



Dig Digital.

Work Digital. Think Archive. Create Access.

A guide to managing digital data generated from archaeological investigations.

Infosheet #1 – FAIR principles and archaeology

The Dig Digital resource sits within the wider context of digital data management in research projects, ensuring that the standards of archaeological research undertaken across all sectors is in line with current practice and expectations. Academic research funding bodies, such as the Arts and Humanities Research Council, require the projects they support to include articulated plans for data management.

The FAIR guiding principles for data management and stewardship ([Wilkinson et al 2016](#)) provide a useful definition of what good data management is: Findable, Accessible, Interoperable and Reusable (FAIR) – see www.go-fair.org/fair-principles/. These foundational principles are easily mapped onto our professional imperative as archaeologists, being consistent with the requirements of ClfA Standards and guidance that archaeological archives are ordered, stable and accessible. The FAIR principles therefore provide a useful platform from which to approach digital archives.

This infosheet summarises how the FAIR guiding principles relate to archaeological projects, standards and good practice, and archives management. Each element is included below in a simple table that includes the Go Fair definition, a practical example and modes of compliance. If you fancy exploring the FAIR principles further, we'd recommend taking a look at [A FAIRy Tale](#).

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

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

Findable – how to make your archive easy to find

Go Fair definition	Components	In practice	Monitoring
<p>F1: (Meta)data are assigned globally unique and persistent identifiers</p>	<p>Your data archive has a unique identifier, such as a <u>DOI</u> (Digital Object Identifier) assigned by the repository, which won't be used for anything else.</p> <p>The identifier can be cited in the same way a publication can, and data can be located by people and machines.</p> <p>The identifier will be persistent, which means your data and metadata can be found (and re-found) for a very long time.</p>	<p>Update the project details on relevant reporting forms, such as OASIS, with the archive identifier to enable stakeholders to easily locate the archive.</p> <p>Use <u>Orcid Identifiers</u> for individuals and research organisations.</p> <p>Ensure metadata meets international standards such as <u>Dublin Core</u> and uses <u>standardised language</u>.</p>	<p>Does the project have an up-to-date OASIS record that signposts the location of the data archive?</p> <p>Does the data archive have a unique and persistent identifier?</p> <p>If not, is this intended?</p>
<p>F2: Data are described with rich metadata</p>	<p>Data is discoverable with machine readable metadata, which describes the context, quality and condition, and characteristics of the data.</p>	<p>Provide metadata that richly describe the research and archive, in line with standards provided by a trusted digital repository (TDR).</p> <p>Collect metadata as the project progresses, to make archive preparation easier.</p> <p>Use standardised and explicit language to describe the research and the resulting archive.</p> <p>Ensure other heritage information services (eg Historic Environment Record) and archive repositories (eg museums) are provided with a rich description of project, site and archive.</p>	<p>Has metadata been collected using TDR standards?</p> <p>How will users understand the extent of the research, as well as its intent, results and archive content?</p> <p>Is, or will, the data be easily located from available information?</p> <p>Does the intended TDR have a searchable index?</p> <p>How does the metadata support discoverability of the dataset, and clearly cite the identifier for the archive?</p>

Go Fair definition	Components	In practice	Monitoring
<p>F3: Metadata clearly and explicitly include the identifier of the data they describe</p>	<p>The metadata and data can be easily linked by signposting the identifier within the metadata.</p>	<p>Ensure metadata cross references all identifiers for the project, including site codes, national and regional list numbers, and museum accession numbers.</p> <p>Museums should ensure use of an <u>MDA code accession number</u>.</p>	<p>Does the project cross reference other relevant identifiers in planning documentation, such as the data management plan (DMP)?</p>
<p>F4: (Meta)data are registered or indexed in a searchable resource</p>	<p>The repository index is publicly accessible and searchable by people and machines.</p>	<p>Deposit the resource with a publicly accessible TDR.</p> <p>Ensure that the TDR provides a unique global identifier and that it is cited within the OASIS (or equivalent) record.</p> <p>Ensure metadata can be easily referenced by users.</p>	<p>Does the intended TDR have a publicly searchable index?</p> <p>Is the intended TDR index searchable by machines?</p> <p>Is the metadata easy to reference clearly?</p>

Accessible – how to make your archive easy to access

Go Fair definition	Components	In practice	Monitoring
<p>A1: (Meta)data are retrievable by their identifier using a standardised communication protocol</p>	<p>The data and metadata should be retrievable without the need for special tools or communication methods.</p> <p>For example, internet users can retrieve (meta)data by clicking a link to a webpage or other online location.</p>	<p>Ensure both the metadata and, ideally, the data archive itself is simple to access.</p> <p>Check that the intended TDR provides access to the meta(data) without the need for additional tools or barriers, such as specialised or propriety tools.</p> <p>If data is sensitive or embargoed, to remain FAIR, provide details of with whom access to the data can be discussed.</p> <p>Ensure the protocol for access is clear and explicit in the metadata.</p>	<p>Will the data and associated metadata be easily accessible via the intended TDR?</p> <p>If access to the meta(data) is not immediately possible, is it clear who can be contacted to discuss access?</p>
<p>A1.1: The protocol is open, free and universally implementable</p>	<p>The means of access (the protocol) should be free and open, meaning that anyone with a computer and the internet can retrieve at least the metadata at no cost.</p>	<p>Choose Open Access where possible.</p> <p>This means that both the metadata and, ideally, the data archive itself should be free and open to access.</p> <p>In addition to meeting FAIR principles, ClfA Standards require that the archaeological archive is accessible, which includes both metadata and data.</p>	<p>Is the method of access free and open?</p> <p>Is access to the data and metadata provided at no cost?</p>
<p>A1.2: The protocol allows for an authentication and authorisation where necessary</p>	<p>Conditions of access to data are made clear.</p> <p>Where research data is embargoed, both people and machines can understand the requirements.</p>	<p>Ensure the archive and metadata are easy to access, and conditions of that access clearly stated.</p> <p>Where data is embargoed, ensure metadata remain accessible, providing details about the archive, its creators and contents.</p> <p>Ensure any embargoes are reasonable.</p>	<p>Are the conditions of access clearly identified?</p> <p>Where research data is embargoed, are the conditions of access made clear?</p>

