SSHOC
social sciences & humanities open cloud

Research and Innovation Action

Social Sciences & Humanities Open Cloud

Project Number: 823782 Start Date of Project: 01/01/2019 Duration: 40 months

Deliverable 5.15 Report on opening access to research data in the archaeology domain

Dissemination Level	PU
Due Date of Deliverable	28.02.2022, M38
Actual Submission Date	29.04.2022
Work Package	WP5 Innovation in Data Access
Task	Task 5.6 Issues in providing Open Data in Heritage Science and Archaeology
Type	Deliverable
Approval Status	Approved by EC - 04 May 2022)
Version	V1.0
Number of Pages	p.1 – p.50

Abstract: Deliverable attempted to create an assessment workflow to address FAIR data quality, using the ADS as a case study. The qualitative assessments resulted in recommendations for improvement that include an automated assessment of the FAIRness of ADS using the F-UJI tool developed by the FAIRsFAIR Project. This deliverable also sought to contextualise the archaeology case study by synthesising recent, proximal work undertaken in collaboration with ADS that is highly relevant.

SSHOC, "Social Sciences and Humanities Open Cloud", has received funding from the EU Horizon 2020 Research and Innovation Programme (2014-2020); H2020-INFRAEOSC-04-2018, under the agreement No. 823782

History

Version	Date	Reason	Revised by
0.1	05.10.2021	Structure and initial drafting	Holly Wright, Ray Moore (UoY ADS)
0.2	10.02.2022	Discussion and editing	Tim Evans (UoY ADS)
0.3	16.02.2022	Editing	Julian Richards (UoY ADS)
0.4	16.02.2022	Editing	Katie Green (UoY ADS
1.0	21.02.2022	Final editing, sent to Coordinator	Holly Wright (UoY ADS)

Author List

Organisation	Name	Contact Information
UoY ADS	Holly Wright	holly.wright@york.ac.uk
UoY ADS	Ray Moore	ray.moore@york.ac.uk
UoY ADS	Tim Evans	tim.evans@york.ac.uk

Executive Summary

Archaeology is one of the leading proponents of Open Data in the arts and humanities, and already exhibits broad interest in FAIR, but the diversity of data types and methods used by archaeologists means adoption of FAIR will pose significant challenges, further necessitating urgent collaboration around best practice. A recurring theme is the amount of time and effort it takes to do the kinds of work that allows data to be made FAIR, by both the data creators and the repository. Archaeology is an exemplar of why FAIR will be more difficult to implement for Social Sciences and Humanities.

This deliverable attempted to create an assessment workflow to address FAIR data quality, using the ADS as a case study. The qualitative assessments resulted in recommendations for improvement that are fed back to those who advise the data creators (data managers) and at a policy level. This deliverable also includes an automated assessment of the FAIRness of ADS using the F-UJI tool developed by the FAIRsFAIR Project. It was found to be an incredibly useful way to both see where ADS data is not FAIR in the ways it was expected, and for the explicit way in which it specifies the form the tool expects. Just as important is that F-UJI finds the ways where FAIRness can be improved at a technical level and fed back to technical staff.

This deliverable also sought to further contextualise the archaeology case study by synthesising recent, proximal work undertaken in collaboration with ADS that is highly relevant, such as the comprehensive international survey of repository practices undertaken by Geser (2021) for the ARIADNEplus project, and the work of 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). Taken together these elements constitute a comprehensive report on opening access to research data in the archaeology domain regarding implementation of the FAIR Principles.

Abbreviations and Acronyms

EOSC	European Open Science Cloud
FAIR	Findable Accessible Interoperable Reusable
ADS	Archaeology Data Service
SEADDA	Saving European Archaeology from the Digital Dark Age
RDA	Research Data Alliance
PID	Persistent Identifier
DMP	Data Management Plan
IFDS	Internet of FAIR Data and Services
DOI	Digital Object Identifier
OMS	Object Management System
UID	Unique Identifier
ARK	Archival Resource Keys
DCMES	Dublin Core Metadata Element Set
TGN	Thesaurus of Geographic Names
LCSH	Library of Congress Subject Headings
NERC	Natural Environment Research Council
MEDIN	Marine Environmental Data and Information
FDP	FAIR Data Point
WAF	Web Accessible Folder
FISH	Forum on Information Standards in Heritage
HE	Historic England
HES	Historic Environment Scotland
RCAHMW	Royal Commission on Ancient & Historical Monuments of Wales
PREMIS	PREservation Metadata: Implementation Strategies
GAMS	Geisteswissenschaftliches Asset Management System
ARCHE	A Resource Centre for the HumanitiEs
ADCH-CH	Austrian Centre for Digital Humanities and Cultural Heritage of the Austrian Academy of Sciences

Table of Contents

1. Introduction

The publication of FAIR Data Principles (Wilkinson et al. 2016) brought into focus the need for infrastructures, policies, and procedures to ensure that the research data, remains Findable, Accessible, Interoperable, and Reusable. Yet, while the FAIR acronym is widely recognised and supported, there remains some disagreement over the specifics of the principles, who should be responsible, how best to apply these to existing data management infrastructures, the metrics to assess FAIRness and the form of any assessment. This uncertainty makes a conclusive statement of FAIRness difficult; indeed, it may be several years before any definitive measure, whether through informal assessment or explicit certification, can be made. However, it is important to start. Archaeological data is extremely diverse and is therefore a useful case study that can be used to extrapolate similar 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 explore this as our contribution to SSHOC. By engaging with this process in an in-depth way in SSHOC Task 5.6, both through comprehensive internal assessment, and through external collaboration with the FAIRsFAIR Project, this research represents a 'worked example' of FAIRness for the Social Sciences and Humanities.

This deliverable includes an overview of the FAIR landscape, including larger European and international alignments such as EOSC, and then continues with an assessment of the current FAIRness of ADS. This includes an external qualitative assessment (current status), an internal qualitative recommendation (future enhancements), an automated assessment using the F-UJI tool developed by the FAIRsFAIR project, and then a discussion for each of the FAIR Principles. This deliverable is also an exploration of how workflows may be developed to address FAIR data quality within a CoreTrustSeal accredited repository. To carry out this research, ADS undertook both an internally and externally facing FAIR audit of their archiving practice. This type of external qualitative assessment is useful for providing transparency to both users and depositors as to the specific ways data archived within a CoreTrustSeal repository complies with the FAIR principles. At the same time, an internal qualitative recommendation is useful as a strategic document, with recommendations for how to change workflows and best practice to improve FAIR outcomes in the future.

Both assessments were undertaken by a senior digital archivist within the ADS, who was very familiar with ADS workflows and protocols. As the purpose of FAIR is to provide a framework for good practice around making data machine-discoverable and actionable, these assessments only gave part of the picture. In collaboration with Robert Huber of Bremen University, an automated assessment was undertaken using the F-UJI tool (Huber and Devaraju 2021), which was developed as part of the FAIRsFAIR ("FAIRsFAIR" n.d.) project. The purpose of the F-UJI automated assessment was to test if the assumptions undertaken in the qualitative assessments could be verified as truly machine-discoverable and actionable. The F-UJI tool also provided suggestions on how to address and improve the FAIR score for each Principle.

This was followed by a plan to then assess the FAIRness of the same ADS data aggregated within the ARIADNE Portal, essentially continuing to follow the FAIR trajectory of the same data when included in a large-scale international aggregation portal. Meetings were undertaken between the F-UJI developers, ADS and the CNR-Institute of Information Science and Technologies in Pisa; the partners who manage the Linked Open Data for ARIADNEplus ("ARIADNEplus" n.d.). Much was learned about the complexity of undertaking such an audit at the aggregation level, and it was deemed a worthy undertaking, particularly if it could be extrapolated to provide feedback to other ARIADNE metadata providers about the FAIRness of their (meta)data, but it required additional capacity from ARIADNE partners not available within the timeline of this deliverable.

Therefore, the overall intention of this research was to create an assessment workflow that can address FAIR data quality, by addressing as many aspects of the data lifecycle as possible. This workflow included:

- Qualitative assessment resulting in recommendations for improvement that are fed back to those who advise the data creators (data managers) and at a policy level.
- Quantitative assessment with F-UJI finding the ways where FAIRness can be improved at a technical level and fed back to technical staff.
- Quantitative assessment of the data within an infrastructure such as ARIADNE, to assess the value of data aggregation for enhancing FAIRness, and finding issues when data goes 'into the wild' so that they can be identified and fed back to the data managers.

To situate the findings at ADS and describes the work implemented in SSHOC Task 5.6, this deliverable also includes a discussion of *FAIR Implementation for Archaeological Data in Europe*. During SSHOC, ADS has been actively involved as both Deputy Coordinator of ARIADNEplus, the second four-year development phase of the European aggregation portal for archaeological data. ADS also provides the Chair of the SEADDA COST Action, which is a European networking project to 'Save European Archaeology from the Digital Dark Age' by creating a community of archaeologists and digital specialists working together to secure the future of archaeological data across Europe. Both ARIADNEplus and SEADDA conducted surveys to assess the level of FAIRness in the archaeology domain, albeit indirectly. ADS was also a partner in E-RIHS, the European Research Infrastructure for Heritage Science, and produced the key deliverable *Data Curation Policy* (Wright et al. 2020). This deliverable reviewed the issues concerning data curation for heritage science and provided a policy framework to be implemented by E-RIHS, but the report is designed to be of use to all those with interests in data within the heritage science domain. The report followed the structure of the FAIR Principles, interpreting them in the context of heritage science, and this research also informed the direction of the current deliverable for SSHOC Task 5.6.

