Archaeology and the FAIR Principles

A case study from the Archaeology Data Service

Holly Wright, Archaeology Data Service, University of York, UK

SSHOC archaeological case study Workshop
25 May 2021
Virtual

This project is funded from the EU Horizon 2020 Research and Innovation Programme (2014-2020) under Grant Agreement No. 823782

SSHOC
social sciences & humanities open cloud

What are the FAIR Principles?

"In 2016, the '**FAIR Guiding Principles for scientific data management and stewardship'** were published in *Scientific Data*. The authors intended to provide guidelines to improve the **F**indability, **A**ccessibility, **I**nteroperability, and **R**euse of digital assets. The principles emphasise machine-actionability (i.e., the capacity of computational systems to find, access, interoperate, and reuse data with none or minimal human intervention) because humans increasingly rely on computational support to deal with data as a result of the increase in volume, complexity, and creation speed of data."

GOFAIR: https://www.go-fair.org/fair-principles/

This project is funded from the EU Horizon 2020 Research and Innovation Programme (2014-2020) under Grant Agreement No. 823782

SSHOC
social sciences & humanities open cloud

What are the FAIR Principles?

Findable: The first step in (re)using data is to find them. Metadata and data should be easy to find for both humans and computers. Machine-readable metadata are essential for automatic discovery of datasets and services.

Accessible: Once the user finds the required data, she/he needs to know how can they be accessed, possibly including authentication and authorisation.

Interoperable: The data usually need to be integrated with other data. In addition, the data need to interoperate with applications or workflows for analysis, storage, and processing.

Reusable: The ultimate goal of FAIR is to optimise the reuse of data. To achieve this, metadata and data should be well-described so that they can be replicated and/or combined in different settings.

This project is funded from the EU Horizon 2020 Research and Innovation Programme (2014-2020) under Grant Agreement No. 823782

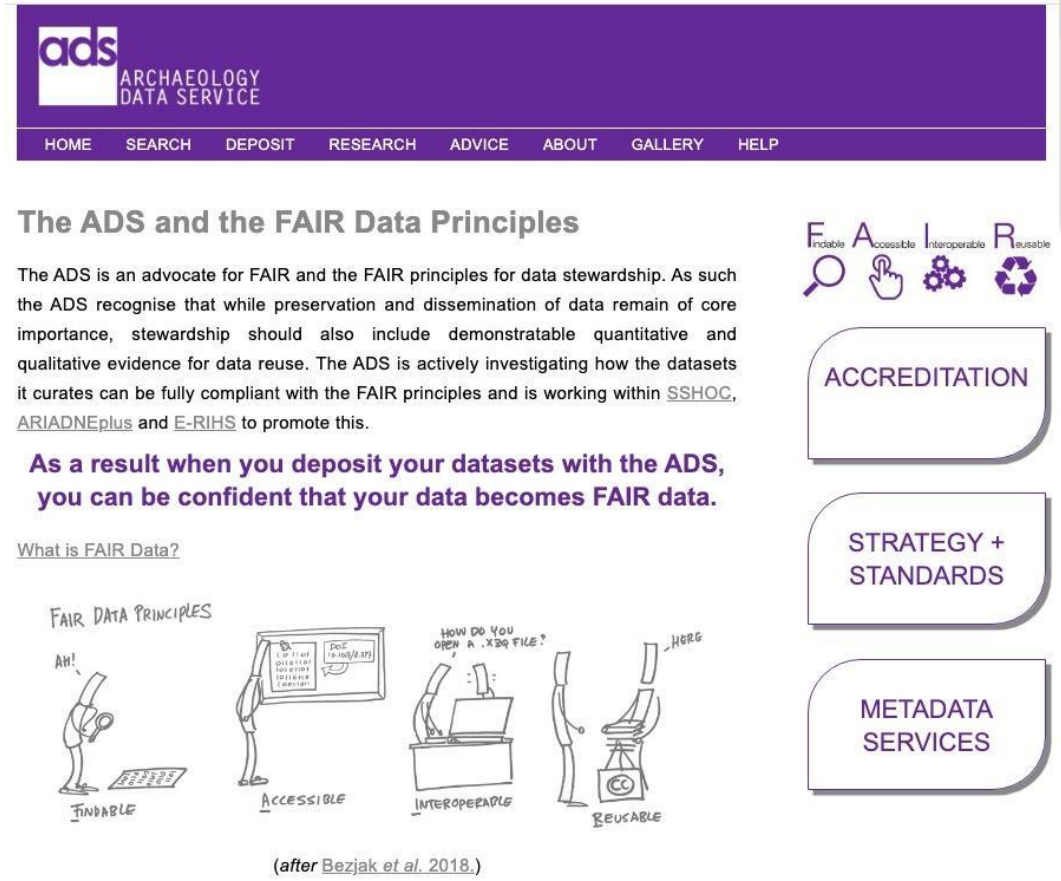
SSHOC
social sciences & humanities open cloud

