# Digital Chronological Data Reuse in Archaeology: three case studies with varying purposes and perspectives

Bryony Moody[1], Tom Dye[2], Keith May[3], Holly Wright[4], Caitlin Buck[1]
[1]School of Mathematics and Statistics, University of Sheffield,
[2] Department of Anthropology, University of Hawaii,
[3] Historic England
[4] Archaeology Data Service, York

August 2021

**Abstract**

A wealth of digital data are produced during an archaeological excavation and because so much of the fieldwork is unrepeatable, once the site is fully excavated, the digital records must be archived in a manner that best facilitates reuse. This paper presents three case studies of users wishing to reuse digital archaeological data from online repositories, with a specific focus on absolute and relative dating evidence. We discuss the problems encountered and how they reflect the wider issues of the reuse of digital archaeological data. Additionally, we provide recommendations specific to chronological data that seek to address the problems encountered.

***Keywords***— reuse, data preservation, chronological modelling, digital repositories

## 1 Introduction

Archaeological excavation typically produces a wealth of data. Increasingly, much of this data are born-digital or ends up in digital format by the time it is archived with a repository. Whilst digital archives are both faster and easier to access than physical ones, the digitisation of excavation data presents its problems, particularly if one seeks to reuse such data. Of course, physical archives present their own issues with the reuse of the materials held within them. However, for the case studies presented within this paper, only digital archaeological data were required for reuse and so we omit any discussion of physical archives from this point.

Ensuring that digital data are reusable in any discipline is important; it allows for the reproducibility or reanalysis of results and data respectively. This is arguably of utmost importance in archaeology. For the most part, archaeological excavation is destruction (Wheeler 1956). The process of gathering data destroys the source of that data permanently. Therefore, these data must be preserved and accessible for future analyses. Even if the source of the data are not destroyed, such as samples used for scientific dating, this source is finite and the process of obtaining the data cannot be repeated indefinitely. The adoption of FAIR principles (Wilkinson et al. 2016) goes some way to addressing these requirements but in this paper, we will examine how Findability, Accessibility and Interoperability do not in themselves automatically enable reuse.

Further, a lack of reusable data in archaeology can prevent advancements in software and methodologies, since well understood existing data are needed for testing and showcasing innovations.

In this paper, we will discuss common reasons for the reuse of archaeological data and the value that this adds to excavation and post-excavation analyses. We outline some circumstances that prevent the reuse of archaeological data. This is supplemented by specific case studies in which the user wished to reuse archaeological data but was hindered. Finally, we discuss how the specific problems encountered in the case studies exemplify issues with data reuse in archaeology in general, the implications of this for archaeological research, and the current research that is being undertaken to improve the situation.

## 1.1 Digital archiving of archaeological data

Computing power has significantly increased in recent decades, resulting in the ability to produce, manipulate and store digital data during the excavation itself and in post-excavation research. The data may be archived within private or public archives, or kept by the excavating organisation and never archived. Moreover, the timeline of the deposition of digital data is determined by the excavating organisation, meaning that they may deposit data in stages, or wait until the end of a project.

The Archaeology Data Service (ADS) is a public digital repository for heritage data, which has made significant progress in ensuring digital archaeological data are archived in such a way that ensures the longevity and access of the files (Richards 2008). Since its formation in 1996, the ADS has provided the leading public digital archive for archaeological data in Great Britain. They accept a wide range of archaeological data, including but not limited to: databases, computer-generated images, photographs, audio files and reports in PDF format (*Guidelines for Depositors* July 2020).

The ADS furthered their work with the public launch of the OASIS project in 2004. The OASIS (Online AccesS to the Index of archaeological investigationS) project is managed by the ADS in partnership with public heritage bodies in England, Scotland, (and to lesser degrees Northern Ireland and Wales) and provides a central digital repository for archaeological grey literature.

Despite making significant progress in the preservation of digital data, the ADS Director and colleagues acknowledge that more work is needed to ensure such data are reusable. The ADS director, Richards (2017), describes how digital fieldwork archives seldom have all data that are required for further analysis or reuse. The minimum data required to allow for reuse depends on the type of data itself. In section 5.1 we discuss specifically the data required for relative and absolute dating evidence to be reusable. What is clear, is that defined standards for all types of data that can be produced during excavation are required to ensure reuse is possible. These standards should take into account all stages of the data cycle, as specified in Yakel et al. (2019), from collection and recording to management and reuse. Yakel et al. (2019) also demonstrated that when reuse is considered, this positively influences other stages of the data cycle too. Richards (2017) recognises that cost is a key factor in preventing the deposition of digital data with the ADS. However, he argues that the deposition of a comprehensive digital archive upon completion of any project should be part of the standard workflow.

## 1.2 Data management standards in Archaeology

Previous research exploring issues with the reuse of digital data in archaeology, such as Faniel et al. (2013) and Huggett (2018), agree that although standards have been established for storing and preserving digital data, reuse issues still need to be addressed. An overriding issue is the lack of standardisation in how archaeological data are recorded and managed digitally. Archaeology in the UK is carried out in both the research and commercial sector, by at least ten major archaeological contracting organisations, a multitude of smaller operators, along with many individual archaeological consultants and specialists. This leads to considerable variation in practice during excavation, post-excavation, analysis and publication and, hence, to a wide range of methods and mechanisms for producing and managing digital archaeological data. Further, variation in data outputs is exacerbated by variations in the many different types of archaeology encountered (e.g. multi-periods, scale of excavation, complexity of site stratigraphy, investigation methods, project type, resources available). Consequently, although the ADS provides a central repository for depositing archaeological data, the data are not always in an optimal format for reuse, if deposited at all.