In the case of ARIADNEplus, a comprehensive international survey of repository practices was undertaken (Geser 2021). Survey participants included digital repositories that were operative or in development that store and provide access to results from archaeological work, so answers were meant to be both practical and aspirational. The survey reflected the experiences of around 60 repositories, 43 operative and 17 currently in the process of becoming operative. The survey covered a range of topics and a comprehensive survey on the FAIRness of repositories would have been too complex in this

instance. The survey questions were instead created to cover important aspects addressed by the FAIR Principles in ways that were easily answered by respondents without referring to the Principles directly. This included identifiers, metadata richness, vocabularies, data discovery, and licensing. Data discovery included both repository search interfaces and external aggregation platforms.

As part of SEADDA Working Group 1: *Stewardship of Archaeological Data*, a survey on *Digital Archiving in Archaeology: The State of the Art* was undertaken in the form of an open access, peer reviewed, special issue published in Internet Archaeology, where 19 different countries, nations, regions and institutions reported on the current State of the Art as they saw it (Richards et al. 2021). These papers describe everything from legal frameworks to interoperability standards, and include national and regional contributions from Austria, Belgium, Bulgaria, Czech Republic, France, Germany, Greece, Hungary, Italy, Netherlands, Norway, Poland, Portugal, Serbia, Slovakia, Slovenia, Spain, Sweden and the UK. Countries outside of Europe also contributed articles on the State of the Art in their countries, including Argentina, Israel, Japan, Turkey and the United States. This special issue represents another recent dataset for understanding the level of engagement with data management best practice across much of Europe, including engagement with the FAIR Principles. A synthesis of *National FAIR Progress for Archaeological Data* with representation from a range of European countries completes the research for this deliverable and is followed by a concluding discussion.

2. The Current FAIR Landscape: Considerations for Repositories

The FAIR acronym (Findable, Accessible, Interoperable and Reusable) has become a recognisable feature in the data landscape, framing wider discussions amongst data creators, consumers, repositories and even funders about notions of best practice, data management and data stewardship, but while the acronym itself feels fixed and recognisable, understanding and defining the specifics of these Principles remains somewhat contested and open to (mis-)interpretation. Similarly, while there is consensus that the FAIR Principles themselves are an important part of the data ecosystem, there also remains disagreement about the best approaches to integrate them within current workflows and practices. In part, there is a sense that this is an 'active' landscape with a proliferation of projects investigating FAIR implementation (e.g., FAIRsFAIR, GOFair, OpenAire, etc.), each having a slightly different interpretation of the principles, or emphasising aspects, whilst foregrounding certain areas within the data lifecycle where FAIRification could best be addressed.

The FAIR data vision, created during the 'Jointly Designing a Data FAIRport' workshop (2014), subsequently enhanced by a Force11 community working group (Martone, 2015) and later formalised as 'The FAIR Guiding Principles for scientific data management and stewardship' (Wilkinson et al. 2016), intends to optimise data sharing and reuse by both humans and machines. Some projects and organisations charged with developing standards, such as the Research Data Alliance (RDA) have been explicit and exhaustive in their designation of FAIR metrics, while others, such as GoFAIR, have chosen a more simplified and concise form. Some are more selective in formulating metrics, such as the European Open Science Cloud (EOSC) and FAIRsFAIR, whilst others rank the principles to provide granularity (RDA). Attempts to make FAIR metrics simpler for wider general consumption are to be commended, but this can also foster confusion for those attempting implementation.

Such discrepancies are not simply restricted to overarching metrics but extend to the core of data principles themselves. As an example, some choose to emphasise parity in the treatment of both data AND metadata, while others are more selective. A key feature of the FAIR principles is the use of persistent identifiers (PIDs), but while some promote the use of PIDs simply for data (FAIRsFAIR, EOSC), others advocate they be used for both data AND metadata (GoFAIR, RDA). For many data creators this probably seems inconsequential, as often data and metadata are stored/disseminated in the same 'notional' place, but for many repositories, including the ADS, data and metadata may not be stored in the same place (e.g., metadata may be stored in a separate file, or even another repository) so the ability to refer to data AND metadata consistently and persistently becomes important. To compound the issue, there seems to be consensus against what is called 'PID saturation', as Lehväslaiho et al. (2019) contend "the use of PIDs needs to be mindful and use cases should be clear and nuanced".

There is also disagreement about responsibilities with regard to FAIR, the best methodologies and approaches to support adherence and implementation of the Principles, and the role of data repositories in the FAIR data landscape. FAIRsFAIR, building upon the *Turning FAIR into Reality* report (Directorate-General for Research and Innovation 2018) along with its own analysis of the research data practice landscape (Whyte et al. 2019), produced a series of recommendations for practical actions to support the realisation of a FAIR ecosystem. Molloy et al. (2020) set out this list of recommendations relevant for digital repositories:

A. *Develop and implement data sharing and interoperability frameworks: The recommendations made here encourage collaboration on research community-specific and inter-domain agreements and uptake of standards, shared vocabularies, ontologies, and metadata schemas to stimulate new research collaborations based upon cross-community discovery, retrieval and reuse of research datasets and other digital research outputs.*
B. *Ensure data management is supported by data management plans (DMPs): This section encourages the enhanced use of data management plans by researchers as dynamic, human-and machine-readable research support resources.*
C. *Develop professional support for FAIR data: Here, we recognise the invaluable contribution of emerging data professional roles such as the data steward and research software engineer and advocate for the emergence of related training and qualification opportunities, career path development, and professional societies. We also recommend the development of a self-assessment framework for research institutions and infrastructures to gauge progress towards building their FAIR data stewardship capabilities.*
D. *Ensure trusted curation of data: These recommendations encourage research communities to explore the use of trusted digital repositories and to find out more about how they can help their research by providing suitable, secure places of deposit for all types of research data including sensitive data.*

Trusted digital repositories have a key role to play in undertaking and facilitating adherence to the FAIR data principles, but other key players propose slightly different routes towards the production of FAIR data. EOSC advocates a top-down approach in which its membership, research funders and policy makers fund and incentivise (financial or otherwise) FAIR practice amongst communities. This, they hope, will support development within communities allowing them to respond to their own needs and requirements (European Commission, Directorate-General for Research and Innovation 2020a). They acknowledge that data repositories play a key role in making data FAIR, and regard this as one of the wider communities it seeks to support. In contrast GoFAIR takes a bottom-up, stakeholder-driven approach towards FAIR implementation, with close relations with EOSC and as part of a broader global Internet of FAIR Data and Services (IFDS). This will be achieved through the creation of an Implementation Network (IN) committed to defining and creating specific materials and tools, to support the development of FAIR infrastructure.

The proliferation of projects investigating FAIR implementation (e.g., FAIRsFAIR, GOFair, OpenAire, etc.) serve to highlight that approaches towards the implementation of the FAIR Principles are both variable and inconsistent, and at this stage it is difficult to discern which 'vision' of FAIR data will gain primacy. A

mechanism to facilitate the implementation of the FAIR Principles that is often mooted is assessment and certification. FAIRsFAIR seems to be leading the way here, arguing that certification "offers recognition and demonstrates trustworthiness to data depositors, users and funders" (L'Hours et al. 2020a). This will take the form of self-assessment and maturity modelling in FAIR certification, while also advocating a closer alignment with existing accreditation systems, specifically the CoreTrustSeal. The working group looking into accreditation (WP4: FAIR Certification of Repositories) sought to align any developments with the work in RDA and EOSC (specifically its FAIR Data Working Group) (L'Hours et al. 2020b).

Similarly, EOSC recognises the need for assessment, acknowledging the role of formal certification approaches (e.g., CoreTrustSeal), but advocates a more data-oriented approach, using automated evaluation throughout the data lifecycle (as opposed at specific locations, or with an institutional focus). The use of automated solutions they hope will address the issue of scalability in assessment. As they observe:

It is too early to recommend 'the software solution' for FAIRness evaluation. … Current automated methods are only applicable to already FAIRified data sets which follow certain standards. As FAIR transformation will be widespread, automated evaluators will be required to ensure scalability, but they will never be able to evaluate data sets still to be developed (European Commission, Directorate-General for Research and Innovation 2020b).

As noted above, there remain discrepancies in the metrics employed by each project/group to make such assessments of FAIRness. A common misunderstanding, however, is the connection between FAIRness and openness. Some see the relationship as synonymous, but from the outset this was a facet that early advocates acknowledged would be difficult to support.

…the use of end user licences or secure data services in the social sciences should not prevent data sets in such fields from obtaining equivalent FAIR scores to those where open access to data is not contentious (Directorate-General for Research and Innovation 2018).

There remains some disagreement as to how far FAIR should be integrated within, for example, the CoreTrustSeal. Some advocate a separate form of certification, others a 'bolt-on' (CoreTrustSeal+FAIR), and discussions also call for a more formal integration of FAIR within the accreditation process. For those already undertaking the process of becoming a CoreTrustSeal repository, such as ADS, an integration would be most welcome.

3. Archaeology Data Service FAIR Compliance

The following section reports the results of the External Qualitative Assessment (current status), the Internal Qualitative Recommendation (future plans) based on the FAIR Data Maturity Model (FAIR Data Maturity Model Working Group 2020), the results of the F-UJI Automated Assessment tool developed by the FAIRsFAIR project, followed by a discussion for each Principle.

3.1 Findable

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

External Qualitative Assessment

- 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.

Internal Qualitative Recommendation

- Recommendation F1.1: A clear policy and set of procedures for the use of DOI/PIDs.
- Recommendation F1.2: An assessment of whether DOIs offer the best solution for the creation of PIDs at the object level.
- Recommendation F1.3: Implementation of globally unique PIDs at an object level for data AND metadata.

