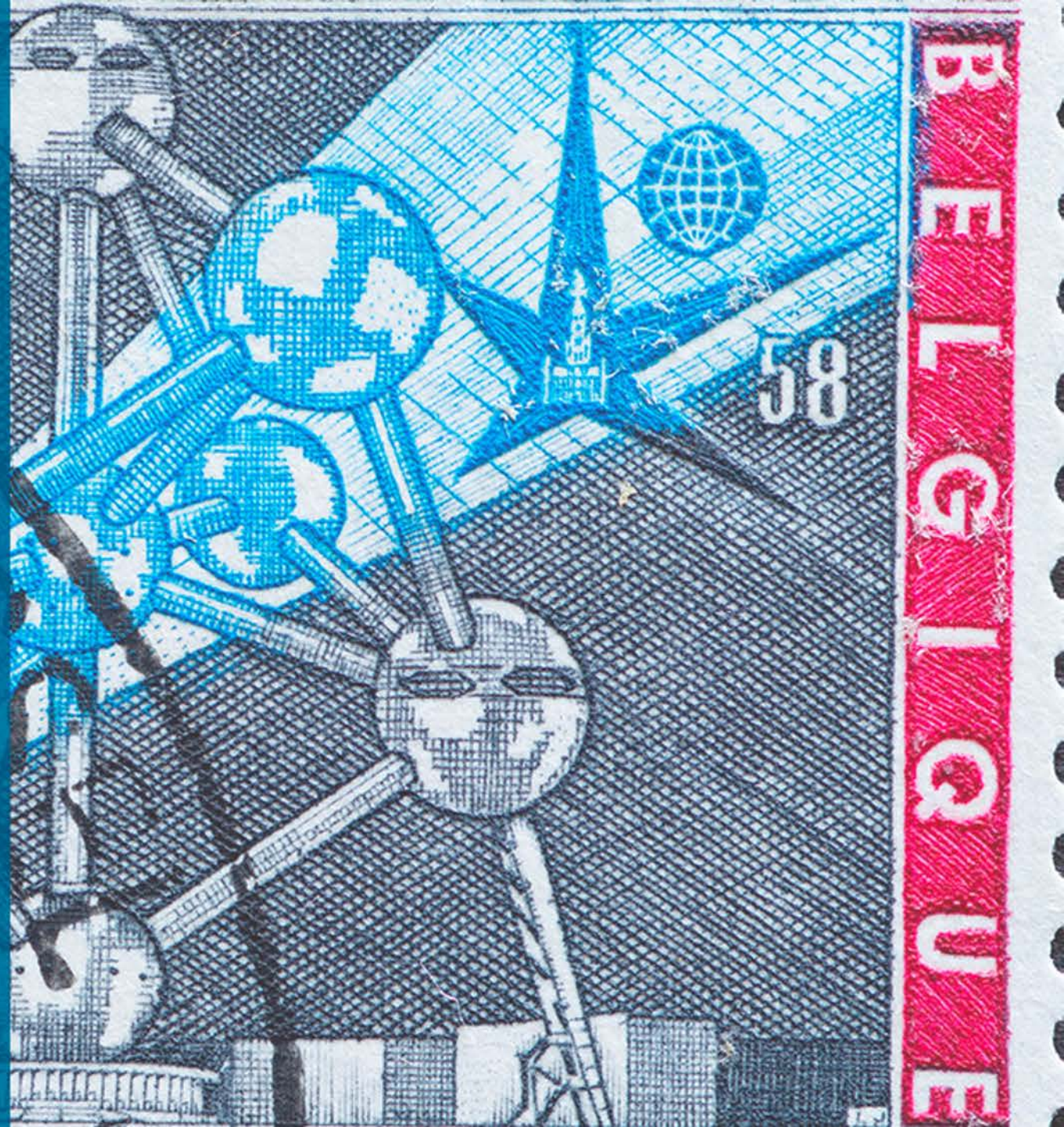
SSHOC

social sciences & humanities open cloud

Advancing SSH Research with SSHOCingly good and sustainable resources

6 & 7 April 2022
Brussels & online



Opening access to research data in the archaeology domain

Archaeology as a case study for FAIR

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

SSHOC Final Event
7 April 2022
Brussels, Belgium

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 Archaeology domain

Task 5.6: Examination of the issues and challenges faced in providing FAIR access to archaeological data, and review of the solutions adopted across Europe