There are several different stages, from the start of the excavation to the final excavation report, during which data may be recorded, managed or stored in a way that facilitates or hinders reuse. There is a lack of formal training in database creation and management provided to most archaeologists (Faniel et al. 2018). Furthermore, there is little incentive for archaeologists to improve this since training can be costly and time-consuming with no immediate benefit to those working on any given excavation. However, without sufficient training, it is difficult for those who are collecting and recording data to see how their decisions impact the other stages in the life cycle of data.

Faniel et al. (2018) discuss the data management practices at two excavations in Europe. They find that the aforementioned issues were present in their case studies, highlighting that a high turnover of staff contributes to the lack of incentive to train staff thoroughly in database management. They also note that humans do not "think like a database", as a result, it is difficult for them to understand how data gathering practices can cause problems with data reuse. For example, humans can readily identify that the two identifiers "A203" and "a203" might relate to the same thing. However, a digital search algorithm may interpret these as two unique identifiers. Even if a human is interpreting the data, if they are not familiar with the original excavation, they may not be confident that the identifiers represent the same thing. It is typically subtle issues like this that accumulate and (taken together) result in data that are not suitable for reuse without significant effort.

Guidelines, such as Historic England's recording manual, seek to ensure data collection practices are uniform within excavations conducted by their staff. However, this does not guarantee standardisation between excavations by different organisations; it is the responsibility of those carrying out the excavations to decide how they wish to record and manage their digital data. Moreover, even if data are collected and managed well on-site, management and manipulation of data during post-excavation analysis can also prevent reuse if, at the very least, a record of precisely how the data were used is not archived. Experts often clean, tidy or process data in non-replicable, ad-hoc ways to help answer specific research questions. As a result, the methods are hard to describe fully and results produced cannot readily be documented, revised or augmented, should additional data or methods of analyses and/or data become available. In summary, formal workflow documentation protocols are typically limited or non-existent and are certainly not routinely archived with the data or reports.

Ideally, the raw digital data (as collected on-site) would always be archived within a digital repository, along with work-flow summaries, models, algorithms, resulting outputs and notes for future users linking archived materials to any formal reports and/or grey literature. Although, this would preferably be deposited with an accredited repository such as the ADS, digital archives provided by individual organisations such as the Çatalhöyük online digital archive (Grossner et al. 2014) can provide valuable sources of data, so long as they are archived in a manner suitable for reuse. We discuss using the Çatalhöyük online digital archive in the case study in Section 4.

Finally, even if digital data have been managed well during both the excavation and post-excavation analysis, they need to be deposited in a non-proprietary format, such as plain text files. If data are stored in a proprietary format, this can prevent reuse, particularly if the software required to access the data incurs a cost. The ADS open standards (*Guidelines for Depositors* July 2020) seek to prevent this by allowing only specific file formats to be deposited. However, this is not guaranteed if data are kept with a private organisation, even if they choose to make the data publicly available. Additionally, if data are kept only by the individual organisations who generate them, or by an individual specialist, it limits reuse. These issues prevent reuse, particularly for early-career researchers who may not have the necessary personal contracts or funds to locate the data they require.

## 1.3 Value in the reuse of relative and absolute dating evidence

The authors' present interesting in reusing absolute and relative dating evidence in archaeology derives from our interest in improving the understanding of the chronology of an archaeological site or landscape using Bayesian inference. Utilising Bayesian methods in chronology construction allows for the combination of relative and absolute dating evidence in a rigorous statistical framework to provide improved estimates of chronologies. This theory is well established and widely utilised within the archaeology community (Buck et al. 1996, Bayliss 2009).

Multiple scenarios may arise in which it is useful to revisit archived chronological data. For example, an increase in computing power can enable more complex analyses to be carried out, with revised or new models provided. One might also wish to revisit data from an excavation if further scientific dating occurs, e.g. after new excavations have been undertaken or when research focus changes in a follow-up phase of excavation. Any new dating augments the original data and allows for a revised chronology to be provided, such as in Marciniak et al. (2015), where a new dating program changed the understanding of the Late Neolithic community of Çatalhöyük.

Given the uncertainty associated with modelling the chronology of an archaeological site, experts may not agree on the best model to use. Therefore, it is important to have data, associated models and methods available in public archives in case there is a need to revise or revisit previous work, see Discamps et al. (2015), Higham & Heep (2019).

A further example is Bayliss et al. (2014) who remodelled the chronology of Buildings 1 and 5 from the north area of Çatalhöyük, originally explored by Cessford et al. (2005). The Cessford and Bayliss models are substantially different leading to quite different archaeological conclusions. Dye & Buck (2015) used this scenario to illustrate why it is beneficial to consider multiple chronological models for any given archaeological project. They propose using mathematical graphs to semi-automate the construction of such models. To develop the concepts discussed in Dye & Buck (2015), a PhD project was funded in 2018 and Moody appointed as the postgraduate researcher. This project seeks to provide prototype software that will improve the management and modelling of relative and absolute dating evidence for Bayesian chronology construction.

A consistent standard for the management and archiving of absolute and relative dating evidence is not currently well established. As outlined in Figure 1, there are multiple stages where data might be collected, augmented or used as model input; it is the responsibility of those excavating to decide what and when data are deposited. An additional problem presents itself since chronological models of archaeological sites are often built by hand. In such situations, data move from digital to paper form, and so do not appear in the digital archive.