F-UJI Automated Assessment

Table 1 Data is assigned a globally unique identifier

Result	Comments	Next Step
Score: 1.0-1.0 of 1	OK All data sets have a DOI	

Table 2 Data is assigned a persistent identifier

Result	Comments	Next Step
Score: 1.0-1.0 of 1	OK All data sets have a DOI	

Discussion

This is often regarded as the 'key' metric for the creation of FAIR data and the ADS can legitimately state that it uses them. The use of PIDs/DOIs for collections of data (i.e., archives deposited under a single licence) is consistent across the ADS. In addition, each unique bibliographic object, such as a journal article, monograph, or grey literature report is also assigned a DOI. However other data objects are normally not given this granularity for persistent identification. The exception is where granularity (for citation) has been specifically requested by the depositor. For example, for the excavation archive from Burdale: An Anglian settlement in the Yorkshire Wolds, each component data object or group of objects such as the excavation plan[1] and geophysical survey [2] were assigned DOIs. Examples such as this, and the bibliographic data represents how the use of PID's can look with regard to FAIR, with both data AND metadata utilising PIDs (in this case the same DOI).

It should be clearly noted that the ADS does not create DOIs for all its data objects, currently well in excess of 3.5 million unique entities. The rationale for this is primarily cost. The ADS are part of a larger consortium within DataCite headed by the British Library, with an annual subscription charge covering a set number of DOIs to be minted per annum (at the time of writing c.10,000). The ADS accession over 100,000 objects over the same period. Creating a DOI for each would thus require a direct agreement with DataCite (i.e., leaving the British Library consortium and the community of support and knowledge exchange therein) and a significantly higher subscription charge. Any such increase in annual budget for DOIs would thus have to be passed onto the depositor or funding bodies. Aside from cost, there is also a practical consideration. The storage of more DOIs–each referencing the digital object they represent–is not a significant technical challenge, but the human definition of what constitutes a defined and citable object is rather more problematic. In the Burdale example cited above, ambiguity was resolved by the depositor, but in most cases (particularly smaller collections) clear boundaries may become difficult to establish. Furthermore, in many cases this granularity (for DOIs) is not helpful. For example, an archive with 20 raster images and a spreadsheet from a small evaluation may not need a DOI for every image. One could even argue that too many DOIs appearing on a page could lead to confusion for the end user.

That is not, however, to discount the need for persistent and globally unique identifiers at the object level. As demonstrated, the infrastructure for a wider implementation of PIDs at an object level already exists. This is via the ADS implementation of the PREMIS model of data objects and intellectual entities (Library of Congress, n.d.) which has been built into a formal database structure known as the ADS Object Management System (OMS). Each object, from something as simple as a raster image to a more complex

[1] https://doi.org/10.5284/1021543 [accessed 21/02/2022]
[2] https://doi.org/10.5284/1021549 [accessed 21/02/2022]

set of data representations (files), such as a 3D model or geophysical survey, is assigned an internal Unique Identifier (UID).

It would be useful to be able to turn current object IDs into persistent, globally unique identifiers. While DOI is by far the most common form, FAIR metrics do not stipulate what form these identifiers should take. Certainly, alternative PID infrastructures already exist and there are others in the pipeline (European Commission, Directorate-General for Research and Innovation 2020b). One example the ADS has considered, but not implemented, is the use of Archival Resource Keys (ARKs), non-paywalled, decentralised persistent identifiers used by many libraries, data centres, archives, and museums. However, it may be argued that combining two types of (external) persistent identifiers within the same UI still presents a dilemma to the lay user, and arguably too much room for misunderstanding for those requiring a simple citation. Any such system that required a general user to decipher a long manual on how to cite, depending on the level of granularity within an archive, arguably works against the intentions of FAIR.

The alternative to large numbers of DOIs, or a hybrid system, may thus lie with the repository itself. In the examples cited above the object is resolved to a web page via a standard URL. A move to a system of persistent internal URIs which resolve to a page where both the metadata and data object can be accessed would be envisaged. The repository, which itself should be based on persistence of data in line with current accreditation requirements, thus takes on the role of hosting permalinks or PURLs. Such an approach, especially for an organisation with a large amount of legacy data such as the ADS, still has its drawbacks. Systems architecture must be in place to ensure that data objects migrated within the archive are still resolvable, and workflows in place to ensure that staff are aware of the requirements of such a system.

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

External Qualitative Assessment

- 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).
- All metadata is displayed alongside data, with technical metadata downloadable in open formats.

Internal Qualitative Recommendation

Recommendation F2.1: Increase the amount of metadata supplied to DataCite, and focus on increasing consistency and interoperability for creators by adding ORCIDs wherever possible.

Recommendation F2.2: Augmentation/enriching of metadata by ADS staff and research into machine learning including NLP and NER should also be pursued.

F-UJI Automated Assessment

Table 3 Metadata includes descriptive core elements (creator, title, data identifier, publisher, publication date, summary, and keywords) to support data findability

Result	Comments	Next Step
Score: 1.0-2.0 of 2	Some data sets do not have sufficient metadata. The example below offers publisher, summary, title only. This may be due to a broken DataCite connection. Only limited metadata is given on the landing page (OpenGraph). Therefore FAIRness largely relies on FAIR enabling services (DataCite)	Rec.: Add more metadata to the landing page for example using schema.org/Dataset

debug message	count
Not all required core descriptive metadata elements exist, missing	**6**

Discussion

This can be considered a great success for the ADS, that rich metadata are available for digital assets, and unlike many repositories, this extends beyond simple resource discovery metadata into detailed technical metadata and own preservation metadata. As highlighted later in the discussion of R1, much of this metadata is now directly linked to the data object itself. Limitations are usually due to the ADS requirement that metadata be supplied by the depositor (usually, but not always, the data creator). Practicality, circumstance, and pragmatism sometimes lead to metadata not being as fully defined (i.e., rich) as the ADS would prefer, and often there is a lack of time for archivists to create this themselves.

That said, all Digital Archivists at any level within the ADS have a background in archaeology, most at higher degree-level, and this is critical for ensuring metadata is as fully defined as possible. This may take the form of understanding the domain well enough to fix or create missing metadata or communicate issues with the metadata to the depositor and/or data creator. Archaeology is complex, uses an incredibly wide range of technologies and methodologies, and is therefore one of, if not the most, challenging domain in the Social Sciences and Humanities. Making pragmatic decisions about how to create useful and FAIR metadata for archaeological data is nearly impossible without domain expertise.

More recently, ADS procedures for accession and appraisal have been revised to establish standard minimum requirements for metadata, and for collections metadata to encourage and provide the tools for archivists to supplement it where possible. These interventions may be as small as ensuring that subject terms use the required (Linked Open Data) vocabularies such as the Library of Congress Subject Headings (LCSH) or the Getty Thesaurus of Geographic Names (TGN), or that metadata allows re-use/interoperability by external aggregation portals such as ARIADNE or Europeana. A key avenue for future research will be to revisit efforts for machine enrichment of metadata. With nearly all descriptive metadata (e.g., abstracts/descriptions) now held in the ADS OMS, and objects themselves are in open formats conducive to machine learning-based approaches, the opportunities for significant metadata augmentation are now a reality.

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

External Qualitative Assessment

- 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.

Internal Qualitative Recommendation

Recommendation F3.1: Investigate how the use of DOIs/PIDs can be incorporated into existing object-level metadata schemas and practices (as a short-term) solution.

F-UJI Automated Assessment

Table 4 Metadata includes the identifier of the data it describes

Result	Comments	Next Step
Score: 0.0-0.0 of 1	None of the data sets seems to expose information on e.g., downloadable files in the metadata.	Rec.: Add data file info which is shown on the web pages after the user clicks on 'Downloads'

debug message	count
Data (content) identifier is missing.	500

Discussion

Certainly, at a resource discovery level this is achieved, with DOIs prominently displayed within archive interfaces (alongside the metadata display page). It should be noted however that DOIs are not embedded in the formal metadata rendered on an individual metadata page, displaying instead in the header with the creator and publication date, and that in some instances the DOI may not always be prominent to a user.

ADS DOIs are included as standard within all external metadata portals and harvesters, including DataCite itself. DOIs may be missing in certain cases of grey literature where the ADS holds a metadata record but not the digital object itself; most of these cases relate to early unpublished fieldwork report records where a user has recorded that the object is held elsewhere (i.e., an external website). It is ADS policy not to issue a DOI or any form of persistent identifier where ADS do not physically hold the digital object, and certainly not where the metadata only links to an external website.

The finding that there are lacunae concerning information on downloadable files in the metadata is perhaps subjective and open to interpretation. Each metadata page, and thus record, lists the contents of each archive, or in the case of a single bibliographic record the number of elements which make up that object (usually one). Each download is accompanied by an information page which provides technical metadata (including file type and size) so that the user can understand precisely what it is they are downloading. Perhaps this issue is more related to, as noted in F1, the historic lack of a persistent identifier for objects that has led to a small degree of disconnect between the collection metadata (higher level) and the objects themselves. As noted above, any such exposition is achieved by listing the contents of the archive by data type and number of objects, for example “Image: 10 objects”. As stated in the requirement “[t]he association between a metadata file and the dataset should be made explicit by mentioning a dataset’s globally unique and persistent identifier in the metadata” (GoFAIR). This should also be considered as part of Recommendation F1.2: An assessment of whether DOIs offer the best solution for the creation of PIDs at the object level and Recommendation F1.3: Implementation of PIDs at an object level for data AND metadata above.

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

External Qualitative Assessment

- ADS datasets are findable through the repository's own indices 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.

Internal Qualitative Recommendation

Recommendation F4.1: A stronger commitment to the exposure and availability of ADS metadata and a more proactive stance on profile and dissemination.

F-UJI Automated Assessment

Table 5 Metadata is offered in such a way that it can be retrieved programmatically

Result	Comments	Next Step
Score: 2.0-2.0 of 2	OK Since all tested datasets have a DataCite registered DOI this test has been passed even though Mendeley was not available. However, since metadata is mainly offered through FAIR enabling services (DataCite) this can lead to reduced search engine coverage.	See above: offer metadata via landing page.

debug message	count
Mendeley Data API not available or returns errors: HTTPSConnectionPool(host='api.datasearch.elsevier.com', port=443): Read timed out. (read timeout=1)	501
Mendeley Data API not available	499
DataCite API not available or returns errors	84

Discussion

ADS datasets are findable and indexed through our own indices, catalogues and resources, and readily harvestable through our own metadata services. It may be worthwhile investigating the creation of a FAIR Data Point (FDP) as suggested by the GoFAIR project to publish ADS metadata in ways that are more machine readable and actionable.

