2 | Repository Options for Research Data

Katherine McNeill

Data are fundamental in virtually all forms of research yet time-intensive to collect and generate. Many research questions can be answered by using secondary data (that collected by another researcher), and thus data sharing has become of growing interest to funders and publishers. Effective data sharing depends upon repositories for long-term storage of and access to research data. In the context of this volume on the role of institutional repositories (IRs), various types of repositories are available for locally produced data: institutional repositories, domain repositories for specific types of data, and more. What options are available? How do researchers select a repository for deposit? What might institutions recommend to their researchers? How does the IR fit into this landscape? This chapter will answer these questions and share the experience of the library system of one research-extensive university in the United States, the Massachusetts Institute of Technology (MIT).

CONTEXT AND LITERATURE REVIEW

Data repositories serve a pivotal role in the data life cycle. The secondary use of data, if shared, enables further investigation and is almost always more efficient than collecting one's own data. The past 10 years have seen a dramatic increase in attention to this issue in many fields, building on the robust and long history of data sharing in some disciplines; for example, ICPSR (the Inter-university Consortium for Political and Social Research) has been preserving and providing access to quantitative data in the social sciences since 1962.[1] Data sharing has become of growing interest to funders worldwide, who aim to extend the impact of their funding, and publishers, who desire reproducibility of the research that they publish. Requirements of these bodies have raised the profile of data sharing and the role of data repositories.

There are numerous long-term benefits and significant time saved by managing data well, yet researchers find it challenging to invest time in managing, documenting, and sharing their data, and need support (Akers & Doty, 2013; Carlson, Fosmire, Miller, & Nelson, 2011; Housewright, Schonfeld, & Wulfson, 2013; Tenopir et al., 2011). The investment of time in preparing data for deposit in a repository, largely spent preparing the data and documentation for public use, generally still outweighs the perceived benefits to researchers of sharing their data.

What can be done at the institutional level to enable researchers to most effectively manage and share their data? What is the role of repositories? While researchers are experts in their academic fields, librarians bring skills in the management, organization, and preservation of information (Erway, 2013; Tenopir, Birch, & Allard, 2012) and can provide services alongside those of other units at their institution (Fearon, Gunia, Pralle, Lake, & Sallans, 2013; Hofelich Mohr & Lindsay, 2014). Librarians experienced in the discovery and use of data are equipped to advise researchers about the form that data and documentation should take to make them independently understandable for public use at the end of the data life cycle (McNeill, 2011). In addition, librarians can provide services for checking and preparing data for sharing (Peer, Green, & Stephenson, 2014). Moreover, librarians generally have well-developed connections with academic departments across their universities and are well positioned to work upstream in the research life cycle and enable the “last mile” of the research data management infrastructure (Gabridge, 2009).

Despite the resources and support needed to prepare data for deposit in a repository, the benefits to researchers in the long run are significant: having a researcher’s data in a repository makes it more readily discoverable, often relieves the data producer of the need to serve users, and can support a university’s ability to comply with sharing requirements and verify its research results. Moreover, select repositories provide curation features to enhance access and long-term preservation of the data.

DATA REPOSITORIES

Repositories for research data fit within a broader set of institutional research data management (RDM) services. Institutional repositories designed specifically for research data are neither the sole answer to RDM services, nor required for robust data management, as discussed below. Rather, data repositories, whether based at one's university or elsewhere, are key components of technological services that, along with consultative services, contribute to a robust university RDM infrastructure (Rice et al., 2013; Rice & Haywood, 2011; Soehner, Steeves, & Ward, 2010; Tenopir et al., 2012).

What kinds of repositories are available for data? How do storage needs for data differ from those for other types of materials? Data, in their varied forms (e.g., quantitative, qualitative, geospatial, images, models, binary files, code, and more), have different requirements in archiving than do most textual publications. File format obsolescence can be a significant challenge, given varied, complex, and rapidly changing data formats. The quality of data must be verified for effective reuse (Peer et al., 2014). Simple access to data alone is insufficient for public use; research data in any form are rarely self-describing and thus must be accompanied by documentation that adequately states the provenance, context, and content of the data files (Mauer & Watteler, 2013).