Go Fair definition	Components	In practice	Monitoring
<p>A2: Metadata should be accessible even when the data is no longer available</p>	<p>Metadata that relates to the archive remains accessible, even when the data itself is no longer available.</p>	<p>Ensure that the TDR supports long-term preservation and access to the archive holdings, maintaining the data and metadata in perpetuity.</p>	<p>If the data is no longer accessible, can the metadata be retrieved?</p>

Interoperable – how to make your data understood

Go Fair definition	Components	In practice	Monitoring
<p>I1: (Meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation</p>	<p>Digital data and metadata are uploaded to a TDR that can exchange and use information from other platforms, facilitating data aggregation and cross-searching.</p> <p>Researchers should be able to use the research data, understand key terms used and easily integrate with other datasets.</p>	<p>The words and terms we use can often be ambiguous so, wherever possible, use shared vocabularies and data standards.</p> <p>Where available, use standardised <u>Heritage Data vocabularies and thesauri</u>.</p> <p>Use common and future-proof formats, avoiding use of proprietary software and making data available in open formats where possible.</p>	<p>Does research use and cite standardised vocabularies?</p> <p>Does the intended TDR support interoperability through use of standardised vocabularies and data exchange?</p>
<p>I2: (Meta)data use vocabularies that follow the FAIR principles</p>	<p>The standardised vocabularies used are well documented and accessible to anyone who uses the dataset.</p>	<p>Use community recognised and published terminologies, identifying the source and citing clearly.</p> <p>Check the intended TDR supports the use of sustainable and open vocabularies.</p>	<p>Does the research data and metadata use vocabularies that meet FAIR principles, where possible?</p>
<p>I3: (Meta)data include qualified references to other (meta)data</p>	<p>If digital data relate to, or utilise other datasets, including metadata, references and links should be explained informatively, clearly citing relevant data by using a persistent identifier.</p>	<p>Ensure the TDR supports qualified referencing with and between publication datasets and resources, so that both machines and humans can understand how data relate to each other.</p>	<p>Where other datasets are used or linked to the research, are they identified in the project data and metadata?</p>

Reusable – how to make your data easy to use

Go Fair definition	Components	In practice	Monitoring
<p>R1: (Meta)data are richly described with a plurality of accurate and relevant attributes</p>	<p>Data are easier to reuse for both humans and machines where the context of the data are clear.</p> <p>Similar types of information should be organised in a standard way and use well-established and sustainable file formats.</p> <p>Documentation should use metadata that follow a common template and a common vocabulary.</p> <p>A full description of the data and where they came from must be provided to enable users (both machine and human) to decide whether the data are useful.</p>	<p>Follow guidelines for data and metadata compilation provided by the TDR.</p> <p>Include descriptions of data, as well as the relationship between archive elements and any data processing steps taken.</p> <p>Ensure the dataset is fully indexed and its relationship with physical data is understood.</p> <p>Maintain data integrity – ensure the selected digital data archive that will be transferred to the TDR is complete and authentic.</p>	<p>Is the archive fully described and documented with optimal metadata created to the standards of the intended TDR?</p> <p>Is the context of the data clear and described?</p> <p>Are you able to understand how the data have been created?</p> <p>Are the data and metadata complete?</p>
<p>R1.1: (Meta)data are released with a clear and accessible data usage license</p>	<p>The conditions under which the data can be used should be clear to machines and humans.</p> <p>The data archive should have clearly defined terms of access and reuse, within collection-level and technical-level metadata.</p>	<p>Follow guidelines from the TDR on the appropriate licence for your archive, which will often be part of the Creative Commons suite.</p> <p>Ensure that you've considered and clearly described the legal usage rights and licence needed to support re-use of (meta)data.</p> <p>Ensure all data owners/copyright holders agree to the terms of deposition with the repository.</p> <p>If the archive includes data from other people, ensure it is clear who created the data and how it fits into your project archive.</p>	<p>Is it clear how the data can be used?</p>

<p>R1.2: (Meta)data are associated with detailed provenance</p>	<p>Both machines and humans will need to understand whether the data are reusable in a different context.</p> <p>This includes being able to ascertain how the data has been created, who created it, what equipment was used and whether it has been processed.</p>	<p>Make sure that the origins of the data are clear, in terms of both technical provenance and the context of the project.</p> <p>Include information about the data creators, as well as how the data should be cited and acknowledged.</p>	<p>Is it clear how the data has been created, who created it, and how it can be cited?</p>
<p>R1.3: (Meta)data meet domain-relevant community standards</p>	<p>Standard ways for organising and describing data are used.</p> <p>The (meta)data standards outlined by the TDR are adhered to.</p>	<p>Use standards that are common to the archaeology and heritage profession, and that are relevant to your project archive.</p> <p>Not every aspect of your project will have a suite of standards or common vocabularies, so it may not be possible in all cases.</p> <p>Where community standards don't exist, provide an explanation and cite any data or sources used that are relevant to how the data are described.</p>	<p>Are the community standards used accessible and appropriate to the dataset?</p> <p>Are the data easy and suitable for others to reuse?</p>