Approach

- Archaeological data is extremely diverse, and therefore a useful case study to extrapolate best practice analysis across many Social Sciences and Humanities domains.
- As the leading domain-specific archive for archaeological data, the Archaeology Data Service (ADS) was an ideal test case to undertake an in-depth exploration for SSHOC.

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

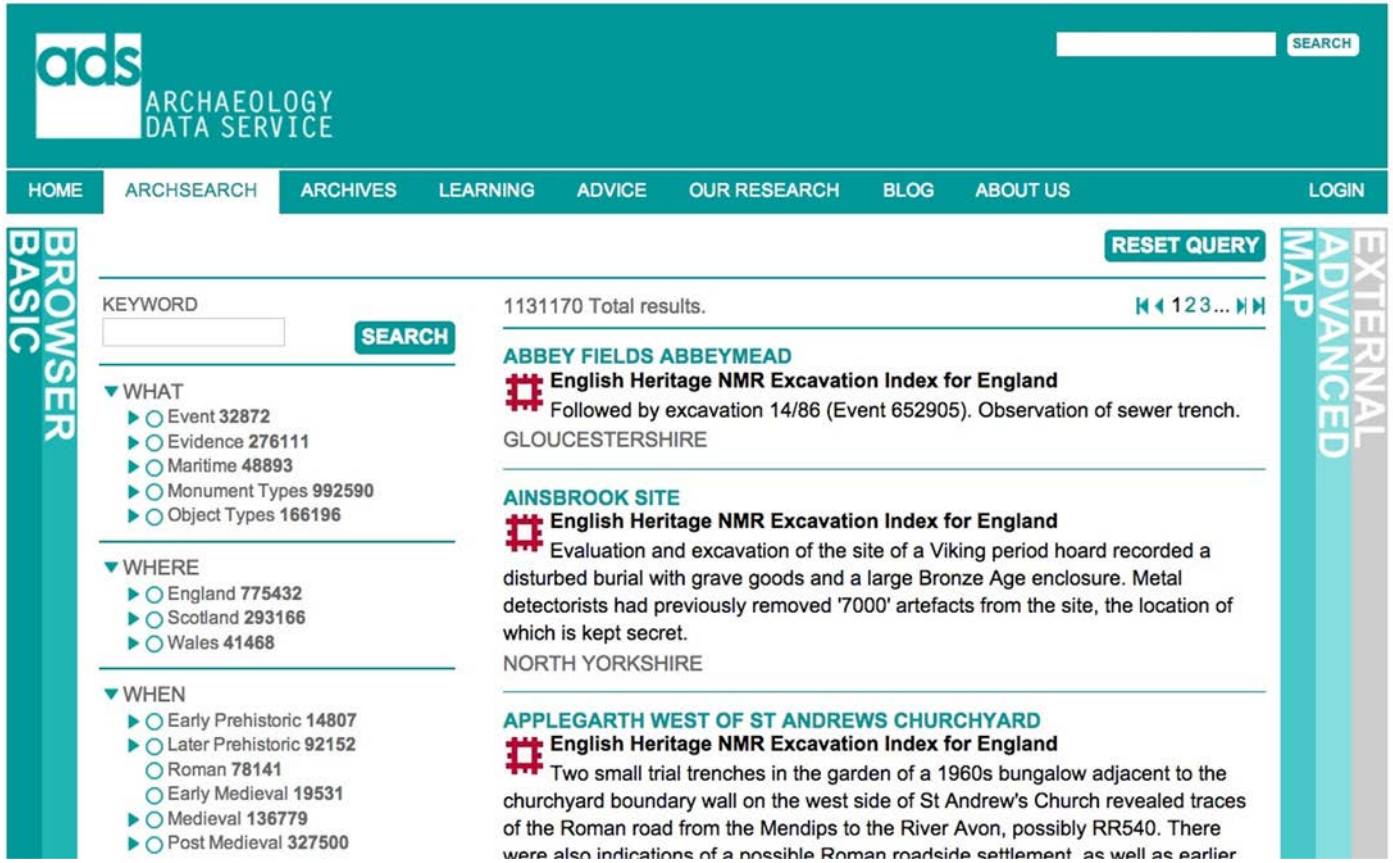
- Digital preservation
- Free online access to data
- Guidance and support for data creators
- Research