Other options for increasing the availability of ADS metadata would be to create a Web Accessible Folder (WAF) with automatically generated XML of all archive metadata mapped to Dublin Core. Similarly, an API for archives metadata could be created to allow a range of users, including other interested repositories that may want to link to ADS resources, simple access to a metadata 'endpoint' that is not SPARQL or OAI-PMH. Finally, a move to creating Open Government Licence compliant services for all collections could provide opportunities for sharing metadata with spatially driven metadata portals (such as INSPIRE).

3.2 Accessible

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

External Qualitative Assessment

- 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.).

Internal Qualitative Recommendation

Recommendation 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).

F-UJI Automated Assessment

Table 6 Data is accessible through a standardized communication protocol

Result	Comments	Next Step
Score: 0.0-0.0 of 1	This could not be assessed since links to data are not exposed in metadata.	See comment above

Discussion

At an archive level the ADS includes information on accessing data (RDA-A1-01M), but it might be expedient to include a clearer expression of the 'terms of access' (much like licences) initially at a collection-level and, in future, at an object-level. It may also be useful to include in ADS metadata schema (at an object-level). The requirement only stipulates that "This can be included in the metadata or in some other place, for example on a landing page of the digital object" (RDA-A1-01M, see RDA FAIR Data Maturity Model Working Group 2020: 16).15 It should then, at the least, be added to the 'object page', but this should be easily achievable.

It can be certainly stated that data and metadata available through 'human intervention 16' (RDA-A1-02M and RDA-A1-02D) using standardized protocol (i.e., HTTPS) (RDA-A1-04M and RDA-A1-04D).

Automatic access (RDA-A1-05D) is a little more complex in that it states that: This indicator can be evaluated by resolving the link to the data, e.g., by resolving the persistent identifier and verifying that the data is reached. In the common case that the identifier is an HTTP URI, this can be done using the HTTP GET method. The evaluator or evaluation tool may also want to verify that the resolution delivers the correct data (European Commission, Directorate-General for Research and Innovation 2020b).

The example implies that PIDs should resolve directly to the object, but the ADS uses an intervening 'landing page' (library record, object page, etc.), so consequently data cannot be accessed automatically/programmatically. However, even with the landing page it should still be possible to access data directly using the method outlined in the example. This may be re-examined as part of the wider consideration under Recommendation F1.1 above. RDA-A1-05D is regarded as 'Important' and not 'Essential' to the FAIR data principles so may not be of immediate concern.

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

Table 7 Metadata is accessible through a standardized communication protocol

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

Discussion

ADS data and metadata is freely accessible using the HTTPS protocol, but as noted above, the exception is large datasets which are available 'on request'. Historically, this has been *ad hoc* provision of data through 'external' services (DropBox, OneDrive, GoogleDrive, etc.) whose 'openness' could be questioned. The University of York DropOff service may provide a more 'open' alternative to the sharing of these datasets, although long-term ADS FTP service would be re-examined. The notes for this requirement make specific reference to any protocol being "free of charge", but as it is known services can be monetarily 'free' but not open.

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

External Qualitative Assessment

- 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.

Internal Qualitative Recommendation
None

F-UJI Automated Assessment

Table 8 Metadata contains access level and access conditions of the data

Result	Comments	Next Step
Score: 0.0-0.5 of 1	Some data sets indicate licenses only. None of the datasets offers information on the access conditions (public, restricted etc.)	Include this information in the metadata e.g., using schema.org. Use appropriate vocabularies such as the EU publications access rights terms

debug message	count
Unable to determine the access level	500
Access condition looks like license, therefore the following is ignored	85
NO access information is available in metadata	85

Discussion

Historically the ADS used a single form of license which predated the formal definitions of Creative Commons. These conditions of access have been recorded at a higher collection level and recorded as a statement such as "Subject to ADS Terms and Conditions of Use". Since the ADS' adoption of Creative Commons licenses in 2016, and the need for forms of Open Government Licence for such datasets, the ADS collections database has been expanded to differentiate between licenses. This is usually included in metadata for external aggregators such as ARIADNE and Europeana.

It is interesting to note that the DataCite metadata schema only requests a rights statement, with definition of the licence or identifier scheme URI ("DataCite Schema" n.d.). Thus, are *both* the licence, if clearly identified and understandable (such as the ubiquitous CC-BY), and an access statement required? If so, the ADS would suggest that clearer examples from the repository community on best practice and consistency would be useful outputs.

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

External Qualitative Assessment

- 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 supporting identifiers (DOIs), maintaining resource discovery metadata, and updating current information on resources.

Internal Qualitative Recommendation

Recommendation A2.1: A clear statement on 'end-of-life scenarios' for data. An extension of the Appraisal and Deaccession Policy to include library records, digital objects, etc. With a clear commitment to the continued support of both metadata and PID.

F-UJI Automated Assessment

Not applicable

Discussion

The ADS Appraisal and Deaccession Policy (version 2.0) has been updated to acknowledge the continued support of DOIs and metadata following deaccession, but this focuses primarily on collections, and should be extended to include the deaccession/deletion of individual objects/files. Historically, files were deleted and a landing page created to ensure the DOI continues to resolve. This is unsatisfactory form with regard to FAIR. Essentially, metadata support should be continuous for any published data (collection or file). This should also include a clear statement that the data has been 'removed'.

One interesting possibility includes the preservation processes recorded for every archive within the collection metadata. Based on PREMIS, every normalisation, edit, and removal of objects held by the ADS would be recorded. The rationale for this is internal auditing, and to facilitate the transfer of the archive and metadata to another repository as part of succession planning. This may also help users to understand the structure of prior versions of the data. While objects may no longer exist, evidence that it did exist is retained within the OMS schema.

3.3 Interoperable

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

External Qualitative Assessment

- 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.

Internal Qualitative Recommendation

Recommendation I1.1: An extension of the current resource-discovery infrastructure to that on an object level.

F-UJI Automated Assessment

Table 9 Metadata is represented using a formal knowledge representation language

Result	Comments	Next Step
Score: 1.0-1.0 of 2	This test was passed using DataCite FAIRenabling services (JSON-LD content negotiation via DOI). This will not work in case DataCite services are down. No SPARQL endpoint was found or listed in re3data.	Rec.: Embed JSON-LD in landing page or use microdata

debug message	count
NO SPARQL endpoint found through re3data based on the object URI provided	500

Discussion

All resource discovery metadata is made available using the Dublin Core Metadata Element Set (DCMES), plus DCMI recommended qualifiers, and disseminated (in machine understandable way) through dedicated portals (in in RDF/XML). The repository makes extensive use of controlled vocabularies, ontologies, and thesauri (see Recommendation I2.1 and I2.2 below). The ADS ensures data is preserved and disseminated in appropriate, standardised, open formats (RDA-I1-01D).

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 uses recognised technical vocabularies to denote and categorise preservation activities, namely PREservation Metadata: Implementation Strategies (PREMIS)
- Getty metadata types

Internal Qualitative Recommendation

Recommendation I2.1: An investigation of FAIRness of vocabularies used by the ADS. Where there are issues use its position to 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, along with further investigation and the creation of a programmatic methodology to relate current terms to thesauri.

Recommendation 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 make it easier to highlight the FAIRness of data. Active encouragement of use of controlled vocabularies within the Guidelines for Depositors (2021a) and Guides to Good Practice (2021b).

F-UJI Automated Assessment

Table 10 Metadata uses semantic resources

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 to 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**

Discussion

The ADS makes extensive use of a number of controlled vocabularies within its metadata (see Linked Data Strategy and, more generally, Curatorial Strategy). These policies recognise the importance of these vocabularies in producing FAIR data, but could, perhaps, take a more critical approach to the vocabularies themselves in terms of FAIRness. UK Heritage thesauri certainly meets most of the requirements for FAIR. There could also be wider consideration of other vocabularies, and linkages to other persistent identifiers.

The use of vocabularies within the current workflow(s) of data into the ADS is variable and based on the method of delivery. For example, files (objects) deposited through the automated ADS-easy system, or other forms such as Dropbox and FTP, have all been historically accompanied by a defined metadata template with space for human entry of metadata terms. ADS guides and resources all ask for these to be standard Linked Open Data vocabularies, such as the ones listed above. However, taking just the English monuments thesaurus, which has over 4000 entries alone (Forum on Information Standards in Heritage 2021), a human's ability to consistently and accurately record multiple metadata terms *by hand* for each object is often complicated by mistakes and mis-understandings.

As above, an element of 'metadata fatigue' from ADS depositors is well documented, with users often on tight deadlines with minimal resources to devote the time needed to create rich, thesaurus consistent metadata for large archives. For the UK heritage vocabularies, in particular, there is also the specialised knowledge needed to differentiate between concepts such as an 'Amco Pillbox' and a 'Croft Pillbox'. A

pragmatist would of course argue that both are simply a Pillbox, but a user unfamiliar with these distinctions and the concepts of knowledge organisation may soon become confused, or even disinclined to commit the time to clarify. In addition, metadata forms (spreadsheets) require a human to use the correct spelling of terminologies and/or to include the URI so that the concept can be defined by the archive. As with the terms themselves, the knowledge of how such vocabularies is structured and what a URI is, is not widespread in UK archaeology, and arguably does not need to be. However, for a repository the end result is often a record that is not in a form that can store the string literal and URI in ADS databases.