D5.15 Report on opening access to research data in the Heritage Science and Archaeology domain

Examination of the issues and challenges faced in providing FAIR access to archaeological and archaeological science data, and review of the solutions adopted across Europe

As part of this research, ADS undertook an audit of FAIR compliance for archaeological data, to use as a baseline for further work on the deliverable

https://archaeologydataservice.ac.uk/about/adsFAIR.xhtml



(*after* Bezjak *et al.* 2018.)

This project is funded from the EU Horizon 2020 Research and Innovation Programme (2014-2020) under Grant Agreement No. 823782

SSHOC
social sciences & humanities open cloud

Intro to the Archaeology Data Service

Domain Specific Digital Archive
Set up in 1996
Based at the University of York

Mission: Support research, learning and teaching with free, high quality and dependable digital resources

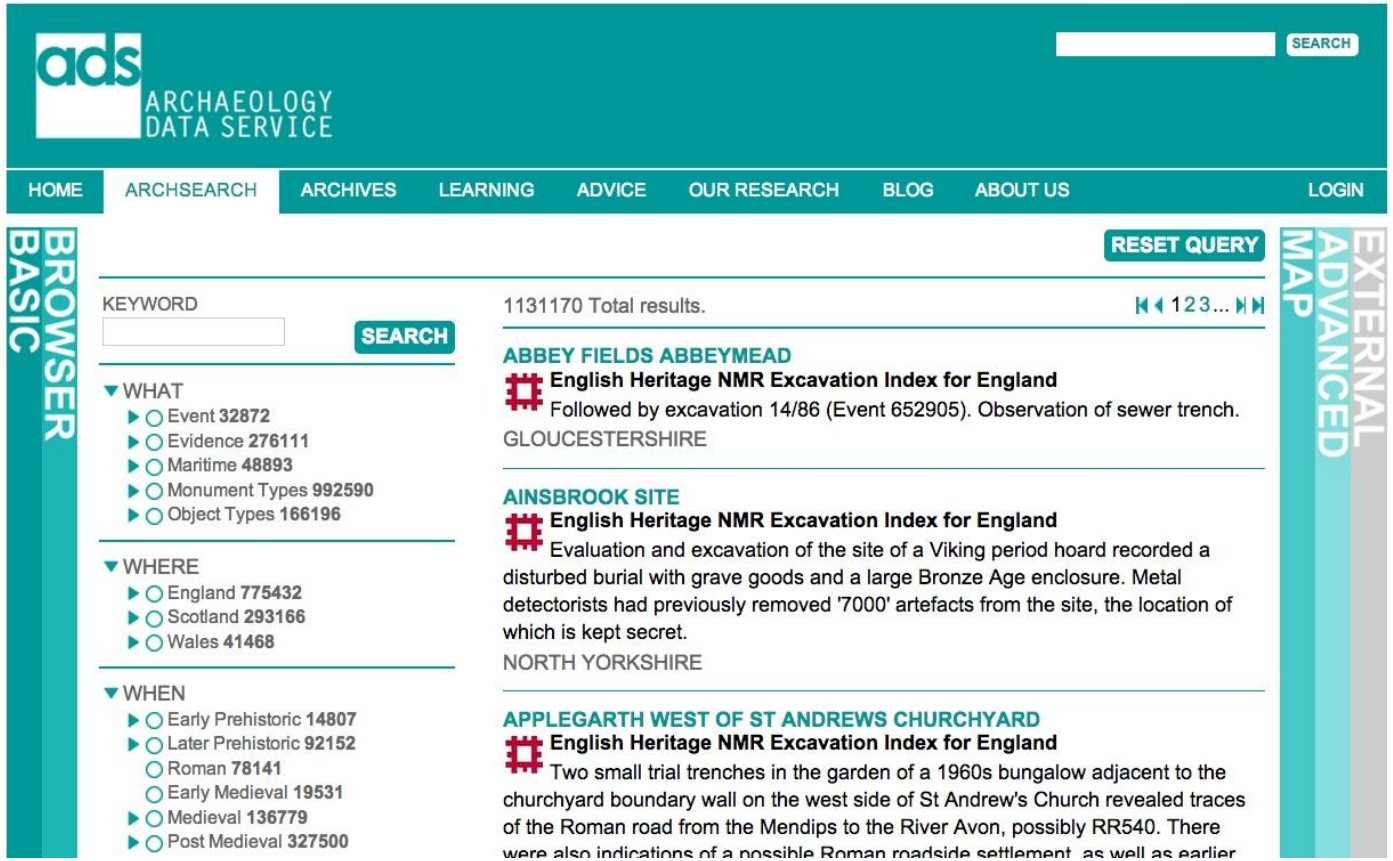
- Digital preservation
- Free online access to data
- Guidance and support for data creators
- Research

