Where Can I Find & Share Data?

Jonathan Wheeler[*]   Karl Benedict[†]

Repositories of various types are increasingly important in the preservation, discovery, access, and reuse of research data. This workshop is focused on fostering a discussion around the various interests that are in play in the development, use, and selection of specific repositories - both as sources for discovering and accessing data for reuse, but also for sharing and preservation of research project products.

Repositories: Open Data Infrastructure

Open Data: Why?

Policies and drivers include federal and other funder requirements:

- White House Office of Science and Technology Policy memo, *Expanding Public Access to the Results of Federally Funded Research*[1]
- Gates Foundation Open Access Policy[2]

The list goes on - NSF, NIH, DoE, DoD. . .

Policy (sticks) aside, there is likewise research into the cultural, scientific, and social benefits of making data open and reusable (carrots). The Fair Guiding Principles[3] provide a nice overview as well as a pretty useful acronym of what *open* means regarding data:

Findable **A**ccessible **I**nteroperable **R**eusable

We note the NSF is currently funding projects around research and data re-use.

Classes of Repositories

Repositories may be categorized in a number of different ways, depending upon needs:

Compliance with standards that define trustworthyness, measures of reliability, or other predefined characteristics, such as the

- Trusted Digital Repository (TDR) Checklist[4], or its successor
- Trustworthy Digital Repositories - ISO 16363[5]/CCSDS 652.0-M-1[6]
- Data Seal of Approval[7]

---

[*]UNM Research Data Services, jwheel01@unm.edu

[†]UNM Research Data Services, kbene@unm.edu

[1]White House Office of Science and Technology Policy (OTSP) (2013). *Expanding Public Access to the Results of Federally Funded Research.* https://obamawhitehouse.archives.gov/blog/2013/02/22/expanding-public-access-results-federally-funded-research

[2]Bill and Melinda Gates Foundation. *Open Access Policy.* https://www.gatesfoundation.org/how-we-work/general-information/open-access-policy

[3]Nature Scientific Data (2016). The FAIR Guiding Principles for scientific data management and stewardship. https://www.nature.com/articles/sdata201618

[4]RLG-NARA Task Force on Digital Repository and Certification (2007). *Trustworthy Repositories Audit & Certification: Criteria and Checklist.* Version 1, February 2007. Robin L. Dale and Bruce Ambacher, ed. The Center for Research Libraries & OCLC Online Computer Library Center, Inc. http://www.crl.edu/sites/default/files/d6/attachments/pages/trac_0.pdf

[5]International Organization for Standardization (ISO) (2012). *ISO 16363:2012 (CCSDS 652.0-R-1) Space data and information transfer systems – Audit and certification of trustworthy digital repositories.* https://www.iso.org/standard/56510.html

[6]The Consultative Committee for Space Data Systems (2011). *Audit and Certification of Trustworthy Digital Repositories Recommended Practice* CCSCS 652.0-M-1. https://public.ccsds.org/pubs/652x0m1.pdf

[7]Data Seal of Approval website (2018). https://www.datasealofapproval.org/en/
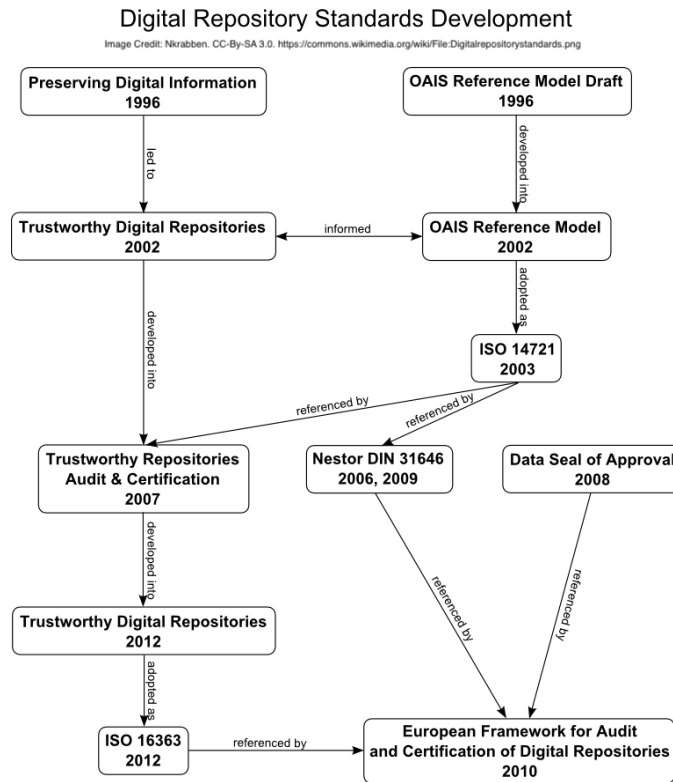
Digital Repository Standards Development

Figure 1: Diagram illustrating the development of digital repositories standards

Types of content that those repositories focus on:

- **Disciplinary repositories** that specialize in content produced by specific research disciplines such as ICPSR[8] for social science data, the Archaeology Data Service[9] for archaeological data, and GenBank[10] for genetic sequence data.
- **General or Interdisciplinary repositories** that contain content that cross disciplinary boundaries, and may also contain multiple types of content including data, documents, and multi-media files.

