation is the process of keeping the data's content, its findability, accessibility and reusability, including the ethical and legal framework agreed upon at the deposition for at least ten years. Another term within this publication, "repository", is used here as a hypernym for any service that hosts research data and outputs, regardless of an existing commitment to long-term preservation. Therefore, in some (broader) contexts, it can be used interchangeably with digital archives, while in others, especially when it comes to digital preservation practice, the differentiation is crucial (Rivers Cofield *et al*. 2024).

A need for this publication was recognised during the four years (2019-2023) of the activities of the COST action CA18128 *SEADDA - Saving European Archaeology from a Digital Dark Age*. SEADDA was - and after its formal end in 2023 still is - a collaborative best practice network of archaeologists, information scientists, librarians, archivists, and administrative staff (Richards *et al*. 2021) from Europe and beyond (Geser *et al*. 2022). The network aims to highlight the challenges stemming from the advancement of digitalisation in the field of archaeology and the handling and preservation of digital data, as well as to build capacity for digital archiving in archaeology. The activities include encouraging the creation of infrastructures for managing and preserving archaeological data, establishing common standards, and thereby facilitating future aggregation of datasets e.g. by the ARIADNE Portal.

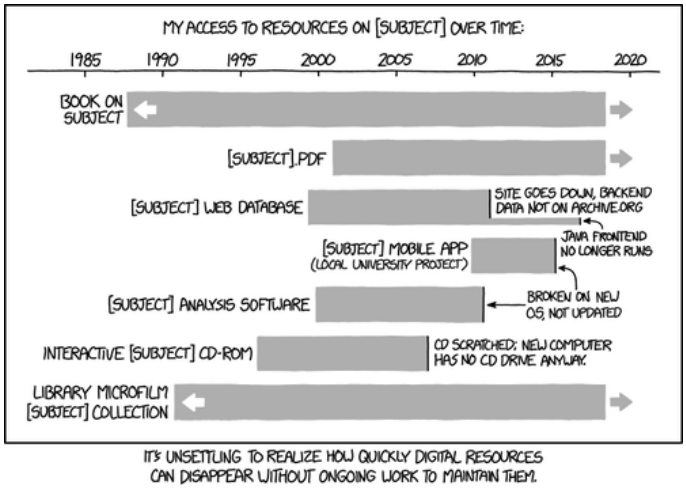

Figure 1. Digital preservation aims to extend the lifespan of digital data. "Digital Resource Lifespan" by Randall Munroe, via xkcd. CC BY-NC 2.5.

Within SEADDA, four working groups were established, tasked with broader topic areas that roughly correlate with the levels of development of digital archiving in the partner countries (Richards *et al*. 2021). Working Group (WG) 1 assessed the state of data archiving and dissemination in the countries and regions participating in SEADDA or ARIADNE Plus (Jakobsson *et al*. 2021; Jakobsson *et al*. 2023), which underlined the disparity in archaeological data archiving practices among the participants, with some of them completely lacking appropriate digital archives.

WG2 built on the findings of WG1 and was tasked with identifying the practical and technical considerations when creating a sustainable digital archive for archaeological data. The group gathered archaeologists, archivists, and specialists in long-term digital preservation who focused on the more technical issues related to the planning for archiving. Among the topics discussed were management structures, hardware and software solutions for archiving, and training of digital archivists. A workshop held in Vienna in December 2019 kicked off the work of the group. The group actively participated in enriching the Community Owned Digital Preservation Tool Registry and in the training provided by the Digital Preservation Coalition.

WG3 focused on examining the state of current international best practice in the areas of digital archiving and dissemination, as well as how it is implemented by existing services. Investigating on how the use and reuse of archaeological data can be improved was the task of WG4.

This publication is organised as a problem-solving guide and will present several options available at the time of writing. While some technical solutions mentioned here might become obsolete or unavailable in the future, we believe that the general line of the decision-making process when planning for digital preservation will serve the reader for some time to come (also see Figure 5).

It includes five sections providing guides to answer the following questions:

- What is digital preservation?
- Why should digital research data be preserved?
- How to find a trusted service capable of preserving digital data that fits individual needs, like discipline or data types?
- What can be done if a suitable infrastructure is not available, and how should the data be prepared for longevity and reusability?
- What should be taken into consideration by an institution when planning for or building up an infrastructure for digital preservation?

2. What is digital preservation?

The primary aim of digital preservation, also referred to as long-term preservation, is to preserve digital materials so that they can be opened, read, edited, searched and ultimately reused effectively extending their lifecycle (UK Data Service Research Data Management s.v. 'Data lifecycle'). The Consultative Committee on Space Data Systems, in its Reference Model for an Open Archival Information System (OAIS Reference Model), formally defines long-term preservation as "the act of maintaining information, independently understandable by a designated community, and with evidence supporting its authenticity, over the long term" (Consultative Committee on Space Data Systems 2012 , 1-13).

Figure 2: Preserving and sharing data ensures that the data life cycle is ongoing. Martina Trognitz. CC BY 4.0.

"Long term" is an indefinite period in which access to at least the content of digital materials is ensured (DPC 2015 Glossary s.v. Long-term preservation). This period extends beyond technological and socio-cultural changes and necessitates ongoing monitoring of emerging media and data formats (Consultative Committee on Space Data Systems 2012 , 1-12; forschungsdaten.org). A digital archive's retention period for digital materials depends on factors such as its funding model, the longevity of the preserving institution, and its preparedness for future disruptions.

At the core of digital preservation is the preservation of files in their full integrity and authenticity — in essence, preserving an exact copy of the digital file, bit for bit. This process, known as bitstream preservation, involves creating multiple copies and regularly checking both the copies and the storage media for data integrity (Brown 2013 , 218-228). Alternatives include the use of offline media such as microfilm (Neuroth *et al*. 2010 , sections 8:32-8:33) or other specially treated film (Sabliński *et al*. 2021).

Beyond preserving the files themselves, digital preservation also focuses on maintaining their accessibility, a concept termed logical preservation (Brown 2013 , 228). Three main logical preservation strategies exist, each with its own advantages and disadvantages: preserving the original software and hardware; emulating software environments; and migrating file formats to preserve the content and meaning of the original (Brown 2013 , 208-214).

Preservation of the original software and hardware environments requires setting up and maintaining a computer museum. This would involve keeping computers, storage media (such as floppy disks), and the devices needed to read the media in an operable condition. Doing this may not be possible in the long term (Neuroth *et al*. 2010 , sections 8:24-8:31).

Preservation by emulation is a strategy widely used to preserve computer games and early software (Brown 2013 , 212) but it is also used for highly specialised datasets like those published on the CERN Open Data Portal for which a dedicated virtual machine or bespoke software tools are offered (CMS open data group). However, as technology changes, and both software and hardware changes, the effort required to develop new emulators increases (Neuroth *et al*. 2010 , 8:16- 8:23).

Additionally, software licensing restrictions and associated costs can sometimes render emulation impossible.

Migrating file formats involves converting outdated or obsolete formats into current supported formats. Instead of preserving the original environment and file, this approach migrates and adapts a file's content and functionality to current technology standards (Neuroth *et al*. 2010 , 8:11- 8:15; Brown 2013 , 209-212). A common example is the change in the default Microsoft (MS) Word file format: with the release of MS Word 2007, the .doc format was replaced by the XML-based .docx format. Consequently, archives that had preserved files in .doc format needed to transfer (migrate) them to .docx.

In this contribution, we will focus on archiving research data through format migration due to its versatility and practicability. Software preservation presents a wide range of technical and organisational challenges that go beyond the scope of this publication (Morrissey 2020). To preserve a file's content and functionality, ensuring it can be opened, read, edited, searched and reused, the following key data management planning areas must be considered: selection for archiving, folder structure, file names, file formats, metadata, and legal and ethical considerations, including access and usage modalities of the data and licensing information.

Planning for data management from the outset of a project, encompassing its creation through its entire lifecycle to archiving, ensures a well-organized dataset and simplifies the archiving process. All data management tasks and issues are documented in a data management plan, covering the task areas mentioned above. Section 5 provides more detailed practical guidance on each of these areas.

3. Incentives for digital preservation

Why should data be preserved? Data preservation is crucial for maintaining the data lifecycle and facilitating the sharing of data that underpins published research, enabling future reference and reuse. With the increasing use of digital methods and the explosion of digital-born data, digital preservation is of particular importance for archaeology. Archaeological research often involves the destruction of the original resource, lacks paper surrogates, and deals with a particularly wide variety of data types (Richards *et al*. 2021). Furthermore, data are vulnerable to loss due to the unreliability of unsuitable storage media such as CDs and flash drives (Bánki *et al*. 2019).

International organisations and institutions, such as the United Nations Educational, Scientific and Cultural Organization (UNESCO) or the European Union (EU), recognise the growing importance of digital data preservation. This is because digitisation has become an important tool for preserving and providing access to cultural heritage for both the public and professionals (Bánki *et al*. 2019). The *Charter on the Preservation of the Digital Heritage* acknowledges that information and documentation of the world's cultural heritage are "increasingly produced, distributed, accessed and maintained in digital form, creating a new legacy - the digital heritage" (UNESCO 2009 , 2), which requires preservation. The EU has also published its *Recommendation on the digitisation and online accessibility of cultural material and digital preservation* (European Commission 2011) and has created a set of policies to support the promotion of Europeana as a platform for digitised cultural heritage, all of which are part of broader European efforts towards digital heritage preservation.

Publishing and sharing data are currently recognised best practice for research integrity (ALLEA 2023). Funding agencies and research institutions increasingly demand sustainable, preferably open, data publishing or archiving, as evidenced by various lists (Hahnel 2015; Calkins *et al*. 2022; forschungsdaten.info 2023; Jisc Open Policy Finder). The slow but steady increase in the number of researchers depositing, archiving, and openly sharing their data in recent years likely reflects the measures taken by funding agencies and publishers (Digital Science *et al*. 2018; 2019; 2022).

Publishing and sharing data, especially through open access, are research practices advocated by the open science paradigm (FOSTER Consortium 2018; European Commission 2019). This calls for greater transparency in all stages of the research process (e.g. open methodology, open source), including its outputs (e.g. open data, open access, and open peer review) (Bezjak *et al*. 2018) and

represent an important driver for digital preservation. Only through preservation can data be sustainably shared and research outputs remain repeatable, reproducible and traceable over time. This in turn fosters knowledge transfer, making research more effective and sustainable by preventing duplicated studies and enabling the reuse of prior work (Abadal 2021), thereby accelerating innovation, as highlighted in the United Nations (UN) Sustainable Development Goals (United Nations 2015; Aleixandre-Benavent *et al*. 2020). The European Research Data Landscape report indicates that "support for open science values and benefits, such as the acceleration of scientific research/public benefit" and "support for openness in science" are seen as incentives for users to deposit data in repositories (European Commission: Directorate General for Research and Innovation *et al*. 2022 , 26).