In what follows, we illustrate the problems that routinely arise in the reuse of digital archaeological data with three case studies where the aim was to reuse absolute or relative dating evidence. For the first case study, the goal was to locate within the ADS all reusable relative and absolute dating evidence. For the second, the goal was to obtain specific data that was reported to have been archived with the ADS. The final case study discusses the use of a digital repository provided by a higher education institution, as opposed to the ADS.

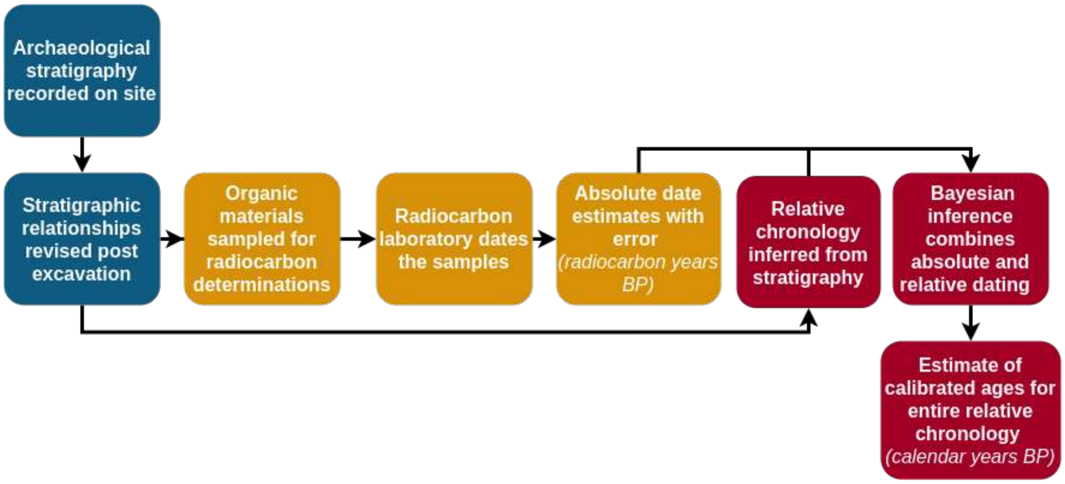
Archaeological stratigraphy recorded on site
Stratigraphic relationships revised post excavation
Organic materials sampled for radiocarbon determinations
Radiocarbon laboratory dates the samples
Absolute date estimates with error (radiocarbon years BP)
Relative chronology inferred from stratigraphy
Bayesian inference combines absolute and relative dating
Estimate of calibrated ages for entire relative chronology (calendar years BP)

Figure 1: Diagram of the end-to-end process from the collection of absolute and relative dating evidence to completion of Bayesian chronology construction. The blue and yellow boxes represent the process of obtaining relative and absolute dating evidence, respectively. The red boxes outline the process of Bayesian inference.

## 2 Case Study 1: seeking dating evidence

The primary objective for the project described in this case study was to seek all absolute and relative dating evidence within the ADS archive and OASIS online repository that might be used to test prototype software being built as part of Moody's PhD research. Specifically, we sought stratigraphy and radiocarbon dates as examples of relative and absolute dating evidence, respectively. A secondary aim was to review the potential for reuse of digital archaeological data stored within the ADS. It would not have been practical to review the potential for reuse of all types of archaeological data within the ADS; indeed, that would have been a PhD project in itself. Instead, the review provides a snapshot of the quality and quantity of data held by the ADS that might be used or reused in chronology construction.

To be potentially useful for Moody's PhD the absolute and relative dating evidence needed (as a minimum) to consist of:

1. A table containing mutually consistent pairwise statements of the stratigraphic relationships between contexts (stratigraphic units) as they were observed in the field.
2. A table containing information about samples (or objects/finds) taken from specified contexts on the site, which might be suitable for scientific dating or that have already been dated.
3. A table containing phases and the contexts attributed to them. The relationships between any phases, declared using some form of specified descriptors such as Allen operators (Allen 1983, May 2020), was also desirable.

To facilitate our exploration of the ADS Archive and OASIS repositories, Tim Evans and Jenny O'Brien of the ADS first provided us with 4,820 digital files in varying formats deposited in the ADS archive. These files all had metadata containing at least of one the key-phrases (text strings or combination of text string) in Table 1 suggesting that absolute or relative dating evidence was present within them. Secondly, they provided 37,320 files (that had been deposited via OASIS) in Portable Document Format `PDF` containing unpublished reports of excavations, both in the research and commercial sector. For the remainder of this paper, these two collections will be referred to as files within the ADS archive and OASIS files respectively.

The 4,820 files within the ADS archive were manually searched by the present authors to see if they held the required data and if they were in a format that would facilitate ready and rapid reuse. We also converted the 37,320 OASIS files from `PDF` to plain text format and carried about a computational search of the content within the documents (using a technique know as `Bash` scripting, see GNU (2007), within the `unix` computer operating system) for key-phrases, to identify the documents most likely to contain the data we required. See Table 2 for the key-phrases used.

| **Type of data sought** | **Key-phrase (combination of text strings)** |
|---|---|
| Stratigraphy | *matr* OR *c14* OR *context* OR *phas* |
| Carbon-14 dates | *radiocarbon* OR *c-14* |
| Phasing | *phasing* OR *phase* |
| Scientific dates | *dating* OR *date* |

Table 1: A list of the key-phrases used to search the metadata of files deposited in the ADS archives to identify files that might contain the absolute and relative dating evidence needed Moody's PhD research.

| **Key-phrase label** | **Text string/combination of text strings** |
|---|---|
| **Stratigraphy** | *stratigraph* |
| **Phasing** | *phase OR phasing* |
| **Stratigraphic diagrams** | *harris matrix* OR *stratigraphic matrix*<br>EXCLUDING<br>*watching brief* OR *matrix will* |
| **Carbon-14 dates** | *carbon date* OR *carbon dating* OR *radiocarbon*<br>AND<br>*Cal BC* OR *AD* OR ± |

Table 2: Table defining key-phrases and the text string/combination of text strings they consist of. Each key-phrase has a label, used to summarise the kind of evidence being sought.

