Dig Digital.

Work Digital. Think Archive. Create Access.

A guide to managing digital data generated from archaeological investigations

Infosheet #4 – metadata and documentation

Metadata and documentation are two key factors that are intrinsic to digital data management. They can also create barriers to successful archive management and deposition and are often perceived to be both complex and unnecessary.

An archaeological project will normally include different levels of metadata. Often described as *data about data*, metadata is the information that allows machines and people to fully understand an archive. Types of information relating to a project will normally include different types of metadata, which can be broadly defined as

- collection level – information that describes the archive itself, including project details and a summary of the data included in the archive

- contextual or reference level – information about how the data is described, such as the definition of field names within a table, which can be relevant to groups of data objects.

- descriptive or file level – information that describes technical and non-technical elements at a file level, such as the camera used or the image caption, which can therefore be unique.

Documentation refers to supporting information that provides a description of processes, policies or standards that influence how data has been collected, processed or managed. These might be internal documents, such as an organisational site manual, or a record of the type of processing that a dataset has undergone, such as for geophysical survey data.

The Dig Digital resource is an Archaeological Archives Forum guidance document that supports CIfA Standards and guidance. It was created by DigVentures, in partnership with CIfA, and funded by Historic England.

You can find the full resource online at: https://www.archaeologists.net/digdigital

Data without metadata cannot be understood

The images below provide a visual summary of why we need information in the form of metadata alongside the data we include in digital archives.

C	09/06	1999	12.00	5	0
D	09/06	1999	15.00	2	1
D	12/06	1999	09.00	9	1
C	14/06	1999	10.30	4	0
C	14/06	1999	17.15	16	0

Type	Date	Year	Time	Age	m/f
C	09/06	1999	12.00	5	0
D	09/06	1999	15.00	2	1
D	12/06	1999	09.00	9	1
C	14/06	1999	10.30	4	0
C	14/06	1999	17.15	16	0

Type: 'C' cat, 'D' dog
Date is in dd/mm format
Year is in yyyy format
Time uses 24-hour clock
Age is in years at last birthday
m/f is sex: '0' female, '1' male

What is metadata?

Metadata is the information that makes digital archive materials accessible to people and – importantly – to machines. The inclusion of structured information about the digital data that makes up an archive supports FAIR principles for data management, making archives *findable, accessible, interoperable*, and *reusable*. Providing metadata ensures that the project will be discoverable, securely archived and usable for future research. As such, metadata is something that the trusted digital repository will require on deposition of an archive and it is important to understand what information is needed and how, so that workflow processes support collection of metadata throughout the project.

There are three common types of metadata found in archaeological datasets and, although people use different terms for them, they tend to refer to similar groups. Each of the groups can include different types of information, some complex and some relatively simple.

- **Collection level** – information that describes the archive itself, including project details and a summary of the data included in the archive. Collection-level metadata is important because it enables discoverability, identifies not just the project but also the data creators and provides a summary of what users should expect. This data will be included within the data management plan (DMP), and might include

    - project title and any identifiers (such as site code)

    - description of the main aims of a project and the summary of the contents of the dataset

    - keywords that describe the content and use standardised terms, such as the FISH Monument Type Thesaurus

    - project team, including who may have created the data and including project stakeholders, such as funding agencies

    - copyright information, including the name of the copyright holder – often the organisation that delivered the project

    - resource type, indicating what types of information the archive includes, such as primary data, processed data or reports – or all the above!

- **Contextual or reference** – information about how the data is described, such as the definition of field names within a table. This can be information about an individual data object, but is often relevant to groups of data objects. This information is important as it provides clarity about what the data is and what it means, facilitating reuse of the archive and supporting interoperability. This type of metadata might include

    - description of how the data collector defines basic recording terms, such as 'context', 'feature', 'colour' or 'compaction'

    - explanation of how measurements or quantities have been recorded, such as 'width', 'thickness', 'volume' or 'length'

    - detail and signposting of how data might be formatted or inputted, such as use of standardised terms, such as 'material type', 'orientation' or 'period' selected from a drop-down list, or a free-text description of 'interpretation'

    This information can be strongly linked to organisational approaches and might be described in an excavation manual or organisational policy.