The open science movement has also inspired another influential set of principles: the *FAIR Guiding Principles for scientific data management and stewardship* (Wilkinson *et al*. 2016). While these principles don't mandate open data (Mons *et al*. 2017), they provide further impetus for data preservation. FAIR stands for Findable, Accessible, Interoperable, and Reusable data and metadata, emphasising how they should be published and archived. The principles aim to ensure that "research objects are reusable, and actually will be reused, and so become as valuable as is possible" (Mons *et al*. 2017) while also requiring clear and transparent information about access and reuse of (meta)data. Applying the FAIR principles should facilitate access and reuse of (meta)data not only for humans but also promote machine-readability. In the context of digital preservation, the FAIR principles can be summarised as follows:

- **Findable**: data and metadata should be easily discoverable by humans and computers. This involves indexing them in a searchable electronic source and providing rich machine-readable metadata that facilitates automated discovery. Crucially, (meta)data must be assigned a globally unique and persistent identifier (PID) for referencing and linking.
- **Accessible**: data and metadata should be accessible through a standardised communication protocol. While this does not necessarily mean that the datasets must be publicly available, the retrieval protocols themselves should be free and open for others to implement.
- **Interoperable**: using formal and standardised formats, knowledge representation languages and FAIR vocabularies ensures the exchange and interpretation of data and metadata.
- **Reusable**: data and metadata should be richly and accurately described to enable future reuse. This includes using clear usage licenses, providing detailed information about data provenance and adhering to community standards.

The emphasis on data reusability highlights digital preservation as a critical component in the provision of FAIR (meta)data. Conversely, the FAIR principles serve as a guide for creating sustainable, persistent digital sources with rich metadata (Nicholson *et al*. 2023).

Complementing the FAIR principles are the CARE principles for Indigenous Data Governance (Carroll *et al*. 2020), which focus more on the content and the individuals associated with the data. These principles - Collective Benefit, Authority to Control, Responsibility, and Ethics - seek to protect the rights and interests of Indigenous peoples within the open data movement, safeguarding their data sovereignty. The CARE principles highlight power imbalances in data sharing and data reuse within digital infrastructures (Carroll *et al*. 2020) and can contribute to "building capacity in digital methods" and "data practices [...] in digital archaeology research" (Gupta *et al*. 2023 , 77) when implemented by infrastructures tasked with data preservation.

At national level, various initiatives are driving digital preservation efforts, as can recently be seen in Austria, Hungary and Portugal. The Austrian Federal Ministry for Arts, Culture, the Civil Service and Sport (BMKOES) launched a strategy (2023) to preserve and further develop cultural heritage by digitising collections and improving accessibility (Strategie Kulturerbe digital [Digital cultural heritage strategy]). This strategy is supported by the Kulturerbe Digital funding programme which has long-term preservation as one of its quality assessment criteria. In Hungary, the Ministry of Human Resources has published a White Book (Bánki *et al*. 2019) to provide comprehensive guidance for digitisation and publication in the field of cultural heritage. The White Book is expected to help strengthen cooperation between cultural institutions and unify documentation procedures. The authors argue that while original objects are irreplaceable, digital preservation offers the possibility of creating (to a certain extent) a digital twin of the object in case of damage or destruction, given that

public collections are vulnerable to natural disasters and human-caused destruction (Bánki *et al*. 2019). In Portugal, the Portuguese Foundation for Science and Technology (FCT), through its National Scientific Computing Unit (FCCN), began implementing POLEN in 2023. POLEN is a pilot project comprising a data management plan system, a research data repository service, and a community support service, designed to promote open science best practices and to ensure the management, sharing and preservation of research data from publicly funded projects. The FCT is currently developing a policy for the management and sharing of research data.

In summary, digital preservation contributes to the creation of sustainable data, which in turn increases the impact and visibility of research, providing a further incentive for researchers (European Commission: Directorate General for Research and Innovation *et al*. 2022 , 26).

4. Finding a digital archive

Digital preservation, as discussed in Section 2, is a multifaceted process requiring specialist knowledge. The variety of tasks involved require not only a specialist and dedicated team but also an organisation committed to providing the necessary resources. The OAIS Reference Model (Consultative Committee on Space Data Systems 2012) refers to the organisation responsible for digital preservation as a "digital archive". While often used synonymously, the term "repository" has a broader meaning, referring to any organisation responsible for maintaining information for access and use (Research Libraries Group 2002 , 59), regardless of whether it has long-term data preservation measures in place. In this article, "repository" is used as a hypernym for any service hosting research data and outputs, irrespective of its commitment to long-term sustainability or reusability.

There are now a significant number of domain/discipline-specific digital archives and trusted repositories suitable for the long-term preservation of research data, which can be found via dedicated registries. Finding and contacting a suitable service at the outset of a project is the first step on the long road to long-term preservation. Early contact with the service allows researchers to take aspects like fees and requirements for the project's data management into account, and to streamline the data transfer to the designated archive or repository.

This section will introduce repository registries and discuss the criteria to consider when selecting a suitable repository for the long-term preservation of archaeological data.

4.1 Registries of repositories

There are several registries, portals, and sources (e.g. Jahn *et al*. 2023) through which digital archives and repositories for research data can be browsed and searched. The following are excellent starting points for the process.

re3data The Registry of Research Data Repositories is a global registry that provides entries indexed by content type, subject (research discipline), and country. The platform administrators collect and review additional metadata about each repository including information on certificates, license types, and PID availability. This helps researchers determine a repository's suitability before contacting it.

OpenDOAR The Open Directory of Open Access Repositories is a service by Jisc that lists repositories with an Open Access policy that meet the requirements set in the Technical Guidance and Requirements of Plan S.

ROAR The Registry of Open Access Repositories, has a broader scope of repository types including e-publications, e-thesis databases, educational materials, and other online research-related resources, even those not explicitly designed as repositories.

FAIRsharing Based at the University of Oxford, FAIRsharing is a community-driven platform that curates and hosts information on standards, databases (including repositories) and data policies across all disciplines with a focus on the FAIR Principles. Resources listed on FAIRsharing are also marked as "deprecated" when no longer in active use.

The above-mentioned services not only list digital archives and repositories suitable for digital preservation but also cover a broader range of platforms and applications, such as repositories for publications only and online databases. Therefore, it is essential to evaluate the potential repositories to determine their suitability for your specific data.

4.2 Search criteria for finding a digital archive

A good starting point to finding a suitable digital archive is re3data.org, where the free-text search (e.g. for "archaeology") and the filters "Subjects" (e.g. "Humanities" or "Ancient Cultures", we recommend using both terms), "Countries", and "Content Types" help in getting a good overview. The search can be continued in other registries if the initial search results are insufficient.

The nature of the data (e.g. type, formats, content, metadata, legal status), specific requirements (e.g. access modalities, security level, legal conditions, retention period), and the desired level of service (e.g. FAQ, online support, human support, curation, self-depositing) all influence repository selection. It is crucial to check whether your institution or research funder has any requirements on this matter or if they recommend specific repositories. When evaluating the available options, the questions below should be considered in order to identify the most suitable service. More criteria to be considered are also discussed by the Digital Curation Centre (Whyte 2015).

Do your data align with the collection scope? That is, do the data match the repository's collection strategy and the specialised focus of the service? A wide variety of repository types exist at national and international levels, including general, journal supplementary, institutional, domain-specific and even project-specific repositories (Whyte 2015; Geser 2019, 25) all with varying visibility. Domain- or discipline-specific repositories are most likely to have the best visibility within a particular field and usually offer tailored data management support and domain expertise. However, they may often have stricter standards regarding data and metadata preparation (Whyte 2015).

Are your data types and file formats supported? Not all repositories support all data types or file formats nor have the capabilities for preserving them. When dealing with specialised data types, such as 3D or GIS, it is important to confirm that the archive supports them and, ideally, already has some expertise in managing them. Support and expertise regarding specific data types and suitable file formats are essential for long-term preservation, as various file formats require normalisation, monitoring, and migration to avoid obsolescence. Digital archives typically provide lists of preferred file formats and may reject files that do not meet their criteria (Whyte 2015). Examples of lists with recommended, preferred, and accepted formats are provided by the Archaeology Data Service, Data Archiving and Networked Services, the Swedish National Data Service or the Digital Archaeological Record.

Does the repository provide support with preparing and managing data for archiving? Digital long-term preservation requires careful planning and heavily influences a project's data management. Some repositories provide information pages to assist researchers in preparing their data for long-term preservation and also point to relevant but often overlooked aspects like licensing and the legal status of data. Future depositors may want to contact curators if they have more specific questions, which is not always possible for all repositories. Some repositories also provide more practical support with transforming file formats or enriching metadata during curation before the data are ingested into the repository.

Does the repository have a mandate for long-term preservation? Does the service have a mission statement clearly stating that long-term archiving is an essential part of the service? A repository wholeheartedly committed to long-term preservation, with all the required knowledge and a dedicated team, will state this. This statement must then be scrutinised for its trustworthiness, which is part of the next question.

Is the service reliable and trustworthy? A digital archive or repository suitable for long-term preservation must reliably and sustainably safeguard the entrusted materials over time (Research Libraries Group 2002), including their storage, migration and providing access (OCLC and CRL 2007 , 2). Repositories with a reliable and sustainable infrastructure composed of a suitable

organisational framework and governance, as well as transparent and comprehensive policies, are called Trustworthy Digital Repositories (TDRs) (Lin *et al*. 2020).

A common denominator for repositories and digital archives committed to long-term preservation is conformance to the OAIS Reference Model (Consultative Committee on Space Data Systems 2012). However, trustworthiness also consists of aspects such as the organisational structure, governance, available resources, and security risk management. To demonstrate trustworthiness, repositories must provide evidence, which requires openness and transparency about their practice. Regular audit processes and certificates (DPC 2015; Lin *et al*. 2020) are now becoming common practice to assess the trustworthiness of a repository. The European Commission even recommends storing data in a certified repository (European Commission: Directorate-General for Research & Innovation 2016 , 7). Therefore , checking whether the repository is certified with one of the following is the easiest way to assess a repository's suitability for long-term preservation: CoreTrustSeal (CTS), nestor Seal/DIN 31644:2012-04, and ISO 16363:2012 (ISO TRAC). Also see Section 6.4 for more details.

Not all repositories are certified, so this requires prospective data providers to conduct their own due diligence. The *Practical Guide to the International Alignment of Research Data Management* offers a list of criteria for identifying a trustworthy repository (Science Europe 2021 , 11-14) along with supporting explanations (Science Europe 2021 , 26-30).

The TRUST Principles for digital repositories published in 2020 (Lin *et al*. 2020) provide another less technical framework for assessing a repository's suitability for long-term preservation. TRUST stands for Transparency, Responsibility, User focus, Sustainability, and Technology and serves as a "mnemonic to remind data repository stakeholders of the need to develop and maintain the infrastructure to foster continuing stewardship of data and enable future use of their data holdings" (Lin *et al*. 2020). The TRUST Principles are linked to the FAIR principles, as preserving FAIR data necessitates storage in a Trusted Digital Repository.