This project is funded from the EU Horizon 2020 Research and Innovation Programme (2014-2020) under Grant Agreement No. 823782

Intro to the Archaeology Data Service

What do we hold?

- ArchSearch: Online catalogue indexing over 1.3 million metadata records including ADS collections
  - 1000+ Project Archives
  - 62,000+ Grey Lit Reports
- Metadata harvested from over 30 UK national and regional historic environment inventories.



This project is funded from the EU Horizon 2020 Research and Innovation Programme (2014-2020) under Grant Agreement No. 823782

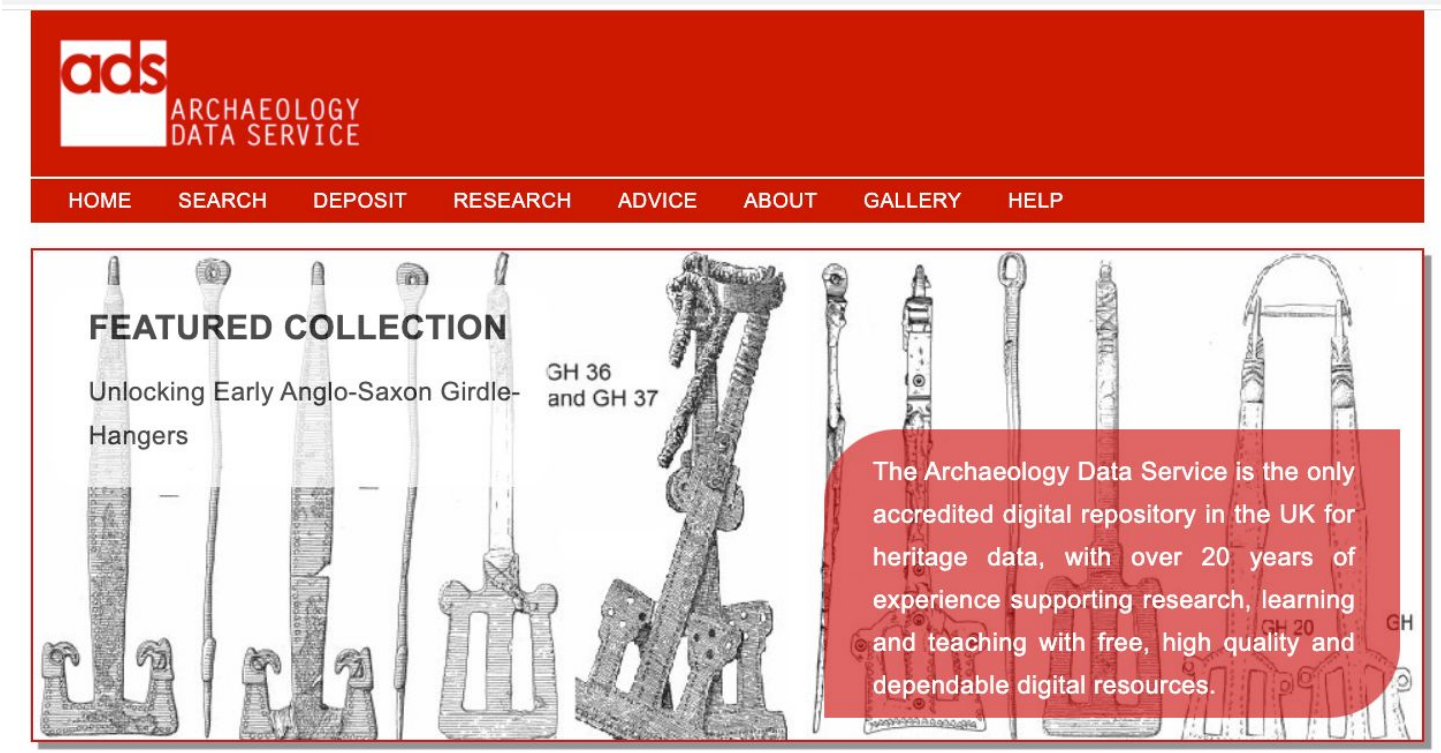
ads ARCHAEOLOGY DATA SERVICE

SSHOC social sciences & humanities open cloud

Intro to the Archaeology Data Service

What do we disseminate?

- Disseminate data we hold
- Provide integrated resource discovery across data we aggregate
- Also disseminate to other aggregators
  - OAI-PMH
  - Linked Data
  - Via Export