Academic institutions have available a range of repository options in order to track, store, preserve, and share research data created by their researchers. Within those options, data repositories differ along several key characteristics:[2]

- Association with an institution
- Specialization in a particular type of data
- Business model
- Levels of professional curation and unmediated deposit models

This final characteristic merits some discussion. Repositories—within and among the categories listed below—vary widely in the extent to which staff members manage data through activities such as accepting, depositing, reviewing, enhancing, managing, and preserving data and associated documentation (Peer et al., 2014). As data storage does not equal preservation, differing repository procedures dictate how well the data can be used in the future. Some repositories have extensive professional involvement, whereas others have an entirely unmediated deposit process and rely exclusively on the depositor to check the quality of the data and documentation. Some such as Dataverse rely upon software features for their curation and preservation.[3] A process of data quality review enables data to be "independently understandable for informed reuse," yet many repositories lack the services necessary to do so, placing that burden upon the researcher (Consultative Committee for Space Data Systems, 2012; Peer et al., 2014, p. 264). Moreover, only some repositories provide for long-term preservation beyond bit-level management, through activities such as emulation or migration of formats, sustainability, technology watch, and activities for usability over time (Choudhury, Palmer, Baker, & DiLauro, 2013; Treloar, Groenewegen, & Harboe-Ree, 2007). Review procedures necessarily place greater requirements on the depositor (e.g., for thorough data documentation) but doing so assures more usable data into the future.

University Institutional Repositories (IRs)

IRs are designed to house the scholarly output of researchers based at that institution, including data, and are at a close distance to the researcher (Baker & Yarmey, 2009). The major use case for an IR is for researchers, and universities, looking for a single common location for data regardless of subject and format, especially in cases where a suitable domain repository does not exist. However, given that data require significantly more management than do publications, IR administrators must consider what if any resources will be deployed to ensure data usability over time. Many IRs have an unmediated deposit process for datasets, but some universities have dedicated workflows for depositing and managing data within their IR (Awre & Duke, 2013; Johnston, 2014; Pink, 2012; Tarver & Phillips, 2012).

Local Data Repositories

A few universities—such as Johns Hopkins, Princeton University, Purdue University, the University of Bristol, and the University of Edinburgh—have created dedicated repositories exclusively for locally produced data.[4] Some even have been custom-designed for the work of a particular research group (Peer & Green, 2012). The major use case for such a repository is for those universities that want local archiving of data and are able to invest additional resources in a system tailored for this format.

National or Government Data Repositories

Researchers in select countries and regions (particularly in Europe) can avail themselves of extensive government infrastructures for storing and sharing research data, such as ReShare from the UK Data Service or Zenodo from the European Union.[5] Some may be specialized in nature and function as domain repositories. The major use case for such repositories is for a researcher with an eligible national or funder affiliation.

Domain Repositories

Data in domain repositories[6] are housed with similar data deposited by researchers from many institutions, which often improves the discovery of data in a particular realm. Such repositories focus on data from a particular subject realm (e.g., ecology, astronomy) and/or format type (e.g., quantitative data, qualitative data, images). Moreover, some repositories may provide particular features for working with or analyzing the particular data type; examples include ICPSR, Research Collaboratory for Structural Bioinformatics Protein Data Bank, National Snow and Ice Data Center, U.S. Virtual Astronomical Observatory, and the Qualitative Data Repository.[7] Another characteristic that varies: the business model of domain repositories can influence who is eligible to deposit, how open the data are for public use, and the curation services the repository can provide (Marcial & Hemminger, 2010). The major use case for a domain repository is for researchers who would like their data to be collocated with those in their subject field and utilize additional features or services provided to manage, access, or preserve that particular type of data.