- **Descriptive and technical** – information that describes technical and non-technical data elements at a file level – such as the camera used to take a picture, the image caption or location of the subject – that is connected to the individual data object and can be unique. This file-level information is important as it will signpost how you can access the data (eg, the file type and software used) as well as provide a description of the individual file (eg the filename and caption).

Metadata for machines

Making your data findable, accessible, interoperable and reusable (the FAIR principles) is what machines are really good at, if the right information is available. That information can be very complex or very simple and will tell machines different things about the archive.

The soup can analogy might perhaps be taken further to illustrate the range of information that could be classed as metadata. The range goes from a can with no information at all (A), through to summary information about the collection itself (B), to covering the technical details of its contents (C) and some additional information how the data can be used (D). For people, the labels on food provide information about the ingredients, where to food comes from and who made it, and how we might eat it. In the same way that the quality of information on a can of soup might help us decide if we want to eat it for lunch, the quality of metadata accompanying an archive allows machines to determine how useful the data itself is and how it can be used.

A	B	C	D
No data *similar to* No metadata!	Summary information *similar to* Collection level?	Nutritional information *similar to* Technical and descriptive metadata?	Recipe idea *similar to* Metadata about use and re-use?

Even where the data from an archive cannot be made available, perhaps because the site itself is sensitive or the data embargoed, metadata can still be deposited. This means that the research project is not invisible to everyone, but there is control over how people can discover and then access the data. By depositing the metadata with a trusted digital repository, your archive will remain both discoverable and accessible to the level required. Where data is embargoed until a certain date, or conditions of access are limited in some way, the information included within the metadata makes that clear to machine readers.

Standards and documentation

The metadata that accompanies an archive allows machines to understand what the data is, what it might relate to and how it can be accessed. When it comes to specific disciplines, descriptive metadata will aid data discovery, aggregation and re-use, and can often be linked to standards, policies or supporting information that is specific to sectors, organisations or even people. Supplemental documentation can be included that acts as a user guide for the data and helps future users understand and trust the project data, ultimately supporting re-use.

A trusted digital repository will support data sharing and aggregation, meaning that information about the data archive is machine readable and/or shared with other catalogues. You can see how this idea of standardised terms and accessible or aggregated data works with resources like Heritage Gateway, Canmore, Ariadne, the Archaeology Data Service or the Portable Antiquities Scheme database. When machines are able to read and understand information from multiple digital archives, search engines become even more powerful – fuelled by all that metadata. By creating information that works for machines, archives work more effectively for people too.

Controlled vocabularies. By using controlled vocabularies or standardised terms within metadata, machines can look for patterns and connections between information sets that we might be interested in – linking digital archives to create a far bigger dataset and aggregated resource. In the case of archaeology, this might include terms used to describe monuments, periods, landscape features, etc. The digital archive and aggregated resource is only ever as powerful as the information it stores, but the good news is that the heritage sector has things in place that already support this, such as the heritage standards available as part of the *Forum on Information Standards in Heritage* (FISH) resource. As these standards may be complex, a signpost from the metadata to the standard being used is helpful – or including the document itself where possible. When using archives, this extra layer of information allows us to see the behind the scenes, so we know what the data creators mean when they use certain terms.

Specialist typologies. A pottery specialist may develop a typological sequence for a specific site to analyse the assemblage. That typology is published and accessible, but remains site specific. The archive would contain the pottery report along with spreadsheets that include the raw data. The accompanying metadata would describe who created the data, as well as the technical information such as field headings in the spreadsheet. The typology might be published or provided as supplementary documentation and the metadata should also identify this as a key source accessible to users. Users can follow up those references and perhaps visit the museum to look at specific examples of each type, comparing them to other sites and making interesting research connections. Machines might read the metadata and understand the type of collection, but are unable to use the pottery type series to link to comparable sites as the series is unique.

Where assemblage types are applicable widely then a standardised typology would be used, referring to forms and fabrics that can occur on many different sites. That typology would also be available and accessible, meaning that others can use the series in the same way. In this case, a machine can understand the data and quickly make links and connections between datasets.