4.3 Alternative options for digital long-term preservation

Ideally, digital archaeological data should be preserved in a certified, domain-specific repository or digital archive that supports all data types and formats intended for deposition and takes full responsibility for the preservation of the records it holds. However, such repositories are still uncommon, as illustrated by the situation in Germany. re3data.org only lists two repositories with CoreTrustSeal certification covering the subject "Ancient Cultures" - Edition Topoi Repository at the FU Berlin and Edmond, the Open Research Data Repository of the Max Planck Society. Given that 2019 appears to be the last year of publication for collections in Edition Topoi Repository, it seems that this is not a real option and leaves just one certified repository. Since this repository is institutional, it isn't available to all researchers in Germany. This scarcity is also evident in other countries and regions (Jakobsson *et al*. 2021; 2023). Because certification is a relatively recent development and the process of preparing for and acquiring a certificate takes time, more repositories are likely to become certified in the future.

However, a repository without a certificate may still be suitable in certain cases, as there are various service levels for archiving (Whyte 2015) and, as discussed in the previous section, trust can also be established through other mechanisms such as assessment lists and criteria (Yakel *et al*. 2013; Whyte 2015; Lin *et al*. 2020 ; Science Europe 2021).

A suitable service does not necessarily have to be domain-specific. A multidisciplinary or generalist repository, such as Zenodo, Harvard Dataverse, B2SHARE and many others, can be a suitable option for the long-term preservation of archaeological data (Stall *et al*. 2023). Findability of archaeological data in generalist repositories can be improved by providing as much information about the dataset as possible when making the deposit and using available features that could help classify the data as archaeology related. In Zenodo, for example, uploaded datasets can be submitted to suitable communities focusing on archaeology. Within a community, members can curate the description of the datasets.

If the data owner is willing to take on some of the archiving tasks and responsibilities, the goal of long-term preservation can also be reached with a repository with fewer capabilities. The required tasks are described in the next section.

5. Preservation without an archive

The main aim of long-term preservation of digital data is to preserve a file's content and functionality for future reuse over a long period, i.e. ten years or longer. A digital archive or repository suitable for long-term preservation takes full responsibility for this aim and will ensure the data entrusted to it are in good shape. However, as previously mentioned, preservation is possible even with less ideal facilities, provided that the data are properly prepared. This requires the consideration of the following key task areas: selection for archiving, folder structure and file naming, file formats, documentation and metadata, as well as legal and ethical considerations, including access and usage modalities of the data and licensing information. The first place to seek guidance on these issues will be an institutional data steward, a person or a department at your institution in charge of helping researchers with (research) data management. University library service centres are another valuable resource.

The remainder of this section provides more in-depth practical advice on each task area, including pointers to further guidance and information on how these tasks can be integrated into data management planning using a data management plan (DMP). Institutional research data policies can also offer more specific guidance but may not yet be widely available.

Further guidance:

- The Guides to Good Practice by the Archaeology Data Service (ADS) and Digital Antiquity (2011) is an extensive guide on how and what to archive when it comes to archaeology. The guides establish preferred formats for preservation and provide advice on data selection and basic data management practices, such as file naming conventions and metadata standards.
- Best Practice Guides for Data Preservation and Archiving by ARIADNEplus is a set of updated guides based on the Guides to Good Practice by the ADS. Also available in Swedish, hosted by the Swedish National Data Service (SND).
- The Standard and Guide to Best Practice in Archaeological Archiving in Europe (EAC Guidelines 1) by the Europae Archaeologiae Consilium (EAC) (Perrin *et al*. 2014) presents high level principles to be followed when planning and preparing for the digital preservation of archaeological documentation. The guide includes checklists and technical recommendations. The guide is also available in other languages.
- The Archaeology Data Primer by the Data Curation Network (Arteaga Cuevas *et al*. 2023) is a concise collection of information on curating archaeological data. It is hosted on GitHub and can be updated and extended by anyone.
- ADAPt Toolkit (Historic England). The Archaeological Digital Archiving Protocol Toolkit (ADAPt) is a ready-to-use set of guidelines and recommendations to integrate data preservation practices during the production of data.
- Dig Digital Directory (CIfA, Chartered Institute for Archaeologists) is a concise guide with data management essentials for archaeologists.
- The FDM-Empfehlungen für den nachhaltigen Umgang mit digitalen Daten in den Altertumswissenschaften by IANUS (2014) is a comprehensive guide for digital preservation of archaeology data available in German with advice on suitable formats for preservation and general data management tasks.
- Research Data Management in Archaeology is a guide for research data management tailored for archaeology at Leiden University.
- The Digital Preservation Handbook by the Digital Preservation Coalition (2015) provides a comprehensive overview of digital archiving and data preservation for novice and expert users. It is not specific to archaeology.
- How-To Guides by the Digital Curation Centre offer guides and checklists for some of the digital archiving steps. They are not specific to archaeology.

5.1 Selection for archiving

Selection for archiving, i.e. appraisal, is a decision-making process that results in some documents being preserved for the future while others are discarded. This step requires much care and thought because at the end of the process stands the "lasting legacy of an unrepeatable event", which should represent "much of the significance of the site or monument studied" (Oniszczuk *et al*. 2021). The selection criteria for digital archiving share similarities with those used in traditional archiving, particularly concerning content and legal aspects. However, digital archiving also introduces technical criteria. To maintain consistency and traceability within a project, the applicable selection criteria should be documented in a dedicated selection strategy.

Key selection criteria to consider are (Whyte *et al*. 2010; and see Archaeology Data Service):

- Relevance: The data must be retained to fulfil project goals or comply with legal or contractual obligations.
- Scientific or historical value: Data should be assessed for their scientific, social or cultural significance to determine whether to retain or discard them.
- Uniqueness: Prioritise data that are not found anywhere else and where data are at risk of disappearing or becoming inaccessible. Examples of non-unique data include early drafts of a final document, duplicate images, and data already securely stored elsewhere.
- Potential for redistribution: Can the files be redistributed and used by others? Files for which you lack the necessary rights (e.g. copyright, an open license or the express permission of the rights holder) cannot be archived and published. Any legal restrictions arising from the content of the files, e.g. ethical considerations or personal rights, must also be addressed. From a technical viewpoint, files in formats that are inaccessible or unusable to others due to very specific and costly software and hardware requirements might also not be kept.
- Non-replicability: Can the data be replicated from already existing sources or would this be too costly or impossible? Data easily recreated from existing datasets do not need to be retained. Where the process in the creation of a file e.g. a 3D model, involves several stages and intermediate datasets, then these may be kept depending on the resources invested.
- Economic case: Is sufficient funding available to keep all the desired data?
- Full documentation: Are the data sufficiently documented and enriched with metadata to enable and enhance reusability? This includes information about the context, methods used, specific technical requirements and records of any legal and ethical restrictions.

Further guidance:

- The DCC How-to Guide *How to Appraise and Select Research Data for Curation* (Whyte *et al*. 2010) is a guide aimed at developing an informed process for appraising and selecting data for long-term preservation. Not specific to archaeology.
- Selection and appraisal of data by the Archaeology Data Service is an adaptation of the DCC How-to Guide geared towards archaeology.
- Five steps to decide what data to keep: a checklist for appraising research data by the DCC (Whyte 2014) is a complement to the DCC How-To Guide in the form of a checklist. Not specific to archaeology.
- Guidance on selection in archaeological archiving (EAC Guidelines 3) by the Europae Archaeologiae Consilium (EAC) (Oniszczuk *et al*. 2021) aims at providing standardised information for selection in archaeological archiving, both digital and analogue. Practical examples are included. The guide is also available in other languages.
- Toolkit for Selecting Archaeological Archives by the Chartered Institute for Archaeologists is a practical guide on how to create a selection strategy for archaeological projects.
- Data Preservation. What to keep, what to delete? by DATACC is a concise and general guide with further references. Not specific to archaeology.
- Selection of Digital Materials for Long-term Retention within the Digital Preservation Handbook (DPC 2015) is an interactive decision tree from the perspective of a digital archive. Not specific to archaeology.

5.2 Folder structure and file naming

Establishing clear rules for file and folder naming, along with a logical folder structure, is essential for finding, understanding, and using data effectively.

Folder structure depends on the specifics of each project, so there is no single correct approach. The structure can be organised by topic, location, material, year, method, file types, individual workflow steps or other criteria. Generally, the file tree should avoid excessive depth, and folder names should be concise and easy to understand. For optimal cross-system compatibility, the total path length (including all folder names and the file name) should not exceed 256 characters. Paths longer than this can cause issues in MS Windows environments. Some organisations may have specific folder structure requirements to adhere to.

File naming rules should be established and documented early in a project to ensure consistency. This documentation should include any abbreviations used. Effective file names are short and descriptive. Ideally, file names should be unique within a dataset, regardless of capitalisation (e.g. "readme" is considered the same as "README"). For optimal cross-compatibility, file names should only consist of alphanumeric characters from the English alphabet (a-z, A-Z and 0-9) and avoid any other special characters, except for hyphens (-) and underscores (_). A full stop/period (.) should only be used to separate the file name and the extension. Using leading zeros for numbered files (e.g. 005 instead of 5) improves sorting and readability. File and folder versions can be indicated by using a version number (e.g. v02) or by attaching the date in ISO format (e.g. 2024-08-16). To avoid using too many hyphens and underscores in file names, capitalised words (camelCase) can be used instead e.g. aLongFileName instead of a_long_file_name or a-long-file-name.

Further guidance:

- ADAPt Toolkit (Historic England) suggests file naming conventions and folder structure based on the MORPHES project stages provided within the ADAPt Toolkit.
- Dig Digital Directory (CIfA) contains file naming and folder structure essentials.
- FDM-Empfehlungen (recommendations) by IANUS is a guide available in German tailored for archaeological projects, which includes Empfehlungen für eine Ordnerstruktur (Recommendations for a folder structure) with pointers to existing folder specifications by German and Austrian authorities (IANUS 2014 s.v. Empfehlungen für eine Ordnerstruktur) and a section about file names.
- ADS Help & Guidance Data management gives pointers for file naming and versioning
- ADS Guides to Good Practice, Dendrochronology. File naming convention suggest folder structure and file names for dendrochronology based on the TRiDaS data model.
- File organisation for the protocol for employing three-dimensional representations in archaeology (PETRA) provides an exemplary folder structure for 3D data and discusses processing workflows, file naming and metadata/paradata (Grimaud *et al*. 2019 , 10-12).
- Reflectance Transformation Imaging (RTI) folder structure is a chapter in the FDM-Empfehlungen by IANUS including a suggested folder structure for RTI data (IANUS 2014 s.v. Reflectance Transformation Imaging (RTI)).
- Grundlagen des Datenmanagements (Data management basics) gives practical pointers for naming and structuring data within folders with further links. Available in German (Trognitz 2022).
- The 5S method for organizing data files, while not specific to archaeology, is a useful practical guide on how to (re-)organise folders and files with the 5S - sort, set in order, shine, standardise, and sustain.

5.3 File formats

Preserving digital files for long-term access, readability, searchability, and ideally editability requires careful file format selection. Choosing the right file format is especially important if the data are stored in a repository without a long-term preservation strategy like software emulation or format migration.

A file format is defined by a specification that determines how the information is encoded within the file. Numerous file formats exist for different data types. For example, images can be stored in jpg, png or tiff format.