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
  - 80,000+ Grey Lit Reports
- Metadata aggregated 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

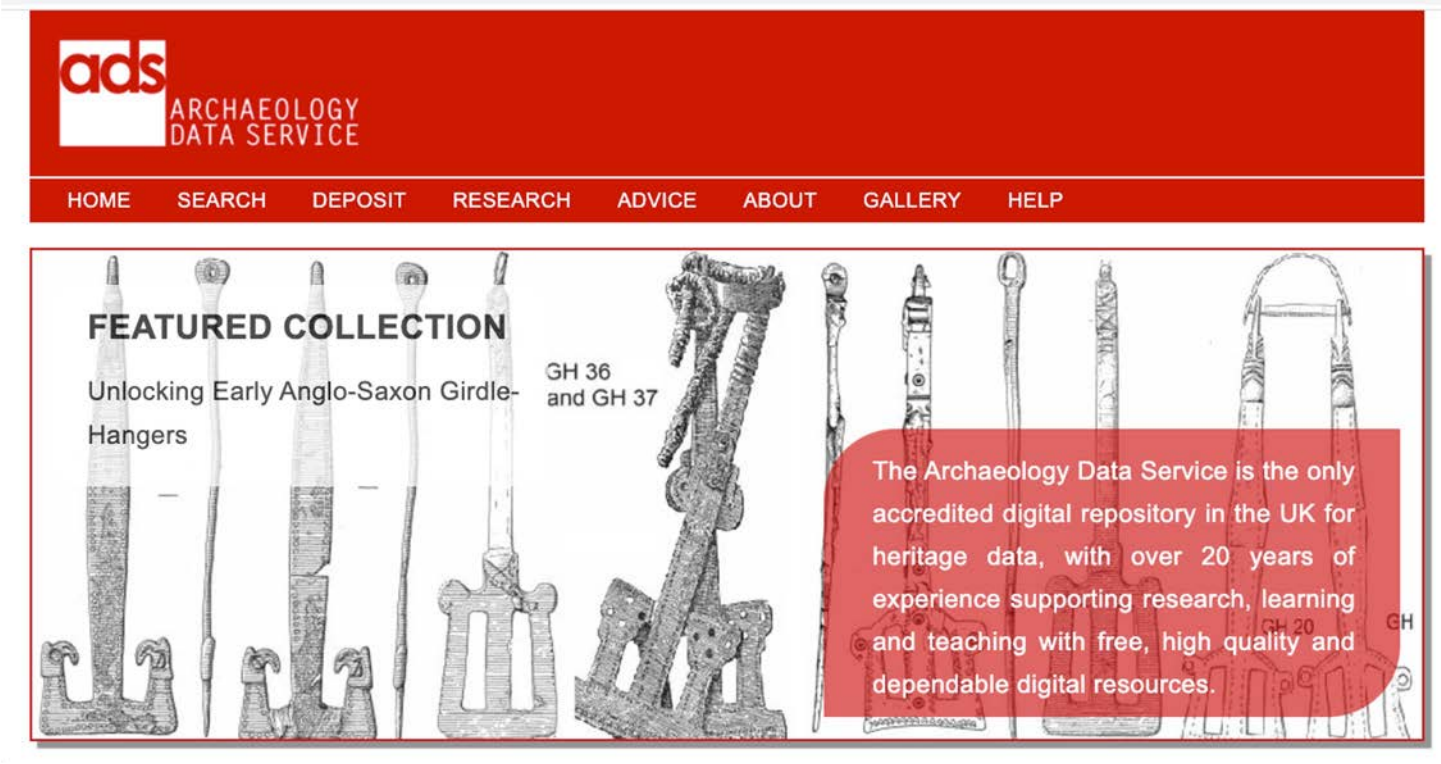