Or, the organizations that host/manage the repositories - frequently as a designated repository for data products produced by or with funds from those organizations. For example:

- Data.gov[11] - the discovery portal for data generated and provided by US Federal Government agencies, and some states, municipalities, and universities
- NASA's Distributed Active Archive Centers (DAACS)[12] - the 12 distributed data centers supported by NASA that provide discovery and access to different thematic collections of Earth science data.
- UNM's digital repository[13] - the designated *institutional repository* for the University of New Mexico.

Open Data: Where?

Requirements for repository services will vary from case to case, and there is no one-size-fits-all solution. Relevant factors include:

- Trusted repository status
- Capacity and commitment to preservation
- Provision of permanent identifiers (DOI)
- Format support
- Federation, discovery, and interoperability with other systems

---

[8]Inter-University Consortium for Political and Social Research (ICPSR). https://www.icpsr.umich.edu/icpsrweb/landing.jsp
[9]Archaeology Data Service (ADS). http://archaeologydataservice.ac.uk/
[10]GenBank. http://www.ncbi.nlm.nih.gov/genbank/
[11]Data.gov (2018). https://catalog.data.gov/dataset
[12]NASA Distributed Active Archive Centers (DAACS) (2018). https://earthdata.nasa.gov/about/daacs
[13]UNM's Institutional Repository (2018). http://digitalrepository.unm.edu

Often these and other repository characteristics and policies are documented. Resources for identifying and assessing repositories include:

- PLOS ONE Data Availability policy (includes a list of recommended repositories[14])
- Nature Scientific Data Recommended Data Repositories[15]
- Scientific Data recommended repositories[16]
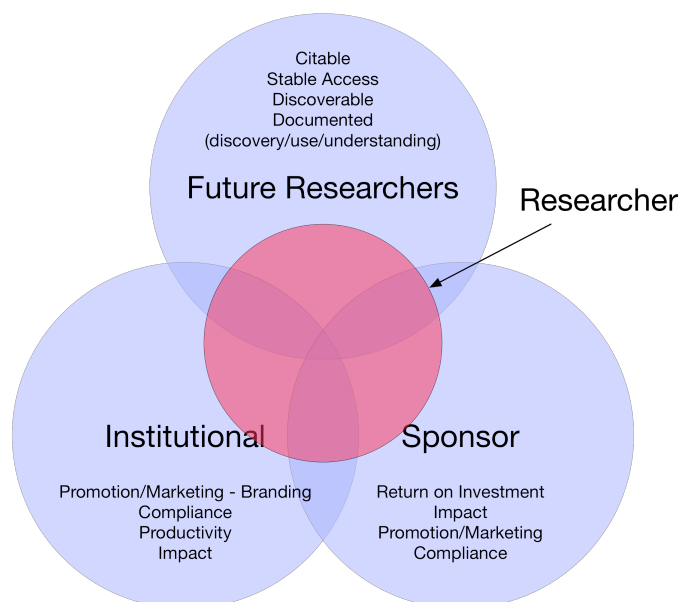- Registry of Research Data Repositories (re3data.org)[17]

A Synthesis

Figure 2: Intersecting Interests

Download this document: https://unmrds.github.io/bb-discovery/bb-discovery.pdf
Github Repository: https://github.com/unmrds/bb-discovery

Further Reading

- Baker, Karen S., and Lynn Yarmey. "Data Stewardship: Environmental Data Curation and a Web-of-Repositories." *International Journal of Digital Curation* 4, no. 2 (2009): 12-27. https://doi.org/10.2218/ijdc.v4i2.90
- Beagrie, Neil, and John Houghton. "The Value and Impact of Data Sharing and Curation: A Synthesis of Three Recent Studies of UK Research Data Centres." London: JISC, 2014. http://repository.jisc.ac.uk/5568/1/iDF308_-_Digital_Infrastructure_Directions_Report%2C_Jan14_v1-04.pdf

---

[14]PLOS ONE (2014). *Data Availability.* http://journals.plos.org/plosone/s/data-availability
[15]Nature Scientific Data (2018). *Recommended Data Repositories.* https://www.nature.com/sdata/policies/repositories
[16]Scientific Data (2018). *Scientific Data Recommended Repositories.* doi:10.6084/m9.figshare.1434640.v11
[17]Registry of Research Data Repositories (2018). https://www.re3data.org/