The key principles for choosing a file format for long-term preservation include:

- Suitability: The format must be capable of storing the intended content or information.
- Popularity: Widely used formats are generally preferred. Their prevalence increases the likelihood of continued software support and long-term usability.
- Openness: Open file formats are favoured over proprietary file formats. Proprietary format specifications are controlled by companies and often kept secret, tying file usage to specific software or hardware. Open file formats have publicly available specifications, enabling developers to create compatible applications.
- Compression: Uncompressed files are preferable to compressed files. Compressing files to reduce file size can result in information loss, such as reduced image quality. Lossless compression is acceptable if the compression algorithm is open source and patent-free.
- Encryption: Avoid encryption or password protection that could hinder future access and reuse of the file(s).
- Compliance: Formats should align with standards and requirements within your specific field, as well as any requirements imposed by funding bodies or other relevant third parties.

Resources with recommended formats for long-term preservation of archaeological data:

- ADS Help & Guidance Data Management contains a data requirements table with preferred and accepted formats.
- ADAPt Toolkit (Historic England) lists file formats based on the recommendations of the Guides to Good Practice.
- Preferred file formats by DANS gives a concise table of preferred and non-preferred formats for the Dutch National Centre of Expertise and Repository for Research Data (DANS).
- Choosing a file format by SND provides some pointers on choosing file formats and a table with preferred formats by the Swedish National Data Service (SND).
- FDM-Empfehlungen (recommendations) by IANUS (2014) includes a table of recommended formats for various data types in German.
- The Guides to Good Practice by the Archaeology Data Service (ADS) and Digital Antiquity (2011) provide an overview of common file formats for each data type along with consideration for long-term preservation.
- The Archaeology Data Primer by the Data Curation Network (Arteaga Cuevas *et al*. 2023) provides a short table with recommended formats per data type.
- Community Owned Digital Preservation Tool Registry (COPTR) documents tools for converting formats. Tools for specific formats (input or output) can be found via a table of file formats and metadata formats. COPTR is not specific to archaeology.
- Digital Preservation Handbook, File formats and standards (DPC 2015) gives a comprehensive introduction, not specific to archaeology, on standard formats and their importance for data preservation.
- Library of Congress Recommended Formats Statement 2024-2025 provides an overview of formats deemed most sustainable.
- Sustainability of Digital Formats by the Library of Congress (Arms *et al*. 2017) gives a more in depth and technical analysis of individual file formats with detailed sustainability factors.

5.4 Documentation and metadata

Extensive documentation and machine-readable metadata are important for preserved data to become findable, interoperable, and reusable. Many documents and files are not understandable on their own and knowledge about their context has to be documented (Huvila 2022). Documentation can be thought of as a "package insert" or a manual that enables others to find, understand, and reuse the data. Comprehensive documentation aims to provide a complete overview of the data's context, including general project information, specific processes, applied methods, and tools used. The documentation itself does not have to meet any technical specifications regarding form, content, length, and structure. Documentation for a dataset can come in the form of README files, logs, manuals, and reports (IANUS 2014 s.v. Dokumentation; Arteaga Cuevas *et al*. 2023).

At the very least, the documentation should provide (IANUS 2014 s.v. Dokumentation):

- information about the research question and material studied
- information about the project managers
- a summary of the results
- a description of workflows and methods, particularly data collection, processing and quality assurance
- a document type list (e.g. diaries, reports, lists, photos etc.)
- descriptions of the used standards, project-specific conventions, thesauri, numbering systems, etc.
- an inventory of the technical devices and programs used
- a bibliography with relevant publications and secondary literature
- important correspondence, contracts, proposals, etc. (anonymised if necessary)
- information about how the data may be used, e.g. by providing a license

Metadata provides a more technical and granular way to describe information resources at different levels of aggregation, from the data collection level down to individual files. Metadata is defined as "data used to describe other data" (Caplan 2003 , 1). Library catalogues containing bibliographic records that describe books offer a clear example of metadata (Gartner 2016 , 4, 29). Metadata for data collection or individual files are typically collected through forms provided by a repository. Because they are machine-readable, metadata are crucial for finding archaeological information, enabling computers to quickly and efficiently search them (Wise and Miller 1997).

Metadata interoperability is enhanced by using metadata standards (also referred to as metadata schemas or ontologies) that describe the structure, scope, and elements used to describe a record. The scope and relevance of collected metadata can vary significantly between repositories, although core metadata elements matching the Dublin Core Metadata Element Set (or "Dublin Core") are commonly used (Kim *et al*. 2019).

Every digital resource should be described by a minimum set of metadata elements based on the Dublin Core elements: Creator, Title, Subject, Description, Publisher, Contributor, Date, Type, Rights, Coverage (e.g. time and place), Relation (e.g. to other data), Language, Format, Identifier, and Source (see Figure 3).

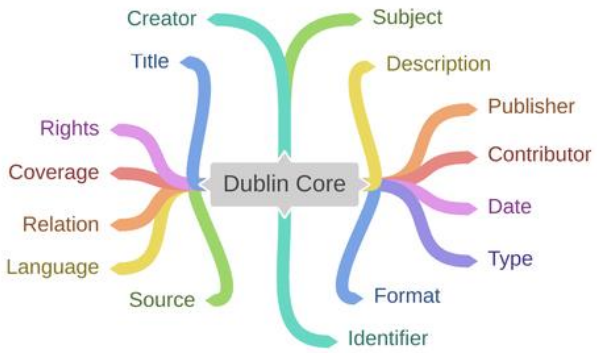

Figure 3. The Dublin Core metadata elements Creator, Title, Subject, Description, Publisher, Contributor, Date, Type, Rights, Coverage, Relation, Language, Format, Identifier, and Source. IANUS, CC BY-SA 3.0 (DE). Created with coggle.it.

Beyond this basic set of elements, other metadata fields can be used to enrich the documentation, either drawing from the updated Dublin Core Metadata Initiative terms (DCMI Metadata Terms), or from other metadata schemas relevant to archaeology listed below. The *Archaeology Data Primer* (Arteaga Cuevas *et al*. 2023) identifies the following archaeology-specific metadata elements:

- Site names and site numbers
- Site types

- Geographic site location
- Cultural keywords
- Materials
- Investigation types
- Period dates
- Project dates
- Subjects

The comparability and interoperability of (meta)data can further be enhanced by recording the information in each metadata element consistently and in a standardised way (Zhang *et al*. 2009). This can be achieved by using controlled vocabularies (also referred to as taxonomies, thesauri or authority files), which are essentially lists of predefined terms (words or phrases) that limit the range of possible values to be used as a value for a metadata element (Harpring 2010; Hedden 2010).

Different vocabularies are readily available to be used to describe cultural heritage or archaeology data and are listed below.

For further guidance, see:

- Why Metadata Matters in Archaeology (Wise and Miller 1997) is a basic introduction into what metadata are and how they apply to archaeology.
- Project documentation (Niven 2011a) & Project metadata (Niven 2011b) concern documentation and metadata in the ADS Guides to Good Practice (Archaeology Data Service and Digital Antiquity 2011).
- Metadata and vocabularies for archaeological datasets sets out a table with resources aimed at introducing metadata and controlled vocabularies and the concepts behind them.
- FDM-Empfehlungen (recommendations) by IANUS (2014) is the German guide which includes a chapter on documentation (Dokumentation) with a subsection about metadata (Dokumentation mit Metadaten).
- Metadata and Vocabularies (CARARE) consists of training material provided by CARARE with a focus on 3D data.
- FAIRsharing.org (the FAIRsharing Community *et al*. 2019) is a community-driven list that includes metadata standards and controlled vocabularies relevant to archaeology

Lists with metadata standards relevant to archaeological research data:

- The RD Metadata Standards Catalog: a collaborative and open directory of metadata standards that can be browsed, among other options, by subject, e.g. Archaeology.
- ADS Guides to Good Practice: ADS Project metadata and tDAR General metadata
- ADS Metadata standards: a list of metadata schemas and vocabularies used by the ADS.
- FDM-Empfehlungen (recommendations) by IANUS (2014): the German guide includes a list of metadata schemas for archaeology in the chapter on structured metadata (Strukturierung von Metadaten).
- Numismatic Description Schema (NUDS): a metadata schema for describing numismatic collections. It is implemented at and published by Nomisma.org.
- Kerameikos.org Ontology: a metadata schema for ceramic data.

Controlled vocabularies relevant to archaeology:

- Basic Register of Thesauri, Ontologies & Classifications (BARTOC): Aggregates vocabularies in SKOS format from multiple sources.
- Getty Art & Architecture Thesaurus (Getty AAT): Widely used vocabulary with general terms.
- Forum on Information Standards in Heritage (FISH Vocabularies): Vocabularies for recording heritage assets developed in the UK.
- List of other terminology resources by FISH: Other sources for controlled vocabularies compiled by the FISH.
- Pleiades (Simon *et al*. 2016): Gazetteer for ancient place names.

- GeoNames: Community-built geographic database.
- PeriodO (Golden *et al*. 2016): Gazetteer of scholarly period definitions.
- iDAI.chronOntology (Schmidle 2021): Gazetteer of periods derived from objects.
- Wikidata (Vrandečić *et al*. 2014): Open knowledge base with items suitable to be used as vocabulary terms, which are also interlinked to other controlled vocabularies.

5.5 Legal and ethical aspects: access to data, licensing, sensitive data

Legal aspects of research data boil down to two questions: "Will others be allowed to use your data?" and "How?". Similarly, when reusing and publishing third-party data, the questions "Are you allowed to use existing data?" and "Are there requirements and restrictions for publishing the data?" have to be addressed.

Answering these questions for the majority of cases requires basic knowledge of intellectual property rights (IPR), especially copyright, public domain, and licenses. Furthermore, the applicable legislation and relevant contracts and policies, such as employment contracts, excavation permits, institutional research data management policies, must be observed. If any doubt arise, especially concerning third-party data, legal advice from an IPR specialist is recommended. A good starting point will be to contact your organisation's legal department.

"Copyright legislation is part of the broader body of law known as intellectual property (IP), which refers broadly to the creations of the human mind" (WIPO 2016) and is organised into two areas: industrial property and copyright. Industrial property relates to inventions and related rights like patents, trademarks, while copyright applies to the rights on original creative work, the "literary and artistic work". What can be considered as such a work is clarified within Article 2 of the Berne Convention: "The expression 'literary and artistic works' shall include every production in the literary, scientific and artistic domain, whatever may be the mode or form of its expression" (WIPO 1979). Most of the data dealt with in archaeology falls within this definition, and includes actual finds and objects (Farmer *et al*. 2024).

Copyright protection begins upon the creation of the work and lasts, depending on the country of jurisdiction, for 50 to 70 years (or even longer in some countries) after the creator's death. When a work is created by a group, the copyright ends with the death of the last author and the expiry of the applicable term. After copyright protection ends, the work enters the public domain and can be used, copied, remixed, etc., without any restrictions.

Licenses and waivers are effective tools for providing potential reusers with clear information about usage modalities. Both can only be granted by the rights holder or someone acting on their behalf. A license is used to grant others the rights to use the licensed work under certain conditions. A waiver relinquishes all rights to a resource (Ball 2014).