ads ARCHAEOLOGY DATA SERVICE

SSHOC social sciences & humanities open cloud

Intro to the Archaeology Data Service

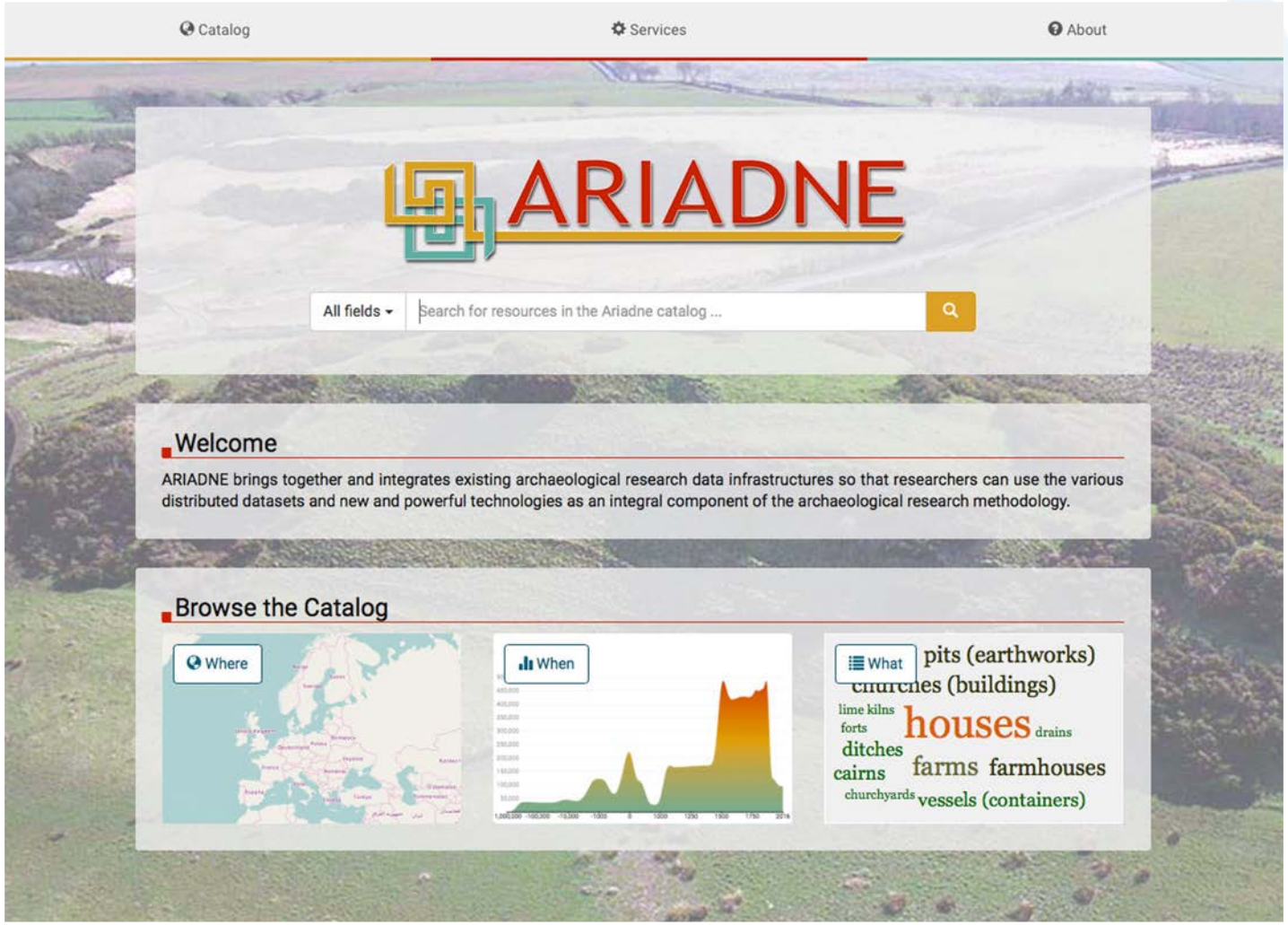
What do we disseminate?

