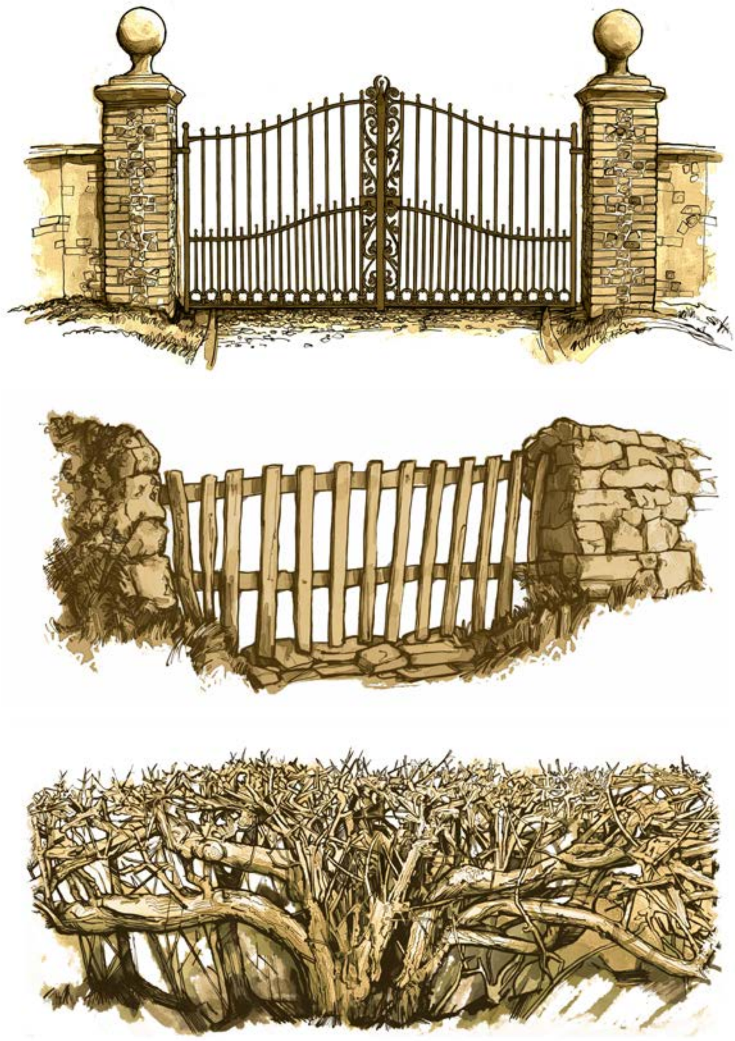
Barriers to the Production of Metadata for Archaeological Datasets

Claire Tsang

Discovery, Innovation and Science in the Historic Environment



Research Report Series no. 245/2020

Barriers to the Production of Metadata for Archaeological Datasets

Claire Tsang

ISSN 2059-4453 (Online)

The Research Report Series incorporates reports by Historic England's expert teams and other researchers. It replaces the former Centre for Archaeology Reports Series, the Archaeological Investigation Report Series, the Architectural Investigation Report Series, and the Research Department Report Series.

Many of the Research Reports are of an interim nature and serve to make available the results of specialist investigations in advance of full publication. They are not usually subject to external refereeing, and their conclusions may sometimes have to be modified in the light of information not available at the time of the investigation. Where no final project report is available, readers must consult the author before citing these reports in any publication.

For more information write to Res.reports@HistoricEngland.org.uk or mail: Historic England, Fort Cumberland, Fort Cumberland Road, Eastney, Portsmouth PO4 9LD

Opinions expressed in Research Reports are those of the author(s) and are not necessarily those of Historic England.

SUMMARY
This paper reviews what stops the production of high-quality metadata for archaeological archives, reducing the intellectual accessibility of data. It discusses difficulties resourcing metadata production, the quality of existing data standards which metadata is produced from, templates for deposition, and lack of guidance of what should be included to enrich metadata. It puts forward principles for the improvement of metadata for fieldwork datasets and applies that to an exemplar metadata template

ACKNOWLEDGEMENTS
The author would like to acknowledge the help and support of Duncan H. Brown, John Vallender, Hugh Corley, Manda Forster, Edmund Lee, Archaeology Data Service, attendees of the CIfA metadata edit-a-thon 2019 and those who provided access to their recording manuals.

FRONT COVER
'Barriers' by John Vallender

ARCHIVE LOCATION
Historic England, Portsmouth

CONTACT DETAILS
Historic England, Fort Cumberland, Fort Cumberland Road, Portsmouth, PO4 9LD
Claire Tsang, 02392 856757, claire.tsang@historicengland.org.uk

CONTENTS

INTRODUCTION

The digital archaeological archive constitutes a '*knowledge base which will be essential in any revisiting of the conclusions and interpretations in the light of new data collected by future archaeologists'* (Perrin *et al*, 2014, p6).

Every archaeological project must therefore aim to produce a stable, ordered and accessible archive that can be assimilated easily into the collections of recognised repositories (Brown, 2011).

These are the statements that guide the principles upheld in this report; archaeologists have a social responsibility to work with scientific integrity in the creation of data so that research contributes to understanding, results can be recreated, reused and reinterpreted. The term 'accessible' used in Brown (2011) is understood here to mean accessible not just in terms of discoverability, but also in terms of understanding and reusability; to enable reuse it is important that data are defined by and described with rich metadata, as upheld by the FAIR Principles (https://www.force11.org/group/fairgroup/fairprinciples), for which the intention is to improve the findability, accessibility, interoperability, and reuse of digital assets.

At the Chartered Institute of Archaeology Conference 2019, Historic England and the Chartered Institute for Archaeology (CIfA) Information Management special interest group (IMSIG) held a metadata edit-a-thon. As the attendees came into the room, one whispered to me that they really hated creating metadata and I nodded in agreement; for anyone who did not produce the original data, it can be hard to express the meaning behind field terms, although that isn't to say it is not hard for those who produced the data. However, if data standards exist in recording manuals, report methodologies, and similar documents, why is metadata creation difficult, and what are the barriers to creating it?

The archiving of digital data creates three main new challenges; the increased requirements and expectations of documentation to greater detail than physical documentary archives, continuous data management and the need to fund the deposition of the digital as well as the physical archive.

As we become more familiar with managing digital data our understanding of what is needed to comprehend that data becomes more developed and early discussion of archaeological metadata, such as Wise and Miller (1997), focuses on high level, collection related metadata required for retrieval, although Wise and Miller also state that:

'*To describe archaeological data adequately (with the goal of making it faster and easier for other people to discover it) we have to understand what people will want to know about it*.'
This applies to all types of metadata, not just at collection level; understanding *what people will want to know about data* is imperative to creating rich metadata and reusable data.

As well as having social responsibility to ensure archaeological data is reusable, publicly funded organisations are guided by government policies, which are developing in regards to improving access to data; see the UK Government Open Data white paper (httpszwww.gov.uk/government/uploads/system/uploads/attachment_data/file/78946/CM8353_acc.pdf), and European Commission H2020 funding guidance (https://ec.europa.eu/research/participants/docs/h2020-funding-guide/index_en.htm).

Recent reports by Donnelly-Symes (2019) and Parker Wooding and Forster (2018) demonstrate that there are low rates of deposition of digital archaeological archives with appropriate repositories. Continued acceptance and deposition of digital archives with museums as identified in the Society of Museum Archaeology (SMA) annual surveys (Boyle *et al*, 2016-2018), highlight a sector-wide lack of understanding of the requirements of managing and preserving digital archives. These reports also show that there are misunderstandings of what is needed to reuse and curate data, indicating that digital archiving skills are still new and not embedded in mature working practices. At the same time, there is a continuous move towards creation of digital data and digital publication and this has exacerbated the issue.

Awareness of low deposition rates coincides with the Department of Culture, Media and Sport (DCMS) accepting, in 2018, the recommendation of Historic England, written in response to a government procured independent review of English Museums (the Mendoza Review (2017), that museums should be relieved of the responsibility of curating digital archaeological archives:

'*DCMS should welcome and endorse guidance from key archaeological organisations that, as soon as practicable, relieves museums of the expectation that they should attempt to curate digital archive material from archaeological projects, in favour of their deposition in a Trusted Digital Repository that will guarantee the preservation and accessibility of digital material, such as the Archaeological Data Service.*' (Recommendation 6) https://historicengland.org.uk/whats-new/news/new-plan-englands-archaeology-archives-challenge/

So, in England (at least) there are low rates of deposition of digital data and a need for strong guidance to deposit data with a trusted digital repository. This means it is imperative that barriers to metadata production are understood now, and that metadata standards are defined.

To attempt to mitigate the issue of low deposition rates of digital data, Historic England is funding a project to create sectoral standards and guidance for managing digital data generated from archaeological investigations; Work Digital/Think Archive (Project 7796), which is being undertaken by Digventures. This will provide a definition of digital archaeological archives together with data management guidance. This paper complements that project, discussing the issues surrounding metadata creation and looking at contextual metadata; what is necessary to understand and describe individual data fields. It focuses on the Archaeology Data Service (ADS) file level metadata templates (https://archaeologydataservice.ac.uk/advice/FilelevelMetadata.xhtml#File-level) as the ADS is the only trusted digital repository curating heritage data in the UK. The discussion centres on metadata for tabular data, spreadsheets and databases as each field (column heading) requires description, but the issues surrounding understanding data through better documentation are applicable to all archaeological data, such as image content, CAD layers and drawing conventions.

Metadata

Metadata, or data about data, is a broad term and is therefore broken down into further definitions, the emphasis of which can be affected by the discipline in question. Gregory *et al* (2009) break metadata down to four types:

- *Structural metadata describes the structure of data sets, whether these are tabular in nature or simply files of raw data or microdata. Which variable's value appears in which column? Which row represents which case? Are there hierarchical relationships? Etc.*
- *Reference metadata consists of what is often thought of as "footnote" metadata, whether this is about methodology, sampling, quality measurements, production notes, etc. This is a very broad term, which can cover a range of information, regarding everything from single data values to entire collections of data.*
- *Administrative metadata are the data which is created by the process of administering data, in its collection, production, publication, or archiving.*
- *Behavioural metadata (also known as "paradata") is information about the reaction and behaviour of users when they are working with data, and respondents while data is being collected (in this case, it is paradata about a collection instrument).* (Gregory *et al*, 2009, 2)

Reference metadata are more commonly called contextual metadata within the archaeological discipline. This paper mainly focuses on this type of metadata, because that is what gives meaning to archaeological information; yet there is limited guidance for describing this, while quality and content are not prescribed by repositories.

Metadata for an archaeological dataset are required at multiple hierarchical levels, just as they are for the cataloguing of the physical documentary archaeological archive. At the higher levels metadata are needed to describe the event, site and entire archive, then to describe individual digital objects, potential sub-parts (such as tables in a database) down to individual field definitions within a database. Therefore, there is a greater expectation and need for metadata at the file and item levels of the ISAD hierarchy (ISAD, 2000), than for physical archives. This is because of the increased need to describe data structure and technology (structural metadata), but also increased expectation to document the definitions of data fields, so that there is clarity of understanding to enable reuse (contextual metadata). This level of description was either not described or embedded in other documents such as recording manuals and report methodologies, but it now needs to be transferred into standardized metadata templates as required by the digital repository. In addition, item level description for images can exist within an archaeological archive itself, i.e. photographic indexes, and it is an additional task for this information to be transferred and enhanced to the repository's prescribed format.

The hierarchical requirements of metadata are represented in metadata templates required for deposition with the Archaeology Data Service. Collection level metadata provides a detailed overview of the collection and provides the repository with the necessary information to feed into a collections management system that allows users to find and retrieve the archive: https://archaeologydataservice.ac.uk/advice/DatasetlevelMetadata.xhtml#Collection-level. This meets the Dublin Core Metadata specifications (http://dublincore.org/).

File level metadata required by the ADS is categorised by type of data, including:

- Documents
- Databases, Spreadsheets and Statistics
- Raster Images
- Geophysics and Remote Sensing
- CAD and Vector Images
- Geographical Information Systems
- Video and Audio
- Virtual Reality
- Photogrammetry

For databases and spreadsheets, information is needed to describe the parts of the file, such as individual tables and sheets (databases also need the relationships documented), and further still, detailed contextual metadata that defines the content of individual fields (e.g. tabular column headings). The ADS guidance for this is 'field descriptions including units of measurements, description of codes, etc' used within each field (https://archaeologydataservice.ac.uk/resources/attach/ADS_spreadsheet_metadata_template.xlsx, accessed 27/02/2020). The ADS metadata templates for these also have a field to list and link to other supporting documentation, showing that what can be included in the metadata template is not always enough to describe the data. This level of contextual metadata is also required for CAD files, in the form of layer descriptions. The information to be included in 'field description' to ensure data can be reused and understood is defined by the data creator, although guidance is provided for each data format in an example template; Table 1 shows the example provided by the ADS (accessed 30/09/2019) https://archaeologydataservice.ac.uk/advice/FilelevelMetadata.xhtml#Databases, Spreadsheets and Statistics.