Creative Commons (CC) licenses are among the most common encountered licenses. These internationally recognised and widely used licenses allow authors to easily label their work with specific usage rights. A CC license clearly states the conditions under which the licensed work can be used. For example, CC BY 4.0 means that a work is licensed with a Creative Commons license version 4.0 and may be used, modified and distributed by others as long as the author and the license are credited. Figure 4 shows common combinations of CC modules organised by their degree of openness.

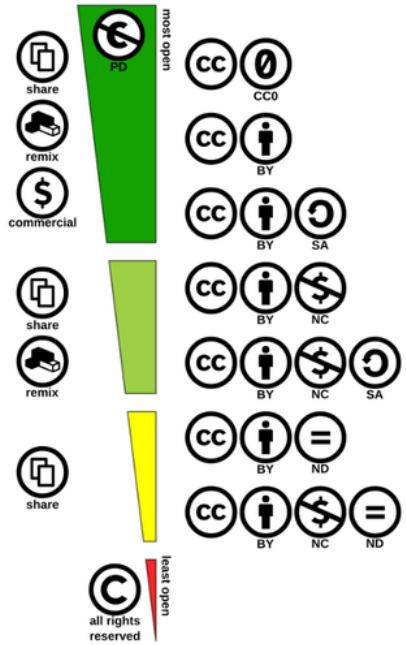

Figure 4. Combination of Creative Commons licenses arranged according to their openness between public domain at the top, to all rights reserved at the bottom. Author Shaddim: original CC license symbols by Creative Commons, via Wikimedia Commons, CC BY 4.0.

CC licenses are suitable for licensing creative works, such as texts, images or other work where copyright applies. For resources like software or databases, other licenses are more suitable (Guibault *et al*. 2013 , 149-150; Kreutzer 2014). The Open Data Commons licenses, which are maintained by the Open Knowledge Foundation, are open licenses bespoke to databases (Ball 2014).

In line with open science principles, an open license should be used whenever possible. Data in the public domain can be labelled as such by using the Creative Commons Public Domain Mark (CC PDM). Furthermore, the Europeana Public Domain Charter (Europeana 2010) advocates for public domain works to remain in the public domain. This means that digitisations of public domain works should ideally also be published as public domain, to avoid situations where, for example, photographs of cultural heritage objects can't be used without permission, as is the case with the Nebra Sky Disk (Ostendorff 2024).

Two additional issues require consideration before submitting data to a repository and they concern sensitive data: ethics and the protection of personal data. For personal data, such as photographs of excavation participants, name lists, etc. the European General Data Protection Regulation (GDPR), which only applies to living persons, requires informed consent prior to publication. Another way of dealing with personal data is anonymisation.

Consultation with an ethical board might also be necessary whenever there is a danger of potential misuse of research outputs, especially when dealing with human remains. Generally, codes of conduct for research integrity e.g. (ALLEA 2023), and the CARE principles for Indigenous Data Governance (Carroll *et al*. 2020) should be followed.

Further guidance:

- Understanding Copyright and Related Rights (WIPO 2016): Brochure with essential and comprehensible information on copyright and related rights.
- How to License Research Data (Ball 2014): A guide to licenses with several examples and use cases.
- Toolkit for Researchers on Legal Issues from OpenAire (Margoni *et al*. 2018): An accessible guide to common legal issues in research with recommendations.

- Open Content - A Practical Guide to Using Creative Commons Licences (Kreutzer 2014): Practical oriented guide for the use and application of open content licenses. Available in multiple languages.
- Ethics and Legality in the Digital Arts and Humanities (ELDAH): A working group within DARIAH concerned with ethics and legal issues, in particular with regard to open science.
- GLAM-E Lab: Resources to guide towards institutional open access. Includes a handbook for copyright clearance (Farmer *et al*. 2024).
- Europeana Copyright Community: A group within the Europeana Pro network for sharing resources and legal advice on copyright issues in the context of digital European cultural heritage. Tools include various flowcharts, also for US legislation.
- How to select an accurate rights statement by the Europeana Copyright Community: A guide for choosing an accurate rights statement for data to be shared on Europeana or elsewhere.
- Public domain calculators: Flowcharts for various countries to help decide if a work is in the public domain.
- CLARIN license chooser tools: A list with links to various tools for selecting the appropriate license for your work.
- Forschungsdaten veröffentlichen? [Publish research data?] (Schleußinger *et al*. 2019): A decision tree to guide through important legal aspects when publishing research data. Available in German only.
- Amnesia Anonymization Tool by OpenAIRE: An online service for anonymisation of data.

5.6 Data management plan(ning)

The creation of a dataset should be preceded by thorough data management planning that is documented in a bespoke plan - the data management plan (DMP). Just as product design is essential for manufacturing, or research design for conducting an experiment, or program specification for software development, a DMP ensures that all stages in the data lifecycle are taken into account within a project and that the data can be reused in the future. Data management planning at the outset of a project will ensure an organised dataset and a simplified archiving process at the end of a project.

In its simplest form, a DMP is a document that describes, within a specific context (e.g. project, research question), the type of data being created (e.g. datasets, documents, structured text, code, etc.), the methods used in its creation, who created the data and who is responsible for it, as well as where (e.g. storage type and location) the data can be found, how it can be accessed (e.g. applicable licenses, restrictions) and what requirements, standards, regulations, and obligations must be observed and complied with. The documentation should detail applied workflows and processes, address legal and ethical considerations, and outline roles and responsibilities throughout the project lifecycle. A DMP is a living document that can and should be reviewed and adapted over the course of a project, with all changes documented.

Recognising the importance of proper data management, a growing number of funding bodies now require DMPs as part of research project proposals. In the *Practical Guide to the International Alignment of Research Data Management*, Science Europe (2021) provide a set of core requirements for managing research data, which are now widely accepted and implemented across multiple stakeholders. Based on these requirements, a set of tools for data management planning for archaeology was developed as part of ARIADNEplus (Doorn and Ronzino 2022a ; 2022b). Furthermore, a growing number of universities and organisations provide templates for and assistance with DMPs.

Further guidance:

- ARGOS by OpenAIRE and EUDAT: An open platform for creating and managing research DMPs that comply with the FAIR and Open Science principles. ARGOS can fetch information from various sources and already comes with various templates, including the DMP template by ARIADNE.

- DMPonline by the Digital Curation Centre: Allows the creation of public or private DMPs based on various templates tailored to the requirements of various EU and UK funding bodies.
- ezDMP: A DMP tool tailored to the requirements of the National Science Foundation (NSF) in the US.
- Data Stewardship Wizard: Tools and smart questionnaires to guide through the DMP creation process depending on the project's phase and earlier answers.
- ARIADNEplus Data Management Plan Tools: Includes various DMP templates developed within the AriadnePlus project specifically for archaeological datasets (Doorn and Ronzino 2022a).
- Dig Digital Directory (CIfA): Data Management Planning.

6. Working towards a digital archive

The incentives for digital preservation, as outlined in Section 3, and the increasing number of funders requiring research data to be deposited in trusted or certified repositories, have led to a growing number of researchers archiving their data (Digital Science *et al*. 2018 ; 2019; 2022; Geser 2019 , 31-40). However, the number of suitable digital archives or trusted repositories is still far from meeting the demand for digital preservation (Jakobsson *et al*. 2021; 2023). Therefore, any new addition to the list of certified repositories suitable for long-term preservation is warmly welcome.

Building up a robust digital archive from scratch requires careful planning, specialised (and often hard-to-find) staff, and a firm commitment from the host institution to persevere through challenges, especially the ever-present issue of underfunding. Digital preservation is a continual process and requires steady maintenance work, updates and improvement to hardware, software, and workflows.

The following section focuses on the more technical aspects of digital preservation essential for modern digital archives: planning, software, certification, aggregators, workflows, and staff capacity and expertise.

6.1 Planning for a long-term preservation service

Planning for the implementation of a digital archive is a significant step that requires time and careful consideration of the complexities involved in a long-term preservation service. Once the commitment of the designated hosting institution has been established, the requirements can be defined. The necessary requirements can be determined on the basis of various materials but should at a minimum include the OAIS Reference Model (Consultative Committee on Space Data Systems 2012), the TRUST Principles for digital repositories (Lin *et al*. 2020) (see Section 4.2), and relevant standards (not only, but also metadata and controlled vocabularies as described in Section 5.4). In addition, if certification is intended, the relevant requirements should be considered from the outset and included in respective policies. Planning for the long term should include finding reliable partners and allow for any future changes.

The planning horizon for a digital archive is typically measured in years rather than weeks or months. For example, planning for the archaeological database and repository of the Hungarian National Museum started in 2012 as part of the ARIADNE project and progressed on a smaller scale every year until 2016 when the site was launched (Kreiter 2019; Péter 2023). Similarly, the German research data centre IANUS was planned during its first funding phase, 2011-2014, before implementation began in a second funding phase and subsequently launched in 2016.

Some resources for planning a digital archive include:

- Reference Model for an Open Archival Information System (OAIS) (Consultative Committee on Space Data Systems 2012)
- TRUST Principles for digital repositories (Lin *et al*. 2020)
- Digital Preservation Handbook (DPC 2015)
- Practical Preservation Book (Brown 2013)

- Fach- und Organisationskonzept zum Betrieb eines nationalen Forschungsdatenzentrums für die Archäologien und Altertumswissenschaften in Deutschland [Technical and organisational concept for the operation of a national research data centre for archaeology and classical studies in Germany] (IANUS 2016)
- The questionnaire for surveying archiving institutions and/ or archiving software in the Annex, which is based on questionnaires published by IANUS in 2012 (now unavailable) but provides an extensive list of questions useful in gathering requirements.

6.2. Software for digital archives

Given the wide array of repository software solutions available, there is no single "best" software for a digital archive or a repository for long-term preservation. The reason for this is that long-term archiving is a process with many different components, concepts and workflows (see Section 2 and Section 3). The requirements established during the planning phase will vary from one institution or discipline to another (Van Garderen 2006). Furthermore, new solutions become available while established solutions can become outdated or obsolete. Technological advances may require a change of the underlying software of existing digital archives as it is being done for DANS with a transition from their custom built system EASY to the software Dataverse or for ARCHE, where performance issues led to abandoning Fedora for the custom build ARCHE Suite (Żółtak *et al*. 2022).

Technically, various approaches are viable: from developing a bespoke and fully customisable software stack, installing and hosting a customisable, off-the-shelf product, partnering with existing services, to outsourcing to a reliable but non-customisable service. It all depends on the requirements, which have to be identified up front to avoid future problems. Provided the surrounding workflows and procedures are sound, the software solution can be relatively simple, even file-based with an accompanying database for metadata storage and querying, as is the case for the Hungarian archaeological database mentioned above (see Section 6.1).

Beyond the back-end of the digital archive, the front-end or the Graphical User Interface (GUI) presented to the user deserves attention. The GUI should provide a user-friendly experience for searching and viewing of the archived content. For archaeological data, this includes not only text and images, but also 3D or geospatial data. Displaying geospatial data, for example, requires a map view capable of rendering points, polygons and layers.