Self-Deposit Independent Repositories

An emerging type of repository is one designed around self-deposit and self-management models, such as Dataverse and Figshare.[8] Developed and maintained by Harvard University, Dataverse is an open source software system for storing and providing access to quantitative data; Harvard makes its local installation of Dataverse open to deposit by any researcher group worldwide.[9] Such repositories provide researchers with a high level of control of the deposit process and collections and place little if any requirements on depositors. Information professionals, however, caution that while systems for self-archiving may appear to the researcher to have great ease of use, the lack of professional review makes them likely to result in inadequate documentation and ultimately the loss of usable data (Peer et al., 2014). Therefore, the use case for such a system is the researcher who either is willing and able to do a thorough review of data and documentation in advance of deposit, or values the independence of such a system above the assurances of future usability that a more professionally managed system would provide.

Journal Replication Data Archives

Journals increasingly require sharing the data that underlie a publication, as well as the computer code to generate the findings, in order to enable replication and further research. Journals vary in how they direct researchers to store and share their data, including on-demand requests of authors, journal Web sites, established journal data repositories, or deposit in a subject or domain repository.[10] Journal requirements thus necessarily influence the researchers' chosen mechanism for data sharing. Moreover, the policies and practices, and the method by which they are enforced, significantly affect the availability and ultimate usability of such data (McCullough, McGeary, & Harrison, 2008). In addition, journals with policies generally only require sharing of the data to reproduce tables in the paper and do not foster access to the full set of data generated in the research.

Staging Repositories

Complementing the options above, a select number of institutions have established formal systems for researchers to store, document, and work with data more systematically during the active phase of research, in order to facilitate ultimate deposit into a preservation repository (Steinhart, 2007; Smithsonian Institution, n.d.; Treloar et al., 2007). Similarly, scientific workflow systems are used by some researchers to structure their information during the active phase of research and could potentially be deployed in a centralized way (Littauer, Ram, Ludäscher, Michener, & Koskela, 2012; Lyle, Alter, & Green, 2014). Such systems have great potential value to institutions that have the resources and the organizational culture that would benefit from centralization of data management during the active phase of research.

The Role of the IR

What is the role of an IR in the context of this array of repositories? University IRs can serve as a fallback location for storing their researchers' data in the absence of a domain repository. Promoting the IR as the preferred repository for locally produced data can provide simplicity for researchers and service providers alike and potentially enable a more systematic transfer of data from active data storage. If universities have formal policies for RDM, they can align the characteristics of an IR with those policies and local user needs. In addition, universities leveraging their IR to meet funder requirements for open access to publications can examine its suitability to support data sharing requirements as well. Universities wanting to ensure long-term access and usability of their researchers' data, however, will need to consider what combination of services, policies, and quality assurance will be required.

SUPPORTING DATA REPOSITORY SELECTION: EXPERIENCES AT THE MASSACHUSETTS INSTITUTE OF TECHNOLOGY (MIT)

The main components of the MIT Libraries' RDM service are a Web guide,[11] instruction workshops, consulting, and repositories for long-term data storage and access. Our consultants advise researchers on a range of issues, including writing funder-required data management plans; storing, organizing, and documenting data files during the active phase of research; creating data documentation for public use; managing and sharing sensitive data; and selecting a repository for long-term public access to data.

REPOSITORIES AVAILABLE TO MIT RESEARCHERS

For data storage in the earlier phase of the data life cycle, MIT has no systematic centralized system and no services or tools for moving data from active storage to long-term repositories. Anecdotal evidence shows that MIT researchers store active data in centralized or departmental research computing systems, hard drives and other removable media, commercial cloud services, and Git-based repositories, among others. If and when they choose to share their research data and store it long-term after the conclusion of the research,[12] MIT researchers, like those at other institutions, have at their disposal a number of possible repositories, discussed in the following paragraphs.