Sheet Name	Sheet Description	Number of rows	Field Name (please start a new cell for each row)	Field Description (inc. units of measurements, description of codes, etc used within each field).
contexts	List of contexts	30	context_id	unique context id
			context_description	description of context
			context_type	context type
			context_drawn	if context appears on plan: Yes/No
finds	List of small finds at King's Manor	1304	small_find_id	Unique trench code
			find_type	Trench length in metres
			find_material	Trench width in metres
			context	Trench start date
samples	List of environmental samples	20	sample_id	Links to Site table
			context	Land use, e.g. wood, close, meadow, open field
			sample_type	Sample type, for example 'bulk'
			sample_volume	Sample volume (in litres)
			Notes	any description of sample
plans	List of drawn plan	12	drawing_number	unique id for plan
			Scale	scale of plan
			illustrator	name of person who drew plan
			contexts	list of contexts present on plan
			Notes	any notes

Table 1: Merger of the ADS exemplar template for spreadsheet metadata and ADS guidance to their content.

The previous absence or inaccessibility of contextual metadata are potentially represented in a lack of reuse of archaeological archives. With digital technology democratising discovery and access, the increased ability to reuse and manipulate digital data leads to a greater expectation of reuse and increased public value, and the deposition standards of trusted digital repositories, as well as an improved understanding of ethical responsibility, means that it is no longer acceptable not to document fully all data. High-quality metadata signals professionalism and digital competence and can generate trust in the data, and this is to the benefit of the data creator.

Reuse

There has been limited research into the requirements for archaeological contextual metadata, what is necessary to express so that data can be reused, and currently there is limited guidance on the content of contextual metadata. There is a lack of standards for metadata in the UK, partly because there is no single standard for recording archaeology, although there are seminal works and patterns in the organic influence of recording manuals that have contributed to informal standardisation. These include (and this is in no way an exhaustive list) the Wheeler/Kenyon method, Atkinson (1946) who published the first detailed manual on field archaeology, Hirst (1976), standards and guidance such as the Munsell Soil Colour Book (2009 and previous editions), Principles of Archaeological Stratigraphy (Harris, 1979), Soil Science and Archaeology (Limbrey, 1975) and the Soil Survey Field Handbook (Hodgson, 1974). Publication of recording standards, the accessibility of manuals and the mobility of staff have also contributed to this (for example the influence of London recording systems on the Northamptonshire 1978 manual, following the movement of personnel from the Department of Urban Archaeology to the county archaeology service). The Museum of London Archaeology Service 1994 (MOLAS 1994) manual has been highly influential for recent UK recording practice and created a certain degree of informal standardisation. This can be attributed to its quality, atomisation of fields and accessibility through publication. However, the lack of standardisation for archaeological recording represents a major gap in existing standards

Studies into data reuse have been conducted; these are mainly from the sciences and quantitative social sciences, but have found that data re-users need to know the context in which data were produced in order to evaluate those data (Faniel et al, 2013). To attempt to understand what is needed to reuse archaeological data, Faniel *et al* undertook a study into the reuse of archaeological data between 2011 and 2012. They interviewed 22 archaeologists, asking them to describe their data reuse experiences. This found that lack of context of the data was a persistent issue, although responders often found ways to ‘make do’ with the contextual data they did receive (Faniel *et al*, 2013, 5). Faniel *et al* found that archaeologists place the most importance on procedures, but the reputation and scholarly affiliation of the archaeologists who conducted the original field studies, the wording and structure of the documentation created during fieldwork and the repository where the data are housed also affect how data is valued for reuse.

Faniel *et al* call for more contextualisation of metadata in the terms of reputation and scholarly affiliation of the archaeologist, to increase trust in the quality of the data. This is a request for behavioural metadata (paradata) that we do not currently record in the UK and has not been expanded on in this paper; however, when researching recording manuals it was identified that Surrey Archaeology Society included fields to define confidence, highlighting the need of those working with data collected by community groups to understand levels of accuracy within those records. Nothing similar exists so explicitly in the recording manuals of either Historic England or the commercial units that were reviewed for this paper, although expression of doubt is often requested in discussion fields, and patterns in data quality can be traced through initials for recorder and checker. What this demonstrates is that archaeologists require paradata not just for reuse, but also during their original research and it is thus included within data. In addition, archaeological data itself contains strong information regarding methodology and influences on excavation and the excavator – weather conditions, contamination, method of excavation are all regularly covered by recording manuals - the equivalent of expressing the equipment calibration in science metadata.

Faniel *et al*’s call for greater contextualisation of data is matched by Huggett (2012) who argues ‘that just as data need metadata to make them discoverable, so they also need provenance metadata as a means of seeking to capture their underlying theory-laden, purpose-laden and process-laden character.’ Huggett considers that Neil Postman's 1993 prediction has come true: "*… the tie between information and human purpose has been severed, i.e., information appears indiscriminately, directed at no one in particular, in enormous volume and at high speeds, and disconnected from theory, meaning, or purpose.*" (Postman, 1993).

Recording Manuals and Contextual Metadata

Although it has been identified that there are limited standards for the content of contextual metadata for archaeological data, the data definitions and methodologies needed to create metadata should exist within archaeological recording manuals and any methodological statements within reports and similar documents.

It is through their recording manuals that archaeological organisations document their over-arching philosophy and influences, methodology, procedures, data standards, terminology, field names and definitions. One of the reasons recording manuals are so important to archaeology is that the 'experiment' of archaeology is undertaken by multiple people at once so that every record needs to be sufficiently extensive and detailed to ensure standardisation of data. Design and content can, however, be affected by the need for manuals to be usable in the field. This need to describe data for multiple data creators is reflected in how often recording manuals only include indexes for finds and environmental data, not covering further data created by individual specialists, such as find recording, sample recording, sample processing and ecofactual retrieval. This can help the aims of the document, keeping it streamlined for the field and ensuring data is created by the appropriate people. However, all datasets need some form of recording manual to define the data and describe the methodology.

The majority of metadata should, therefore, be retrievable from existing documents. In a hypothetical scenario to complete an ADS spreadsheet metadata template the following would occur:

- At file level, the file name, row count, sheet name, and software are copied in
- Additionally, at file level, the overall content of the document and the sheets/tables are defined
- The column headings can be copied into the metadata spreadsheet field name (in Microsoft Excel this can be semi-automated using the transpose feature)
- Then the corresponding definitions of the data can be copied in from the recording manual
- Field names or descriptions can be enhanced with site specific idiosyncrasies
- Finally, bibliographic references to any further information needed to understand the data can be added to the supporting documentation file name(s) field.

This also means that if you have standardised recording systems, you can implement a metadata library which can be duplicated for multiple datasets, making the task highly efficient, and ensuring standardisation of metadata. This approach is a core philosophy of Historic England's Archaeological Archives Team data management tool ADAPt (Archaeological Digital Archiving Protocol; https://historicengland.org.uk/research/methods/archaeology/archaeological-archives/adapt-tookit/).

Each sub-set of data; e.g. soil description, stratigraphic relationships, inclusions, drawing conventions, which build up the complexity of archaeological data, potentially have a data standard. However, recording manuals, do not regularly reference originating standards directly, but acknowledge the recording manual that influenced recording at the start and often exclude bibliographies. Popular for reuse are diagrams from the MOLAS 1994 manual, although referencing in the manuals is variable, and it is more likely that diagrams will be directly referenced (likely because of copyright); however, terminology and thesauri are similar between recording manuals, which signals reuse although origins are not directly referenced. Thus, standards are reused, directly or with adaptation, with limited attribution to the original standard or source.

This creates multiple issues; it is difficult to be clear what standards are used by the sector, it risks standards being changed or developed ad-hoc without knowledge of the effect, it makes it difficult to document standards in metadata, metadata cannot be enriched with the bibliographic references, and it makes it difficult to create Linked Data. As an example of this, while researching field definitions in recording manuals, it was found in Steve Roskams' book 'Excavations' that the flow diagram for assessing deposit composition (Roskams, 2001, 179) is attributed to the Scottish Urban Archaeological Trust. However, it is also in the MOLAS 1994 manual, of which Steve Roskams is an author. The diagram, which is the visual metadata for understanding how soil descriptions were attributed, does not have a reference to the MOLAS manual although for other diagrams there are references to Hodgson (1974), Powers (1953), and Folk (1988). It appears likely that it originates from MOLAS and this shows the difficulties that can be caused by not appropriately citing each standard, risking copyright breach and data standards.

THE CIFA SESSION

At the Chartered Institute for Archaeologists Conference 2019, Historic England, with the CIfA Information special interest group (IMSIG), held a metadata edit-a-thon workshop. The session had two aims; to act as a training session to familiarise attendees with the issues surrounding metadata and to work towards creating exemplar metadata templates for the sector. The concept behind this was based on an examination of whether Historic England's ADAPt project, which aimed to develop tools and workflows that ensured efficiency and standardisation, could be applied to the wider sector.

Preparatory Work

Preparatory work for the session reviewed a selection of recording manuals and tested whether there was enough similarity in the definitions of data to identify and list common fields that archaeologists record so that they can be defined, which there was, despite methodology and data being recorded differently . The recording manuals were reviewed in turn for data related to contextual records, drawings, photographs, registered and bulk finds, and field names and their definitions, or the most relevant text, were copied into a Microsoft Excel spreadsheet.

The recording manuals included Historic England 2018, Albion Archaeology 2017, Wessex Archaeology 2015, Museum of London Archaeology Service 1994, BAJR Guide 23, Surrey Archaeological Society 2010, as well as anonymous Recording Manual 1. The field names in recording manuals and on proforma could cover multiple types of data, i.e. 'method and conditions'. Where this occurred the fields were atomised, as it was important to represent likely computer system structures, rather than physical systems, and because, whether atomised in the recording system or not, each data type needs a definition, and splitting the data highlighted where definitions were either missing or unclear in recording manuals. Anonymized Recording Manual 1 was picked as the leading document because of its clarity of description, similarity to the MOLAS 1994 manual and its inclusion of the proforma. It has to be noted that without the proforma being included it was often unclear what the data structure was behind recording manuals, because of the amount of data being described within paragraphs and without headings. The other recording manuals were then mapped to it, with new fields being added to create an overarching field name list with alternatives names and the multiple field description on single rows within the spreadsheet. Because there were so many fields for 'Context' these needed classification so that the spreadsheet could be manipulated, which separated the data into the following classes: General; Attributes (colour, texture etc); relationships (above, below, same as etc) Interpretation (discussion, phase etc), Other Records (small finds, photos drawings), this had the benefit of being able to state on the day that the 'other records' were a low priority, as they would be defined by the other groups at the session.

The number of fields increased greatly with atomisation of data fields for contexts, so field names were grouped into classes to make the spreadsheet manageable, including 'identifier'; 'location'; 'attribute'; 'attribute: dimension'; 'attribute: deposit'; 'attribute: inclusions'; 'attribute: stoniness'; 'attribute: cut'; 'attribute: levels'; 'method and conditions'; 'stratigraphic'; 'audit'; 'cross-referencing'; 'interpretation'; 'intervention'. This also gave the opportunity to consider the data, highlighting that physical data includes fields that would be defined as a relationship in a database.

The preparatory work identified the following issues and questions:

- Recording manuals are mainly designed towards paper-based recording, meaning data was not fully atomised.
- Focus on paper-based recording meant design included cross-referencing data, while digitally this would be a relationship in a database; should these fields be included?
- As well as discussion and interpretation text fields, there are multiple 'Comments' fields relating to wherever archaeologists felt further information was needed, i.e. cut

comments, deposit comment, dimension comments. It was decided that these would not be included, however. The descriptions for these often highlighted that they were used to record further data that could be atomised, such as extent of excavation, and this was atomised.

- When information in the recording manuals was atomised this massively expanded the number of fields to be described, particularly for contexts, for which there was an increase from about 40 per proforma to over 120, creating a much larger workload than anticipated.
- There was resistance from some archaeological units to allow their recording manual to be used for the session.

The session was arranged so that separate groups of participants, sat at individual tables, dealt with different types of data – drawings, photographs, finds and bulk finds, and contexts. The session provided attendees with physical copies of various recording manuals. The number of context fields meant that the context table was provided with the master list of field names and definitions to be altered and added to, whereas the other groups were asked to create a new list of field names and field definitions on an A1 sheet from the recording manuals.

The Metadata-edit-a-thon

The session started with a warm-up exercise of working through the common fields, such as site code, which demonstrated what was required in the workshop, while also ensuring that each group did not need to define those fields separately.