When looking for tools and applications for specific digital preservation tasks, such as format migration, metadata extraction, or handling a specific file type, the Community Owned Digital Preservation Tool Registry (COPTR 2021) is a useful source. This wiki-based registry lists tools that can be searched by various facets of digital preservation e.g. stage in the lifecycle, function, content type, or file format. Additionally, COPTR is a platform for sharing workflows of various stages of digital archiving. COPTR is an open community, and anyone can contribute to the collection of tools and workflows.

6.3 Certification

Funders are increasingly demanding that data should be deposited in certified repositories (e.g. European Commission: Directorate-General for Research & Innovation 2016 , 7). While certification is a demanding and time-consuming process and requires the involvement of the entire team, it offers significant benefits. It brings greater transparency to internal procedures and policies, and provides an opportunity for reflection on how things are done and how they can be improved. A certificate signifies that the holder is a reliable and sustainable service - technically, financially, and legally trustworthy.

Currently, three certificates are relevant to the field of digital preservation:

- CoreTrustSeal: Established in 2017 (L'Hours *et al*. 2019), the CoreTrustSeal (CTS) is the result of the harmonisation of the Data Seal of Approval (DSA) and the World Data System (WDS). At the time of writing, the CTS has 16 Trustworthy Data Repositories Requirements (CoreTrustSeal Standards and Certification Board 2022). The certification is achieved through self-assessment followed by a review process that should be renewed every three years.

- nestor Seal/DIN 31644:2012-04: The nestor Seal for Trustworthy Digital Archives, adopted by the German Standards Committee (DIN) as DIN 31644 *Information and documentation - Criteria for trustworthy digital archives*, is an extended self-assessment based on the Catalogue of Criteria for Trusted Digital Repositories (nestor Working Group Trusted Repositories - Certification 2009)
- ISO 16363:2012: Space Data and Information Transfer Systems - Audit and Certification of Trust-worthy Digital Repositories (ISO TRAC) was influenced by the development of the OAIS Reference Model, which itself is an ISO standard (ISO 14721:2012 - *Space Data and Information Transfer Systems - Open Archival Information System (OAIS) - Reference Model*). Certification with ISO 16363 can only be achieved with an external audit (PTAB, Primary Trustworthy Digital Repository Authorisation Body Ltd) and the requirements and criteria were previously published (Research Libraries Group 2002; RLG-NARA Task Force on Digital Repository Certification 2005; OCLC and CRL 2007) with an update in 2011 (Consultative Committee on Space Data Systems 2011).

Based on the experiences of certified repositories in Austria (Ernst *et al*. 2020), here are some recommendations for the certification process:

- Allow ample time, especially for first-time certification submissions.
- Do not rush into certification before the service is up and running and the relevant workflows are established.
- Evaluate existing openly available certifications (e.g. CTS) to get an idea of the requirements and use them as a guide.
- Be prepared that the documentation and other publicly available documents and policies might have to be improved or created.
- The certification process should be planned as a project in its own right with a key person responsible for coordinating various staff members and institutional departments.

6.4 Dissemination of data and aggregators

While the core of digital preservation is keeping digital files intact (see Section 2), their findability and accessibility is crucial for a sustainable and FAIR digital archive. Given the growing number of available repositories, simply publishing content via an online GUI is no longer enough to guarantee the visibility of the data beyond the service (Bollwerk *et al*. 2024).

This broader visibility can be achieved by ensuring the data is properly indexed by search engines and actively pushing metadata to, or making it available for collection by, relevant aggregators. In addition to national aggregators, discipline-specific aggregators, such as the ARIADNE portal (ARIADNE Research Infrastructure) and Europeana should be considered to boost visibility and help in providing a single place to look for archaeological data (Richards 2023). Another EU-based aggregator, open to all scientific disciplines and all types of research output (data, publications, software, etc.) globally, is OpenAIRE Explore, which aggregates (and deduplicates) resources from various sources including Zenodo.

Disseminating metadata via an aggregator requires an interoperable infrastructure, compatible with other services and resources. This means adhering to international standards wherever possible, e.g. when selecting the data model, metadata schema, and controlled vocabularies. The current standard for the provision of metadata to aggregators is the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) (Lagoze *et al*. 2002), which allows metadata to be provided in different metadata formats. Disseminating data and metadata as Linked Open Data (LOD) (May *et al*. 2015; Schmidt *et al*. 2022) also increases reusability by enabling integration into computational and quantitative research endeavours.

6.5 Staff: find, train & network

Digital preservation is a complex endeavour that requires a team of IT specialists, data stewards, and digital curators, each providing the essential skills. IT specialists must possess the expertise to build and maintain the technical framework. They also must monitor technology in terms of sustainable file

formats and the evolution of technical standards. Besides data management skills, data stewards and digital curators will require an understanding of the content of the material being archived and knowledge of the legal implications of data ownership and access. Additional archivist training can also be helpful. Formal frameworks for the education and training of digital curators or digital archivists are still scarce, and most people working in the field come from related disciplines and are trained on the job, learning from their colleagues and through professional networks. These networks and community-based initiatives are vital sources of information about professional opportunities for both practitioners and employers seeking staff.

- The Digital Preservation Coalition (DPC), the organisation behind the *Digital Preservation Handbook* (DPC 2015), is one of the most resourceful organisations regarding professional advice, training and personal development of digital preservation professionals. DPC members are eligible to undertake training modules provided by the coalition team and designed for various professional levels. Additionally, the DPC website lists the UK institutions that offer digital preservation education in their curricula and advertises digital archiving job vacancies.
- The Society of American Archivists (SAA) is another large body that provides a variety of courses for digital archivists and advertises jobs in the field.
- The Digital Curation Centre (DCC): A competence centre for curation of digital data. It has a focus on capacity building for professional research data management.
- The Digital Curation group is an open and active discussion list for all things digital preservation and curation. It is a good starting point for specific questions beyond your institution.
- The Research Data Alliance (RDA): A community-driven initiative focusing on the goal of sharing and reusing research data. Thus, many of the RDA's guiding principles and the areas of interest of its members coincide with those of digital archiving professionals. The RDA offers a free individual membership, which allows for excellent networking conditions.
- The University of Oxford offers the Digital Archivist Graduate Training Scheme - a two-year program based at the Bodleian Libraries, in cooperation with the Masters degree in Archives and Records Management at Aberystwyth University.
- Data Steward training is offered by the Universities of Vienna and Graz as a postgraduate qualification course. The programmes include modules about long-term preservation.

7. Conclusion

This article has explored and summarised the concepts, principles, and tasks involved in the digital preservation of research data. In summary, digital long-term preservation is a process with bitstream preservation at its core, augmented by curation workflows and active data management to ensure that digital materials can be opened, read, edited, searched, and ultimately reused by future generations of researchers.

Adhering to the FAIR and CARE principles throughout a project from its inception, is crucial for creating sustainable and reusable data, both during the interim or transitional phases of a project, right through to its final deposit in a digital archive.

We have argued that digital preservation of research data is possible even when a trusted archive is not readily available. Using a less-than-ideal alternative service and proactively taking care of the curation of the data, preservation remains possible. However, this approach requires a significantly greater investment on the part of data owners and creators to ensure that data is deposited in appropriate formats, accompanied by sufficient documentation and compliant with all legal requirements.

For institutions willing to commit themselves to the implementation of a digital archive, we have presented the essential technical requirements and considerations necessary for a modern service. The range of requirements that a digital archive must consider can be found in the annexed questionnaire, which may still not be exhaustive in every detail.

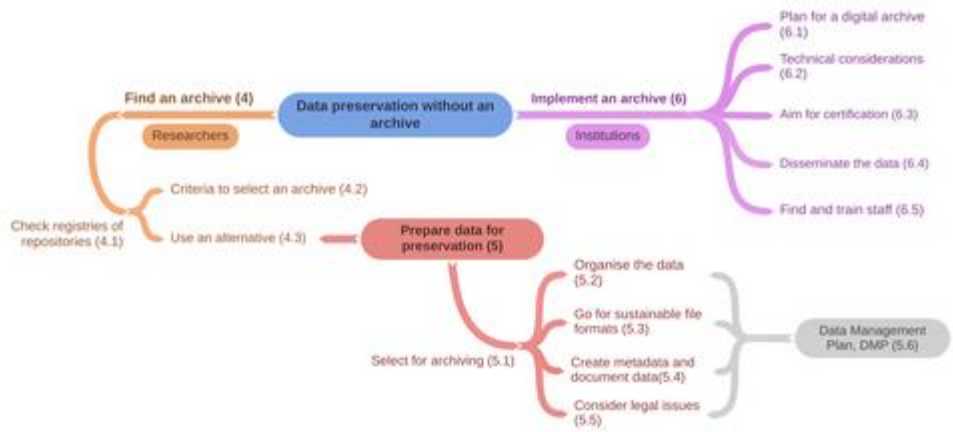

Figure 5. Visual summary and flowchart of the sections within this publication. Sveta Matskevich and Martina Trognitz. CC BY 4.0. Created with coggle.it.

Annex [online only]: Questions for surveying or planning a digital archive

Bibliography

Abadal, E. 2021 'Ciencia abierta: un modelo con piezas por encajar', *Arbor* **197**(799), a588. https://doi.org/10.3989/arbor.2021.799003

Aleixandre-Benavent, R., Vidal-Infer, A., Alonso-Arroyo, A., Peset, F., & Ferrer Sapena, A. 2020 'Research Data Sharing in Spain: Exploring Determinants, Practices, and Perceptions', *Data* **5**(29). https://doi.org/10.3390/data5020029

ALLEA 2023 *The European Code of Conduct for Research Integrity*, Revised Edition 2023. ALLEA. https://doi.org/10.26356/ECOC [Accessed September 17, 2023].

Archaeology Data Service & Digital Antiquity 2011 *Guides to Good Practice*, University of York. http://guides.archaeologydataservice.ac.uk/g2gp/Main [Accessed February 27, 2024].

Arms, C.R., Fleischhauer, C., & Murray, K. 2017 *Sustainability of Digital Formats: Planning for Library of Congress Collections*, Library of Congress. https://www.loc.gov/preservation/digital/formats/index.shtml [Accessed August 21, 2024].

Arteaga Cuevas, M., Fernandez, R., & Wittmann, H. 2023 *Data Curation Network Data Primers. Archaeology Data Primer*, Data Curation Network. https://github.com/DataCurationNetwork/data-primers/blob/master/Archaeology%20Data%20Primer/archaeology-primer.md [Accessed February 27, 2024].

Ball, A. 2014 *How to License Research Data. DCC How-to Guides*, Digital Curation Centre. https://www.dcc.ac.uk/guidance/how-guides/license-research-data [Accessed December 10, 2024].

Bánki, Z., Fonyódi, K., Káldos, J., Kómár, E., Ráduly, G. and Szatucsek, Z. 2019 *Fehér könyv* (White Book), Emberi Erőforrások Minisztériuma. https://kds.gov.hu/feher-konyv/ [Accessed August 24, 2024].