Documenting processes. A more technical example of documentation would be a description of the processing raw data has undergone as part of analytics and interpretive activities, or to create final

outputs. This form of project documentation might be relevant to geophysical data or images used for creation of photogrammetry models. Essentially, some elements of a digital archive might need to include a raw data set and a processed dataset. In these cases, the project archive would benefit from the inclusion of documentation describing what has happened to the data as it travelled through the project pipeline from collection to output. This level of documentation supports the authenticity of the derived dataset and describes its provenance, making it transparent that data has undergone certain processes and what they are.

Planning for metadata

Archaeologists are not new to collecting and sharing data about our archives in a way that supports use of site records by colleagues, researchers, or museums. The various registers used for context records, artefacts, photographs and drawings document the process of archaeology and make it accessible for others. Working with physical archive elements is pretty much embedded within our everyday project delivery, and the information we collect along the way forms part of that process.

So, as with all aspects of the project archive, considering the requirements of the repository from the project planning stages means that the whole team can support the process. For digital data within a working project archive, this may mean using specified filename formats and versioning from the outset, keeping certain types of files in particular folders, or ensuring that data provided by team members outside of an organisation (such as project specialists) are asked to provide digital data and supporting information on submission of reports.

Planning how the metadata required for a project can be collected effectively will support the resources needed to prepare a digital archive for deposition. Ensuring that all team members are fully aware of the need for metadata, in the right structure and format, will also prove effective for both time and budget management.

- ✓ Do not leave metadata to the end of the project – documenting what, how, why, who and when for data within a project can be time consuming and difficult if left to the very end.
- ✓ Do make metadata creation part of the research process – document as you go and make every team member part of the workflow.
- ✓ Create a data management plan before the project delivery stage and data collection starts, and implement the plan.
- ✓ Organise your data – use conventions for file naming and version control, structure your data files and use open file formats from the outset where possible.
- ✓ Create or source templates for metadata creation that will work with the intended trusted digital repository – they may have templates or requirements.
- ✓ Create an organisational-level template for contextual or reference metadata types. This Dig Digital resource includes some example definitions of record field names to kick this process off.
- ✓ Identify how, when and by whom metadata tables should be populated for each data type.
- ✓ Be aware of supplemental documentation that will support the digital archive – such as data processing documentation, links to thesauri used, or the organisation's excavation manual.

Metadata examples

The examples below provide an explanation of common field names found within archaeological archive. The examples provide the source for the term (eg the register), the field name, description given and include an example of the information provided. Each example refers to the appropriate documentation or controlled terms where relevant.

Context number

Register	Field name	Field description	Example
Context descriptions	Context number	The sequential unique numerical identifier given to the unit of stratification to identify cuts, layers, fills, masonry, skeleton, and timber. The unique ID number consists of the trench number (eg '1') followed by a three digit number starting at '001'.	1001
Documentation	2021 ABC Ltd - Excavation manual.		
Controlled terms	n/a		

Notes the convention of the *how* the identifier is created, as well as the nature and meaning of the field contents. Documentation includes the organisational excavation manual, which will be deposited with the archive.

Colour

Register	Field name	Field description	Example
Context descriptions	Colour	Used to describe contexts identified as layers and deposits. This field notes the colour of the deposit/layer. The colour refers to the majority of the context and uses the Munsell colour chart to identify the colour (see ABC Ltd Excavation manual 2021 on methodology of recording the deposits colour in the field). The colour is noted as the code as identified in the Munsell colour chart and its corresponding name.	2.5Y 4/2 Dark greyish brown
Documentation	2021 ABC Ltd - Excavation manual.		
Controlled terms	Munsell Color (Firm). Munsell Soil Color Charts: with Genuine Munsell Color Chips. Grand Rapids, MI: Munsell Color, 2010.		

Notes the convention of the *how* the colour code/name is created, and links to external references. Documentation includes the organisational excavation manual, which will be deposited with the archive. Controlled terms refer to the Munsell colour chart and a full bibliographic reference is given.

Interpretation

Register	Field name	Field description	Example
Context descriptions	Interpretation	Interpretative comments describing the character and function of the context (free text) made by the excavator.	Layer of clay at base of stone lined box, probably to pack the cut to support the stone slab 2334. Perhaps the stone lined box held water and the clay was to waterproof the box.
Documentation	2021 ABC Archaeology Ltd - Excavation manual.		
Controlled terms	n/a		

Notes the nature and meaning of the field contents. Documentation includes the organisations excavation manual, which will be deposited with the archive.