During the session, metadata templates for common fields, photographs and finds were completed; drawings had a list of terms, but not full descriptions. The group working on context data worked through the majority of context attribute fields, but not through relationship fields. The finds table breached the aims of the session; rather than pull the fields from the recording manuals, the attendees identified that the recording manuals merged material type and object type, so they separated those attributes and generated new field names and descriptions.
The conclusions from the session were:

- The text in recording manuals is not always easy to create metadata from because it is spread out or phrased inappropriately for metadata
- There is variation in the level of detail in the recording manuals for each field
- There were difficulties defining fields that seemed ‘obvious’ without guidance for what would be appropriate
- Clarity is needed as to whether the exemplar metadata templates need to cover cross-referencing fields, that would otherwise be a relationship in a database
- Despite expertise and familiarity with archaeological terminology, producing definitions took time and consideration.

It is believed the session was successful in completing its primary aim of familiarising attendees with the issues surrounding metadata, and went some way towards creating exemplar metadata templates for the sector. It also showed how a metadata hack-a-thon session functions, which will be useful for improving any potential subsequent sessions. However, the session relied on descriptions in the recording manuals to be collated, refined or added, and mainly what the session highlighted was that there was not a clear prescription for how data should be defined, nor any other information needed for reuse, and without this, we were able to enhance field definition but did little to expand on it to consider what else is needed to understand data.

BARRIERS TO THE CREATION OF METADATA

The preparatory work and the Metadata-edit-a-thon highlighted many issues around the complexity of producing metadata, showing that its production is difficult, despite the expectation that the information needed to complete it is already in existence. These findings altered this research, from attempting to improve, standardise and enhance the efficiency of metadata production, to understanding the barriers to metadata production. The hope is that investigating these would provide a greater understanding of how to enable metadata production, with the overarching aim of improving the quality of documentation and archive deposition rates.

Barrier One: Resource Barriers

When archiving physical documentary archaeological archives, the archiving is normally completed to Item Level in the Manual of Archival Description (MAD3; Procter and Cook 2000), where Item Level is equivalent to the ISAD 'file' level of description (ISAD, 2000). The MAD3 model was modelled to archaeology in 2000 by Perrin and Brown (Perrin and Brown, 2000), which adopted the model for archaeological archives, adapting Level 2: Group from provenance organisations or individuals to archaeological project (see figure 1).

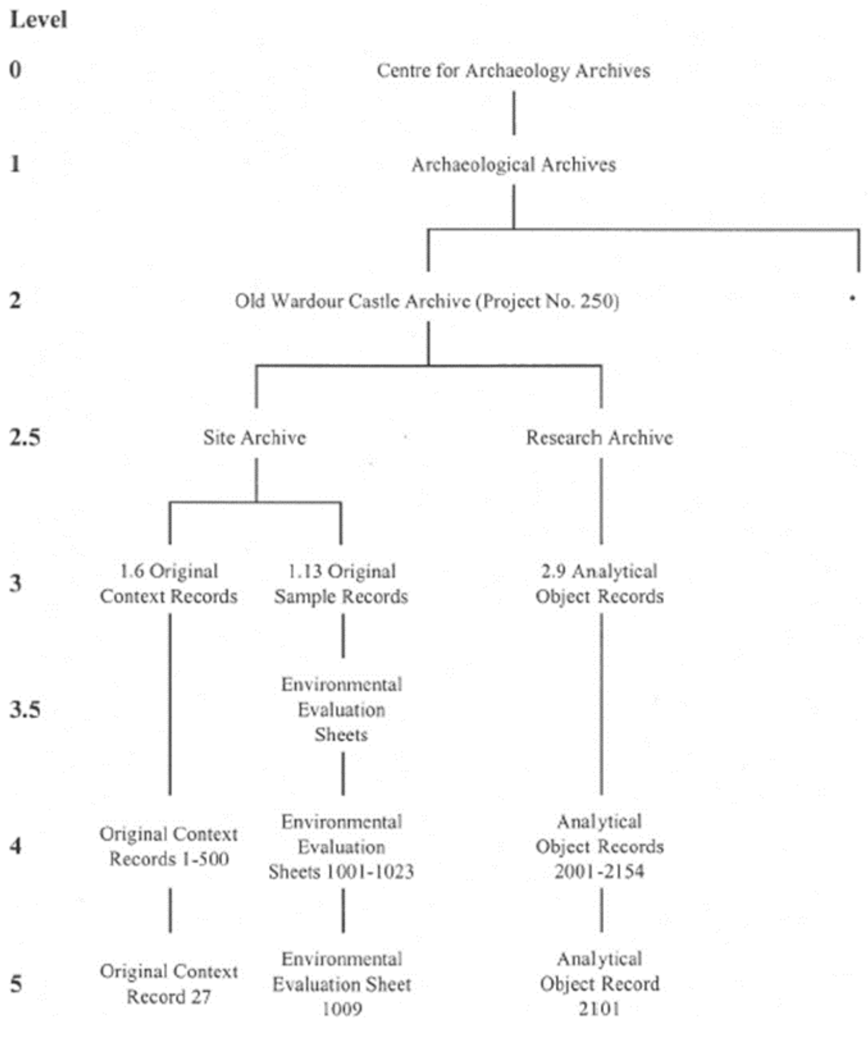


Figure 1: levels of implementation of the Model of Archival Description from Perrin & Brown (2000)

At item level, components of the archive are recorded in groups; for example, pro-forma are catalogued as blocks and reports as singular objects. The lowest level of the hierarchical structure is 'piece' and describes the component parts of the Item. Piece level description is usually embedded in the archive itself in the form of index proforma, such as photographic and drawing indexes. This relieves the archivist of the task of describing each sheet or photograph, allowing them to collate the archive into number runs and catalogue them in blocks of a practical size. However, to prepare a digital archive for deposition, each digital file needs description and thus what was a single item, such as a report, now has multiple parts with the text, figures and tables created in different formats. Those need recording at piece level, which greatly expands the cataloguing task. This example accepts the best practice of archiving the text, supporting tables and images separately. The expansion of archiving tasks can also be exacerbated by portions of documents being created in multiple files - for example where each appendix is an individual text file.

For digital data, such as spreadsheets and databases, metadata are needed at item level (MAD3) and file level (ISAD), to describe the content of the digital file as a whole, and it is also required for describing the individual data fields. This again is the equivalent of recording to piece level (MAD3) or Item level (ISAD), and represents a further expansion of the amount of cataloguing to be completed.

Digital image files, such as site photographs, although documented in the archive, need documenting again in the form of metadata, as required by the repository. The information required and recorded about archaeological data in the archive and what it is necessary to record for deposition are not always aligned. A site photograph will typically be documented with the following fields: description; facing direction; scale; taken by and date taken. The data needed by the ADS is caption; subject; period; period date; creator; copyright holder; location; grid references - longitude, latitude and OSGB Easting Northing; creation date; software; software version. For a small, single trench or period project a certain amount of this can be filled in quickly, however, even for a small site the caption needs to be written and the subject known. This information needs to be pulled from other records and this is made more time consuming if the data is recorded in disconnected spreadsheets, rather than a relational database.

For 'caption' the ADS requires: "Add a short caption for each image which should describe what the image depicts. This metadata will appear next to the image within the interface, and will be used by those searching your collection." https://archaeologydataservice.ac.uk/advice/FilelevelMetadata.xhtml#RasterImages.
This raises the question as to whether what is written in the description field on site will include enough interpretative information to complete the aim of enabling re-use, and thus a like for like filling in of the field from the site description may not be enough. Here, the argument for ensuring that as much archival information as possible is front-loaded into data creation does not work, because the information required in the metadata template is interpretive, and that is not always known during data-gathering and is more likely to be finalised during subsequent analysis. It is general practice not to revisit the site photographic records under the staged approach to projects (such as MORPHE), or ethical data recording, which means that the site archive is not amended to enhance the record, so that becomes a new and additional task. Without clarity of how much time it takes to complete the metadata template, it is difficult to calculate the resources needed and under-resourcing the task becomes a barrier to metadata. However, there is limited published research into the amount that digital data expands the working project archive in terms of numbers of files generally or the quantities of files identified for inclusion in the archive after selection.

Research into archaeological photography practice at Çatalhöyük has shown that the move from analogue to digital photographs represented a 240% increase in the number of images, with a marked variation of quality 'in terms of resolution, file durability and utility for re-examining the archaeological record' (Morgan, 2016).

This highlights that understanding how digital data has expanded can be achieved when the archive is broken down into types of data, and further research is needed in this area to understand increasing archive volumes so that inadequate resourcing does not become a barrier to metadata production.

Metadata production to a greater level of detail equates to a much larger task than documenting the physical documentary archive, although this task has also extended the ability to copy and reuse data. The concept that data is ‘cheap’ to produce therefore greatly increases the number of files in both the ‘working’ and preserved archives. Again, there are limited figures for how much archives have expanded. Historic England, although able to generate figures for the working archives for projects carried out before data management procedures were put in place, has identified that the figures are highly idiosyncratic for each project, affected by the archaeology found, the length of the project, research techniques used and the behaviours and skills of staff. It is acknowledged that both the working and preserved digital archives are much larger than the physical documentary archives; there is also a greater inclusion of data that would previously have been excluded from the physical documentary archive because of its digital form, and increased likelihood of analytical data to be included as well as reports.

Failure to prepare data for archiving stems mainly from a lack of awareness of the requirements for digital archiving, and often it is therefore poorly resourced. This topic slips easily away from barriers to metadata production, towards covering transfer to archive-appropriate file formats, and although this does increase the resources needed for preparing for archive, this is not covered here. However, attempting to produce metadata can show that the dataset has not been created with the archive in mind and this can hinder metadata production, leading to, for example:

- Data with multiple column headings
- Data formatted as a table or for display
- Multiple headings in multiple rows
- Merged cells
- Multiple data types in one cell
- Using formatting and text colours to relay information
- Summary data/titles in the spreadsheet (disconnected from the dataset)
- Data in inappropriate formats, such as tables in Microsoft Word.

These all mean that data have to be changed for archiving, and what is needed to do this is a full understanding of the implications for the dataset; for example, Table 2 had a spreadsheet which had multiple headings, which meant that, rather than just sample number, column headings had sample number, phase and context in each row. Having this information visible was useful to the data creator when producing the data, but in the archive the data needs to be defined by a single row of column headings. The options were; to collate the data into a single row, which takes time, or to remove it so that there was a single definable field/column title. ‘Context’ can be easily removed, as this is already cross-referenced within the archive, however, ‘Phase’ could have been from the final phasing codes or initial interpretation, so has to be checked against the final report, because working to previous phase information could affect interpretation.

Had this data been more ambiguous it would have been best if the archivist had either sent the data back to the data creator, if contactable (given the length of some archaeological projects) or acted with the permission of the data creator.

Although seemingly a relatively minor task, this continuous checking with data creators and amending data within multiple files has a cumulative effect. Had the multiple column headings been merged, the field definition would have been awkward to describe and so idiosyncratic that no standardised metadata template could be used to make the task efficient.

Sample	**3001**	**3004**	**3000**	**1001**	**3002**	**3003**	**1000**	**7807**
Context	**324**	**343**	**322**	**182**	**319**	**330**	**173**	**8023**
Date range	**11th-15th**	**11th-14th**	**11th-15th**	**LBA-R**	**11th-15th**	**11th-14th**	**11th-14th**	**12th**
Spot date	**L11th-13th**	**11th-14th**	**13th,poss14th**	**LIA,prob R**	**13th-e14th**	**L11th-14th**	**11th-14th**	
Common name								
cf uncharred Beet perianth		1						
Corncockle								2
Mineralised Polygonaceae								1
Black-bindweed				2				
Knotgrass				1				
Knotgrasses				1				
Docks	1		2					
Parsley's-piert			1					
cf Pea			3					

Table 2: Example data with multiple headings and merged data in rows

Sample	Species	Number	Bone/Scales	Side
7801	Pleuronectidae	1	precaudal vertebra 3	
7801	Gadidae	1	caudal vertebra	
7801	Unidentified	1	scales	
Hand collected	Gadus morhua	1	Cleithrum	R
Hand collected	Belone belone	1	Dentary	L
Hand collected	Conger conger	1	Premaxilla	L
Hand collected	Conger conger	1	Dentary	R
Hand collected	Gadus morhua	1	Caudal vertebra 5	
Hand collected	Gadidae	1	Cleithrum	R
Hand collected	Unidentified	1	Subopercular	
7800	Clupea harengus	1	articular	R

Table 3: Unnecessary coding within 'Side' column

Coding can be used to compensate for the limitations of screen size, where there is a need for multiple columns to ensure data is narrow. It can be particularly irritating for data creators if the column headings, or only a few individual cell values, are much longer than the rest of the data in that column, sending data off-screen. This can drive the use of coding or shortening of entries so that data fits neatly onto the screen or page. The use of codes will also quicken data entry.

However, all codes require definition in the metadata. Data are difficult to use when having to cross-refer to metadata, while describing coding extends the length of time it takes to produce metadata. For example in Table 3 L and R are used to define left and right bones of fish , but this is not defined in the methodology. It would however, have been a simple task for the data creator to replace the terms on completion of the dataset with the full word, and consider the potential audience, such as those for whom English is not their first language.