Bezjak, S., Conzett, P., Fernandes, P.,Görögh, E., Helbig, K., Kramer, B., Labastida, I., Niemeyer, K., Psomopoulos, F., Ross-Hellauer, T., Schneider, R, Tennant, J., Verbakel, E., Clyburne-Sherin, A. 2018 *Open Science Training Handbook*, Facilitating Open Science in European Research. https://zenodo.org/records/2587951 [Accessed February 10, 2023].

Bundesministerium für Kunst, Kultur, öffentlichen Dienst und Sport 2023 *Strategie Kulturerbe Digital. Digitaler Aktionsplan Austria*, Wien. https://www.bmkoes.gv.at/dam/jcr:418639e8-05d5-44b4-a5e0-fe12b517a742/Strategie-Kulturerbe-digital.pdf [Accessed March 25, 2025]

Bollwerk, E., Gupta, N., & Smith, J. 2024 'A Systems-Thinking Model of Data Management and Use in US Archaeology', *Advances in Archaeological Practice* **12**(1), 53-59. https://doi.org/10.1017/aap.2023.41

Brown, A. 2013 *Practical Digital Preservation: A How-to Guide for Organizations of Any Size*, Facet. https://www.cambridge.org/core/books/practical-digital-preservation/E7E951A42FFA5BC14D412FEEA0972367 [Accessed November 27, 2023].

Calkins, H., Condon, P., Petters, J., Woodbrook, R., & Boehm, R. 2022 *SPARC Data Sharing Resource Update 2020*, Open Science Framework. https://osf.io/c4pd8/ [Accessed January 4, 2024].

Caplan, P. 2003 *Metadata fundamentals for all librarians*, Chicago: American Library Association. http://archive.org/details/metadatafundamen0000capl [Accessed August 22, 2024].

Carroll, S.R., Garba, I., Figueroa-Rodríguez, O.L., Holbrook, J., Lovett, R., Materechera, S., Parsons, M., Raseroka, K., Rodriguez-Lonebear, D., Rowe, R., Sara, R., Walker, J.D., Anderson, J. and Hudson, M. 2020 'The CARE Principles for Indigenous Data Governance', *Data Science Journal* **19**(1). https://datascience.codata.org/articles/10.5334/dsj-2020-043

Chartered Institute for Archaeologists (CIfA) 2019 *Toolkit for Selecting Archaeological Archives*. https://www.archaeologists.net/selection-toolkit/toolkit-overview [Accessed February 28, 2024].

Chartered Institute for Archaeologists (CIfA) 2022 *The Dig Digital Directory*. https://www.archaeologists.net/sites/default/files/downloads/selection-toolkit/7796_DigDigital_Directory_V2.4_0.pdf [Accessed February 28, 2024].

Consultative Committee on Space Data Systems 2011 *Audit and Certification of Trustworthy Digital Repositories*, Washington (DC): CCSDS Secretariat. https://public.ccsds.org/pubs/652x0m1.pdf [Accessed February 5, 2024].

Consultative Committee on Space Data Systems 2012 *Reference Model for an Open Archival Information System (OAIS)*, Washington (DC): CCSDS Secretariat. https://public.ccsds.org/pubs/650x0m2.pdf [Accessed February 28, 2024]

CoreTrustSeal Standards and Certification Board 2022 *CoreTrustSeal Requirements 2023-2025*. https://zenodo.org/record/7051012 [Accessed February 5, 2024].

Digital Science, Hahnel, M., Fane, B., Treadway, J., Baynes, G., Wilkinson, R., Mons, B., Schultes, E., Olavo Bonino da Silva Santos, L., Arefiev, P. and Osipov, I. 2018 *The State of Open Data Report 2018*, Digital Science. https://digitalscience.figshare.com/articles/report/The_State_of_Open_Data_Report_2018/7195058/2 [Accessed November 30, 2023].

Digital Science, Fane, B., Ayris, P., Hahnel, M., Hrynaszkiewicz, I., Baynes, G. and Farrell, E, 2019 *The State of Open Data Report 2019*, Digital Science. https://digitalscience.figshare.com/articles/report/The_State_of_Open_Data_Report_2019/9980783/2 [Accessed November 30, 2023].

Digital Science, Goodey, G., Hahnel, M., Zhou, Y., Jiang, L., Chandramouliswaran, I., Hafez, A., Paine, T., Gregurick, S., Simango, S., Miguel Palma Peña, J., Murray, H., Cannon, M., Grant, R., McKellar, K. and Day, L. 2022 *The State of Open Data 2022*, Digital Science. https://digitalscience.figshare.com/articles/report/The_State_of_Open_Data_2022/21276984/5 [Accessed November 30, 2023].

DIN. DIN 31644:2012-04 *Information und Dokumentation_ - Kriterien für vertrauenswürdige digitale Langzeitarchive*. https://www.beuth.de/de/-/-/147058907 [Accessed February 5, 2024].

Doorn, P. and Ronzino, P. 2022a *ARIADNEplus Data Management Plan Tools*. https://vast-lab.org/dmp/index.html [Accessed August 28, 2024].

Doorn, P. and Ronzino, P. 2022b *Guide for Archaeological Data Management Planning*. https://training.ariadne-infrastructure.eu/dmp-guidance/ [Accessed August 28, 2024].

DPC 2015 *Digital Preservation Handbook*, Digital Preservation Coalition. https://www.dpconline.org/handbook/institutional-strategies/audit-and-certification [Accessed February 5, 2024].

Ernst, D., Novotny, G. and Schönher, E.M. 2020 '(Core Trust) Seal your repository!', *Mitteilungen der Vereinigung Österreichischer Bibliothekarinnen und Bibliothekare* **73**(1), 46-59. https://doi.org/10.31263/voebm.v73i1.3491

European Commission 2011 *Commission Recommendation of 27 October 2011 on the digitisation and online accessibility of cultural material and digital preservation (2011/711/EU)*. https://eur-lex.europa.eu/eli/reco/2011/711/oj/eng

European Commission 2019 *Open Science*. https://research-and-innovation.ec.europa.eu/strategy/strategy-research-and-innovation/our-digital-future/open-science_en [Accessed December 8, 2024].

European Commission: Directorate General for Research and Innovation 2022 *European Research Data Landscape - Final report*, Publications Office of the European Union. https://data.europa.eu/doi/10.2777/3648 [Accessed January 4, 2024].

European Commission: Directorate-General for Research and Innovation 2016 *H2020 Programme: Guidelines on FAIR Data Management in Horizon 2020*, Version 3.0. https://www.oceanbestpractices.net/handle/11329/1259 [Accessed February 5, 2024].

Europeana 2010 *The Europeana Public Domain Charter*. https://pro.europeana.eu/post/the-europeana-public-domain-charter [Accessed December 12, 2024].

Farmer, F., Wallace, A. and Weinberg, M. 2024 *Copyright Clearance Handbook for Public Domain Publications of Digital Collections*, glam-e-lab. https://glamelab.org/products/copyright-clearance-handbook-for-public-domain-publications/ [Accessed December 12, 2024].

forschungsdaten.info 2023 *Funder Guidelines*. forschungsdaten.info. https://forschungsdaten.info/praxis-kompakt/english-pages/funder-guidelines/ [Accessed January 4, 2024].

forshungsdaten.org. *Langzeitarchivierung*. https://www.forschungsdaten.org/index.php/Langzeitarchivierung [Accessed November 29, 2023].

FOSTER Consortium 2018 *What is Open Science?* https://zenodo.org/record/2629946 [Accessed December 10, 2024].

Gartner, R. 2016 *Metadata: Shaping Knowledge from Antiquity to the Semantic Web*, Cham: Springer International Publishing.

Geser, G. (ed) 2019 *ARIADNEplus Community Needs Survey 2019*, ARIADNEplus. http://ariadne-infrastructure.eu/wp-content/uploads/2019/11/ARIADNEplus-Survey-2019-Report.pdf [Accessed February 1, 2024].

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

Golden, P. and Shaw, R. 2016 'Nanopublication beyond the sciences: the PeriodO period gazetteer', *PeerJ Computer Science* **44**. https://doi.org/10.7717/peerj-cs.44

Grimaud, V. and Cassen, S. 2019 'Implementing a protocol for employing three-dimensional representations in archaeology (PETRA) for the documentation of neolithic funeral architecture in

Western France', *Digital Applications in Archaeology and Cultural Heritage* **13**, e00096 https://doi.org/10.1016/j.daach.2019.e00096

Guibault, L. and Wiebe, A. (eds) 2013 *Safe to be open: study on the protection of research data and recommendations for access and usage*, Göttingen: Göttingen University Press. https://doi.org/10.17875/gup2013-160

Gupta, N., Martindale, A., Supernant, K., & Elvidge, M. 2023 'The CARE Principles and the Reuse, Sharing, and Curation of Indigenous Data in Canadian Archaeology', *Advances in Archaeological Practice* **11**(1), 76-89. https://doi.org/10.1017/aap.2022.33

Hahnel, M. 2015 *Global funders who require data archiving as a condition of grants*. https://figshare.com/articles/dataset/Global_funders_who_require_data_archiving_as_a_condition_of_grants/1281141/1 [Accessed January 4, 2024].

Harpring, P. 2010 *Introduction to Controlled Vocabularies Online*, J. Paul Getty Trust. https://www.getty.edu/research/publications/electronic_publications/intro_controlled_vocab/index.html [Accessed April 20, 2022].

Hedden, H. 2010 'Taxonomies and controlled vocabularies best practices for metadata', *Journal of Digital Asset Management* **6**(5), 279-284. https://doi.org/10.1057/dam.2010.29

Huvila, I. 2022 'Improving the usefulness of research data with better paradata', *Open Information Science* **6**(1), 28-48. https://doi.org/10.1515/opis-2022-0129

IANUS 2014 *FDM-Empfehlungen für den nachhaltigen Umgang mit digitalen Daten in den Altertumswissenschaften*. https://dx.doi.org/10.13149/000.111000-a

IANUS 2016 *Fach- und Organisationskonzept zum Betrieb eines nationalen Forschungsdatenzentrums für die Archäologien und Altertumswissenschaften in Deutschland*. https://ianus-fdz.de/files/Konzept-IANUS_v0-95_2016-01-12.pdf [Accessed December 16, 2024].

Jahn, N., Laakso, M., Lazzeri, E. and McQuilton, P. 2023 *Study on the readiness of research data and literature repositories to facilitate compliance with the Open Science Horizon Europe MGA requirements*, Zenodo. https://zenodo.org/records/7728016 [Accessed August 24, 2024].

Jakobsson, U., Novák, D., Richards, J.D., Štular, B. and Wright, H. (eds) 2021 'Digital Archiving in Archaeology: The State of the Art', *Internet Archaeology* **58**. https://intarch.ac.uk/journal/issue58/index.html [Accessed April 18, 2022].