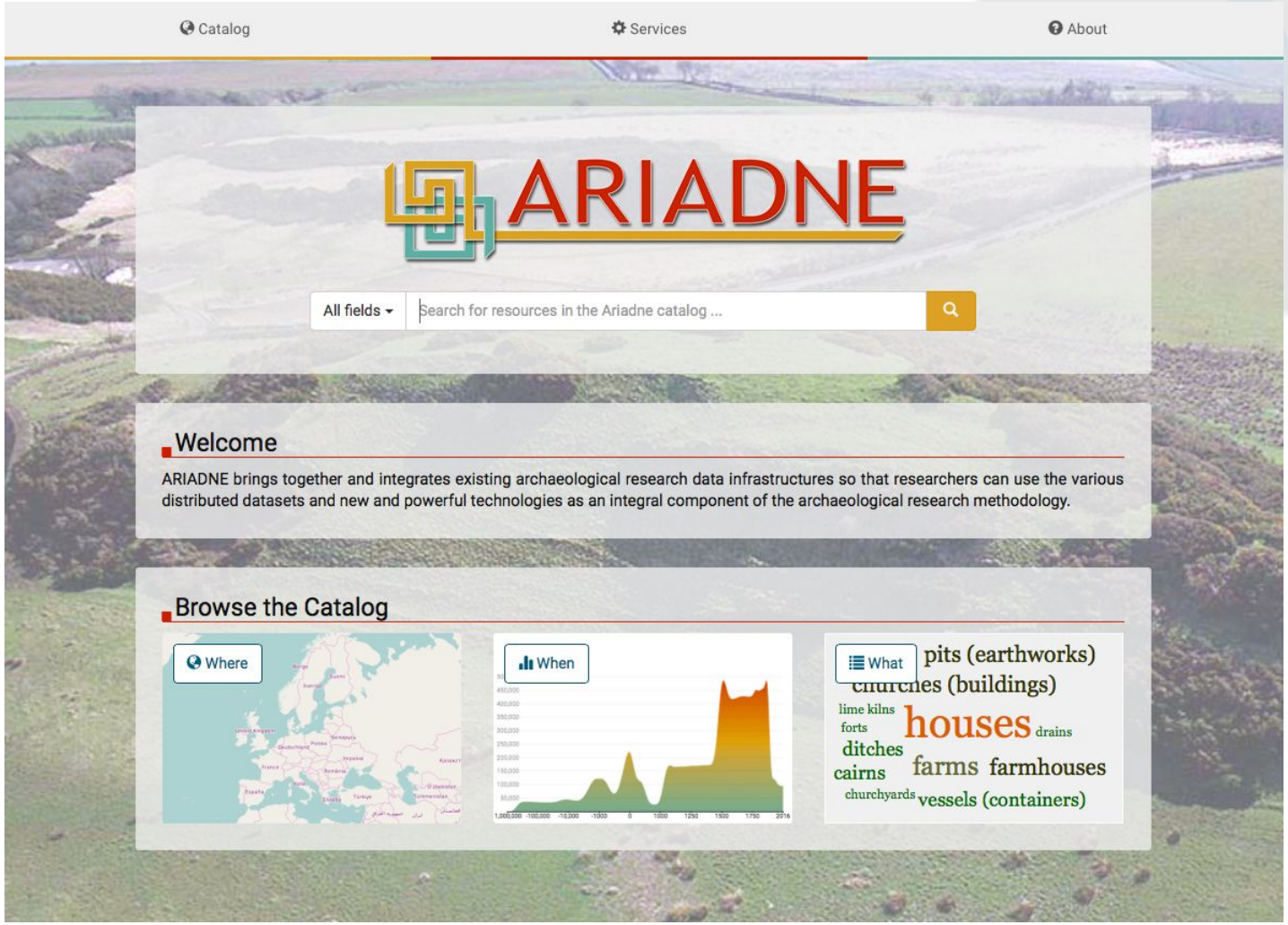


This project is funded from the EU Horizon 2020 Research and Innovation Programme (2014-2020) under Grant Agreement No. 823782

ads ARCHAEOLOGY DATA SERVICE

SSHOC social sciences & humanities open cloud

Intro to the Archaeology Data Service



This project is funded from the EU Horizon 2020 Research and Innovation Programme (2014-2020) under Grant Agreement No. 823782

Intro to the Archaeology Data Service

How is archaeological data different to other SSH data?

- Archaeological research is often non-repeatable (e.g. excavation destroys the archaeological site) so the data becomes primary data.
- Digital data is also fragile and requires long-term stewardship.
  - This is why people who work with archaeological data are obsessed with preservation and data persistence.
- Archaeological data is typically very heterogeneous and difficult.
- Archaeologists will use any kind of research tool or methodology if it helps to answer their research questions, so they are digital early adopters of a huge range of data types.

This project is funded from the EU Horizon 2020 Research and Innovation Programme (2014-2020) under Grant Agreement No. 823782