DSpace@MIT

MIT's IR is based on the DSpace software that the MIT Libraries codeveloped (Smith, 2002). Like many IRs, DSpace@MIT[13] can house any file or material type produced or sponsored by MIT faculty, including data in any form (e.g., quantitative, images, audial, textual, and more), and the MIT Libraries have created policies for accepting datasets.[14] While staff involvement is required to create a collection, subsequent deposit by local researchers is unmediated, and library staff members do not review the data or documentation for completeness. In addition, assurances of long-term access depend upon the file format provided by the depositor. MIT researchers who have selected DSpace as their data repository of choice have stated reasons such as the association with the institution and collocation with other of their publications that may end up in the IR.

Harvard Dataverse Network

MIT researchers have selected Dataverse as their repository of choice for features such as public access to data, deposits accepted from any researcher based at any institution, collections can be individually branded, researchers maintain a high level of control of both the deposit process and their collections, and the historic relationship between MIT and Harvard.

Domain Repositories

As with researchers at other institutions, those at MIT can deposit in domain repositories. MIT researchers have selected such repositories for features such as collocation and discovery along with similar data, added features for their data type(s) (e.g., tailored metadata, analysis features, and preservation services), and specialized services that are beyond the expertise or resources of most university IRs or self-publishing systems (e.g., ICPSR can enable restricted access to sensitive data).[15]

THE SELECTION PROCESS

Given these various options, the MIT Libraries' RDM consultants, knowledgeable librarians, play a key role in assisting researchers who need a repository to select one, as part of broader data management conversations. Consultants provide each researcher with a menu of options, rather than suggesting a generally preferred repository. This practice arose from the experience of service providers, rather than being formally established. Library staff members articulate to the researcher differences among repositories and the risks and benefits of each, as the optimal choice depends upon the needs and preferences of the researcher. Even universities that have established a dedicated institutional data repository present it as one option among others (Johns Hopkins University Libraries, 2014; McGinty, 2014).

The MIT Libraries' research data management consultant will consider the data at hand and perform activities such as the following:

- Present the leading repository options, their most significant differences, and any requirements for preparing the data for deposit.
- Communicate with a potential repository to better understand its services.
- Coach the researcher to select a repository.
- At times, facilitate deposit of the researchers' data through activities such as these: communicate with repository staff to help the user prepare for the deposit process, advise on file formats, and review and provide feedback on data documentation for public use (McNeill, 2011). Services for quality review of data and documentation have not been widespread or integrated into most MIT consultations; doing so would be more involved and improve the quality of data deposited (Peer et al., 2014).

In summary, data repository technology, however essential, is not a service that can stand on its own. Local consultants, who can help the researcher select a repository and prepare appropriately, are critical to what researchers have termed the consultative and technological infrastructure required for RDM (Tenopir et al., 2012, p. 3).

NO SINGLE SOLUTION

For the foreseeable future, researchers at MIT and elsewhere will continue to choose varying repositories for their needs; some like the assurance that their data are being stored at their institution, others prefer the control and flexibility of self-archiving models such as Dataverse, and still others opt for a more full-service domain repository providing curation services to ensure that the data will be usable into the future. Some institutions will choose to prefer and urge deposit in their IR, whereas others will guide researchers through the optimal selection on a case-by-case basis. In this context, enabling data discovery among different repositories is of growing importance. Some institutions have developed formal mechanisms to track the data assets created at their university, regardless of location of deposit (Rice et al., 2013; Rice & Haywood, 2011; Rumsey & Neil, 2013; Wright et al., 2013). Such registries also are being created at the national level (Australian National Data Service, 2014; Molloy, 2014). RDM professionals also are working to develop complex solutions for interoperability among repositories in order to both facilitate data discovery across locations and link publications with their underlying data; efforts include the DataCite Metadata Search and activities of the Research Data Alliance/World Data System Publishing Data Services Working Group (Plale et al., 2013).[16] Moreover, repositories need not necessarily be in competition with one another for researchers' deposits, but rather could engage in complementary partnerships in support of data discovery, access, and preservation (Green & Gutmann, 2007; Lyle et al., 2014).