- 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

Intro to the Archaeology Data Service



https://portal.ariadne-infrastructure.eu/

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, including scientific data (so not only a good SSH case study).

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 Archaeology domain

Task 5.6: Examination of the issues and challenges faced in providing FAIR access to archaeological data, and review of the solutions adopted across Europe

Implementation

- Overview of the FAIR landscape, including larger European and international alignments such as EOSC.
- Assessment of the current FAIRness of ADS.
- An automated assessment using the F-UJI tool developed by the FAIRsFAIR project.
- Discussion of how workflows may be developed to address FAIR data quality within a CoreTrustSeal accredited repository.

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

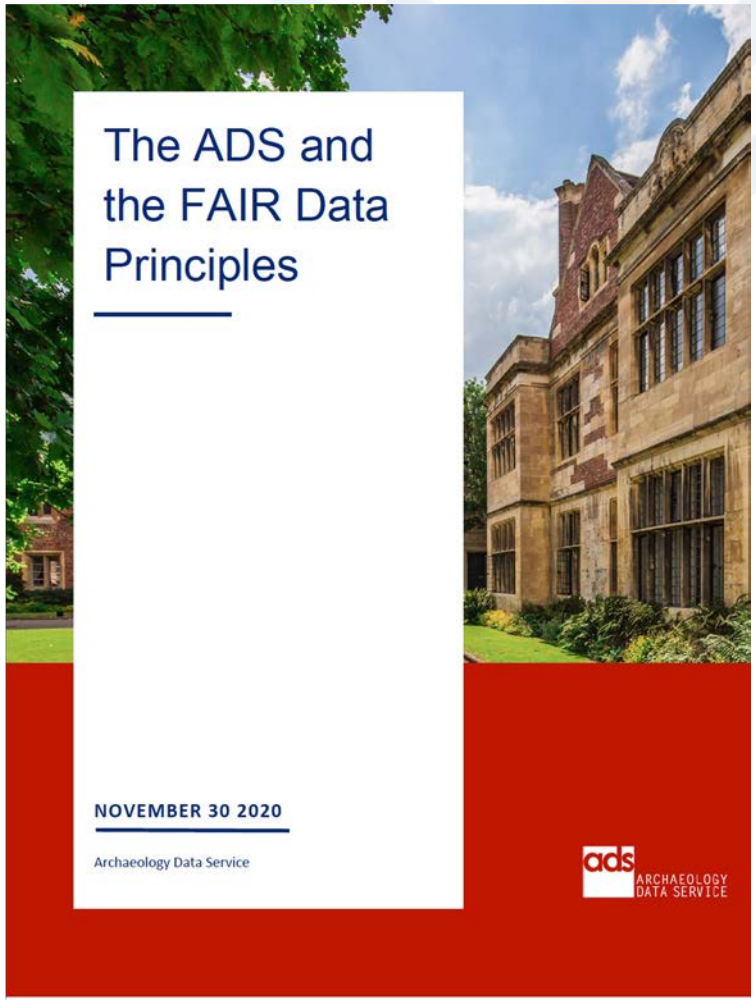
SSHOC
social sciences & humanities open cloud