Recently, ADS has devoted a proportion of time for each archive to cleaning this object metadata into a state that permits the use of FAIR vocabularies to describe the collection and the object. However, time spent on the latter still has to be limited, both in terms of practicality (updating many thousands of objects) and cost. As with Findable, the probable solution is an increased emphasis on machine learning technologies to generate these controlled terminologies for the user and repository. This would represent a new generation of ingest systems which use Natural Language Processing and Named Entity Recognition to prompt the user to select the terms identified from an abstract or user keywords, for example "excavation of a Roman enclosure" becomes: Excavation[3], Roman[4], Enclosure[5].

From a user and data management perspective, this lessens the immediate workload load for the creator and repository and could lead to increased emphasis on the quality of standard elements, such as description of what was found within an archaeological intervention.

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

External Qualitative Assessment

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

Internal Qualitative Recommendation

Recommendation I3.1: A clear policy on the use of references within the CMS and OMS. Also increased guidance on use of citation within the Guidelines for Depositors.

[3] http://purl.org/heritagedata/schemes/agl_et/concepts/145129 [accessed 21/02/2022]
[4] http://purl.org/heritagedata/schemes/eh_period/concepts/RO [accessed 21/02/2022]
[5] http://purl.org/heritagedata/schemes/eh_tmt2/concepts/70354 [accessed 21/02/2022]

F-UJI Automated Assessment

Table 11 Metadata includes links between the data and its related entities

Result	Comments	Next Step
Score: 0.0-1.0 of 1	Only a few data sets seem to contain relations to other entities. This seems to be missing in the metadata since related publications are shown on the landing page HTML e.g. 10.5284/1088084	Rec.: Include relations to other entities e.g., related publications in metadata.

Discussion

The inclusion of '(qualified) references to other (metadata) data' (for example RDA-I3-02D) within data has been a core remit of the ADS. The ADS metadata schema does include space for recording relations, and more recently citation of related resources, and uses controlled terminologies to differentiate (e.g., IsCitedBy, isVersionOf and cites), and these are recorded where supplied. As with I2, connections to other resources such as publications have historically been dependent on the creator/depositor supplying this within the metadata at the point of deposition. The ADS encourages and actively supports the use of reference and citation within deposited data, and part of the accession workflow is to check and validate any such metadata supplied. A recurrent theme through the ADS response to FAIR has been the limitation placed on staff resources to allow a wider review and collation of resources that have not been included in the depositor metadata but should be added. For example, a depositor may neglect an ORCID ID or publication reference, but the repository staff knows one exists. This is achievable on a small case-by-case basis but is not scalable for the increasingly large numbers of datasets ADS now ingests.

As previously discussed, there should be an increase in the use of technology around persistent identifiers to identify and collate references to related (meta)data. One avenue the ADS has explored recently is the use of the DataCite and CrossRef APIs to capture citation events referencing ADS DOIs. Such a process is however a constant one, with such citations having to be continually updated and incorporated back into the repository metadata. At a much higher level is the level of connectivity and infrastructure required by repositories to cross-reference and share. For example, the ADS may hold metadata by scholar X as defined by an ORCID, and another organisation may also hold information by the same individual but defined by an ISNI, WikiData Q code, or some other form of identifier. Achieving some of these FAIR subcategories is a collective effort for a designated community or group of repositories, not a single repository alone.

3.4 Reusable

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

External Qualitative Assessment

Addressed in R1.1-2

Internal Qualitative Recommendation

Recommendation R1: Consider a wider use of standards registries (e.g., the RDA-endorsed FAIRsharing) and the inclusion of ADS standards.

F-UJI Automated Assessment

Table 12 Metadata specifies the content of the data

Result	Comments	Next Step
Score: 1.0-1.0 of 4	Only the resource type (dataset) is given. No links to downloadable files are given in metadata.	See comment above

debug message	count
Measured variables given in metadata do not match data object content	500
NO data object content available/accessible to perform file descriptors (type and size) tests	500
NO measured variables found in metadata, skip 'measured variable' test.	500

Discussion

This metric "concerns the quantity but also the quality of metadata provided in order to enhance data reusability" (European Commission, Directorate-General for Research and Innovation 2020b). It may be worthwhile to investigate whether these are listed in standards registries (e.g., the RDA-endorsed FAIRsharing) and if not, working with data creators to ensure they are included.

R1.1. (META)DATA ARE RELEASED WITH A CLEAR AND ACCESSIBLE DATA USAGE LICENSE

External Qualitative Assessment

- 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).

Internal Qualitative Recommendation

Recommendation R1.1.1: Reconsider use and display of licencing information. More prominently situated in collection interfaces and particularly within resource metadata page. Also consider granularity of licence information (could it be stored with other object metadata) and perhaps its inclusion in object page and file-level metadata.

Recommendation R1.1.2: Include links to machine-readable (vis-à-vis human readable) versions of licence when exporting or exposing metadata through the various ADS portals. Also consider linking to a machine-readable version of the licence within metadata and interfaces (e.g. RDF expressions of Creative Commons licences).

Recommendation R1.1.3: Consider 'updating' ADS licences to a more standard form of licence.

F-UJI Automated Assessment

Table 13 Metadata includes licence information under which data can be reused

Result	Comments	Next Step
Score: 1.0-2.0 of 2	In most cases the available proprietary license or access terms could not be recognized or parsed by F-UJI.	Use a domain agnostic vocabulary to indicate licences as well as access conditions. Use appropriate metadata elements e.g. dc.rights, dct.licence

debug message	count
NO SPDX license representation (spdx url, osi_approved) found	**415**

Discussion

All datasets have a reuse license (RDA-R1.1-01M) which is available at a dataset/resource level. Whether this is 'clearly' expressed at collection level has been debated above, but it is noticeably lacking from file-level metadata. This should be consistent with 'terms of access' noted in Recommendation A1.

While recent (and future) collections are disseminated using a standard reuse licence (RDA-R1.1-02M), historic collections utilise 'a locally defined licence' which could be problematic in terms of FAIRness, but it should perhaps be emphasised that this is regarded as an 'important' and not 'essential' facet of the RDA requirements (European Commission, Directorate-General for Research and Innovation 2020b). That said this might be a useful driver to 'entice' depositors to sign a 'new' licence.

Unfortunately, ADS metadata does not resolve to a machine-understandable expression of the conditions of the license (only the human-readable summary). As contended:

Ambiguity could severely limit the reuse of your data by organisations that struggle to comply with licensing restrictions… The conditions under which the data can be used should be clear to machines and humans (GoFAIR 2017a).

Again, it should be noted that machine readability is regarded as 'important' and not 'essential' according to RDA metrics, but it could be a relatively simple fix at the application level to include the URL/identifier

of the more modern terms of CC and Open Government Licences used within the ADS. Much work has already been done to contact depositors and replace the ADS standard reuse licence with a CC licence, but this continues to be a long and difficult process for an organisation that is 26 years old and long predates the Creative Commons initiative. Until this is resolved ADS ability to satisfy this Principle in an automated system such as F-UJI will not be possible.

R1.2. (META)DATA ARE ASSOCIATED WITH DETAILED PROVENANCE

External Qualitative Assessment

- 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.

Internal Qualitative Recommendation

None

F-UJI Automated Assessment

Table 14 Metadata includes provenance information about data creation or generation

Result	Comments	Next Step
Score: 1.0-1.0 of 2	No formal ontology to express prov e.g., PROV-O is used.	

debug message	count
Formal provenance metadata is unavailable	**500**

Discussion

This should specifically focus on "information about the origin, history or workflow that generated the data, in a way that is compliant with the standards that are used in the community for which the data is curated" (European Commission, Directorate-General for Research and Innovation 2020b). GoFAIR goes further and contends that "Ideally, this workflow is described in a machine-readable format" (GoFAIR 2017b). ADS does receive some provenance information as part of collection and file level metadata, whether this is according to a 'community-specific standard' could be debated. Can something like the Guides to Good Practice (2021b) be regarded as a 'community-specific standard'?

Other provenance information, such as details on workflow, data collection or processing of data, is not formally collected (and definitely doesn't follow a standard). At the same time this is not machine readable (as GoFAIR outlines). Detailed provenance information is collected to a standard RDA-R1.2-01M. This is listed as 'fully implemented' but may require more investigation for a definitive statement. Perhaps, because this is a more difficult aspect of FAIR, this metric is listed as 'important' and not 'essential'.

Although not specifically mentioned it is tempting to link with provenance information that ADS does document, specifically the information produced during preparation for preservation and dissemination, once data has reached the repository. Assessing RDA-R1.2-02M is more complex and the documentation is a little unclear about what specifically it is attempting to assess, consequently this is left as 'not being considered'.

R1.3. (META)DATA MEET DOMAIN-RELEVANT COMMUNITY STANDARDS

External Qualitative Assessment

- 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.

Internal Qualitative Recommendation

Recommendation R1.3.1: producing a DCMI XML metadata target.

Recommendation R1.3.2: clarify compatibility with MIDAS Data Standard on internal processes and r3data entry.

F-UJI Automated Assessment

Table 15 Metadata follows a standard recommended by the target research community of the data

Result	Comments	Next Step
Score: 1.0-1.0 of 1	OK, e.g., OAI offers ISO, and MEDIN style ISO. The OAI service seems to be configured for the marine community rather than for archaeological purposes? In re3data e.g., MIDAS is listed but could not be detected by F-UJI	Rec.: include domain specific metadata formats in OAI-PMH or via typed links. For example, MIDAS XML?

debug message	count
NO metadata standard(s) of the repository specified in re3data	**500**

Table 16 Data is available in a file format recommended by the target research community

Result	Comments	Next Step
Score: 0.0-0.0 of 1	Links to data files are not given in metadata see comments above.	

debug message	count
Could not perform file format checks as data content identifier(s) unavailable/inaccessible	**500**

Discussion

The ADS was one of the initial service providers of the Arts and Humanities Data Service (AHDS) established in 1996, and the discovery metadata for archives and collections is therefore derived from the AHDS Common Metadata Format, which was itself based on the Dublin Core Metadata Element Set

(DCMES), plus DCMI recommended qualifiers. With the demise of the AHDS this format has remained unchanged, and compliance with DC often used as a good bridging point to other metadata standards including the Marine Environment Data Information Network (MEDIN) - to which ADS provides metadata for its portal - ISO and DataCite. Latterly, it has been used as a pathway to mapping ADS collection metadata to the ARIADNE Ontology (AO-Cat), an implementation of the CIDOC-CRM, and also the CARARE metadata schema (2013), which is based on the MIDAS Heritage (2012) standard, which is intended to deliver metadata to the CARARE service environment about online collections, monument inventory database and digital objects.