Using these searches, we found very little in the way of data that would be useful for Moody's ongoing PhD project. However, we gained a qualitative snapshot of the suitability (or otherwise) of the ADS and OASIS materials for use in modern chronological modelling projects.

Of the 4,820 files from the ADS Archive whose metadata suggested that they might contain the chronological data we were seeking, about half contained data with potential for reuse, however, 87% of these contained data that required additional effort before it could be reused. Specifically, we encountered many radiocarbon dating certificates in `PDF` format. Despite containing important absolute dating evidence which can be displayed by a computer, the radiocarbon data within the `PDF` cannot be identified without a human extracting the key information first and (manually or with purpose written code) re-tabulating in another format.

Of all the 4,820 ADS archive files, around 25% were in plain text format and thus were directly interpretable by a computer. However, a file being archived in plain text format did not guarantee reusable data. We found 169 files that were either empty plain text files or plain texts files with table headings suggesting that stratigraphic data were intended to be added, but never had been.

Of the 37,320 OASIS files, only 102 contained all the key-phrases we were seeking, outlined in Table 2. Most of the 102 files did contain some form of data useful for chronological modelling. However, recall that all OASIS files are in PDF format, so a human must first interpret the data from the PDF before it can be reused. In addition, the relative dating evidence encountered in the OASIS files (i.e. matrix diagrams), unfortunately, did not represent stratigraphy that was complex enough for Moody's research.

Some files highlighted by the key-phrase search in Table 2 were false positives. These were excavation reports containing the phrase "stratigraphic archive" to describe the location where stratigraphic data are kept. However, no location of this "stratigraphic archive" was identified. Upon investigation, we found that some of the companies who deposited the excavation reports containing the phrase "stratigraphic archive" did not deposit such data with the ADS, but have internal record systems that are not publicly accessible.

On other occasions, missing data seemed to arise from multi-stage projects. According to a computational search of the OASIS files, 8% of the total files contained the phrase "assessment report". Assessment reports are produced after excavation, but before post-excavation analysis. Given that 61% of the files containing the phrase "stratigraphic archive" also contained the phrase "assessment report" these files may be from an early excavation, with more data expected later. However, since the ADS do not require depositors to define at what stage in a project a file has been deposited, we were unable to determine if this is the case.

# 3 Case Study 2: phyletic seriation of beads

We now consider a case study in which specific data were sought. As part of a project to investigate the potential contribution of Bayesian calibration to phyletic seriation, Dye and Buck (with Robert DiNapoli) sought to remodel an innovative dating program, based on an occurrence seriation of Anglo-Saxon graves (Hines & Bayliss 2013). The dating program was attractive for phyletic seriation because the artefacts in each grave were all buried at the same time, the dating materials were selected, dated, and modelled by a recognized expert in the field, and chronologically sensitive artefact types that looked to be good candidates for phyletic seriation were recovered, including seaxes in the male graves and beads in the female graves. The graves were not stratigraphically related to one another, so the phyletic seriation was based solely on the age determinations and the artefact associations.

The Anglo-Saxon grave data were deposited in the ADS archive. They include a clean and well-organized relational database of the finds with 20 tables containing information on the grave, the individual buried in the grave, the artefacts recovered from the grave, and the radiocarbon dating evidence. The database is augmented with full metadata for each table and an entity-relationship diagram illustrating the relational structure. Given these well-organised materials, it was a trivial matter to create a local SQLite database that could be queried reliably. These queries eventually revealed that the database does not include information on two early bead types reported by the original authors, BE1-ConSeg and BE1-ConCyl. However, the database did ably support queries designed to model the chronology of amber beads, which were not modelled by the original authors.

The original authors developed and reported separate chronological models in OxCal for the male and female graves. Nevertheless, the ADS archive contained a single OxCal file that modelled the chronology of the female graves; an OxCal file that models the chronology of the male graves is not included in the archive. The initial attempt to calibrate the OxCal model for the female graves failed because the bespoke calibration curve developed for and used by the project was not deposited in the ADS archive. Fortunately, Buck recalled a discussion about the bespoke curve on the OxCal Google Group in 2010; by an exceptional stroke of good luck, a digital version of the bespoke curve was attached to one of the messages in the discussion. When the bespoke curve was downloaded and installed locally, the OxCal model for the female graves calibrated successfully. Nevertheless, the local calibration results differed from the results reported by the project. A close comparison of the OxCal code deposited in the ADS repository with the published description of the OxCal model used by the project indicated differences in the models. The published model uses the OxCal Phase() and Boundary() commands to estimate the start and end dates of bead currencies, while the chronological model deposited in the ADS repository uses the OxCal First() and Last() commands to estimate the start and end dates of bead currencies. The investigators recognized that the use of Phase() and Boundary() commands introduced unrealistic assumptions to the chronological model (e.g., Bayliss et al. 2013, 359); replacing them with First() and Last() commands alleviates the perceived problem and is thus an improvement to the chronological model. However, the revised OxCal file deposited in the ADS repository complicates efforts to replicate the published analysis.