All this is pointing to there being a hitherto unconsidered task when preparing digital data for deposition. Just as with report writing, it is necessary to include a data reviewing task on completion of research, which will ensure data are fit for publication, archiving and reuse. This is an addition, not a substitute for designing datasets appropriately from the outset. Not acknowledging this means that archive preparation is under-resourced, and this becomes a barrier to metadata production.

The silver lining in this is that it shows that the task of metadata creation is beneficial to the data. It drives effective planning of data structure and atomisation, and it creates the opportunity for thoughtful consideration of expression, and the clarity needed for reuse that will also be beneficial to the overall production of data.

It is clear from the complexity of archaeological data that the person performing the role of creating contextual metadata needs to be fully able to express all the information required, through a clear understanding of the data, techniques, methodologies and terminologies used and any idiosyncrasies or options within documented standards. Because an archaeological project includes multiple types of specialist research, it should be clear that metadata creation cannot be carried out by one person; never before would we have expected a single person (i.e. the archive compiler) to have a detailed understanding of multiple specialist data standards. For this task to be efficient, it must be undertaken by the data creator, either at the time of creation or prior to submitting their data for archiving. Metadata creation should be considered a data standard documentation task, not an archiving task. Delaying metadata creation and deferring it to those with responsibility for collating the archive is itself a barrier to metadata creation, creating a burden on the collator because it is inefficient, likely to lead to inaccuracy or weakly described metadata, and misdirects resources for other archiving tasks such as compilation, quality checking, and data selection.

Barrier Two: The Quality of the Documentation of Data Standards

The most basic barriers to metadata production in terms of content are either that the data standards used are not fully documented, as identified in the preparatory work and the CIfA session, or that the standards are not accessible.

In 2017 the CIfA finds group undertook a project to review the quality of reporting on archaeological artefacts in England. The impetus for the project came from a growing level of concern over the standards of artefact work in developer-led archaeology, as reflected in the quality of reporting on artefact assemblages available as 'grey literature' (Cattermole 2017, 1). Findings showed that:

- Details of sampling and recovery strategies are not routinely included in grey literature reports
- Specialist reports do not routinely make explicit reference to current, accepted standards or to good practice guidance.

Without these, the re-user cannot understand the data, which means that if someone other than the data creator is required to create metadata, which is what happens when metadata creation is considered to be an archiving task at the end of a project, that person may not have access to the information needed to understand it. That can mean that data cannot be defined and could be misinterpreted or left ambiguously documented.

Reasons for a lack of data standards documentation can include:

- They do not exist, so data structure is ad-hoc
- The need for them is not understood
- Misunderstanding what needs to be documented to reuse data
- Over familiarity with the data and the assumption that metadata are not necessary
- The creation of data without re-use in mind
- Attitudes of protectionism over the data - it is not for reuse
- The removal of pre-existing documentation from the archive.

Protectionism is difficult to evaluate, what might be seen as defensiveness could easily be a lack of awareness. Attitudes are certainly changing; twenty years ago there was resistance to digitally publishing the Historic England report series, now the benefits are recognised, and we are discussing disseminating our data. For the CIfA session, there were a couple of examples of contracting units not wishing their recording manual to be used, and this has been the case for a concurrent Historic England funded project. This is despite (or maybe because of) circulation to a non-permanent staff base, and proven similarity with other recording manuals, in particular the MOLAS 1994 manual. Protectionism over recording manuals is likely to be driven by concepts of commercial sensitivity.

Although there is an expectation that attitudes will change, this protectionist stance could be broken down by defining standards for the publication of archaeological data, with the provision of exemplar templates.

The Format of Existing Data Standards

The review of recording manuals during the CIfA workshop and in preparation for it, highlighted issues with creating metadata from the recording manuals, despite the size of the documents and the level of detail included.

Issues identified included:

- Definitions embedded in large paragraphs of procedures
- Not all data was described or defined, there was no atomisation on pro-forma and little discussion of what to record
- Assumptional definitions, with focus on an audience *au fait* with archaeological recording, despite there not being a single, universal recording system in the UK.

Recording Manual	**Definition of 'Context Number'**
Historic England (2018)	Assign each context a unique number as soon as you have defined its extent, using the Context Index (see Section 1 .6 Record Number Allocation). The numbers are often assigned in blocks by trench (site subdivision), and the same number sequence is used for all contexts, whether Cuts, Deposits, Timber or Masonry. There are separate paper recording sheets for Cuts, Deposits, Timber and Masonry contexts. You will also need to create a spatial record for the context when you assign the number. Ideally this will be the full outline, but if that is not possible a single point should be surveyed as a label (see Section 8.4, Total station surveying and the Intrasis database).
Recording Manual 3	Whenever a new cut, layer, deposit or group is encountered and is about to be recorded, a new number should be issued and entered in this column. A sequential number allocation, the index is often used as a register to control and check the allocation of numbers. A number of indices can be operated simultaneously, e.g. each site subdivision may be allocated a block of numbers from the primary site number sequence - cross check with the number record sheet, if applicable. This is taken from a pre-allocated block from the unique number sequence. The context index is customarily used on site as a register to avoid duplication.
MOLAS (1994)	Every unit of stratification is given a number which is entered in this box. The sequence starts at 1 for each site and will normally be continuous. The next context number should be obtained from the context register at the time of taking the context sheet. On some excavations certain types of contexts might be assigned context numbers from a separate part of the numerical sequence; for example, skeletons on a cemetery excavation might be assigned context numbers from 1000 onwards.
Recording Manual 1	A unique consecutive identifier allocated to the context
Recording Manual 2	Context numbers
Dockrill (1997) (Scatness Excavation Manual)	The context number is the unique identifier given during excavation to an archaeological context (stratigraphic unit).
Hirst (1976)	NULL
Surrey Archaeological Society (2010)	NULL

Table 4: Relevant data taken from a sample study of recording manuals and metadata templates that describe/refer to 'context number' (or stratigraphic unit).

Table 4 compares descriptions of the term 'Context Number' from a collated review of six different recording manuals, together with one metadata template deposited with the ADS –

Recording Manual 2. This demonstrates the variety of field definitions: not at all; a short description, such as 'A unique consecutive identifier allocated to the context'; a slightly longer description that includes various attempted definitions, including 'context' (stratigraphic unit), format (sequential and will start at X) and divisions in the number series, when the number is recorded (as soon as extent is defined). This shows the variety within the ways archaeologists, at the time of recording, express their understanding of the same field.

Each of the issues listed above will now be considered under matching headings.

Definitions Embedded in Large Paragraphs

The layout and content of the recording manual affects how easy it is to produce metadata. Aside from the fact that the data descriptions are not fully atomised, the layout of the MOLA 1994 manual was the easiest to use, in terms of extracting field definitions. In part, this was because the document is structured with clear headings and only provides a standard for recording, with minimal procedures.

Having data definitions buried in long sections of text means that it is difficult to pull out the relevant information to produce metadata, requiring some effort, together with an understanding of the data, for the text to be restructured. For example, in (anonymous) Recording Manual 1, the following explains 'grid reference':

'At the start of a project the Project Leader will normally notify field staff whether an overall site grid will be used, what form it will take and if grid refs are required. For most watching briefs, evaluations and for sites surveyed using GPS/TST you should cross through this box, but for excavations using a grid recording system, such as single context recording, the references provide a means to rapidly locate individual contexts and plans, and helps in constructing site matrices and should be recorded here.

The site grid is composed of 5x5 metre squares which are each identified by the coordinates of their south-west corner, written in the format of the Easting followed by the Northing, with the two readings separated by a forward slash for example 110/225. The grid will be laid out across site using a TST and or tapes and grid pegs will be labelled with their co-ordinates. The site grid will normally use the same Ordnance Survey co-ordinate system we use for GPS and CAD, although the first 3 numbers will be missed off each Easting and Northing.

Add the co-ordinates of all the grid squares that your context lies within, e.g. 120- 5/590 indicates that the context is present in grid squares 120/590 and 125/590.

Occasionally a site grid will not use OS co-ordinates, and may even not be aligned to OS grid north, in these cases the grid will have a point of origin (e.g. 100/500) outside of, and to the SW of all the areas of archaeological recording. The grid will use the arbitrary coordinates in the same way as OS grid co-ordinates.

If only a part of a site requires a grid (e.g. an area of intercutting pits or horizontal stratigraphy) then this local grid can either be located on OS grid, or on an arbitrary grid as appropriate to the area requiring it.

If using an arbitrary grid then note that although plans are aligned to the Grid North, there is no change to the direction of site north as far as feature orientations/sections etc is concerned. Whatever the source of your grid values, it is used in the same way, and all plans are always related to the SW co-ordinate of a specific grid square.'

In this example, the recording manual fulfils its role as a guide for site but is not focused on defining the data, while the length of the text makes it difficult to extract information that is relevant to metadata. Here, paragraph two is potentially the relevant text for field definition.

From the second paragraph we can ascertain that 'grid reference' means:
'The coordinates of the south-west corner of a 5x5 metre square, which is part of the site grid, formatted as the Easting followed by the Northing, with the two readings separated by a forward slash for example 110/225. The numbers are the Ordnance survey system with the first 3 numbers missing.'

Missing from this is any definition of what a grid system is, and also any clear explanation that the grid system is aligned with north, which is accepted archaeological practice. This example has also highlighted that the intended audience for the recording manual is made up of skilled practitioners, although that may not be the case with those reusing the data.

Paragraphs four to six demonstrate the variability of the recording manual, where the recording manual also records the *options* for recording grid references, rather than how to decide what to record in the field. That decision needs to be documented and made accessible to enable metadata to be produced. Difficulty finding this information may lead the metadata creator to define the field ambiguously, removing any reference to what system the grid is based on, such as how it is recorded in the recording manual. What is also needed is documentation of how this is applied on site, and recorded, as well as what grid system was used.

Poorly Described Data and Lack of Atomisation

One of the biggest issues for the CIfA workshop, in attempting to create a complete list of the fields we record during excavation, was that in the recording manuals, field names covered more than one data type, such as 'methods and conditions'. The data being entered did not always reflect the field name, especially where singular field names covered multiple data types, such as:

Recorded by	Initials and date
Size	Length, width, height/depth
Colour	Text and Munsell number
Inclusions	Inclusion material, frequency, size, angularity.

The structure of recording manuals reflects the layout of proforma, in which fields are condensed to fit them onto A4 paper, and do not represent the structure of digital recording systems in which the data are more likely to be atomised. Not having the data standards explicitly described in a master document risks misinterpretation of any data definitions, de-standardises the metadata and makes the process inefficient.

The process of producing metadata for the example in Table 2 highlighted the presence of merged information, with the field 'species' also including preservation information. Metadata should be able to tell the re-user what data there are without having actually to access the data. Here, if the re-user was looking for the state of preservation, describing these data as species in the field value definition, as well as the data description would hide this. This merging of data was enabled by there being singular preservation conditions for each species; however, a larger dataset would benefit from atomisation in order to be searchable. Data are being designed ad-hoc, and this will make metadata production inefficient, and make it difficult to compare and reuse data.

Whether the data is atomised or not, each type of data requires description and atomising that definition allows visualisation of the content, while if the structure of the data types remains merged, the definitions can be merged as well, although this is likely to have a negative impact on how the data can sorted and searched.

Gaps in Recording Manuals/Assumptions Made About the Audience

Definitions in the recording manuals can be unclear, e.g.:

Field Name	Definition
Context Number	Write the context here

This is a procedural instruction only and lacks definition of the format of the data to be entered, what a context number is and its relationship to the site or other records.

Field Name	Definition
Level Top	Level AOD of top surface of context

This definition is based on the familiarity of archaeological practice where the assumption is that it is clear what the AOD stands for and how it is measured and that all archaeological data is recorded to the same co-ordinate system. So, if the metadata creator is attempting to create high-quality metadata, then the explanation in the recording manual requires expansion. Improved clarity would benefit the recording manual and the practice of recording on site as well as the metadata.

It is also assumptive to exclude any definition of what data is to be recorded:

Field Name	Definition
Direction of View	This should indicate the direction that the photographer was facing to take the photograph.

This definition looks quite obvious. It is normal archaeological practice that 'facing' or direction of view is recorded to eight cardinal points, so the above assumes the recorder has that knowledge; however, it could be recorded to sixteen cardinal points. There is also no guidance for how the data should be formatted; NE or north-east, which risks non-standardised data entry as well as ill-defined metadata.

As these examples show, it can be difficult to describe some fields because it feels like stating the obvious. For example, with Project Number, it can be sorely tempting to add the description 'the project number' to the field description because otherwise, it can feel as though you are repeating a dictionary definition of the word project (e.g. an 'individual or collaborative enterprise that is carefully planned to achieve a particular aim'), even though it is only because of familiarity with archaeological data that we understand it is not an 'identifier given to an estimate or forecast'.