Arguably a weakness detected by the FUJI assessment is that this metadata, in its simplest form, is not available as a defined service or endpoint with Dublin Core metadata in a suitable format (such as XML), and an issue which is being taken forward as a simple and archivable recommendation.
It should be noted that the reference to MIDAS within the initial ADS assessment is itself something of a misnomer, which has been perpetuated within r3data as the ADS using MIDAS as a distinct standard for its metadata. ADS metadata, through its compatibility with AO-Cat (CIDOC CRM) and CARARE (MIDAS) is compatible with this standard, but an additional formal mapping and endpoint is not required. It should also be noted that the use of 'MIDAS' was an historic term used to indicate the use of a 'picklist' for archaeological periods used within the MIDAS manual itself. This has been superseded by the use of linked open vocabularies for heritage data, which are not mandatory for defining the subject, event and temporal fields.

4. FAIR Implementation for Archaeological Data in Europe

The ADS qualitative and quantitative assessments represent an in-depth case study for FAIRness in archaeology, from the longest-established, domain specific, accredited archive in the world. Much of the best practice seen in other countries, and particularly across Europe is either based on, or developed in collaboration with, ADS. This can be through the use of and/or contribution to the Guides to Good Practice, direct training and consultation, or collaboration across a range of research projects and aggregation initiatives. For example, the ADS participated in the preparatory phase of E-RIHS where its role was to look at the development of Policy for the archiving of Heritage Science data produced in the so-called DIGILAB. This revealed a very fragmented landscape and particular challenges in making Heritage Science data FAIR, which was attempted to be addressed in E-RIHS deliverable *Data Curation Policy* (Wright et al. 2020).

During SSHOC, ADS has been actively involved as both Deputy Coordinator of ARIADNEplus, the second four-year development phase of the European aggregation portal for archaeological data, and Chair of the SEADDA COST Action, which is a European networking project to "Save European Archaeology from

the Digital Dark Age" by creating a community of archaeologists and digital specialists working together to secure the future of archaeological data across Europe.

Both ARIADNEplus and SEADDA had a remit to conduct surveys to assess the level of FAIRness in the archaeology domain, albeit indirectly. In the case of ARIADNEplus, a comprehensive international survey of repository practices was undertaken (Geser 2021). Survey participants included digital repositories that were operative or in development that store and provide access to results from archaeological work, so answers were meant to be both practical and aspirational. The survey reflected the experiences of around 60 repositories, 43 operative and 17 currently in the process of becoming operative (Table 17).

Table 17 Countries and number of repositories that respond to the survey

Countries	***Repositories***		***Countries***	***Repositories***
European countries			Netherlands	1
Austria	3		Poland	3
Belgium	2		Portugal	4
Bosnia & Herzegovina	2		Romania	2
Bulgaria	2		Serbia	1
Croatia	2		Slovakia	2
Cyprus	1		Slovenia	1
Czechia	1		Spain	2
Denmark	1		Sweden	2
Estonia	1		Switzerland	2
Finland	1		United Kingdom	2
France	1		**Other countries**	
Germany	3		Argentina	1
Greece	3		Canada	1

Hungary	1		Israel	2
Italy	3		Japan	1
Latvia	1		Turkey	1
Lithuania	2		United States	1
Malta	1			**60**

The survey covered a range of topics and a comprehensive survey on the FAIRness of repositories would have been too complex in this instance. The survey questions were instead created to cover important aspects addressed by the FAIR Principles in ways that were easily answered by respondents without referring to the Principles directly. Questions concerned (meta)data identifiers, metadata richness, vocabulary in use, (meta)data discovery, and licensing. To situate archaeological repositories within the larger FAIR ecosystem, Geser looked at the international Figshare "The State of Open Data" surveys and found the percentage of researchers familiar with FAIR increased from 15% in 2018 to 20% in 2020. Other respondents had heard of FAIR.

This survey collected results from a range of FAIR-related questions. This included identifiers, metadata richness, vocabularies, data discovery, and licensing. Data discovery included both repository search interfaces and external aggregation platforms. For more detailed information and representative comments from the respondents, see Geser (2021).

(Meta)data identifiers

Survey question: *Are deposited data assigned globally unique and persistent identifiers (e.g., DOI, Handle, URN or other)?*
All 60 respondents answered it, 29 said "Yes", 11 "No", and 20 selected the additional option "Not yet".

Metadata richness

Survey question: *Are deposited data described with rich metadata, i.e. many descriptive attributes?*
All 60 respondents answered it, 47 said "Yes", 13 "No".

Vocabulary support

Table 18 Survey question: What vocabulary does the repository support?

International vocabulary (for example, Getty Art and Architecture Thesaurus)	19
National vocabulary (e.g., a thesaurus of a national authority or association)	25
Own standardised vocabulary (e.g., an own thesaurus)	35
Own list of terms	25
Keywords given by depositors	17

The majority of respondents (39 out of 60) said that their repository uses two or more of the categories defined above. These are used for metadata records of single items (e.g., publications, fieldwork or laboratory reports) or records of project archives with different types of content (e.g., various components of an excavation or other type or archaeological intervention). Most repositories use their own list of terms (which may include terms or keywords given by depositors), some use a national vocabulary, and a few use both.

Repository search interfaces

Survey question: Does the repository provide a metadata search interface?
All 60 respondents answered it, 36 said “Yes”, 10 “No”, and 14 “Not yet”.

External search platforms

Survey Question: Does the repository make metadata available to external search platforms/engines?
All 60 respondents answered it, 25 said “Yes”, 26 “No”, and 9 “Don’t know”.

Many said their repository does not/they did not know if their repository shared metadata with external search platforms. This is important, as it indicates the latter do not see the importance of making their holdings findable via external aggregation platforms. This may be due to having a user base that is well known and not expected to increase, lack of access to a suitable external platform, or a legacy metadata management system that does not support metadata harvesting.

Licence frameworks

Table 19 Survey question: Which licence frameworks does the repository support?

Public Domain Dedication, e.g., CC0, PDDL or other.	16
Users must only give attribution, e.g., CC BY, ODC-By or other.	22
Users must share new work under the same license, e.g., CC BY-SA, ODC-ODbL or other.	12
Do not allow commercial use, e.g., CC-BY-NC or other.	17
Do not allow derivative works, e.g., CC BY-ND or other.	9
Own terms and conditions, incl. some restrictions (e.g., non-commercial, no derivatives or other).	29
All or most works are fully copyright protected.	20

The survey answers showed four approaches to licensing:

1. Restricted approach: 19 repositories – 8 with "All or most works are fully copyright protected"; 11 in also apply "Own terms and conditions, incl. some restrictions" or state commercial use or derivative work is not allowed.
2. Open approach: 16 repositories – 4 Public Domain, 6 Attribution, 3 both, 1 also Share-Alike, 2 only Share-Alike.
3. Mixed approach: 8 repositories – All hold Public Domain data or data defined by their own terms and conditions or standard licenses.
4. Various restrictions or non-commercial use: 17 repositories – 12 only "Own terms and conditions, incl. some restrictions (e.g., non-commercial, no derivatives or other)"; 5 indicated commercial use of content is not allowed.

Enabling open data access

In addition to the questions meant to correlate with FAIR Principles, Gesar et al. included a range of related, and equally important questions on open access policies, how to improve data access, and how to demonstrate data reuse.

Support of open data policies: Repositories need policies and guidelines from heritage authorities so they can support open data access and reuse; 39 repositories required regulations and 36 clear guidelines by the authorities. Many cited barriers to creating and using open data (29) for example, licensing or fears that data might be misused. Training of repository staff to support new policies on open/FAIR data was also deemed important (28) and that the sharing of good practices and technical workflows could help support open data access and reuse policies.

Control of data access: A range of approaches were found; 24 repositories had a fully open access approach, (no registration required), 15 had data that can be accessed without registration and other data accessible with specific permissions, and 21 had data accessible only accessible with specific permissions.

Improving data access: The most cited means for improving data access were to improve or replace their existing data management system (30 respondents), improve metadata quality (34), provide metadata to external search platforms/engines (27), use Linked Data to interlink own and other (meta)data (26).

The 17 repositories still in a preparatory phase cited wanting to improve their data management system (11), aligning their internal vocabulary with external vocabularies (e.g., an international or national thesaurus) and/or use advanced ontologies (e.g., CIDOC-CRM), wanting their data to be discoverable by external aggregation platforms, possibly using Linked Data.

The operational repositories (43) cited wanting to improve or replace existing data management systems (19), around half cited better access to complex or high-volume data objects (e.g., 3D models, LiDAR data). Others wanted to improve metadata quality and replace or align their internal vocabulary with others and to provide metadata to external aggregation platforms and possibly use a Linked Data approach.

Data re-use is difficult to demonstrate: reuse takes place outside of what repositories can easily track and measure. Asked whether the repository collects information about date re-use (e.g., references in publications or other sources) only nine of 56 respondents said "Yes". This is another significant finding, as reuse is the purpose for which the rest of the FAIR principles exist.