CONCLUSION

Future developments will influence the array of repository options at MIT. The MIT Libraries are collaborating more with departments on campus whose services relate to RDM, which may spawn new projects related to data repositories; for example, working with IT to streamline the storage and movement of data throughout the life cycle, partnering with our scholarly publication department to advance repository services for supporting federal public access requirements and linking data and publications, and more. In addition, future services and activities at MIT certainly will be influenced by developments in the field at large. It is vital to look outward at how other institutions are enabling long-term data storage and access and use those ideas to continually evolve local services.

Institutions considering developing or enhancing data repository services should consider several important issues. Universities should not assume that they must have a single solution for housing their researchers' data, or even that a university must house the data produced by its researchers, but local repositories can play key roles. If an institution is to accept data in its IR, or consider the creation of a local data repository, it must decide how data will be accepted and processed, what level of mediation will be suitable for the deposit process, what level of quality assurance of data and documentation are desired, and what workflows and communications to establish with researchers. When considering repositories as a service, institutions should spend equal effort considering the consultative services that will enable researcher selection and preparation for deposit into local or remote repositories. How will this process work and how can one best communicate the array of options to researchers? Will services (beyond guidelines and consultation) be provided to help researchers prepare data for sharing and long-term preservation?

In conclusion, data repositories play a vital role in enabling the storage, sharing, and secondary use of research data. Data repository options vary, and individual researchers need support finding the appropriate solution for their needs. Institutions must consider what array of options works best in their local context.

NOTES

I would like to thank Patsy Baudoin, Ellen Duranceau, and Ann Green for reviewing and providing helpful feedback and ideas for this chapter.

1. See http://www.icpsr.umich.edu/icpsrweb/content/membership/about.html.
2. The re3data.org schema describes repositories along a range of characteristics, including subject, content types, countries, type (disciplinary/institutional), terms of use and deposit, and more. See http://www.re3data.org/schema/2-1/ and Pampel et al. (2013).
3. See http://thedata.org/.
4. See https://archive.data.jhu.edu, http://dataspace.princeton.edu/jspui, https://research.hub.purdue.edu, http://data.bris.ac.uk/data, http://datashare.is.ed.ac.uk.
5. See http://reshare.ukdataservice.ac.uk/, https://zenodo.org/. Note: Zenodo now is open to researchers worldwide.
6. For a directory, see http://www.re3data.org/ and Pampel et al. (2013).
7. See http://www.icpsr.umich.edu, http://www.rcsb.org, http://nsidc.org/, http://www.usvao.org/, and https://qdr.syr.edu/
8. See http://figshare.com/.
9. Software: http://thedata.org; Harvard's installation: http://thedata.harvard.edu/
10. For examples of these respective practices, see AEA Journal Data and Program Archives: https://www.aeaweb.org/rfe/showCat.php?cat_id=9; Dryad journal integration: http://datadryad.org/pages/journalIntegration; American

Journal of Political Science (AJPS) Dataverse: http://dvn.iq.harvard.edu/dvn/dv/ajps; PKP-Dataverse Integration Project: http://projects.iq.harvard.edu/ojs-dvn; and Nature: http://www.nature.com/sdata/data-policies/repositories or BioMed Central: http://www.biomedcentral.com/about/supportingdata

11. See http://libraries.mit.edu/data-management.
12. For those who chose to do so. Not all researchers—even those under data sharing requirements—formally share their data via repositories; many continue to engage in very limited data management and sharing practices.
13. See http://dspace.mit.edu.
14. See http://libguides.mit.edu/content.php?pid=456907&sid=3741704.
15. See http://www.icpsr.umich.edu/icpsrweb/content/ICPSR/access/restricted.
16. See http://search.datacite.org and https://rd-alliance.org/group/rdawds-publishing-data-services-wg.html.