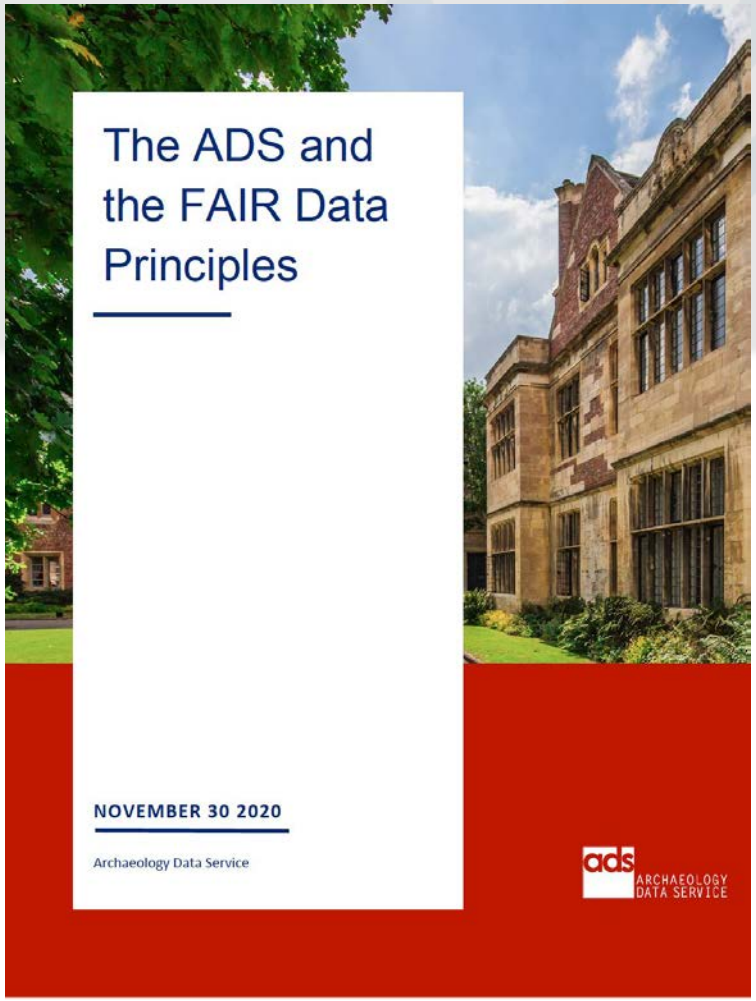
SSHOC
social sciences & humanities open cloud

Archaeology Data Service FAIR Audit

Led by Digital Archivist Ray Moore, who undertook our Core Trust Seal accreditation process, with input from all staff

Determined we should do an audit that would result in internally and externally-facing reports

- Internal report for ADS staff to inform our strategic planning process using the RDA FAIR Data Maturity Model tool, so that our progress can be measured over time
- External report for users/depositors to show how data deposited with ADS is FAIR data



This project is funded from the EU Horizon 2020 Research and Innovation Programme (2014-2020) under Grant Agreement No. 823782

SSHOC
social sciences & humanities open cloud

Archaeology Data Service FAIR Audit

Internal audit was confirmed our data is FAIR, but resulted in 14 recommended actions to prioritise within ADS for specific improvement. For example:

- F1.3 Implementation of globally unique PIDs at an object level for data AND metadata
- A1 Consider a more granular expression of ‘terms of access’ at an object level within metadata (even if it is essentially the same for all digital objects)
- I2.3 Request clearer documentation from depositors where data makes use of controlled vocabularies (for example, in a database). This is not directly requested, but would mean we could highlight FAIRness of data. Active encouragement of use of controlled vocabularies within Guides to Good Practice/Guidelines for Depositors
- R1 Consider a wider use of standards registries (e.g. the RDA-endorsed FAIRsharing) and the inclusion of our own standards

This project is funded from the EU Horizon 2020 Research and Innovation Programme (2014-2020) under Grant Agreement No. 823782

SSHOC
social sciences & humanities open cloud

Archaeology Data Service FAIR Audit

External audit was meant to provide transparency for users and depositors

- Most archaeologists are not familiar with the FAIR Principles
- Helps depositors understand the value of FAIR data and convey it in **funding applications** and as **impact indicators**
- Helps to promote best practice, both for data creators and data users
- Will continue to be updated as ADS works to make its data more FAIR

This project is funded from the EU Horizon 2020 Research and Innovation Programme (2014-2020) under Grant Agreement No. 823782

SSHOC
social sciences & humanities open cloud

Archaeology Data Service FAIR Audit

Findable

F1. (Meta)data are assigned a globally unique and persistent identifier.