Period (MIDAS)

Register	Field name	Field description	Example
Context descriptions	Period (MIDAS)	The period(s) the context dates to using MIDAS FISH thesaurus for chronology (2021) available here: http://www.heritage-standards.org.uk/chronology/	Victorian
Documentation	n/a		
Controlled terms	MIDAS FISH thesaurus for chronology (2021) available here: http://www.heritage-standards.org.uk/chronology/		

Notes the source of the *where* the period terms originate and describes nature and meaning of the field contents. Controlled terms include MIDAS FISH chronology. A URL or DOI can be provided in the metadata to link to the thesaurus.

Feature narrative

Register	Field name	Field description	Example
Feature register	Feature narrative	The interpretative comments (free text) on the narrative of the feature made by the excavator. This is a longer description of how the feature developed and what its character and function was.	This curvilinear stone structure is likely part of the internal structure of the Norman limekiln. It is possible that a symmetrically opposite feature of a similar construction exists to the northern end of F217, but further excavation would be required to prove this. Collectively, these internal 'walls' could have acted as channels for heat to surround the limestone being roasted within.
Documentation	2021 ABC Archaeology Ltd - Excavation manual.		
Controlled terms			

Notes the nature and meaning of the field contents. Documentation includes the organisations excavation manual, which will be deposited with the project archive.

Material type (MIDAS)

Register	Field name	Field description	Example
Finds Register	Material type (MIDAS)	The material the find(s) are made from, using the MIDAS FISH thesaurus for object material (2021) Can be accessed: http://www.heritage-standards.org.uk/wp-content/uploads/2021/02/ObjectMat_class.pdf	Lead
Documentation			
Controlled terms	MIDAS FISH thesaurus for object material (2021) available here: http://www.heritage-standards.org.uk/wp-content/uploads/2021/02/ObjectMat_class.pdf		

Notes the source of the *where* the material type names originate and describes nature and meaning of the field contents. Controlled terms include MIDAS FISH object material thesaurus. A live link is provided to access the thesaurus.

Find type (Other)

Register	Field name	Field description	Example
Finds Register	Find type (Other)	This field notes the type of find(s). The vertebra were identified using identification guides: BoneID; Hillson 2003; 2005. The marine mollusc remains were identified using identification guides: Hayward and Ryland 1995; Marine Bivalve Shells of the British Isles. The fish bone remains were identified using identification guides: Archaeological Fish Resource; Camphuysen and Henderson 2017; Nabone Fish; Osteobase; Wheeler and Jones 1989. The animal bone/ shell analysis was undertaken within guidance outlined in CIfA 2014; Baker and Worley 2019. The pottery analysis was undertaken within guidance and standards outline in Barclay et al. 2016. See *Pottery specialist glossary, 2022* for pottery type definitions.	Midlands Purple-type ware
Documentation		Chartered Institute for Archaeologists (CIfA) 2014. Standard and guidance for the collection, documentation, conservation and research of archaeological materials. Can be accessed: https://www.archaeologists.net/sites/default/files/CIfAS&GFinds_1.pdf Baker, P. and Worley, F. 2019. Animal Bones and Archaeology - Recovery to archive. Historic England Handbooks for Archaeology. Barclay, A., Knight, D., Booth, P., Evans, J, Brown, D., and Wood, I. 2016. A standard for pottery studies in archaeology. Prehistoric Ceramics Research Group / Study Group for Roman Pottery / Medieval Pottery Research Group.	
Controlled terms		BoneID. http://www.boneid.net Accessed March 2020. Hillson, S. 2003. Mammal Bones and Teeth. An introductory guide to methods of identification. London: Institute of Archaeology, University College London. Hillson, S. 2005. Teeth. Second Edition. Cambridge Manuals in Archaeology. Cambridge: Cambridge University Press. Hayward, P. J. and Ryland, J. S. (eds) 1995. Handbook of the Marine Fauna of North-West Europe. Oxford: Oxford University Press. Marine Bivalve Shells of the British Isles. https://naturalhistory.museumwales.ac.uk/British bivalves. Accessed March 2020. Camphuysen, C. J and Henderson, P. A. 2017. North Sea fish and their remains. Royal Netherlands Institute for Sea Research and Pisces Conservation Ltd. Nabone Fish. https://www.nabohome.org/products/manuals/fishbone/fish/fish.html Accessed March 2020. Osteobase. http://osteobase.mnhn.fr Accessed March 2020. . Wheeler, A. and Jones, A. K. G. 1989. Fishes. Cambridge Manuals in Archaeology. Cambridge: Cambridge University Press. Pottery specialist. 2022. Pottery types glossary.	