Among the sampled recording manuals, the following attempted definitions of 'butts' were available:

- What other contexts the context butts
- All contexts this context butts
- The contexts physically touched by a later deposit or wall are recorded here. The deposit [001] is later than [003] and physically butts [003]. On the record card for [001] we would record that it butts [003], though the context stratigraphically below [00l] in the matrix would be [002]
- This relates to structural contexts; where one is built up against an existing feature. For example, in Figure 4.7a the later wall A Butts wall B, and the earlier wall B is Butted By wall A. Similarly, a timber context may be Butted By a later wall or floor surface.

Whether 'butts' is used for all deposits or structural contexts only, is a defining part of how the term is used. However, two of the recording manuals provided very little guidance for use. If recording standards are not explicit in the recording manuals then metadata are unlikely to be comprehensive.

In summary, the most obvious barrier to reproducing data standards as metadata is that the data standard did not exist in the first place. However, where standards do exist, they can be incomplete and difficult to produce metadata from. This is further compounded by a failure to design data appropriately for archiving.

Barrier Three: Metadata Templates

The use of spreadsheets by the ADS for their metadata templates suggests that the information to be entered into a single spreadsheet cell is sufficient to understand the original data, and also that metadata can be created in a simple text format, with no paragraphs or reference to thesauri. This is also the impression given by the exemplar templates provided by the ADS (https://archaeologydataservice.ac.uk/resources/attach/ADS_database_metadata_example.pdf, (see table 1). Although supporting documents can be referred to on the metadata template, this gives the impression that they are of secondary importance. If we understand what has to be expressed to enable the reuse data then the influence of a template format is irrelevant and can be compensated for. However, if metadata template completion is done as part of a 'tick box' exercise, or without understanding, then the templates suggest a short text-based description is appropriate.

One of the most popular and reused parts of the MOLAS 1994 Manual, and to a lesser extent those of Historic England, are the diagrams, which represent visual metadata. These diagrams include:

MOLAS: The percentage composition of inclusions (redrawn after Hodgson 1974); degree of sorting of deposit composition (after Folk 1988); degree of roundness in the shape of pebbles (after Powers 1953); flow diagram for the composition of archaeological sediments; depth of stakehole; break of slope; top and base; shape of base; stone finishing and coursing; brick bonding; brick and walling terms; pointing; timber conversion; common carpentry joints.

Historic England: Pebble shapes; break of slope top and bottom (matches MOLA 1994); fingering texturing methodology; coursing (after MOLA 1994); brick bonding; brick terminology; wall butting; pointing; timber conversion; common types of joints used in timber-framed construction and trusses (after Harris, 1979); arch descriptions; moulding; rib vaults; tracery; window frame parts; timber-frame structures; door parts (all after Etherton 1991).

Without access to the diagrams, re-users cannot be clear of definitions, nor how decisions were made, but diagrams are excluded from metadata templates based on a spreadsheet format. Further to this, if visual metadata are excluded then it needs to be expressed in a form of text that does not already exist and can be difficult to produce, hence the use of diagrams in the first place.

This suggests a continued need to include recording manuals in the preserved archive, because defining data terminology solely within metadata templates, and replacing the richness within recording manuals of methodological description, thesauri, and data definitions, is not enough to facilitate re-use.

Barrier Four: Limited Guidance for How Metadata Needs to be Expressed

Realistically it is difficult to explain why contextual metadata deposited with the ADS can be of a limited quality; it can be affected by the knowledge and skills of metadata creators or by levels of resourcing.

It is not possible to search for types of data on the ADS search engine (https://archaeologydataservice.ac.uk/archive/), searching is only by event details. However, two deposits with the ADS have been found, which include full contextual information that is representative of the information recorded in recording manuals (*it is also noted that attempts to* *find relevant metadata found archives without metadata and no supporting documentation for archives with scanned proforma).*

One organisation, whose recording manual has been collected during this research, had included a metadata table in the archive. Table 5 compares the descriptions in the metadata table with those in the manual.

Metadata Field Name	**Metadata Field Description**	**Recording Manual Field Description**	**Issue**
Context ID	Database Primary key: Unique Identifier assigned to each context excavated and each 'group' number assigned to grouped feature cut numbers.	Whenever a new cut, layer, deposit or group is encountered and is about to be recorded, a new number should be issued and entered in this column. A sequential number allocation, the index is often used as a register to control and check the allocation of numbers. A number of indices can be operated simultaneously, e.g. each site subdivision may be allocated a block of numbers from the primary site number sequence - cross check with the NUMBER RECORD sheet, if applicable. This is taken from a pre-allocated block from the unique number sequence. The CONTEXT INDEX is customarily used on site as a register to avoid duplication.	There is a difference between the manual and the metadata description: the manual says that 'contexts encountered' are to be recorded; the metadata description refers to 'contexts excavated'.
Shape Base	Denotes the shape of the base of the cut	A simple, easily understood descriptive term should be used, and reference should be made to the Field Guide for Context Sheet Recording, page 3. For more complicated shapes, notes can be made in the Additional Description section.	Only the metadata defines the field; the manual highlights further information in a field guide.
Length	Descriptive label denoting the intervention length dimension	The length of the excavated intervention must be recorded. The dimensions of the context must be recorded in metres, i.e. 1.25m, 0.01m etc, not centimetres. If a dimension is variable, record an average and discuss the variability within the additional description section.	The metadata table does not give the unit of measurement.
EX Method	Denotes the excavation method used to investigate feature	Null	Excavation method was not covered in the recording manual; it is possible it is in the field guides.
Soil Colour	Descriptive label of soil colour of excavated context	Colour: The soil description needs to be as accurate as possible, and describe the main colour of the layer. Each description needs: One modifier i.e. Light, mid, dark or very dark Use one colour and one hue per description i.e. Light greyish brown, not light to mid greyish brown. Colour should be identified from moist samples. Variations such as mottling, oxidation and lenses should be noted in the additional description section or on the continuation sheet. Auger: A damp sample of the relevant deposit is smeared across the colour box on the white field sheet. Only at the post-fieldwork stage is the description transferred to the colour box on the blue sheet. At this stage Munsell soil colour charts must be used to ensure uniform terms of reference. Variations such as mottling and oxidation should be noted in the description section or on a continuation sheet.	The metadata description does not identify the Munsell system as the standard used.

Table 5: A comparison of the text in a contractor's recording manual (Anonymous Recording Manual 1) and the field descriptions in their metadata table.

The metadata provided in the recording manual is less descriptive, although the recording manual itself does not hold all the relevant information, some if which is found in field guides. The issue of not providing units of measurements was also a feature of the second example found, as shown in Table 6. Here also they are not defined, despite this being a specific requirement in the metadata template. In Table 6 there is a pattern of limited field definitions, with the field name reused as part of a description that is often limited to a short sentence. Spelling errors shown in the table are original and suggest this task was given limited consideration. The provision of thesauri or word lists is also mixed, with none present for texture, full definition of stone size, and only a partial list for category and inclusions.

Metadata Field Name	**Metadata Field Description**
Subdiv	Site sub division/area context is in
Gridref	Grid reference of context
Context	Context numbers
Category	Type of context (ie Feature, Fill, Layer, Structure)
Keyword	Context type
Direction	Direction of linear features
Length	Length of context
Width	Width of context
Diameter	Diameter of context
Depth	Depth/thickness of context
Leveltop	Level AOD of top surface of context
Levelbot	Level AOD of base of context
LevelNos	Numbers given to levels taken of context
Munsell page	Page of Munsell color chart in Munsell color book
Munsell col No	Number of the color on the grid of the Musell page
Texture	Texture of the soil
Stonesize	Stone size catagories present (1=<6mm; 2=6mm-2cm; 3=2-6cm; 4=6-20cm; 5=20-60cm; 6=>60cm)
Stoneabund	Abundance of stone in the context (1=1-5%;2=6-15%;3=16-35%;4=36-70%;5=>70%)
Inclusions	Inclusions in the context (ie charcoal, slate)

Table 6: Example of metadata (derived from Recording Manual 2) as deposited with the ADS

This pattern of shortened text for field descriptions is also reflected in the results from the CIfA session; Table 7 shows the results of the definitions for the 'general' fields that were at the top of multiple pro-forma.

Field	**Description**
Project identifier	The institution's unique identifier for the project
HER identifier	The Historic Environment Record identifier for the site
Site code	The institution's unique identifier for the site
Site name	The name of the site
Repository Accession code	The unique number given to the archive by the repository
Year	Year of the fieldwork
Site Sub Division	Identifier for a subdivision of a site area which is investigated archaeologically for an archaeological area defined

Table 7: The metadata produced for 'general' fields at the CIfA conference

This demonstrates a lack of clarity and understanding for how archaeological data needs to be described for reuse, and that a commonly defined structure for describing data would be beneficial to the sector.

Further Work on the Recording Manuals

Several issues arise from the experience of producing an exemplar metadata template for the CIfA conference workshop, and indeed from the workshop itself. These include: the benefits of familiarity with the sample practice manuals used; the problems posed by the length of those manuals; text that does not clearly describe data standards. If therefore, a metadata template is to be developed for use throughout the community, then a prototype would need to be produced that included guidance on the content of field descriptions.

Research into recording manuals continued after the session, with Hirst (1976), Dockerill (1997) and Recording Manual 2 added to those previously collected. The increasingly large dataset was transferred to a Microsoft Access database, and relationships were created between the master list of field names and descriptions and the 'raw' recording manual data.

This review looked at other, historical recording manuals, including Hirst (1976), Boddington (1978) and Jefferies (1977). The aim was to identify any influence these earlier manuals may have had, and their impact on the language we use now. Also, as published, they referred to the original standards that were not referenced in the modern recording manuals, guiding us back to the original (or as near as possible) expression of a field. For example, Hirst's was the first published manual to use Limbrey's definitions of Colour, Texture, Structure, and Clarity of Horizon. Like the MOLAS 1994 manual, the earlier manuals focus on data standards, not procedures, making them easier to document. This provided a broader range of examples of data descriptions, while using a database made it easier to compare data descriptions from each manual, where there was enough information in the master for a description to be written.

While working through the recording manuals it became clear that an alternative name field would be needed, to map common archaeological terms against each other. This raised the question of what was the appropriate terminology to use for the field names. It was initially decided that as the MOLAS 1994 manual had been so influential on the development of single context recording in the UK, MOLAS terms would be prioritised. This also increased the likelihood that the language in the other manuals would be broadly similar. Even so, it is worth observing that the older manuals, such as Hirst and Boddington, have better references for standards and language we commonly use now, which have been inherited from guidance and standards we no longer reference, such as Limbrey (1975).

This comparative exercise initially revealed patterns within what archaeologists generally record to understand their data, rather than generic data field descriptions, although it must be acknowledged that these patterns had the potential to help develop principles for an improved contextualisation of data.

Standards for cataloguing archaeological archives also offer potential standards for how to describe data at file level. There are three main standards for this in England: the model for archival description (MAD) (Perrin and Brown, 2000); the FISH thesauri; the terminology used within OASIS based on the recording system within what was the National Monuments Record Catalogue. The OASIS standard is currently being reworked, but both the FISH and MAD models are increasingly outdated by developments in digital recording. The FISH thesaurus also does not have a standardised level of recording pertaining to archaeological data products, for example for reports, although there are thesauri relating to dating and radiocarbon reports, no other specialist analytical reports are included. The MAD model was designed to create a hierarchical structure for archives produced as a product of the MAP2 (Andrews, 1991) project flowline, but continues to function for the Management of Research Projects in the Historic Environment (MORPHE) (Lee, 2015), although this too is currently under review.

Use of the cataloguing terminology within file names, file and sub-file level description and where relevant, elsewhere, provide standardisation and defined terminology, making definition clearer.

As well as providing definitions for archaeological data, cataloguing for physical archives provides further insight into useful ways of describing data; for example, catalogues include definitions of record numbers covered by the dataset (when this comprises the main data). This helps confirm what data are where, but also for understanding what number blocks are used for which data, and this is potentially useful to add to the metadata.

SUMMARY

Archaeological data are complex and varied, being both qualitative and quantitative, and can inherently include methodological information, paradata, and specialist data. Also included are cross-referencing fields, which would be relationships within a database, and that makes data description a more complicated task.

Barriers to the production of metadata include a broad range of issues which can be grouped into five subjects:

- Limited guidance to what metadata should contain
- The resource implications of metadata production and the preparation of archive for deposition
- Inaccessible or non-existent methodologies or data standards for the creation of metadata
- The format and content of existing standards
- The format of the metadata templates themselves.

These barriers are also affected by attitudes towards archiving; something this report has only touched upon, such as protectionism over data and recording standards, as well as assumptions about metadata and archiving being a final task of a project.

Lack of understanding of both how digital technology has increased the volume of an archive and how much more information needs to be documented for it to be preserved and reused means that the task is difficult to resource. These issues are compounded by the way data are produced without consistency or standardisation, or in ways that require the format to be adapted during archive compilation.

Considering metadata production to be an archiving task creates a disconnection that worsens as more time passes since the data were created and the availability of the person who understands those data best. This slows and impedes metadata production, and is compounded by the complexity and variety of archaeological data. Metadata production must not, therefore, become an archiving task because the level of information required for an archaeological archive means data can only be described by their creator.

