John Metz

ORCID: 0000-0002-6631-9810

INSC 590

Assignment #11: Data Preservation Project – Option #2

April 9, 2020

I Explored the re3data repository ( <https://www.re3data.org/>), which is a global registry providing information on and access to more than 2,400 research data repositories (RDR) of all types worldwide. Re3data acts as a clearinghouse to help guide interested parties to the research repositories that best meet their needs and interests. It is also a resource for researchers looking for a repository with which to store their data. Data repositories seeking to be listed by re3data apply and are considered based on the level and type of access allowed to the repository and its data as well as the clarity of their terms of use. Once accepted, an RDR is discoverable through a simple but powerful search interface with 27 possible filters to zero-in on repositories that meet a user’s criteria. I searched for archaeological data repositories using the search term, “archaeology,” and received 32 results. The information returned in the search helps the user to determine which RDR best meets their needs and includes a title, alternate title, a list of subjects covered, a list of the types of content hosted, the country where the repository exists, an abstract, and series of symbols denoting the level and terms of access, use, persistent identifiers, certifications, and policies.

The RDR I found most interesting was the Archaeology Data Service based in the United Kingdom (<https://archaeologydataservice.ac.uk>). The ADS is a digital archive that was founded on sound archival practices in collecting, managing, and making digital archaeological data available to researchers for over twenty years. As the old maxim goes, archaeology is an inherently destructive process, necessitating the comprehensive documentation of an excavation so that future researchers can replicate the context. National policy in England requires that all cultural resources that might be damaged or destroyed must be recorded and the information made accessible to the public. Established in 1996 in response to the rapid increase in born-digital archaeological data, the ADS was initially grant funded but adopted a business model of charging depositors based on the volume of data for the long-term preservation of that material (ADS, Our Work, n.d.). The types of data accepted by the ADS includes raw data, statistical data, databases, graphics, audiovisual content, image files, and text documents.

Data deposited with the ADS is managed within a framework that is Open Archival Information System (OAIS) compliant. This framework spells out the roles and responsibilities of data producers, managers, and consumers. In this case, the ADS receives a Submission Information Package (SIP) from the creator including data-type and file-level metadata, the Archival Information Package (AIP) created from the SIP includes data and metadata, and the Dissemination Information Package (DIP) created from the SIP or the AIP is provided to users as downloadable files. Adhering to the OAIS reference model ensures preservation of and access to data over the long term. ADS employs the DOI persistent identifier system and it supports 5 different metadata standards, including OAI-ORE (Open Archives Initiative Object Reuse and Exchange), Midas (Heritage), DataCite, Dublin Core, as well as their own metadata Schemas. They also provide access to a collections policy, a guide to repository operations, terms of use/access, and a preservation policy (<https://www.re3data.org/repository/r3d100000006>). ADS offers access to data under a Creative Commons Attribution 4.0 International (CC BY 4.0) license.

Like re3data, the ADS interface is also intuitive with links to search engines, depositing data, highlighted research projects making use of data in the repository, advice on topics related archaeological data management and use, a gallery of highlighted graphics from the repository, a help link, as well as an “About” link that provides information on the ADS framework and certification. Searching us done using a general portal or one of three more specific portals; ArchSearch (monument and event records), Archives (data), and Library (journals, books, and reports). A search using the ArchSearch and Archives portals offer four broad filters; what, when, where, and resource (e.g. funder, program). Within these filters are up to four additional levels of sub-filters to allow extremely targeted searching capability. The Library search engine employees a more standard bibliographic/archival search capability with multiple access points (e.g. author, title, publication type, etc.).

While many of the 1,300,000 resources made accessible by the ADS are survey forms and/or publications, 1,824 include data sets. For example, searching the excavations at Eynsham Abbey, 1989-92, takes the user to a landing page where there are links to an introduction (abstract), an in-depth overview of the project/resource, downloads, metadata, and usage statistics (https://archaeologydataservice.ac.uk/archives/view/eynsham\_OAU/. Clicking on the download link reveals text files (PDF), delimited text files (artifact inventory data), drawing files (sections and plans) in a variety of formats (DWG, DXF, PDF), and image files (JPG). Metadata fields include title, location, two types of grid references, subject, period, project dates, and available data types (https://archaeologydataservice.ac.uk/archives/view/eynsham\_OAU/metadata.cfm). The usage statistics provide monthly data on access, including archive visits, file downloads, page views, and cumulative totals for the three types of access.

The ADS is an established digital repository of the highest caliber. While it is not certified by the ICSU World Data System, it does have the Data Seal of Approval (initially awarded in 2011 and renewed in 2013). The Data Seal of Approval certification process is rigorous and judged by compliance in 16 areas of data management and curation. These include mission/scope, licenses, continuity of access, confidentiality, organizational infrastructure, expert guidance, data integrity and authenticity, appraisal, storage procedures, preservation plan, data quality, work flows, data discovery and identification, data reuse, technical infrastructure, and security (Core Trustworthy Data Depositories Requirements, 2019). The Data Seal of Approval uses the *Trustworthy Repositories Audit & Certification: Criteria and Checklist*, which in turn is based on the OAIS reference model (TRAC, n.d.). The ADS is also an associate member of the Digital Preservation Coalition, and it received the Digital Preservation Coalition (DPC) Decennial Award for the most outstanding contribution to digital preservation over the preceding decade in 2010.

References:

Archaeology Data Service (n.d.). Our Work. Retrieved from <https://archaeologydataservice.ac.uk/about/ourWork.xhtml>.

Center for Research Libraries (n.d.). *Trustworthy Repositories Audit & Certification: Criteria and Checklist (TRAC),* Metrics. Retrieved from <https://www.crl.edu/archiving-preservation/digital-archives/metrics-assessing-and-certifying/trac>.

Core trust Seal (2019). Core Trustworthy Data Repositories Requirements. Retrieved from <https://www.coretrustseal.org/wp-content/uploads/2019/02/CoreTrustSeal-Draft_Requirements_2020-2022_v00_01.pdf>.

 re3data.org (n.d.). Registry of Research Data Repositories. Retrieved from <https://doi.org/10.17616/R3D>.