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 made more FAIR



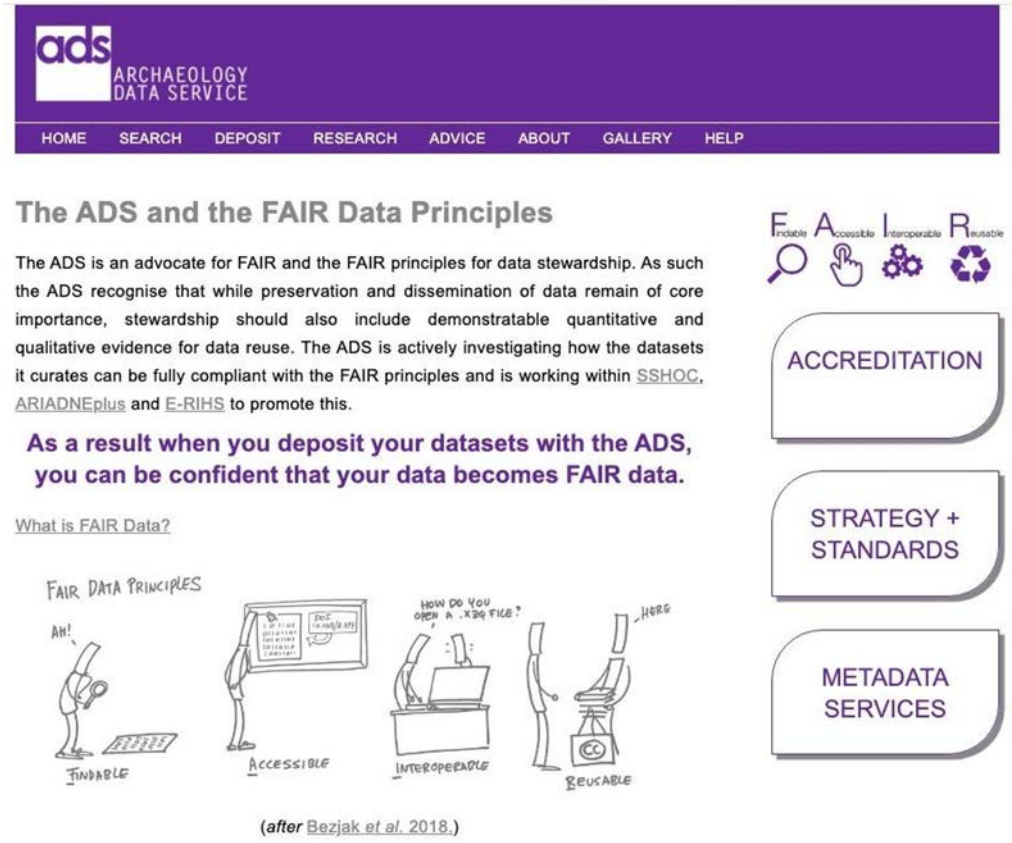
This project is funded from the EU Horizon 2020 Research and Innovation Programme (2014-2020) under Grant Agreement No. 823782

SSHOC
social sciences & humanities open cloud

Archaeology Data Service FAIR Audit

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



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

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

"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

Collaboration with FAIRsFAIR and testing the F-UJI Tool

Pilot Repository	Certification	Subject Areas	Repository Representatives
	CoreTrustSeal WDS Regular Member	Earth and Environmental Science	Uwe Schindler Michael Diepenbroek
PHAIDRA	CoreTrustSeal	Cultural Heritage	Yuri Carrer Cristiana Bettella GianLuca Drago Giulio Turetta
CSIRO	CoreTrustSeal	Multiple disciplines	Mikaela Lawrence Dominic Hogan Cynthia Love
WDC CLIMATE	CoreTrustSeal WDS Regular Member	Earth System Science	Andrej Fast Amandine Kaiser Hannes Thiemann
DataverseNO	CoreTrustSeal	Multiple disciplines	Philipp Conzett (Uit/DataverseNO) Gustavo Durand (Harvard/Dataverse) Julian Gautier (Harvard/Dataverse)
DataverseNL	-	Multiple disciplines	Laura Huis in 't Veld Marion Wittenberg Paul Boon