In the end, the materials available in the ADS repository, with the addition of the bespoke calibration curve retrieved from the OxCal Google Group, proved sufficient to investigate the potential contribution of Bayesian calibration to the phyletic seriation of beads recovered from female Anglo-Saxon graves. Nevertheless, the plan to investigate the phyletic seriation of seaxes from the male graves could not be implemented because the OxCal file with the chronological model for the male graves was not found in the repository.

# 4 Case Study 3: obtaining a complex Harris matrix

For our final case study, we look at an example of using a digital repository aside from the ADS. As part of a project to generate a graphical representation of a chronology derived from a real-world Harris matrix, Dye & Buck (2015) chose an excavation report from excavations at Buildings 1 and 5 on the East Mound of the Çatalhöyük site. The data requirements, in this case, were precisely defined by the software designed for the project, which calls for an exhaustive list of excavated contexts, observations of context superposition, attributes of the contexts, and associations of contexts and dated materials, much like Case Study 1.

The Çatalhöyük online digital archive (Grossner et al. 2014) makes available an impressive array of project materials, including data from building, space, feature, and unit sheets, along with samples, "x-finds" (which are small finds and all artefacts associated with floors or features), and skeletons recorded by the excavator, and daily sketches. These materials can be discovered with a series of pre-defined "simple" and "complex" queries, where simple queries find records that match a value for any one of several pre-defined fields, and complex queries find records that match values for two or more pre-defined fields. Arbitrary queries are not supported. The data required to construct a Harris matrix are stored in the table(s) that contribute(s) to the "Unit sheet" category exposed by the digital archive. The unit sheet carries information on "Unit Stratigraphy" with lists of units observed to be above, below, and equal to it. Presumably, a query might be constructed to return all of the units assigned to Buildings 1 and 5, along with the units excavated immediately outside them, which would satisfy the requirement for an exhaustive list of excavated contexts and their attributes. Another query might return all of the observed instances of stratigraphic superposition. Nevertheless, this was not attempted because the archive will not return a digital copy of query results, but instead returns an error message that points to a configuration error in the website server.

Instead, a decision was taken to forego constructing a new Harris matrix and to start with the published matrix for Buildings 1 and 5, which might be augmented with information from the archive, as needed. The excavations at Buildings 1 and 5 were among the first carried out at Çatalhöyük since it was first excavated by Mellaart in the 1960s. The excavator of both buildings, Craig Cessford, kindly corresponded about his field experience and recording techniques and reports that he experimented with several Harris matrix software applications at the time including the Bonn Archaeological Software Package, in addition to drawing matrices in spreadsheet software (Microsoft Excel), without settling on software that might be adopted by other excavators involved with the multi-year project, once Cessford was no longer working with the data. Further, no record of the software used was deposited within the archive. When Alex Bayliss, who also kindly corresponded with us, joined the Çatalhöyük project and began to work out the chronology of Buildings 1 and 5 she found a binary file from an unknown software application and a hard copy of the matrix, which she converted to PDF format for publication as a digital file on a CD-ROM distributed with her report (Bayliss et al. 2014). The PDF representation of the matrix yielded an exhaustive list of excavated contexts and a chronologically-relevant subset of context superposition observations, which were captured manually and tracked by marking up the PDF in the vector graphics software, Inkscape. Associations of contexts and dated materials were taken from published tables. The last bits of the puzzle—attributes of the contexts, in particular, the building to which the context was assigned—were found in the Çatalhöyük online digital repository. Queries for units assigned to Buildings 1 and 5 were augmented by queries for individual units represented on the Harris matrix, but not assigned to one of the two buildings. Data were once again entered manually with reference to the lists produced by the queries.

The data captured this way proved sufficient to illustrate how a Bayesian chronological model might be generated from a complex real-world Harris matrix. Considerable effort was expended in manual data entry, most of which might have been avoided if the Çatalhöyük project data were more fully accessible. The Çatalhöyük online digital archive is a tremendous resource, but it is poorly designed for the task that faced Dye and Buck. The archive, as it stands, is unsuited to reproduce stratigraphic and chronological analyses carried out by the project, nor would it feasibly support alternative analyses of stratification and chronology. For these types of analyses, a repository built along the lines of the repository for the Anglo-Saxon graves would be welcome.

# 5 Discussion and Recommendations

The case studies examined in this paper illustrate various issues that prevent the reuse of chronological archaeological data. As seen in case study 1, there is evidence that stratigraphic data are not always deposited with the ADS and even when they are archived, case studies 1 and 3 reveal that additional work would be required before the data can be used for computational analyses.

An additional theme in all case studies was insufficient or incomplete data being archived from a project. A lack of documentation detailing how data may have been augmented from its original form, to its current form in the archive, limited reuse. This was evident in case study 2, where published results from the male Anglo-Saxon graves could not be obtained using archived models and data. The incomplete empty or near-empty files found as part of case study 1 offered an interesting juxtaposition to the discussion in Richards (2017), in which cost is cited as a key factor in preventing the deposition of files. Yet empty files are being deposited, at a real financial cost to the depositor.

Further, case study 1 demonstrated a lack of directly reusable complex stratigraphy within the ADS archives. Anecdotally, this might be because large and complex stratigraphic matrix diagrams tend to be managed on paper, due to the limitations of existing software. The prototype software being produced as part of Moody's PhD seeks to provide a solution to this problem. However, somewhat paradoxically, large digital stratigraphic data sets are needed to be able to develop and test the software development.