REFERENCES

Akers, K. G., & Doty, J. (2013). Disciplinary differences in faculty research data management practices and perspectives. *International Journal of Digital Curation, 8*(2), 5–26. http://dx.doi.org/10.2218/ijdc.v8i2.263

Australian National Data Service. (2014). Data discovery and access. Retrieved from http://www.ands.org.au/discovery/discoveryandaccess.html

Awre, C., & Duke, M. (2013). *Storing and sharing data in an institutional repository Hydra@Hull.* Edinburgh: Digital Curation Centre. Retrieved from http://www.dcc.ac.uk/resources/developing-rdm-services/storing-sharing-data-hull

Baker, K. S., & Yarmey, L. (2009). Data stewardship: Environmental data curation and a web-of-repositories. *International Journal of Digital Curation, 4*(2), 12–27. http://dx.doi.org/10.2218/ijdc.v4i2.90

Carlson, J., Fosmire, M., Miller, C. C., & Nelson, M. S. (2011). Determining data information literacy needs: A study of students and research faculty. *portal: Libraries and the Academy, 11*(2), 629–657. http://dx.doi.org/10.1353/pla.2011.0022

Choudhury, G. S., Palmer, C., Baker, K., & DiLauro, T. (2013, January). *Levels of services and curation for high-functioning data*. Presented at the Eighth International Digital Curation Conference, Amsterdam, Netherlands. Retrieved from http://www.dcc.ac.uk/webfm_send/1093

Consultative Committee for Space Data Systems. (2012). *Reference model for an open archival information system (OAIS)* (No. Magenta Book CCSDS 650.0

-M-2). Retrieved from http://public.ccsds.org/publications/archive/650x0m2.pdf

Erway, R. (2013). *Starting the conversation: University-wide research data management policy*. OCLC Research. Retrieved from http://www.conference-center.oclc.org/content/dam/research/publications/library/2013/2013-08.pdf

Fearon, D., Gunia, B., Pralle, B. E., Lake, S., & Sallans, A. L. (2013). Research data management services: SPEC Kit 334. Retrieved from http://publications.arl.org/Research-Data-Management-Services-SPEC-Kit-334

Gabridge, T. (2009). *The last mile: Liaison roles in curating science and engineering research data* (No. 265; pp. 15–22). Retrieved from http://www.arl.org/bm~doc/rli-265-gabridge.pdf

Green, A., & Gutmann, M. P. (2007). Building partnerships among social science researchers, institution-based repositories and domain specific data archives. *OCLC Systems & Services, 23*(1), 35–53. http://dx.doi.org/10.1108/10650750710720757

Hofelich Mohr, A., & Lindsay, T. (2014, June 4). *It takes a village: Strengthening data management through collaboration with diverse institutional offices*. Presented at the International Association for Social Science Information Services & Technology Annual Conference, Toronto, ON, Canada. Retrieved from http://www.library.yorku.ca/binaries/iassist2014/2H/2014_2H_HofelichMohr.pptx

Housewright, R., Schonfeld, R. C., & Wulfson, K. (2013). *Ithaka S+R US faculty survey 2012* (p. 44). Retrieved from http://www.sr.ithaka.org/research-publications/us-faculty-survey-2012

Johns Hopkins University Libraries. (2014). About storing & archiving your research data. Johns Hopkins University Libraries. Retrieved from http://dmp.data.jhu.edu/preserve-share-research-data/preserve-archive/

Johnston, L. (2014). *A workflow model for curating research data in the University of Minnesota Libraries: Report from the 2013 data curation pilot*. University of Minnesota Digital Conservancy. Retrieved from http://hdl.handle.net/11299/162338

Littauer, R., Ram, K., Ludäscher, B., Michener, W., & Koskela, R. (2012). Trends in use of scientific workflows: Insights from a public repository and recommendations for best practice. *International Journal of Digital Curation, 7*(2), 92–100. http://dx.doi.org/10.2218/ijdc.v7i2.232

Lyle, J., Alter, G., & Green, A. (2014). Partnering to curate and archive social science data. In J. M. Ray (Ed.), *Research data management: Practical strategies for information professionals* (pp. 203–221). West Lafayette, IN: Purdue University Press.