Jakobsson, U., Novák, D., Richards, J.D., Štular, B. and Wright, H. (eds) 2023 'Digital Archiving in Archaeology: Additional State of the Art and Further Analyses', *Internet Archaeology* **63**. https://intarch.ac.uk/journal/issue63/index.html

Kim, J., Yakel, E. and Faniel, I.M. 2019 'Exposing Standardization and Consistency Issues in Repository Metadata Requirements for Data Deposition', *College and Research Libraries* **80**(6), 843-875. https://doi.org/10.5860/crl.80.6.843

Kreiter, A. 2019 'The Hungarian archaeology database in the light of ARIADNE' in J. Richards and F. Niccolucci (eds) *The ARIADNE Impact*, Budapest: Archaeolingua. 63-68. https://doi.org/10.5281/zenodo.3476711 [Accessed December 14, 2024].

Kreutzer, T. 2014 *Open Content: a practical guide to using Creative Commons Licences*, Wikimedia Deutschland e. V., German Commission for UNESCO, & North Rhine-Westphalian Library Service Centre (hbz), Bonn: German Comm. for UNESCO. https://commons.wikimedia.org/wiki/File:Open_Content_-_A_Practical_Guide_to_Using_Creative_Commons_Licences.pdf [Accessed December 10, 2024].

Lagoze, C., Van de Sompel, H., Nelson, M., & Warner, S. 2002 *Open Archives Initiative - Protocol for Metadata Harvesting - v.2.0*. https://www.openarchives.org/OAI/openarchivesprotocol.html [Accessed December 18, 2024.

L'Hours, H., Kleemola, M. and De Leeuw, L. 2019 'CoreTrustSeal: From academic collaboration to sustainable services', *IASSIST Quarterly* **43**(1), 1-17. https://doi.org/10.29173/iq936

Lin, D., Crabtree, J., Dillo, I. *et al*. 2020 'The TRUST Principles for digital repositories', *Scientific Data* **7**(1), 144. https://doi.org/10.1038/s41597-020-0486-7

Margoni, T. and Tsiavos, P. 2018 *Toolkit for Researchers on Legal Issues*. https://doi.org/10.5281/zenodo.2574618

May, K., Binding, C. and Tudhope, D. 2015 'Barriers and opportunities for Linked Open Data use in archaeology and cultural heritage', *Archäologische Informationen* **38**, 173-184. https://doi.org/10.11588/ai.2015.1.26162

Mons, B., Neylon, C., Velterop, J., Dumontier, M., Olavo Bonino da Silva Santos, L. and Wilkinson, M.D. 2017 'Cloudy, increasingly FAIR; revisiting the FAIR Data guiding principles for the European Open Science Cloud', *Information Services & Use* **37**(1), 49-56. https://doi.org/10.3233/ISU-170824

Morrissey, S.M. 2020 *Preserving Software: Motivations, Challenges and Approaches*, Digital Preservation Coalition. https://doi.org/10.7207/twgn20-02

nestor Working Group Trusted Repositories - Certification 2009 *Catalogue of Criteria for Trusted Digital Repositories*, 2nd edition. http://nbn-resolving.de/urn:nbn:de:0008-2010030806 [Accessed February 5, 2024].

Neuroth, H., Oßwald, A., Scheffel, R., Strathmann, S., & Huth, K. (eds) 2010 *nestor Handbuch. Eine kleine Enzyklopädie der digitalen Langzeitarchivierung 2.3*, nestor. https://nbn-resolving.de/urn:nbn:de:0008-2010071949 [Accessed November 29, 2023].

Nicholson, C., Kansa, S., Gupta, N., and Fernandez, R. 2023 'Will It Ever Be FAIR?: Making Archaeological Data Findable, Accessible, Interoperable, and Reusable', *Advances in Archaeological Practice* **11**(1), 63-75. https://doi.org/10.1017/aap.2022.40

Niven, K. 2011a 'Project documentation', *Guides to Good Practice*, University of York. https://archaeologydataservice.ac.uk/help-guidance/guides-to-good-practice/the-project-lifecycle/project-documentation/ [Accessed December 6, 2024].

Niven, K. 2011b 'Project metadata', *Guides to Good Practice*, University of York. https://archaeologydataservice.ac.uk/help-guidance/guides-to-good-practice/the-project-lifecycle/project-metadata/ [Accessed December 6, 2024].

Oniszczuk, A., Tsang, C., Brown, D.H., Novák, D. and de LANGHE, K. 2021 *Guidance on selection in archaeological archiving*, Namur: EAC. https://doi.org/10.5281/zenodo.10671359

Online Computer Library Center, Inc. (OCLC) and Center for Research Libraries (CRL) 2007 *Trustworthy Repositories Audit & Certification: Criteria and Checklist*. https://web.archive.org/web/20250123181052/http://www.crl.edu/sites/default/files/d6/attachments/pages/trac_0.pdf [Accessed February 5, 2024].

Ostendorff, S. 2024 'The Nebra Sky Disk and the liberation of cultural goods in Europe', *Europeana PRO*. https://pro.europeana.eu/post/the-nebra-sky-disk-and-the-liberation-of-cultural-goods-in-europe [Accessed December 12, 2024].

Perrin, K., Brown, D.H., Lange, G. *et al*. 2014 *A standard and guide to best practice for archaeological archiving in Europe*, Namur: EAC. https://doi.org/10.5281/zenodo.10664080

Péter, R. 2023 'The ARIADNEPlus Project has Come to an End;, *Hungarian Archaeology* **12**(1), 74-76. https://files.archaeolingua.hu/2023TA/Upload/Peter_E23TA.pdf [Accessed December 14, 2024].

Research Libraries Group 2002 *Trusted Digital Repositories: Attributes and Responsibilities. An RLG-OCLC Report*. https://www.oclc.org/content/dam/research/activities/trustedrep/repositories.pdf [Accessed January 30, 2024].

Richards, J.D. 2023 'Joined up Thinking: Aggregating archaeological datasets at an international scale', *Internet Archaeology* **64**. https://doi.org/10.11141/ia.64.3

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

Rivers Cofield, S., Childs, S.T. and Majewski, T. 2024 'A Survey of How Archaeological Repositories Are Managing Digital Associated Records and Data: A Byte of the Reality Sandwich', *Advances in Archaeological Practice* **12**(1), 20-33. https://doi.org/10.1017/aap.2023.29

RLG-NARA Task Force on Digital Repository Certification 2005 *An audit checklist for the certification of trusted digital repositories*. https://web.archive.org/web/20051126181100/http://www.rlg.org/en/pdfs/rlgnara-repositorieschecklist.pdf [Accessed February 5, 2024].

Sabliński, J. and Trujillo, A. 2021 'Piql. Long-term preservation technology study', *Archeion* **122**, 13-32. https://doi.org/10.4467/26581264ARC.21.011.14491

Schleußinger, M. and Rex, J. 2019 'Forschungsdaten veröffentlichen?' [poster], Zenodo. https://doi.org/10.5281/zenodo.3368292

Schmidle, W. 2021 'ChronOntology, a gazetteer for temporal terms (Abstract)'', *Data for History 2021: Modelling Time, Places, Agents, Berlin 19 May-30 June 2021*. https://d4h2020.sciencesconf.org/data/pages/Schmidle_ChronOntology_1.pdf [Accessed December 10, 2024].

Schmidt, S.C., Thiery, F. and Trognitz, M. 2022 'Practices of Linked Open Data in Archaeology and Their Realisation in Wikidata', *Digital* **2**(3), 333–364. https://doi.org/10.3390/digital2030019

Science Europe 2021 *Practical Guide to the International Alignment of Research Data Management - Extended Edition*. https://doi.org/10.5281/zenodo.4915861

Simon, R., Isaksen, L., Barker, E. and de Soto Cañamares, P. 2016 'The Pleiades Gazetteer and the Pelagios Project' in M. L. Berman, R. Mostern, & H. Southall (eds) *Placing names: enriching and integrating gazetteers*, The Spatial Humanities series. 97-109. Bloomington: Indiana University Press. https://doi.org/10.2307/j.ctt2005zq7.12

Stall, S., Martone, M.E., Chandramouliswaran, I. *et al*. 2023 *Generalist Repository Comparison Chart*. https://doi.org/10.5281/zenodo.7946938

the FAIRsharing Community 2019 'FAIRsharing as a community approach to standards, repositories and policies', *Nature Biotechnology* **37**(4), 358-367. https://doi.org/10.1038/s41587-019-0080-8

Trognitz, M. 2022 *Grundlagen des Datenmanagements*, ACDH-CH Howto. https://howto.acdh.oeaw.ac.at/de/resource/posts/grundlagen-datenmanagement [Accessed February 28, 2024].

UNESCO 2009 *Charter on the Preservation of the Digital Heritage*. https://unesdoc.unesco.org/ark:/48223/pf0000179529 [Accessed December 10, 2024].

United Nations 2015 *Transforming our world: the 2030 Agenda for Sustainable Development*. https://undocs.org/en/A/RES/70/1

Van Garderen, P. 2006 *Does my digital archives need a digital repository system?* https://vangarderen.net/posts/does-my-digital-archives-need-a-digital-repository.html [Accessed December 17, 2024].

Vrandečić, D. and Krötzsch, M. 2014 'Wikidata: a free collaborative knowledgebase', *Communications of the ACM* **57**(10), 78-85. https://doi.org/10.1145/2629489

Whyte, A. 2014 *Five steps to decide what data to keep: a checklist for appraising research data v.1*. https://www.dcc.ac.uk/guidance/how-guides/five-steps-decide-what-data-keep [Accessed February 28, 2024].

Whyte, A. 2015 *Where to keep research data: DCC checklist for evaluating data repositories*. https://www.dcc.ac.uk/resources/how-guides [Accessed September 17, 2023].

Whyte, A. and Wilson, A. 2010 *How to Appraise and Select Research Data for Curation*, DCC How-to Guides. https://www.dcc.ac.uk/guidance/how-guides/appraise-select-data [Accessed February 27, 2024].

Wilkinson, M.D., Dumontier, M., Aalbersberg, I.J., *et al*. 2016 'The FAIR Guiding Principles for scientific data management and stewardship', *Scientific Data* **3**(1), 160018. https://doi.org/10.1038/sdata.2016.18

WIPO 1979 *Berne Convention for the Protection of Literary and Artistic Works*. https://www.wipo.int/treaties/en/ip/berne/index.html [Accessed December 12, 2024].

WIPO 2016 *Understanding Copyright and Related Rights*, Geneva, Switzerland: World Intellectual Property Organization (WIPO). https://doi.org/10.34667/tind.28946

Wise, A. and Miller, P. 1997 'Why metadata matters in archaeology', *Internet Archaeology* **2**. https://doi.org/10.11141/ia.2.5

Yakel, E., Faniel, I.M., Kriesberg, A., and Yoon, A. 2013 'Trust in Digital Repositories', *International Journal of Digital Curation* **8**(1), 143–156. https://doi.org/10.2218/ijdc.v8i1.251

Zhang, A.B. and Gourley, D. 2009 *Creating digital collections: a practical guide*, Oxford: Chandos Publishing. https://doi.org/10.1533/9781780631387

Żółtak, M., Trognitz, M. and Ďurčo, M. 2022 'ARCHE Suite: A Flexible Approach to Repository Metadata Management' in M. Monachini and M. Eskevich (eds) *Selected Papers from the CLARIN Annual Conference 2021*, Linköping Electronic Conference Proceedings **189**. 190-199. https://doi.org/10.3384/ecp18917