In all case studies presented, had the authors not known the appropriate people to contact, reuse would have been hindered further. The authors had time generously given to them by the staff of the ADS to obtain potential sources of data. This is not a process that is reasonable for everyone seeking multiple sources of a specific type of data. Similarly, Dye had to seek the data he required from research contacts when ideally this would have been included within site archives.

Public digital archives provided by private archaeological companies or academic institutions, such as the Çatalhöyük online digital archive, may provide a suitable alternative to depositing with the ADS, provided the data are in a reusable format. However, they require the user to be aware they exist, know where to find them, and that they will remain sustainable beyond the lifetime of any particular project.

The ADS are aware of the lack of standardisation in digital archaeological data, they are working to ensure that all data deposited within their repository satisfy the FAIR (Findable, Accessible, Interoperable and Reusable) principles (Hagstrom 2020). The FAIR principles are designed to ensure that digital data, in general, can be found, accessed and are interoperable and reusable. As such, there is a heavy focus on ensuring the metadata of files are of a high standard. This is incredibly useful for querying databases and extracting files that potentially contain a specific type of data. However, it does not guarantee that the data contained within the file are reusable. This was evident in case study 1, where it was easy to search the metadata for key-phrases, but these documents often did not contain directly reusable data. Huggett (2018) also raises concerns that the FAIR principles fall short of providing plausible solutions for improving the process of reusing archaeological data, highlighting that making data shareable and accessible does not ensure the data are reusable. Huggett (2018) does not seek to provide recommendations for improving the reuse of archaeological data. However, he does pose important questions, such as how the steps in the data reuse cycle shape archaeological practice and which data are being prioritised and which are being left behind.

For chronological modelling, within our research, we recommend data archiving be prioritised as follows, see Figure 2:

1. **Core data identifier**: context number.
2. **Essential data**: stratigraphic relationships, phase number, absolute (e.e.c̃oin or radiocarbon) date, including any unique laboratory identifier and laboratory error.
3. **Valuable data**: context type, physical relationships, supplementary information from a dating laboratory or specialist, additional phase information (date range or temporal operator), textual descriptor, archaeologists notes.

This hierarchy of importance is specific to chronological modelling, however, and certainly does not imply the importance of each type of data in general.

## 5.1 Recommendations

Though we do not seek to suggest deposition standards for all digital archaeological data, the subsequent subsections provide some further specific recommendations for absolute and relative dating evidence.

### 5.1.1 General data standards

All raw data, as well as models, algorithms, resulting outputs and notes used to obtain published results, should be archived within a digital repository. Data should be archived in non-proprietary formats, ideally plain text. Archiving data in a plain text format, such as comma-separated (`CSV`) files, is good practice for multiple reasons. First, it is an extremely widely used format, and so it is highly likely to have longevity. Second, almost all software packages can read and produce text files. Finally, such files are easy to view and edit, even with very modest computer skills.

Furthermore, training to improve the data literacy of those collecting and handling data is recommended, so that they might gain an understanding of why it is imperative that data are formatted and archived in such a way that facilitates reuse. For the remainder of this section, we recommend such appropriate formats.

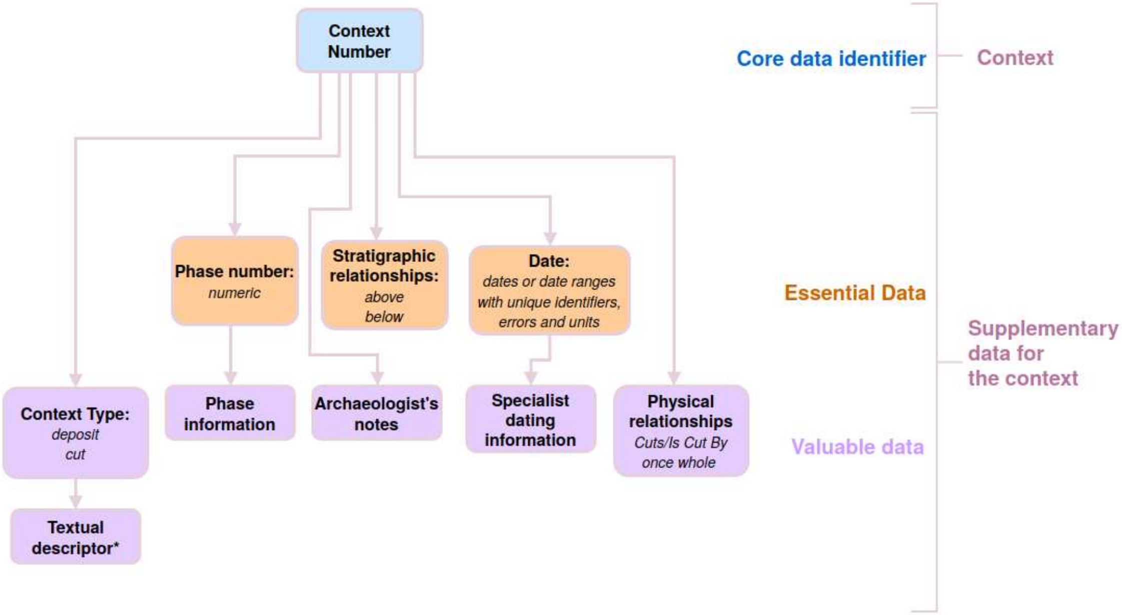
Context Number
Core data identifier
Context
Phase number: *numeric*
Stratigraphic relationships: *above below*
Date: *dates or date ranges with unique identifiers, errors and units*
Essential Data
Supplementary data for the context
Context Type: *deposit cut*
Phase information
Archaeologist's notes
Specialist dating information
Physical relationships *Cuts/Is Cut By once whole*
Valuable data
Textual descriptor*

Figure 2: Diagram showing the types of data we seek in order to carry out chronological modelling. The hierarchical structure indicates relative importance for reusability. *Textual descriptor should be a brief description of the context using controlled vocabularies such as those defined by Historic England (2021).