It is likely that a failure adequately to resource the task of metadata production, or even to perform it effectively, is why archive deposition rates for digital material are currently so low. Inefficient metadata production is likely to drive down standards, diverting resources towards an ad-hoc recording of metadata and the altering of datasets, rather than resourcing compiling, checking and enhancement.

Reliance on the production of metadata from existing archaeological recording standards is flawed. Guidance does not always exist, or the requisite information can be unattainable, or documents are formatted in such a way that metadata writing is difficult. Manuals do not always give full data definitions, or definitions can be missing, while they are more likely to include standards and options promoted across an organisation, rather than specific decisions or idiosyncrasies that have emerged through the course of a project. They also will not include information that was gathered after the recording event, such as the extent of the record and finally, they do not often include bibliographic references to any standards used.

Archaeologists require recording manuals to perform multiple functions, they document data standards, act as site procedures and are for use in the field, therefore there is a need for them to be highly detailed as well as practical for the field and these needs are not always aligned. Manuals often reflect a certain bias among authors, where text assumes a certain level of knowledge or understanding, in both the language used and in missing passages and this is shown in the different levels of descriptions among manuals produced by commercial, academic, and community groups. Although manuals contain levels of detail for some practices,

they do not fulfil the role of being a data standard for digital archaeological information. This failure makes it difficult to link data and impedes reuse.

It is also likely that existing metadata templates, including those provided by the ADS, have the potential to obstruct the production of rich metadata because they encourage short, text-based descriptions, which do not always reflect the complexity of the descriptions required for archaeological data. There is a particular barrier here in the case of using diagrams within metadata. Templates, however, remain only a notional barrier because additional information can be deposited alongside them and a greater problem is the limited understanding of how to contextualise data, and a lack of guidance to support this.

Despite these imperfections, recording manuals can include sufficiently rich descriptions of data and methodologies to enable users to understand and contextualise the archaeological record; as described by Faniel *et al* (2013). Because they include descriptions of procedures, manuals can provide paradata, together with an understanding of quality assurance and auditing responsibilities. The recording manuals reviewed during this research however, do not fulfil the dual role of a site guide and a data standard, and this report has shown that the level of detail found in recording manuals is not transferred to metadata templates, where short descriptions are produced instead. These short descriptions are ambiguous definitions without further consideration of what is needed to understand the data.

The task of completing metadata, which leads to questioning what has to be expressed to enable reuse, has highlighted gaps in the recording manuals and can have a beneficial reciprocal relationship in improving data standards overall. There is a need to improve recording manuals so that data is contextualised both for initial site recording and metadata production and to make metadata collation more efficient. Identifiable improvements include proper descriptions of standards used, including bibliographic referencing, and full atomisation of fields so that descriptions of data are also atomised and clearer.

TOWARDS AN EXEMPLAR METADATA TEMPLATE AND PRINCIPLES FOR THE CONTEXTUALISATION OF ARCHAEOLOGICAL DATA

This research into patterns of data descriptions among existing manuals, standards and methodologies has shown how metadata can be improved following the introduction of defining principles for the contextualisation of archaeological information. Those principles can be applied to an exemplar metadata template, which will illustrate the minimum level of detail required and provide a basis from which archaeologists can build their own metadata.

The absence of a standardised recording system in the UK should not impede the development of a generic metadata template. This is because it is possible to describe the data that we use without defining when to do so (procedural), or the methodology, thesauri and other standards used. A generic metadata template need not be prescriptive in requiring all the fields it contains to be included in a recording system, but it is important to provide the sector with a clear example of contextualised metadata because currently, only limited data are being deposited and metadata are of a poor standard.

The principles for data description, together with a generic metadata template, have the potential to assist with requirements two, four and five, as listed above:

2. Improve understanding of what is required for data to be reused through the provision of an example that can be built upon.
4. Enable efficient metadata production, by developing supporting tools and data standards, which can be modelled on the generic data standard/metadata template by any archaeological organisation.
5. Break down protectionism around existing standards and data, by making definitions of archaeological data freely available and highlighting the similarities between recording systems.

Universally accepted principles for data description can also be used to create metadata for archaeological contexts:

- Use simple language and common archaeological terminology unambiguously to describe fully the meanings and scope of fields
- Include data format definitions, (e.g. integer; floating point number to two decimal points)
- Include bibliographic references to original data standards
- Express the relationships between records, (e.g. identifier numbers being unique to sites, projects or organisations)
- Define structural relationships (e.g. ‘within’ and ‘includes’)
- Define the types of data a record covers, (e.g. contexts or structural groups)
- Define the extent of the data, (e.g. number blocks covered)
- Identify any thesauri used, if published these can be referenced, if not the full thesaurus needs to be included
- Define when data are not to be applied (e.g. a field is used for deposits only)
- Include examples that will guide data entry.

These principles should be developed alongside guidance for data atomization, recommendations for the inclusion of methodologies and visual metadata (i.e. diagrams). Consideration should also be given to the ways in which the introductory text in recording manuals often set out the conceptual basis for approaching a project, which should inform the contextualization of data and the use of paradata.

These draft principles for metadata have been applied to a generic metadata template in Appendix 1. Appendix 2 includes lists of fields which are currently excluded, either because of their idiosyncratic nature or because they are fields that substitute for what would be a relationship in a database.

CONCLUSION

Returning to the guiding principles behind this report, as set out in the introduction, it has been shown that archaeologists are failing to create knowledge bases that can be revisited and are instead creating inaccessible data, mainly due to poor documentation. Alongside low rates of archive deposition, there is no clear understanding of the requirements for depositing metadata so that datasets can be discovered, retrieved and reused.

As things stand, it is difficult not to conclude that Huggett was right, in that Postman's prediction that 'information… is disconnected from theory, meaning and purpose', and that the conclusion of Faniel *et al*, that data re-users need to 'make do', is an understatement. It is clear that if we, as archaeologists, are to uphold our social responsibility to work with scientific integrity in the creation of data, so that research contributes to understanding, and results can be recreated, re-used and reinterpreted, we need to improve the quality of metadata, particularly if we are to ensure that data can be linked, that datasets are open, and that the FAIR principles are upheld.

Currently, there is no single existing product that holds the contextual metadata that must be deposited with a dataset; recording manuals, reported methodologies or metadata templates themselves, do not hold nor can hold, the information required.

There is a need to ensure that metadata are of high quality and resourced and created efficiently. To that end, this report has attempted to identify existing barriers to metadata production so that they can be understood and broken down.

To break down those barriers, the sector needs to:

1. Improve digital literacy in the design of data for archiving and reuse
2. Improve understanding of what is required for data to be reused
3. Improve understanding of how archaeological archives expand in digital format and how the resources for digital archive preparation increase
4. Enable efficient metadata production, by developing supporting tools and data standards
5. Break down protectionism around existing standards and data.

To undertake this, it will be necessary to:

- Provide training for data creators, in metadata content and data design
- Provide practical archaeological examples of metadata, including explanations of metadata types, as well as data designed for archiving
- Research how effective data management impacts on the size of digital archives
- Produce data standards and improved documentation
- Recognise that it is not always possible for recording manuals to be complete data standards, so it may be necessary to produce new, separate data standards
- Preserve data standards within the archive, alongside metadata templates
- Ensure data creators are responsible for metadata creation
- Implement tools to ensure efficiencies, such as producing reusable metadata templates, metadata libraries, and flowlines
- Promote access to existing published recording manuals and where possible promote publication and access to recording manuals and other data standards.

Bibliography

Albion Archaeology, 2017, *Fieldwork Procedures Manual*, Albion Archaeology: Bedford

Andrews, Gill, 1991, *Management of archaeological projects: MAP2*. English Heritage: London

Archaeology Data Service, 2020, Guidelines For Depositors: *File-level Metadata Requirements; Databases, Spreadsheets and Statistics,* Archaeology Data Service https://archaeologydataservice.ac.uk/advice/FilelevelMetadata.xhtml#Databases, Spreadsheets and Statistics.

Atkinson, R. J. C, 1946, *Field Archaeology*, Methuen & Co Ltd: London

Boddington, A, 1978, *The Excavation Record: Part 1 Stratification. Northamptonshire County Council*: Northampton

Boyle *et al*, 2016-2018, *Museums Collecting Archaeology (England) Report Year 1 to 3*, http://socmusarch.org.uk/projects/hesma-annual-survey-of-museums-collecting-archaeology-reports/

Brown, A and Perrin, K, 2000 *A Model for the Description of Archaeological Archives*. English Heritage

Brown, DH, 2011, *Archaeological Archives. A guide to best practice in creation, compilation, transfer and curation*. Archaeological Archives Forum, https://archaeologydataservice.ac.uk/catalogue/adsdata/arch-799-1/dissemination/pdf/AArchives_v2.pdf 2nd eds

Cattermole, A, 2017, *Review of the Standard of Reporting on Archaeological Artefacts in England,* Chartered Institute for Archaeology http://www.archaeologists.net/sites/default/files/7090FinalReport.compressed.pdf

Connolly, D, 2009, *Record Sheet and Report Templates Risk Assessment Forms & Other Guides*, (*BAJR Practical Guide Series*, 23) http://www.bajr.org/BAJRGuides/23.%20Recording,%20Reporting,%20Planning%20Conventions%20and%20other%20templates/23Templates.pdf

Dockrill, SJ, 2007, *Old Scatness Excavation Manual: A Case Study in Archaeological Recording*, Lerwick: Shetland Heritage Publications

Donnelly-Symes, B, 2019, *Planning for Archives: Opportunities and Omissions*. ALGAO: England

Etherton, D, Fleming, J, Honour, H, Pevsner, N, 1991, *The Penguin Dictionary of Architecture*. Penguin: New York 4th eds

European Commission, 23 January 2020, *Horizon 2020 Online Manual: Funding Guide* https://ec.europa.eu/research/participants/docs/h2020-funding-guide/index_en.htm

Faniel, I, Eric Kansa, E, Whitcher Kansa, S, Barrera-Gomez, J and Yakel, E, 2013, “The Challenges of Digging Data: A Study of Context in Archaeological Data Reuse.” JCDL 2013 Proceedings of the 13th ACM/IEEE-CS Joint Conference on Digital Libraries, 295-304. New York, NY: ACM. http://dx.doi.org/10.1145/2467696.2467712

Folk, 1988 (referenced in MOLAS 1994) unable to identify the original, possibly refers to publications in 1968 or 1980

Gregory, A, Heus, P, & Ryssevik, J, 2009, *Metadata.* (RatSWD Working Paper Series, 57). Berlin: Rat für Sozialund Wirtschaftsdaten (RatSWD). https://nbn-resolving.org/urn:nbn:de:0168-ssoar-409531

Harris, E, 1989 *Principles of Archaeological Stratigraphy* (2nd ed), Academic Press: London

Hirst, S, 1976, *Recordings on Excavation I: The Written Record.* Rescue Publication no. 7, Hertford

Historic England, 2018, Recording Manual, Historic England: Portsmouth

Historic England, March 2018, *Mendoza Review: Historic England's recommendations to DCMS on the future of archaeological archives, March 2018.* Historic England https://historicengland.org.uk/content/docs/consultations/he-response-to-dcms-mendoza-review-mar18-pdf/

Historic England, June 2018, *New Plan to Address England's Archaeology Archives Challenge,* Historic England https://historicengland.org.uk/whats-new/news/new-plan-englands-archaeology-archives-challenge/

Historic England, 2018, *Excavation Recording Manual*, Historic England: Portsmouth https://historicengland.org.uk/content/docs/research/historic-england-archaeological-recording-manual-2018/

Historic England, 2019, ADAPt: The Archaeological Digital Archiving Protocol Toolkit (https://historicengland.org.uk/research/methods/archaeology/archaeological-archives/adapt-tookit/)

Hodgson, JM, 1974, *Soil Survey Field Handbook,* Bartholomew Press: London

Huggett, J, 2014, 'Promise and Paradox: Accessing Open Data in Archaeology'. In: Mills, C, Pidd, M, and Ward, E. *Proceedings of the Digital Humanities Congress 2012*. Studies in the Digital Humanities. Sheffield: The Digital Humanities Institute, 2014. https://www.dhi.ac.uk/openbook/chapter/dhc2012-huggett

International Council on Archives, 2000, *ISAD (G): General International Standard Archival 2nd ed.* Subdirecctión General de Archivos Estatales: Madrid. https://www.ica.org/sites/default/files/CBPS_2000_Guidelines_ISAD%28G%29_Second-edition_EN.pdf

Jefferies, JS, 1977, Excavation Records: Techniques in Use by the Central Excavation Unit, Directorate of Ancient Monuments and Historic Buildings Occasional Papers. No.1 DoE London

Lee, E, 2015, *Management of Research Projects in the Historic Environment: The MoRPHE Project Managers' Guide*. Historic England: Swindon

Limbrey S, 1975, *Soil Science and Archaeology*, Academic Press: London

Mendoza, N, 2017, The Mendoza Review: an independent review of museums in England. Department for Digital, Culture, Media & Sport: London. https://www.gov.uk/government/publications/the-mendoza-review-independent-review-of-museums-in-england

Minister of State for the Cabinet Office, 2012 *Open Data White Paper: Unleashing the Potential* www.gov.uk/government/uploads/system/uploads/attachment_data/file/78946/CM8353_acc.pdf

Morgan, C, 2016, Analog to Digital: Transitions in Theory and Practice in Archaeological Photography at Çatalhöyük, Internet Archaeology 42. https://doi.org/10.11141/ia.42.7Munsell Color, 2009, *Munsell Soil Color Book* Munsell.