Notes the source of the *where* the find type names originate and describes nature and meaning of the field contents. Documentation includes CIfA guidance, and published standards and guidance. A live link is provided to access the CIfA guidance. Full bibliographic references are provided for published works. Controlled terms include specialist thesaurus. Live links are provided to access the online references. Full bibliographic references are provided for published works. The specialist glossary will be deposited with the archive.

Quantity

Register	Field name	Field description	Example
Finds Register	Quantity	The quantity of fragments of that material type within the context/sample or included in the special find number.	10
Documentation	2020 ABC Archaeology Ltd - Post-excavation processing manual		
Controlled terms			

Notes the nature and meaning of the field contents. Documentation includes the organisation's post-excavation processing manual, which will be deposited with the archive.

Weight (g)

Register	Field name	Field description	Example
Finds Register	Weight (g)	The weight in grams to 2 decimal places of the material type within the context/sample or of the special find. See ABC Ltd Post-excavation processing manual, 2020 for the make and model of weighing scales used (this will vary depending on the weight of the material).	23.12
Documentation	2020 ABC Archaeology Ltd - Post-excavation processing manual		
Controlled terms			

Notes the unit used, the format and *how* this measurement is created. Notes the nature and meaning of the field contents. Documentation includes the organisation's post-excavation processing manual, which will be deposited with the archive.

Direction

Register	Field name	Field description	Example
Photo register	Direction	The direction in which the photo is looking. Direction should be described as one of eight cardinal points (north, north east, east, south east, south, south west, west, north west).	north
Documentation	2021 ABC Archaeology Ltd - Excavation manual.		
Controlled terms	north, north east, east, south east, south, south west, west, north west		

Notes the nature and meaning of the field contents. Documentation includes the organisations excavation manual, which will be deposited with the archive. There are eight terms (cardinal points) that can be used in this field, these are listed.

Date (photo taken)

Register	Field name	Field description	Example
Photo register	Date (photo taken)	The date the photo was taken. Formatted dd/mm/yyyy.	25/02/2022
Documentation			
Controlled terms			

Notes the nature and meaning of the field contents, and the exact formatting of the data.

Copyright holder organisation

Register	Field name	Field description	Example
Photo register	Copyright holder organisation	The copyright holder for the image (only applicable if an organisation holds the copyright).	ABC Archaeology Ltd
Documentation			
Controlled terms			

Notes the nature and meaning of the field contents.

Condition

Register	Field name	Field description	Example
Sample register	Condition	This notes the preservation conditions of the context the sample is recovered from (as excavated, see *ABC Ltd Excavation manual 2021* on methodology of recovering samples on site). Use: Dry, Moist, Waterlogged	Moist
Documentation	2021 ABC Archaeology Ltd - Excavation manual.		
Controlled terms	Dry, Moist, Waterlogged		

Notes the nature and meaning of the field contents. Documentation includes the organisation's excavation manual, which will be deposited with the archive. There are three controlled terms, these are listed.

Reason for sample

Register	Field name	Field description	Example
Sample register	Reason for sample	Describes why the sample was taken (free text). This should highlight research questions the sample could answer as described in the Project Design.	The layer (2300) capped features F201 and F202. There was burnt material in deposit (2300) and it therefore seemed suitable for radiocarbon dating. Dating deposit (2300) would provide a Terminus ante quem for features F201 and F202.
Documentation	2021 ABC Archaeology Ltd - Excavation manual. 2022 Project Design.		
Controlled terms			

Notes the nature and meaning of the field contents. Documentation includes the organisations excavation manual and the specific Project Design, both will be deposited with the archive.