The FAIR section in the Geser et al. (2021) report also contains a list of conclusions and suggested actions paraphrased below:

Repository support of FAIR

- *(Meta)data identifiers:* State-of-the-art repositories should provide advice on how to assign such identifiers.
- *Metadata richness:* Improvement of metadata quality is an important topic for advice on good practice.
- *Vocabulary support:* Advice is needed on how to standardise vocabularies and/or align them with international standards (e.g., Getty AAT) would be beneficial.
- *Data discovery:* Gather more information from the repositories that stated they do not have a metadata search interface and/or do not share metadata with external search platforms, to better understand the reasons.
- *Licence frameworks:* Advice is needed on copyright clearance and why some restrictions should be reconsidered.

Enabling open data access

- *Support of open data policies:* Clear positions of heritage authorities are needed, along with support and training of repository staff to support new policies on open/FAIR data.
- *Control of data access:* Advice is needed to reduce barriers to data access where mechanisms for not disclosing sensitive data.
- *Improving data access:* Advice is needed on improving or replacing the existing data management systems, improving metadata quality, providing metadata to external search platforms/engines, using Linked Data to interlink own and other (meta)data.

Analysis of data access and re-use

- *Data access:* Advice on tracking data access might allow identifying where access procedures could be improved and better reporting on repository usage.
- Data re-use: Re-use for new research and other purposes demonstrates best that funds for data preservation and access are well invested, but few repositories are able to track this in any substantive way. This is the key area for improvement in the archaeology domain.

Also of interest is that for those repositories that track and analyse their data access all reported increases in access during the COVID-19 pandemic, reporting increases from 5% to over 100%, showing the importance of publicly shared data, data repositories and discovery and access services to archaeologists.

5. National FAIR Progress for Archaeological Data

As part of the work of SEADDA Working Group 1: *Stewardship of Archaeological Data*, a survey on *Digital Archiving in Archaeology: The State of the Art* was planned. This was envisioned as a standard online survey similar to the one undertaken by ARIADNEplus, but due to the limitations in other aspects of the Action imposed by the COVID-19 pandemic in 2020-21, it was possible to reimagine and expand how the survey was implemented.

The result was a more ambitious, open access, peer reviewed, special issue published in *Internet Archaeology*, where 19 different countries, nations, regions and institutions reported on the current *State of the Art* (Richards et al. 2021) as they saw it. These papers describe everything from legal frameworks to interoperability standards, and also provide an indirect current resource for understanding FAIR implementation across Europe. The issue includes national and regional contributions from Austria, Belgium, Bulgaria, Czech Republic, France, Germany, Greece, Hungary, Italy, Netherlands, Norway, Poland, Portugal, Serbia, Slovakia, Slovenia, Spain, Sweden and the UK. Countries outside of Europe also contributed articles on the *State of the Art* in their countries, including Argentina, Israel, Japan, Turkey and the United States.

This special issue represents another recent dataset for understanding the level of engagement with data management best practice across much of Europe, including engagement, or lack thereof, with the FAIR Principles. It will not come as a surprise that the majority of contributors do not reference the FAIR Principles as part of their current practice, and indeed are not working towards specific aspects of FAIR. Most are laying the groundwork however, by making data openly available beyond their own institution or governmental body, by standardising vocabularies and/or clarifying data usage licencing. It is possible to see the direction of travel in the countries highlighted below.

In Austria, while not discussed explicitly, two repositories have capacity for archaeological data, and also hold the CoreTrustSeal, which has many foundational elements in common with the FAIR Principles. These archives are the Geisteswissenschaftliches Asset Management System (GAMS, AMS for the Humanities) which is hosted by the Zentrum für Informationsmodellierung - Austrian Centre for Digital Humanities of the University of Graz and A Resource Centre for the HumanitiEs (ARCHE) at the Austrian Centre for Digital Humanities and Cultural Heritage of the Austrian Academy of Sciences (ACDH-CH). ARCHE was designed for the long-term preservation of data from the Austrian arts and humanities community but holds a significant amount of archaeological data (Trognitz 2021).

While the Czech Republic has data repositories with CoreTrustSeal certification, they are focused on the areas of linguistics and social sciences. The Archaeological Information System of the Czech Republic (AIS CR) is on the Roadmap of Large Research Infrastructures adopted by the Government of the Czech Republic however, making it the national repository for archaeological data. The AIS CR has yet to pursue

certification, but it's position within the Ministry of Education, Youth and Sports, which is in charge of EOSC participation in the Czech Republic, means attention must be given to the FAIR Principles. As is the case in most countries, the majority of archaeological fieldwork is carried out in advance of development projects, and results in some form of fieldwork report. As these reports are subject to the Czech Copyright Act, significant negotiations were undertaken to secure licensing agreements with individual data providers in order to provide the data according to the FAIR principles. Currently, 81 of the 111 organisation licenced to undertake archaeological excavations in the Czech Republic have signed licence agreement, which has allowed the AIS CR to make over 90% of the archived documents openly accessible. The metadata for the remainder are available and searchable, and the copyright holder to be consulted about gaining access. The AIS CR is working towards additional aspects of FAIR compliance (Novák, Kuna, and Lečbychová 2021).

In Flanders, Belgium, archiving guidance and practice has reasonably high FAIR conformance. Much like the Czech Republic, the overwhelming majority of archaeological research in Flanders is undertaken as a mandatory requirement ahead of development. Archaeological research is regulated by a Code of Good Practice which was written taking on board the FAIR Principles without mentioning them explicitly. The resulting data is made available in Archaeoportal which also has significant alignment. For example, Archaeolportal provides permanent, dereferenceable unique identifiers (Hacıgüzeller et al. 2021).

In Greece, there is some progress in FAIR compliance via the aggregation of research metadata with national initiatives such as OpenArchives and SearchCulture which conform to the Open Archives Initiative, allowing ingestion of Greek content in the European Digital Library using the European Data Model (EDM). Nearly all archaeological interventions are undertaken centrally in Greece, and access to state-owned archaeological resources remains limited, but the current National Action Plan on Open Government includes a commitment to provide open access to cultural assets, targeting data homogenisation and data licensing, so there is movement in a FAIR direction (Tsiafaki and Katsianis 2021).

In Hungary, the Archaeology Database of the National Museum provides archaeologists with an accessible online catalogue of archaeological sites, including site metadata and documentation. Open access to archaeological data is still in its infancy in Hungary, and the database provides different levels of access. Short reports of sites published in the volumes of archaeological Investigations in Hungary are openly available to anyone, but no other documents or spatial information on sites are accessible to the public. Professionals may have full access to reports and data so their access is more FAIR, with interoperability through standardised metadata for descriptive concepts, terms and temporal coverage (Kreiter 2021).

In Sweden, digital objects, reports and finds in the National Heritage Board Archive are assigned unique and persistent identifiers and are openly searchable online. Metadata can be retrieved from the archive in a machine-readable way via OAI-PMH, and all digital objects are given Creative Commons licences, and

metadata is released as CC-0 (Löwenborg et al. 2021). For the Swedish National Data Service, depositors must include sufficient documentation/metadata. DOIs are provided for the datasets and archived in accordance with the OAIS-model for data preservation which is compatible with FAIR compliance, but more capacity needs to be created to support and train researchers in best practice to move towards greater FAIRness (Jakobsson 2021).

In France, the Frantiq network, for example, brings together data from both development-led and academic archaeology and has been developing a common bibliographic catalogue, the 'Catalogue Collectif Indexé' as well as PACTOLS, an archaeology-themed thesaurus for the scientific and academic community and is the main French thesaurus in archaeology facilitating FAIR interoperability. The Ministry of Research, which includes the development-led archaeology sector, is responsible for the vast majority of digital data produced in archaeology. Inrap, which is the central institute within this system, is in the process of developing a proactive policy for sharing the data resulting from its research in order to move towards processes and datasets that comply with the FAIR principles (Marx et al. 2021).

For Italy, work has centred on The National Geoportal for Archaeology (GNA) project as part of a broader project for the standardisation of scientific documentation for all archaeological research carried out in Italy. Achieving interoperability between the various databases within the GNA has been difficult due to large amounts of legacy data stored in various archives alongside new data. Digital objects are now geo-referenced using WGS84, as part of a comprehensive metadata description of the dataset (content, methods of acquisition and access, authors and managers, funding) based on shared standards and vocabularies, to increase the Findability, Accessibility and interoperability of the data (Calandra et al. 2021).

In Norway, excavations are mainly carried out by museums, along with the Norwegian Institute for Cultural Research), while archaeological surveys are mainly undertaken at the county level. Riksantikvaren (Directorate of National Heritage) holds the national Historic Environment Records (HER), and archives excavation documentation from churches and medieval cities. The Norwegian university museums run a national repository and make the collections available online. In addition, ADED (Archaeological Digital Excavation Documentation) offers detailed excavation documentation, and the BItFROST infrastructure project contributes to better storage and availability of 3D data. Aspects of FAIR are already in place through the use of vocabulary standards such as the MUSIT and ADED data models, which build on CIDOC CRM. Data openly available online at Unimusportalen, which gives access to all the university museums' collections and can be used under a clear CC-licence. Archaeology will also be included in the FAIR@UiO project at the University of Oslo, which will address how to create sustainable repositories for research data adhering to the FAIR principles (Matsumoto et al. 2021).

The country most engaged with the FAIRness of archaeological data is the Netherlands, not only with regard to data originating in the country, but in developing international best practice. The Dutch government is developing a national Dutch infrastructure for research data based on local and thematic

Digital Competence Centres (DCCs) to help researchers process and deposit their data according to FAIR principles. Archaeology is part of the DCC for the Humanities and Social Sciences, and as a mature discipline in digital research data management, a best practice exemplar for related disciplines. Data Archiving and Networked Services (DANS) is the Dutch national expertise centre and repository for research data and is one of the leading repositories in Europe with CoreTrustSeal certification. The e-Depot for Dutch archaeology is a collaboration between DANS and the Cultural Heritage Agency of the Netherlands (RCE) and enables the digital research data of Dutch archaeologists to remain accessible and usable in the long term, but DANS is launching a new infrastructure based on domain-specific Data Stations. To guarantee long-term and secure storage of the archaeological collection according to the newest standards, the existing archaeological data archive is being migrated to the Data Station Archaeology. DANS has a leading role in policy making, such as creating archaeological data management plans (DMP) and promoting data quality (Hollander and DANS 2021).