### 5.1.2 Relative dating evidence

At the very minimum, stratigraphic data should be archived as three columns in plain text format, two of which contain context numbers and the other an indication of the relationship between them (perhaps using keywords such as "once-whole", or "above" or "below" for contexts that are in superposition and which archaeologists commonly record in the Harris matrix as defined by Harris (1989)), even if the matrix itself is not deposited. Any phasing or dating information must be attributed to the appropriate context(s) or sample(s) via databases or relational documents.

We recognise that stratigraphic data are often far more complex than the aforementioned three columns of data. Any additional data only aids the process of reuse, so long as it is clear which context any additional data are attributed to, and a record is provided of why and how any relative dating evidence has been augmented from the original data collected on-site.

### 5.1.3 Absolute dating evidence

We also recommend archiving a plain text table containing information about the samples (or objects/finds) taken from specified contexts on the site that might be suitable for scientific dating or that have already been dated. The simplest digital form these might take would be:

- several columns detailing any relative chronology information, the details of which would vary by sample type and dating method, but would, as a minimum, contain a context number.
- Several further columns with a description of the sample (or find) and of any sub-sampling or pre-processing, along with any notes from specialist analysis, the scientific dates themselves, their unique identifiers and laboratory errors and units. For specific conventions for recording and reporting radiocarbon dates, see Millard (2014).
- For absolute, but not scientific, dates like those from coins or grave inscriptions, fewer items of data are likely to be needed, but a similar approach should be followed so that contexts, information sources etc. are archived

### 5.1.4 Work flow documentation

More consistency and documentation of the methods used for grouping, dating and phasing of stratigraphic data during analysis, along with appropriate archiving of that documentation, still requires wider agreement across the archaeological sector. This is something May and colleagues are trying to address as part of his AHRC leadership fellowship project (known as The Matrix project), in which he is seeking to ensure data in archaeological archives become more interoperable and re-usable.

We recognise that the cost of the recommendations discussed in this section may provide a barrier to implementation. In particular, our recommendation to provide training in data literacy. Other options, such as the training manuals provided by Historic England during their excavations (though costly to produce initially) can provide low cost training on how to collect data. However, this does not provide insight as to how decisions about data collection and handling can affect data use and reuse in the future.

# 6 Future Research and conclusion

Since 2018, two research projects (the Matrix Project led by May and Moody's PhD) have been funded by the Arts and Humanities Research Council (AHRC) and supported by Historic England. These projects aim to provide software and recommendations to improve the reuse of digital archaeological data. The authors hope that the open research nature of these projects will mean that, since all software produced will be open source and freely available to use, this will allow archaeologists to improve data management and re-usability of such data. Furthermore, the Saving European Archaeology from the Digital Dark age (SEADDA Cost Action 2021) working group are regularly holding meetings and workshops that seek to improve digital data reuse in archaeology. Although the literature and the aforementioned projects demonstrate that the archaeological community recognises there is a need to improve the potential for reuse of archaeological data, this problem will require reform of practices across the whole community, rather than a limited number of projects.

# 7 Acknowledgments

The authors would like to thank the staff of the ADS, in particular Tim Evans, Jenny O'Brien, Holly Wright and Julian Richards, for their generosity in giving their time and expertise to aid the research undertaken in this paper. As indicated in the text, we are also indebted to Alex Bayliss and Craig Cessford, whose advice on a previous project led to the richness of our case study 3

# 8 Funding Sources

This work was supported by the Arts and Humanities Research Council via funding under their grants (ref AH/S001328/1 and AH/T002093/1), without which this research could not have taken place.

# References

Allen, J. F. (1983), 'Maintaining knowledge about temporal intervals', *Communications of the ACM* **26**(11), 832–843.
**URL:** *https://doi.org/10.1145/182.358434*

Bayliss, A. (2009), 'Rolling out revolution: using radiocarbon dating in archaeology', *Radiocarbon* **51**, 123–47.
**URL:** *https://doi.org/10.1017/S0033822200033750*

Bayliss, A., Farid, S. & Higham, T. (2014), Time will tell: practising Bayesian chronological modelling on the east mound, *in* I. Hodder, ed., 'Çatalhöyök excavations: the 2000–2008 seasons', (CA): Cotsen Institute of Archaeology., Los Angeles, pp. 53–90.

Bayliss, A., Hines, J. & Nielsen, K. H. (2013), *Interpretative chronologies for the female graves*, Vol. 33 of Hines & Bayliss (2013), chapter 7, pp. 339–458.
**URL:** *https://doi.org/10.5284/1018290*

Buck, C. E., Cavanagh, W. G. & Litton, C. D. (1996), *Bayesian Approach to Interpreting Archaeological Data*, Wiley, Chichester, England, chapter 9, pp. 201–252.