F-UJI Automated FAIR Data Assessment Tool

F-UJI is a service based on REST, piloting a programmatic assessment of the FAIRness of research datasets

FAIRSFAIR Fostering Fair Data Practices in Europe

SSHOC social sciences & humanities open cloud

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

Examples

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

External Qualitative Assessment

- 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.

Internal Qualitative Recommendation

Recommendation A1.1: A clear policy of sharing large files and datasets using more open services.

F-UJI Automated Assessment

Result	Comments	Next Step
Score: 1.0-1.0 of 1	OK	

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

Examples

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

External Qualitative Assessment

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

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

Examples

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

Internal Qualitative Recommendation

- *Recommendation I2.1:* An investigation of FAIRness of vocabularies used by the ADS. Where there are issues, raise awareness of FAIR with creators/communities, and ideally to leverage increased FAIRness.
- *Recommendation I2.2:* Consider a more wholesale and consistent implementation of these thesauri at an object level.
- *Recommendation I2.3:* Request clearer documentation from depositors where data makes use of controlled vocabularies (for example, in a database). Currently, 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.

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

Examples

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

F-UJI Automated Assessment

Result	Comments	Next Step
Score: 0.0-0.0 of 1	Whereas the service seems to use controlled vocabularies such as http://purl.org/heritagedata it seems not be used in the metadata detected by F-UJI.	Rec.: Use vocabularies in schema.org as discussed here: https://github.com/ESIPFed/science-on-schema.org/issues/27

debug message	count
NO vocabulary namespace match is found	**500**
Vocabulary namespace (s) specified but no match is found in LOD reference list	**500**

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

Examples

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

Discussion

- ADS makes extensive use of a number of controlled vocabularies within its metadata, but could take a more critical approach to the vocabularies themselves in terms of FAIRness.
- UK Heritage thesauri certainly meets most of the requirements for FAIR, but other vocabularies, and linkages to other persistent identifiers could be considered.

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 Archaeology domain

Task 5.6: Examination of the issues and challenges faced in providing FAIR access to archaeological data, and review of the solutions adopted across Europe

Review

- During SSHOC, ADS was actively involved as Deputy Coordinator of ARIADNEplus and Chair of the SEADDA COST Action.
- These relationships were used to contextualise the archaeology case study by synthesising recent, proximal work undertaken in collaboration with ADS that is highly relevant:
  - Comprehensive international survey of repository practices (holding archaeological data) undertaken by Geser (2021) for ARIADNEplus
  - Special issue authored by SEADDA Working Group 1: *Stewardship of Archaeological Data*, and its survey on *Digital Archiving in Archaeology: The State of the Art* (Richards *et al.* 2021)

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 Archaeology domain

Task 5.6: Examination of the issues and challenges faced in providing FAIR access to archaeological data, and review of the solutions adopted across Europe

Conclusion

Undertaking this research was very useful to better understand how, ***in practice,*** FAIRness can be:

Implemented
Assessed
Communicated
Improved

for archaeological data, and therefore SSH data (and beyond).

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

Thank you for your attention!

Geser, Guntram. 2021. *D2.3 Final Report on Community Needs.* https://doi.org/10.5281/zenodo.5647356.

Richards, Julian D., Ulf Jakobsson, David Novák, Benjamin Štular, and Holly Wright, eds. 2021. “Digital Archiving in Archaeology: The State of the Art” 58 (June). https://doi.org/10.11141/ia.58.23.

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