6. Conclusion

Archaeology is one of the leading proponents of Open Data in the arts and humanities, and already exhibits broad interest in FAIR, but the diversity of data types and methods used by archaeologists means adoption of FAIR will pose significant challenges, further necessitating urgent collaboration around best practice. A recurring theme is the amount of time and effort it takes to do the kinds of work that allows data to be made FAIR, by both the data creators and the repository. Archaeology is an exemplar of why FAIR will be more difficult to implement for Social Sciences and Humanities.

This deliverable attempted to create an assessment workflow to address FAIR data quality, using the ADS as a case study. The qualitative assessments resulted in recommendations for improvement that are fed back to those who advise the data creators (data managers) and at a policy level. The addition of the automated assessment of the F-UJI was an incredibly useful way to both see where ADS data is not FAIR in the ways it was expected, and for the explicit way in which it specifies the form the tool expects. This allows ADS to make practical decisions. Just as important is that F-UJI finds the ways where FAIRness can be improved at a technical level and fed back to technical staff. The next step in the workflow should be the quantitative assessment of the same data within an infrastructure such as ARIADNE, to assess the value of data aggregation for enhancing FAIRness, and finding issues when data goes 'into the wild' so that issues can be identified and fed back to the data managers supplying the (meta)data, and data aggregators. It is very important to have these feedback loops that can assess and adjust all along these workflows as work continues at larger and larger scales. Failure to do this is going to result in larger and larger disconnects between data creators, data managers and data users, resulting in data and metadata that is not fit for purpose.

This deliverable also sought to further contextualise the archaeology case study by synthesising recent, proximal work undertaken in collaboration with ADS that is highly relevant, such as the comprehensive international survey of repository practices undertaken by Geser (2021) for the ARIADNEplus project, and the work of 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). Taken together these elements constitute a comprehensive report on opening access to research data in the archaeology domain with regard to implementation of the FAIR Principles.

References

Archaeology Data Service. 2021a. “Guidelines for Depositors.” 2021. https://archaeologydataservice.ac.uk/advice/guidelinesForDepositors.xhtml.

———. 2021b. “Guides to Good Practice.” 2021. https://guides.archaeologydataservice.ac.uk/g2gpwiki/.

“ARIADNEplus.” n.d. ARIADNE PLUS. Accessed February 17, 2022. https://ariadne-infrastructure.eu/.

Calandra, Elena, Valeria Boi, Annalisa Falcone, Valeria Acconcia, Sara Di Giorgio, Flavia Massara, and Paola Ronzino. 2021. “Policy and Practice for Digital Archaeological Archiving in Italy.” *Internet Archaeology*, no. 58 (December). https://doi.org/10.11141/ia.58.27.

CARARE. 2013. CARARE Metadata Schema. 2013. https://pro.carare.eu/en/introduction-carare-aggregation-services/carare-metadata-schema/.

“DataCite Schema.” n.d. DataCite Schema. Accessed February 17, 2022. https://schema.datacite.org/.

Directorate-General for Research, and Innovation. 2018. *Turning FAIR Data into Reality : Final Report and Action Plan from the European Commission Expert Group on FAIR Data*. Publications Office of the European Union.

England, Historic. 2012. “MIDAS Heritage.” MIDAS Heritage. 2012. https://historicengland.org.uk/images-books/publications/midas-heritage/midas-heritage-2012-v1_1/.

European Commission, Directorate-General for Research and Innovation. 2020a. *Six Recommendations for Implementation of FAIR Practice by the FAIR in Practice Task Force of the European Open Science Cloud FAIR Working Group*. Publications Office of the European Union.

———. 2020b. *A Persistent Identifier (PID) Policy for the European Open Science Cloud (EOSC)*. Publications Office of the European Union.

FAIR Data Maturity Model Working Group. 2020. *FAIR Data Maturity Model. Specification and Guidelines*. https://zenodo.org/record/3909563.

“FAIRsFAIR.” n.d. Accessed February 17, 2022. https://www.fairsfair.eu/.

Forum on Information Standards in Heritage. 2021. “Monument Type Thesaurus.” http://www.heritage-standards.org.uk/wp-content/uploads/2021/02/Mon_alpha.pdf.

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

GoFAIR. 2017a. “R1.1: (Meta)data Are Released with a Clear and Accessible Data Usage License.” GO FAIR. November 24, 2017. https://www.go-fair.org/fair-principles/r1-1-metadata-released-clear-accessible-data-usage-license/.

———. 2017b. “R1.2: (Meta)data Are Associated with Detailed Provenance.” GO FAIR. November 24, 2017. https://www.go-fair.org/fair-principles/r1-2-metadata-associated-detailed-provenance/.

Hacıgüzeller, Piraye, Koen Van Daele, Frank Carpentier, and Raf Ribbens. 2021. “Digital Archiving of Archaeological Resources in Flanders (Belgium): A Brief Review.” *Internet Archaeology*, May. https://doi.org/10.11141/ia.58.7.

Hollander, Hella, and DANS. 2021. “Digital Dutch Archaeology: Future Perspectives.” *Internet Archaeology*, no. 58 (December). https://doi.org/10.11141/ia.58.28.

Huber, Robert, and Anusuriya Devaraju. 2021. "F-UJI - an Automated FAIR Data Assessment Tool." January 5, 2021. https://www.f-uji.net/.

Jakobsson, Ulf. 2021. "Digital Archaeological Archiving in Sweden: The Swedish National Data Service Perspective." *Internet Archaeology*, May. https://doi.org/10.11141/ia.58.18.

Kreiter, Attila. 2021. "The Hungarian Archaeology Database." *Internet Archaeology*, May. https://doi.org/10.11141/ia.58.9.

Lehväslaiho, Heikki, Jessica Parland-von Essen, Claudia Behnke, Heidi Laine, Leah Riungu-Kalliosaari, Yann Le Franc, and Christine Staiger. 2019. *D2.1 Report on FAIR Requirements for Persistence and Interoperability 2019*. https://doi.org/10.5281/zenodo.3557381.

L'Hours, Hervé, Ilona von Stein, Frans Huigen, Anusuriya Devaraju, Mustapha Mokrane, Joy Davidson, Jerry de Vries, Patricia Herterich, Linas Cepinskas, and Robert Huber. 2020a. *D4.2 Repository Certification Mechanism: A Recommendation on the Extended Requirements and Procedures*. https://doi.org/10.5281/zenodo.3835698.

———. 2020b. *CoreTrustSeal plus FAIR Overview*. https://doi.org/10.5281/zenodo.4003630.

Library of Congress. n.d. "The PREMIS Data Dictionary for Preservation Metadata." https://www.loc.gov/standards/premis/.

Löwenborg, Daniel, Maria Jonsson, Åsa Larsson, and Johan Nordinge. 2021. "A Turn towards the Digital. An Overview of Swedish Heritage Information Management Today." *Internet Archaeology*, May. https://doi.org/10.11141/ia.58.19.

Marx, Amala, Inrap, Kai Salas Rossenbach, Emmanuelle Bryas, Inrap, and Inrap. 2021. "Digital Archiving and Data Stewardship in French Archaeology." *Internet Archaeology*, no. 58 (December). https://doi.org/10.11141/ia.58.26.

Matsumoto, Mieko, Museum of Cultural History, University of Oslo, Espen Uleberg, and Museum of Cultural History, University of Oslo. 2021. "Curation of Digital Archaeological Data in Norway." *Internet Archaeology*, no. 58 (December). https://doi.org/10.11141/ia.58.29.

Molloy, Laura, Josefine Nordling, Marjan Grootveld, René van Horik, Angus Whyte, Joy Davidson, Patricia Herterich, et al. 2020. *D3.4 Recommendations on Practice to Support FAIR Data Principles*. https://doi.org/10.5281/zenodo.3924132.

Novák, David, Martin Kuna, and Olga Lečbychová. 2021. "Taming the Beast. Approaches to Digital Archiving in Czech Archaeology." *Internet Archaeology*, May. https://doi.org/10.11141/ia.58.5.

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.

Trognitz, Martina. 2021. "Saving Us from the Digital Dark Age: The Austrian Perspective." *Internet Archaeology*, May. https://doi.org/10.11141/ia.58.2.

Tsiafaki, Despoina, and Markos Katsianis. 2021. "Stewardship of Digital Archaeological Data in Greece: A Landscape of Fragmentation." *Internet Archaeology*, May. https://doi.org/10.11141/ia.58.8.

Whyte, Angus, Claudia Engelhart, Daniel Bangert, Gabin Kayumbi-Kabeya, Simon Lambert, Mark Thorley, Ryan O'Connor, Patricia Herterich, and Joy Davidson. 2019. *D3.2 FAIR Data Practice Analysis*. https://doi.org/10.5281/zenodo.3581353.

Wilkinson, Mark D., Michel Dumontier, I. Jsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie

Baak, Niklas Blomberg, et al. 2016. "The FAIR Guiding Principles for Scientific Data Management and Stewardship." *Scientific Data* 3 (March): 160018.

Wright, Holly, Julian Richards, Olivia Foster, Jessica Hendy, Carrie Wright, Paul Northrup, Troy Rasbury, and Ashley Coutu. 2020. *E-RIHS PP D5.3 Data Curation Policy*. https://doi.org/10.5281/zenodo.3946202.

List of tables