For a fuller discussion of the ADS metadata and the use of persistent identifiers see our Metadata Overview page.

- The ADS uses Digital Object Identifier (DOIs) persistent identifiers for all collections.
- The ADS supports the use of ORCID IDs.
- The ADS supports the use of WikiData Q Codes.

F2. Data are described with rich metadata (defined by R1 below).

- All ADS resources are documented using the Dublin Core Metadata Element Set (DCMES) plus DCMI recommended qualifiers.
- The ADS also provides rich qualitative and technical metadata for all digital objects. These are repository specific metadata requirements, derived from domain-specific community standards (i.e. Guides to Good Practice, see also R1.3 below).
- All metadata is displayed alongside data, with technical metadata downloadable in open formats.

F3. Metadata clearly and explicitly include the identifier of the data they describe.

- All persistent identifiers for ADS collections are clearly displayed, alongside data, within each archive interface.
- The ADS supports the use of additional or supplemental identifiers relating to the dataset that link to external repositories, agencies or resources. This includes identifiers for physical, as well as digital, collections.

This project is funded from the EU Horizon 2020 Research and Innovation Programme (2014-2020) under Grant Agreement No. 823782

SSHOC
social sciences & humanities open cloud

Archaeology Data Service FAIR Audit

F4. (Meta)data are registered or indexed in a searchable resource.

- ADS datasets are findable through the repositories own indexes and catalogues.
  - ArchSearch
  - Archives
  - ADS Library
- ADS collections are also available through external catalogues and resources, including:
  - Heritage Gateway
  - DataCite
  - the Keepers Registry
  - Natural Environment Research Council (NERC) data discovery portal
  - ARIADNEPlus Portal
  - Marine Environmental Data and Information Network (MEDIN) data portal
  - Europeana
- ADS catalogues and indexes are searchable and harvestable through a series of OAI-PMH targets, and as linked open data using a SPARQL query web interface.

This project is funded from the EU Horizon 2020 Research and Innovation Programme (2014-2020) under Grant Agreement No. 823782

SSHOC social sciences & humanities open cloud

Archaeology Data Service FAIR Audit

Accessible

A1. (Meta)data are retrievable by their identifier using a standardised communications protocol.

- All ADS datasets utilise the HTTPS protocol to ensure free and open access to resources and to facilitate data retrieval.
- In rare instances, where discrete data objects are too large to support easy exchange using HTTPS, the ADS makes data available 'on request' using free and open exchange services (e.g. University of York DropOff Service, etc.).

A1.1 The protocol is open, free, and universally implementable.

- The ADS uses the HTTPS protocol for the sharing of resources and transfer of datasets. This is widely supported, open, and freely available.
- The repository utilises open and free file-sharing services where files or datasets are too large for easy exchange using HTTPS. Typically the ADS utilises the open and free University of York DropOff Service to share data when this is necessary.

A1.2 The protocol allows for an authentication and authorisation procedure, where necessary.

- The use of HTTPS provides authentication of the ADS website, and ensures the protection of the privacy and integrity of disseminated data. The repository ensures that all server-side digital certificates are current and up to date.

A2. Metadata are accessible, even when the data are no longer available.

- Yes. As an accredited digital repository the ADS supports long-term preservation and access of its holdings, consequently all datasets and metadata are maintained in perpetuity.
- The ADS maintains a clear Appraisal and Deaccession Policy which outlines current practice for datasets removed from the archives holdings. In such instances the ADS is committed to suporting identifiers (DOIs), maintaining resource discovery metadata, and updating current information on resources.

This project is funded from the EU Horizon 2020 Research and Innovation Programme (2014-2020) under Grant Agreement No. 823782

SSHOC
social sciences & humanities open cloud

Archaeology Data Service FAIR Audit

Interoperable

I1. (Meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.