Museum of London Archaeology Service, 1994, *Archaeological Site Manual.* Museum of London: London 3rd eds

Postman, N, 1993, *Technopoly. The Surrender of Culture to Technology*. Vintage: New York

Proctor, M and Cook M, 2016, Manual of Archival Description 3rd Edition, Routledge: London

Parker Wooding, J and Forster, M, 2018, *Creating a sectoral standard and guidance for managing digital data generated from archaeological investigations: Starter for Ten Survey: Summary Report.*

Perrin, K, Brown, D, Lange, G, Bibby, D, Carlsson, A, Degraeve, A, Kuna, M, Larsson, Y, Palsdottir, S, Stoll-Tucker, B, Dunning, C, Rogalla von Bieberstein, A, 2014, *A Standard and Guide to Best Practice for Archaeological Archiving in Europe.* https://archaeologydataservice.ac.uk/arches/Wiki.jsp?page=Main

Powers, MC, 1953, "A New Roundness Scale for Sedimentary Particles", *Journal of Sedimentary Research* (1953) 23 (2): 117–119. https://doi.org/10.1306/D4269567-2B26-11D7-8648000102C1865D

Roskams, S, 2001, *Excavation*, Cambridge University Press, Cambridge

Surrey Archaeological Society, 2010, Excavation Recording Manual, Surrey Archaeological Society https://www.surreyarchaeology.org.uk/content/excavation-recording-manual-0

Wessex Archaeology, 2015, *Fieldwork Recording Manual,* Wessex Archaeology: Salisbury

Wise, A and Miller, P, 1997 *Why Metadata Matters in Archaeology*, Internet Archaeology 2. https://doi.org/10.11141/ia.2.5

APPENDIX 1: EXEMPLAR METADATA FOR CONTEXT DESCRIPTION

Field identifier	Field name	Field description	Class identifer	Data class	Alternative Names	Known references and standards	Known Visual Metadata	Thesauri
1	Context Number	The [sequential] [site or project] unique [numerical] identifier given to the unit of stratification to identify [cuts, layers, deposits and structural groups*] formatted as [a five digit number], covering number blocks [u-v] and [x-z], [include any divisions by site sub division or context type]. *These classes need definition elsewhere	1	identifier	Stratigraphic Unit Number		FALSE	FALSE
2	Site Code	The identifier given to the investigation of archaeological site, the parameters of which are defined within the scope of the project, unique to [insert organisation [repository or creator]], formatted as [three initials and year the event started, e.g. LBP1976] [Include whether part of a wider project]	1	identifier			FALSE	FALSE
3	Site Name	The name given to the investigation of archaeological site, the parameters of which are defined within the scope of the project [the project design/wsi]	1	identifier			FALSE	FALSE
4	Year	The year the investigation was initiated, [formatted as YYYY]	1	identifier			FALSE	FALSE
5	Grid Square(s)	The [South Western corner] co-ordinates for all the squares that the context occupies are included, as [semi colon] separated values; the grid is composed of [5 metre squares], data is formatted as [110/225 [easting/northing]]. The survey grid is documented in [insert doc/drawing reference].	2	location	Grid Reference		FALSE	FALSE
6	Site subdivision	The identifier given to sub division of the site, formatted as a [sequential number starting at 1] given to [trenches, area], trenches [A, B & D] were opened	2	location	Location, Trench		FALSE	FALSE
9	Context Type	Simple term definition of the type of context. Thesaurus: [cut, fill, layer, structural group, masonry, timber, skeleton, coffin] definitions of these terms can be found in the [supporting documentation]	3	attribute	Context Category		FALSE	TRUE
10	Context Simple Name	The simple name of the context is a brief description, such as Pit: fill; Surface: metalled; Layer. [Thesaurus is listed in [file name]	3	attribute	Context sub-type		FALSE	TRUE
11	Context description	Textual description of the context, noting additional useful information about the context not covered by other fields, including uncertainty	3	attribute			FALSE	FALSE
12	Orientation	The orientation of the longest side of the context, recorded to [8 cardinal points] formatted as [NEE -SWW]. If a post or stakehole, give direction of the top of the cut relative to the bottom, (for example; top is NE of base)	3	attribute	Direction		FALSE	FALSE
13	Length	The known length of the longest side of the context recorded in [metres to two decimal places]	3.1	attribute: dimension			FALSE	FALSE
14	Width	The known width of the context recorded in [metres to two decimal places]. Width is measured perpendicular to length.	3.1	attribute: dimension			FALSE	FALSE
15	Height/Depth	The known height or depth of the context recorded in [metres to two decimal places]. Height or depth is the vertical dimension of the context and measured between the lowest to highest points. Sloping stakeholes are measured along the axis of the hole.	3.1	attribute: dimension	depth; height; thickness		TRUE	FALSE
16	Diameter	The known diameter of the context recorded in [metres to two decimal places] for contexts that are circular	3.1	attribute: dimension			FALSE	FALSE
17	Volume	Total estimated volume of context in litres, [rounded to litres], [integer] formatted as [e.g. 16]	3.1	attribute: dimension			FALSE	FALSE
18	Lens	Description of the lens shape of the context in section. Formatted as floating point number to 2 decimal places]. [For example: '0.2 m thick at centre, lensing out to 0.14m']	3.1	attribute: dimension	Thickness & extent:		FALSE	FALSE
19	Extent	Textual description of whether the context was fully established	3.1	attribute: dimension			FALSE	FALSE
20	Horizon clarity	Description of how clearly defined the [lower] boundary of the deposit is against the context below it. Boundaries can either be [Sharp or Merging] and [even, undulating or convoluted], [from Limbrey 1975].	3.2	attribute: deposit	Boundary; Boundary next to the horizon;	Limbrey 1975, Hodgson 1976,	FALSE	TRUE

APPENDIX 1: EXEMPLAR METADATA FOR CONTEXT DESCRIPTION

Field identifier	Field name	Field description	Class identifer	Data class	Alternative Names	Known references and standards	Known Visual Metadata	Thesauri
21	Texture	The results of finger texturing test to determine the relative proportions of particles of different sizes by its mineral components of less than 2mm. Description of texture is made in terms of defined particle size grades and textural classes composed of mixtures of these. Particle size classes are based on [insert standard, i.e. British Standards Institution] for Clay, Silt and Sand. The finger texturing methodology used is : [insert method and/or reference, i.e. Historic England, 2015 appendix 1] to define soils as [insert terminology and bibliographic reference]	3.2	attribute: deposit	Composition/particle Grain size; Textural Class	Use of the Term Texture can be attributed to Hirst's adoption of the Limbrey standard in the published recording manual. Definition	TRUE	TRUE
22	Composition	This refers to the size of individual grains and clasts within the sediment matrix. Each fraction of the deposit which amounts to more than 10% of the whole deposit should be included in this element of the description. This includes clay, silt, sand and gravel particles and tile, bone, mortar, pottery, molluscs and organic material. The thesauri for this is in [the supporting documentation]	3.2	attribute: deposit			FALSE	TRUE
23	Colour	The colour of the deposit is recorded when [moist] and [freshly excavated], using the Munsell colour chart (XXXX). Modifier, hue and colour are all recorded to make up a three term description.	3.2	attribute: deposit	Soil Colour	Munsell	FALSE	TRUE
24	Munsell Number	The identifier for the Munsell colour description, formatted as [10YR 2/3]	3.2	attribute: deposit		Munsell	FALSE	FALSE
26	Compaction	The compaction of the soil was measured when the soil was [moist]. The following terms used; [loose, friable, firm, hard, cemented] are defined in the [supporting documentation].	3.2	attribute: deposit	MOLA		TRUE	TRUE
27	Archaeological Components: material	The material of archaeological components of the deposit, which are the result of human action and which [make up more that 10%] of the whole. The thesauri used is [in the supporting documentation].	3.21	attribute: inclusions	Inclusions: material		FALSE	TRUE
28	Archaeological Components: Frequency	The frequency of the archaeological components of the deposit, which are the result of human action, is recorded as [frequent', 'moderate' or 'occasional]. Frequency is determined from the [frequency chart in the supporting documentation, redrawn from Hodgson 1974].	3.21	attribute: inclusions	Inclusions: frequency	Hodgson 1974	TRUE	TRUE
29	Archaeological Components: size	The size of each archaeological component, which are the result of human action, are described as [Flecks (up to 6mm), small (6mm-20mm), Medium (21mm-60mm), Large (61mm-120mm). If above 120mm, the actual size should be given]	3.21	attribute: inclusions	Inclusions: Size		FALSE	TRUE
30	Archaeological Components: angularity	The shape of each inclusion is described using the following terms: Very angular, angular, sub-angular, sub rounded, rounded, well rounded, the diagram defining this is in the supporting documentation (after Powers 1953).	3.21	attribute: inclusions	Inclusions: Shape;	Powers 1953	TRUE	TRUE
31	Archaeological Components: fragmentation	Description of the fragmentation of the archaeological component described using the following terms: specified as 'flecks' (up to 6 mm), 'small' (6mm-20mm), 'medium' (20mm-60 mm) or 'large' (60 mm-l20 mm).	3.21	attribute: inclusions	Inclusions: Fragmentation	Molas 1994 (so far)	FALSE	FALSE
32	Coarse component: Sorting	The frequency with which particles of the same size occur within the context. For example, if the deposit consists of a single size coarse component it is 'well sorted'. Terms used to define the sorting are: well sorted; moderately sorted; poorly sorted; very poorly sorted, as defined in figure XXX (redrawn after Folk 1988).				Folk 1988	TRUE	TRUE
33	Coarse components: Frequency	The frequency of the elements of the deposit is recorded as ['frequent', 'moderate' or 'occasional]. Frequency is determined from the [frequency chart in the supporting documentation, redrawn from Hodgson 1974.]	3.22	attribute: stoniness	Stoniness: Abundancy		TRUE	TRUE
34	Coarse component: Size	The [range] of the size of the coarse components should be given as a [minimum and maximum dimension (in millimetres), for example 10-40mm]	3.22	attribute: stoniness	Stoniness: size		FALSE	TRUE