Cessford, C., Blumbach, M., Akoğlu, H. G., Higham, T., Kuniholm, P. I., Manning, S. W., Newton, M. W., Ozbakan, M. & Ozer, A. M. (2005), Absolute dating at Çatalhöyök, *in* 'Inhabiting Çatalhöyük: Reports from the 1995–1999 Seasons', Vol. 4, McDonald Institute for Archaeological Research and British Institute of Archaeology at Ankara, Cambridge and London, pp. 65–99.

Discamps, E., Gravina, B. & Teyssandier, N. (2015), 'In the eye of the beholder: contextual issues for Bayesian modelling at the Middle-to-Upper Palaeolithic transition', *World Archaeology* **47**(4), 601–621.
**URL:** *https://doi.org/10.1080/00438243.2015.1065759*

Dye, T. S. & Buck, C. E. (2015), 'Archaeological sequence diagrams and Bayesian chronological models', *Journal of Archaeological Science* **63**, 84–93.
**URL:** *https://doi.org/10.1016/j.jas.2015.08.008*

Faniel, I., Kansa, E., Whitcher Kansa, S., Barrera-Gomez, J. & Yakel, E. (2013), The challenges of digging data: A study of context in archaeological data reuse, *in* 'Proceedings of the 13th ACM/IEEE-CS Joint Conference on Digital Libraries', JCDL '13, Association for Computing Machinery, New York, NY, USA, pp. 295–304.
**URL:** *https://doi.org/10.1145/2467696.2467712*

Faniel, I. M., Austin, A., Kansa, E., Kansa, S. W., France, P., Jacobs, J., Boytner, R. & Yakel, E. (2018), 'Beyond the archive: Bridging data creation and reuse in archaeology', *Advances in Archaeological Practice* **6**(2), 105–116.
**URL:** *https:/doi.org/10.1017/aap.2018.2*

GNU, P. (2007), 'Free software foundation. bash (3.2. 48)[unix shell program]'.

Grossner, K., Hodder, I., Meeks, E., Engel, C. & Mickel, A. (2014), A living archive for Çatalhöyök, *in* 'Computer Applications in Archaeology Proceedings'.

*Guidelines for Depositors* (July 2020).
**URL:** *https://archaeologydataservice.ac.uk/advice/guidelinesForDepositors.xhtml*

Hagstrom, S. (2020), 'The fair data principles'.
**URL:** *https://www.force11.org/group/fairgroup/fairprinciples*

Harris, E. C. (1989), *Principles of Archaeological Stratigraphy*, 2nd edn, Academic Press, London.
**URL:** *https://doi.org/10.1016/B978-0-12-326651-4.50002-X*

Higham, T. F. G. & Heep, G. S. (2019), 'Reply to: 'In the eye of the beholder: contextual issues for Bayesian modelling at the Middle-to-Upper Palaeolithic transition', by Discamps, Gravina and Teyssandier (2015)', *World Archaeology* **51**(1), 126–133.
**URL:** *https://doi.org/10.1080/00438243.2017.1329026*

Hines, J. & Bayliss, A., eds (2013), *Anglo-Saxon Graves and Grave Goods of the 6th and 7th Centuries AD: A Chronological Framework*, number 33 *in* 'Monograph', Society for Medieval Archaeology, London.

Historic England (2021), 'Fish vocabularies'.
**URL:** *http://www.heritage-standards.org.uk/fish-vocabularies/*

Huggett, J. (2018), 'Reuse remix recycle', *Advances in Archaeological Practice* **6**(2), 93–104.
**URL:** *https://doi.org/10.1017/aap.2018.1*

Marciniak, A., Barański, M., Bayliss, A., Czerniak, L., Goslar, T., Southon, J. & Taylor, R. (2015), 'Fragmenting times: Interpreting a Bayesian chronology for the Late Neolithic occupation of Çatalhöyök East, Turkey', *Antiquity* **89**, 154–176.
**URL:** *https://doi.org/10.15184/aqy.2014.33*

May, K. (2020), 'The matrix: Connecting time and space in archaeological stratigraphic records and archives', *Internet Archaeology* **55**.
**URL:** *https://doi.org/10.11141/ia.55.8*

Millard, A. R. (2014), 'Conventions for reporting radiocarbon determinations', *Radiocarbon* **56**(2), 555–559.
**URL:** *https://doi.org/10.2458/56.17455*

Richards, J. (2008), Managing digital preservation and access: The Archaeology Data Service, *in* F. McManamon, A. Stout & J. Barnes , eds, 'Managing Archaeological Resources', One World

Archaeology, Left Coast Press, pp. 173–94.
**URL:** *https:/doi.org/10.1177/1461957027616923 47*

Richards, J. D. (2017), 'Twenty years preserving data: A view from the United Kingdom', *Advances in Archaeological Practice* **5**(3), 227–237.
**URL:** *https://doi.org/10.1017/aap.2017.11*

SEADDA Cost Action (2021), 'Webpage'. Accessed: 2021-03-26.
**URL:** *https://www.seadda.eu*

Wheeler, M. (1956), *Archaeology from the Earth*, Pelican books A356, Penguin Books.

Wilkinson, M., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., Bonino da Silva Santos, L. O., Bourne, P., Bouwman, J., Brookes, A., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C., Finkers, R. & Mons, B. (2016), 'The fair guiding principles for scientific data management and stewardship', *Scientific Data* **3**.
**URL:** *https://doi.org/10.1038/sdata.2016.18*

Yakel, E., Faniel, I. M. & Maiorana, Z. J. (2019), 'Virtuous and vicious circles in the data life-cycle', *Information Research* **24**.