- Yes, all resource discovery metadata is made available using a qualified Dublin Core in RDF/XML through the ADS Linked Data repository
- External services also consume and disseminate metadata (see above, and Metadata Services for a more detailed discussion

I2. (Meta)data use vocabularies that follow FAIR principles.

For a wider discussion on the vocabularies used in ADS metadata see our Strategy and Standards page.

- The ADS uses a variety of sustainable, open vocabularies to qualitatively classify and identify resources and datasets, including:
  - Heritage Data vocabularies, including those provided by the Forum on Information Standards in Heritage (FISH), Historic England (HE), Historic Environment Scotland (HES), and the Royal Commission on Ancient & Historical Monuments of Wales (RCAHMW)
  - Library of Congress Subject Headings (LCSH)
  - Marine Environmental Data and Information Network (MEDIN)
  - Getty Thesaurus of Geographic Names (TGN)
- The ADS also utilises recognised technical vocabularies to denote and categorise preservation activities
  - PREservation Metadata: Implementation Strategies (PREMIS)
  - Getty metadata types (Baca 2016)

I3. (Meta)data include qualified references to other (meta)data.

- The ADS supports the qualified referencing with and between publications, datasets and resources. Where available the repository uses sustainable referencing, e.g. DOIs.

This project is funded from the EU Horizon 2020 Research and Innovation Programme (2014-2020) under Grant Agreement No. 823782

SSHOC social sciences & humanities open cloud

Archaeology Data Service FAIR Audit

Reusable

R1. Meta(data) are richly described with a plurality of accurate and relevant attributes.

R1.1. (Meta)data are released with a clear and accessible data usage license.

- All ADS resources have clearly defined terms of access and reuse within each collection interface, and within metadata records distributed by the ADS or externally. Typically, data is disseminated under the terms of Attribution 4.0 International (CC BY 4.0), but data may also be disseminated under other forms of Creative Commons (see also the ADS Terms of Use and Access to Data).

R1.2. (Meta)data are associated with detailed provenance.

- The ADS provides detailed provenance metadata for all data. At a collection level this is clearly expressed in the archive interface and discovery metadata, but also at a file level within the technical metadata disseminated alongside the data.

R1.3. (Meta)data meet domain-relevant community standards.

- Yes, the ADS utilises a qualified Dublin Core metadata standard for all collection level metadata (noted above). The repository also uses standardised templates to ensure metadata consistency. All data must be accompanied by appropriate, file specific 'technical' metadata, this is derived from recognised community standards (Guides to Good Practice) to ensure consistency. All (meta)data is accepted, preserved and disseminated in sustainable, open formats. These are expressed in the 'Guidelines for Depositors' and the ADS' Data Procedures. The repository employs appropriate vocabularies to qualitatively describe datasets (noted above) and document preservation actions.

This project is funded from the EU Horizon 2020 Research and Innovation Programme (2014-2020) under Grant Agreement No. 823782

SSHOC social sciences & humanities open cloud

Future Work

Take the lessons learned from the ADS FAIR audit and undertake a similar audit for the ARIADNE metadata aggregation infrastructure for archaeological data

- ADS data is FAIR, but ADS also aggregates it's data within the ARIADNE portal, **making it even more FAIR** (e.g. mapped to CIDOC CRM)
- Undertaking a similar FAIR audit for the ARIADNEplus workflow, both internal and external (with DANS)
- Supplemented with existing work on archaeological science data undertaken as part of the E-RIHS PP.

This project is funded from the EU Horizon 2020 Research and Innovation Programme (2014-2020) under Grant Agreement No. 823782

Thank you for your attention!

Join our community

https://www.sshopencloud.eu

@SSHOpenCloud

info@sshopencloud.eu

/in/sshopencloud

This project is funded from the EU Horizon 2020 Research and Innovation Programme (2014-2020) under Grant Agreement No. 823782

SSHOC
social sciences & humanities open cloud