APPENDIX 1: EXEMPLAR METADATA FOR CONTEXT DESCRIPTION

Field identifier	Field name	Field description	Class identifer	Data class	Alternative Names	Known references and standards	Known Visual Metadata	Thesauri
35	Coarse component: Shape	A field assessment of the shape of the particles, defined as ['platy' (flat), 'elongate' (tube like) or 'spheroidal']	3.22	attribute: stoniness	Stoniness: shape		TRUE	TRUE
36	Coarse component: Angularity	The shape of each coarse component is described using the following terms: [Very angular, angular, sub-angular, sub rounded, rounded, well rounded, the diagram defining this is in the supporting documentation (after Powers 1953)]	3.22	attribute: stoniness	Stoniness: roundness	Powers 1953	FALSE	FALSE
38	Cut: Break of slope: top	Description of the degree with which the top surface of the edge of the cut breaks into the sides, using the following terms, ['sharp', 'gradual' or 'imperceptible'], defined using [the diagram] in the [supporting document]	3.3	attribute: cut	Cut: Sides; top		TRUE	TRUE
39	Cut: Gradient of Sides	The gradient of a sloping side is expressed as X in Y, where X is the horizontal distance between the top and the base and Y is the height difference [see figure X in the recording manual], for sides which are 'concave' but meets the base at a sharp angle, give an overall gradient. Gradient is not recorded when the break of slope is not perceptible. For multi-sided cuts each side of the cut is described, identified by [its cardinal point]	3.3	attribute: cut	Sides steepness; sides gradient		FALSE	FALSE
40	Cut: Sides shape	Description of the shape of the slope of the sides of the cut as ['vertical', 'convex', 'concave' or 'stepped']. For multi-sided cut each side of the cut is described, identified by [its cardinal point]. For post-holes whether they [taper or vertically drop] is described.	3.3	attribute: cut	Cut: Side of the cut		TRUE	TRUE
41	Cut: Opposing sides	Description of whether the opposing sides of a cut are symmetrical or asymmetrical	3.3	attribute: cut			FALSE	FALSE
42	Sides: smoothness	Description of the sides in terms of their smoothness or irregularity	3.3	attribute: cut	Cut: Slope of sides		TRUE	TRUE
44	Cut: Corners	The shape of the corners of the cut in plan, whether they are [rounded or squared]	3.3	attribute: cut			TRUE	TRUE
45	Cut: Base	Description of the base of the cut, noting whether it is ['flat', 'concave', 'sloping' (give the direction of the downward slope),'pointed', 'tapered' ('blunt' or 'sharp') or 'uneven']. In the case of post-holes note whether the base has [tapered point, tapered blunt point, tapered rounded point vertical and flat base], see figure x in the recording manual.	3.3	attribute: cut	Properties of base:	MOLAS 1994	TRUE	TRUE
46	Cut: Base; break of slope	Description of the degree with which the side of the cut breaks into the base, using the following terms, ['sharp', 'gradual' or 'imperceptible'], defined using [the diagram] in the [supporting documentation].	3.3	attribute: cut		MOLAS 1994	TRUE	TRUE
47	Cut: Shape in plan	Description of the shape of the top of the cut: [Irregular; Linear; Oval; Rectangular; Square; Sub-circular; Sub-rectangular].	3.3	attribute: cut		MOLAS 1994	TRUE	FALSE
48	Cut: Truncation	Description of whether the cut has its original shape and dimensions, and if it was truncated, what part and what truncated it.	3.3	attribute: cut			FALSE	FALSE
49	Post hole: Inclination of axis	Description of the inclination in the form of a gradient ['Y mm (vertical) in X mm (horizontal)'].	3.3	attribute: cut	Post holes and stake holes: Angle or rake		FALSE	FALSE
50	Post hole: Rake direction	Direction of the rake/angle of stake/post holes, recorded as the [one of 8] cardinal points it points towards, [formatted as NE].	3.3	attribute: cut			FALSE	FALSE
51	Post holes and stake holes: Point shape	The shape of stake or post holes in section: [Tapered; Tapered - blunt point; Tapered - rounded; Straight with flat base].	3.3	attribute: cut			TRUE	FALSE
54	Post Hole: Sides	Description of the side of post and stake holes, whether the sides [taper or drop vertically] to the base (this may help define whether the post was set or driven).	3.3	attribute: cut			TRUE	TRUE
55	Levels: OD Lowest:	Height measured above ordnance survey datum at the lowest point of the feature, relative to the [Newlyn] datum, measured in [metres] to two decimal places (e.g. 1.22 [mOD])	3.4	attribute: levels			FALSE	FALSE
56	Levels: OD Highest	Height measured above ordnance survey datum at the highest point of the feature, measured in [metres] to two decimal places (e.g. 1.22 [mOD])	3.4	attribute: levels			FALSE	FALSE

APPENDIX 1: EXEMPLAR METADATA FOR CONTEXT DESCRIPTION

Field identifier	Field name	Field description	Class identifer	Data class	Alternative Names	Known references and standards	Known Visual Metadata	Thesauri
57	Levels: OD Others	The vertical distance between the survey point of the context and the ordnance datum, for additional survey points to the highest and the lowest, separated by a semi-colon.(e.g. 1.22 [mOD])	3.4	attribute: levels			FALSE	FALSE
58	Conditions	Description of any conditions that would effect recording or investigation, i.e. lighting, weather, ground conditions	3.5	method and conditions			FALSE	FALSE
59	Extent of excavation	Description of the extent of excavation of a deposit, and how that was achieved [half-sectioned, excavated in quadrants or by slots, describing the portions that were excavated and the width of any sections/slots/sondages.	3.5	method and conditions			FALSE	FALSE
60	Method	The method of how the deposit was excavated - whether it was excavated by trowel, mattock/shovel or machine.	3.5	method and conditions			FALSE	TRUE
61	Contamination	An estimation of the possibility of post-depositional contamination of the deposit: These terms are used: [Probable – when disturbance is present, intercutting is noted or boundaries between contexts are particularly difficult to define. Possible - when there is a chance of movement of material between contexts but this is less likely. Unlikely - when dealing with discrete features and well-sealed contexts. This should not include residuality - when material that was already old was unintentionally incorporated as the deposit was formed (for example, Bronze Age flint working debris found in a Romano-British pit fill).]	3.5	method and conditions			TRUE	TRUE
63	Above	The context being described (X) is stratigraphically above (earlier than) the contexts listed (Y). X is above Y. Multiple contexts are separated by a [semi-colon]. Reciprocal relationship: Below.	4	relationship	Above/Earlier than; Stratigraphically Above; below		TRUE	FALSE
64	Below	The context being described (X) is stratigraphically below (later than) the contexts listed (Y). X is below Y. Contexts are separated by a [semi-colon]. Reciprocal relationship: Above.	4	relationship	Later than; above		FALSE	FALSE
65	Same as	The context being described (X) is the considered the same as (equal to) the context(s) listed (Y). X is the same as Y(n). Multiple contexts are separated by a [semi-colon]. No reciprocal relationship.	4	relationship			FALSE	FALSE
66	Filled by	Used for cuts only. The context being described (X) is filled by the deposit context(s) listed (Y). X is filled by Y(n). All the deposits which fill the cut, including linings etc. Multiple contexts are separated by a [semi-colon]. [Deposit, masonry, timber and human remains] context types can all be the fill of a cut. Reciprocal relationship: Fills.	4	relationship			FALSE	FALSE
67	Fill of	Used for deposits only. The context (X) being described is a fill of the context listed (Y). X is a fill of Y. [Deposit, masonry, timber and human remains] context types can all be the fill of a cut. Reciprocal relationship: Filled by.	4	relationship			FALSE	FALSE
68	Cut by	The context being described (X) is cut by the context(s) listed (Y). X is cut by Y(n). Multiple contexts are separated by a [semi-colon]. Used for deposits only, a cut can not cut another cut. Reciprocal relationship: Cuts.	4	relationship			FALSE	FALSE
69	Cuts	The context being described (X) cuts the context(s) listed (Y). X cuts Y(n). Multiple contexts are separated by a [semi-colon]. Used for cuts only, a cut can not cut another cut. Reciprocal relationship: Cut by.	4	relationship			FALSE	FALSE
70	Butts	The context being described (X) butts the context(s) listed (Y). X butts Y(n). Multiple contexts are separated by a [semi-colon]. This is used for structural elements where one is built up against an existing feature only. Reciprocal relationship: Butted by.	4	relationship			FALSE	FALSE

APPENDIX 1: EXEMPLAR METADATA FOR CONTEXT DESCRIPTION

Field identifier	Field name	Field description	Class identifer	Data class	Alternative Names	Known references and standards	Known Visual Metadata	Thesauri
71	Butted by	The context being described (X) is butted by the context(s) listed (Y). X is butted by Y(n). Multiple contexts are separated by a [semi-colon]. This is used for structural elements only, where one is built up against an existing feature. Reciprocal relationship: Butts.	4	relationship			FALSE	FALSE
72	Part of	The context you are describing is part of a [structural/interpretive] group. The contexts are contemporary with each other. Reciprocal relationship: Consists of.	4	relationship	Contemporary with:		FALSE	FALSE
73	Consists of	The context you are describing is a [structural/interpretive] group, which consists of the context(s) listed. Multiple contexts are separated by a [semi-colon]. Reciprocal relationship: Part of.	4	relationship			FALSE	FALSE
76	Contains	The context being described (X) wholly encompasses the context(s) listed (Y). X contains Y(n). Multiple contexts are separated by a [semi-colon]. Examples are a lens of a different material lying in an otherwise homogenous ditch fill; a masonry element included in the build of another (such as a window constructed at the same time as the wall surrounding it), or an infant skeleton lying within an extensive fill or layer with no indications of a grave cut surviving. This relationship must not be used to link cuts to their fills, or to create groups. Reciprocal relationship: Is Within.	4	relationship			FALSE	FALSE
77	Is within	The context being described (X) lies wholly within the context listed (Y). X is within Y(n). Examples are a lens of a different material lying in an otherwise homogenous ditch fill; a masonry element included in the build of another (such as a window constructed at the same time as the wall surrounding it), or an infant skeleton lying within an extensive fill or layer with no indications of a grave cut surviving. This relationship must not be used to link cuts to their fills, or to create groups. Reciprocal relationship: Contains.	4	relationship			FALSE	FALSE
78	Checked by	The name of the person who checked the record, formatted as [First Name Surname]	5	audit			FALSE	FALSE
79	Date Checked	The date the record was checked, formatted as [dd/mm/yyyy]	5	audit			FALSE	FALSE
80	Recorded by	The name of the person who completed the record, formatted as [First Name Surname]	5	audit	Completed by:		FALSE	FALSE
81	date recorded	The date the record was completed, formatted as [dd/mm/yyyy]	5	audit	Date compiled; Date completed		FALSE	FALSE
82	Excavated by:	The name(s) of all the people who excavated the context, formatted as [first name surname], multiple names separated by a [semi colon]	5	audit			FALSE	FALSE
83	Excavated date:	Date excavation of this context finished, formatted as [dd/mm/yyyy]	5	audit	Finished		FALSE	FALSE
98	Interpretation	Interpretative comments describing the context's character and function, free text field, e.g. [XXX]	7	interpretation			FALSE	FALSE
99	Discussion	Reasoning of the interpretation and expression of dismissals and doubts, positive and negative evidence that supports the interpretation, free text field.	7	interpretation			FALSE	FALSE
102	TYPE OF INTERVENTION	A simple phrase to describe the nature of the excavated intervention in relation to the feature, for example: half-sectioned, 1m slot, quadrant, 100% excavated.	8	Intervention			FALSE	FALSE
103	INTERVENTION DIMENSIONS width	The width of the excavated intervention, recorded in [metres to two decimal places, i.e. 1.25m, 0.01m]. The mean average must be recorded if of variable size. Width is measured perpendicular to length.	8	Intervention			FALSE	FALSE
104	INTERVENTION DIMENSIONS thickness/depth	The thickness/depth of the excavated intervention, recorded in [metres, i.e. 1.25m, 0.01m]. The mean average must be recorded if of variable size.	8	Intervention			FALSE	FALSE

APPENDIX 1: EXEMPLAR METADATA FOR CONTEXT DESCRIPTION

Field identifier	Field name	Field description	Class identifer	Data class	Alternative Names	Known references and standards	Known Visual Metadata	Thesauri
105	INTERVENTION DIMENSIONS length	The length of the excavated intervention, recorded in [metres, i.e. 1.25m, 0.01m]. The [mean] average must be recorded if of variable size.	8	Intervention			FALSE	FALSE
106	INTERVENTION DIMENSIONS diameter	The diameter of excavated circular interventions, recorded in [metres, i.e. 1.25m, 0.01m].	8	Intervention			FALSE	FALSE
107	Was context excavated?	Yes/No response to whether the context was excavated.	8	Intervention			FALSE	FALSE
108	Division Of	The context being described (X) is synthetic division of the context listed (Y), to allow deposits to be divided into a number of separate spatial elements to allow the spatial analysis of recovered finds or environmental material.	8	Intervention			FALSE	FALSE
109	Divided into:	The context being described (X) is synthetically divided into the contexts listed (Y), to allow deposits to be divided into a number of separate spatial elements to allow the spatial analysis of recovered finds or environmental material. Multiple contexts are separated by a [semi-colon].	8	Intervention			FALSE	FALSE

APPENDIX 2: IDENTIFIED FIELD NAMES EXCLUDED FROM EXEMPLAR TEMPLATE

Field identifer	Field name	Field description	Class identifier	Data class	Alternative Names	Known references and standards	Visual Metadata Known	Thesaurii
7	Included In		2	location				FALSE
8	Includes		2	location				FALSE
37	Coarse component: Lithology	The lithology of the coarse component	3.22	attribute: stoniness	Stoniness: lithology			FALSE
52	Cut: outline edges		3.3	attribute: cut				FALSE
53	Cut: Profile Overall		3.3	attribute: cut				FALSE
74	Covers		4	stratigraphic				FALSE
75	Covered by		4	stratigraphic				FALSE
84	Bulk Finds		6	cross referencing				FALSE
85	Registered Artefacts:		6	cross referencing	Small finds			FALSE
86	Samples:		6	cross referencing	Environmental Samples			FALSE
87	Drawing numbers		6	cross referencing				FALSE
88	Photograph numbers		6	cross referencing				FALSE
89	Site Book References		6	cross referencing				FALSE
90	Has		6	cross referencing				FALSE
91	Intersects		6	cross referencing				FALSE
92	Is Intersected By		6	cross referencing				FALSE
94	parent context		7	interpretation				FALSE
95	Preliminary Phase		7	interpretation	Provisional phase			FALSE
96	Preliminary period		7	interpretation	Provisional period			FALSE
97	Preliminary date		7	interpretation	Provisional date:			FALSE
100	INTERPRETATIVE CATEGORY		7	interpretation				TRUE
101	Checked interpretation		7	interpretation				FALSE
110	Contemporary: Feature Label		4	stratigraphic				FALSE
111	Idiosyncratic							FALSE
112	Sealed By		4	stratigraphic				FALSE
113	Seals		4	stratigraphic				FALSE