Marcial, L. H., & Hemminger, B. M. (2010). Scientific data repositories on the Web: An initial survey. *Journal of the American Society for Information Science and Technology, 61*(10), 2029–2048. http://dx.doi.org/10.1002/asi.21339

Mauer, R., & Watteler, O. (2013, May 30). *Data are like parachutes: They work best when open*. Presented at the International Association for Social Science Information Services & Technology Annual Conference, Cologne, Germany. Retrieved from http://www.iassistdata.org/downloads/2013/2013_pechakucha04_mauer_watteler.pdf

McCullough, B. D., McGeary, K. A., & Harrison, T. D. (2008). Do economics journal archives promote replicable research? *Canadian Journal of Economics/Revue Canadienne d'Économique, 41*(4), 1406–1420. http://dx.doi.org/10.1111/j.1540-5982.2008.00509.x

McGinty, S. (2014, February 6). Feedback on data storage. Retrieved from http://www.iassistdata.org/topic/other-topics/digital-repositories

McNeill, K. (2011, October 6). *Role of libraries in data stewardship: A view from one university*. Presented at the ICPSR Biennial Meeting of Official Representatives, Ann Arbor, MI. Retrieved from http://libraries.mit.edu/data-management/files/2014/05/role-of-libraries-in-data-stewardship.pdf

Molloy, L. (2014). *JISC research data registry and discovery service phase 1 final report*. JISC. Retrieved from http://www.dcc.ac.uk/webfm_send/1736

Pampel, H., Vierkant, P., Scholze, F., Bertelmann, R., Kindling, M., Klump, J., . . . Dierolf, U. (2013). Making research data repositories visible: The re3data.org registry. *PLoS ONE, 8*(11), e78080. http://dx.doi.org/10.1371/journal.pone.0078080

Peer, L., & Green, A. (2012). Building an open data repository for a specialized research community: Process, challenges and lessons. *International Journal of Digital Curation, 7*(1), 151–162. http://dx.doi.org/10.2218/ijdc.v7i1.222

Peer, L., Green, A., & Stephenson, E. (2014). Committing to data quality review. *International Journal of Digital Curation, 9*(1), 263–291. http://dx.doi.org/10.2218/ijdc.v9i1.317

Pink, C. (2012). *Building a data repository to meet an institution's needs*. Presented at the Institutional Data Repositories Workshop—Roles and Responsibilities, Open Repositories. Retrieved from http://www.dcc.ac.uk/webfm_send/870

Plale, B., McDonald, R. H., Chandrasekar, K., Kouper, I., Konkiel, S., Hedstrom, M. L., . . . Kumar, P. (2013). SEAD virtual archive: Building a federation of institutional repositories for long-term data preservation in sustainability science. *International Journal of Digital Curation, 8*(2), 172–180. http://dx.doi.org/10.2218/ijdc.v8i2.281

Rice, R., Ekmekcioglu, Ç., Haywood, J., Jones, S., Lewis, S., Macdonald, S., & Weir, T. (2013). Implementing the research data management policy: University of Edinburgh roadmap. *International Journal of Digital Curation, 8*(2), 194–204.

Rice, R., & Haywood, J. (2011). Research data management initiatives at University of Edinburgh. *International Journal of Digital Curation, 6*(2), 232–244. http://dx.doi.org/10.2218/ijdc.v6i2.199

Rumsey, S., & Neil, J. (2013). DataFinder: A research data catalogue for Oxford. *Ariadne* (71). Retrieved from http://www.ariadne.ac.uk/issue71/rumseyjefferies

Smith, M. (2002). DSpace: An institutional repository from the MIT Libraries and Hewlett Packard Laboratories. In *Research and advanced technology for digital libraries* (pp. 543–549). Springer. Retrieved from http://link.springer.com/chapter/10.1007/3-540-45747-X_40

Smithsonian Institution. (n.d.). SIdora. Retrieved from http://www.fedoracommons.org/node/60

Soehner, C., Steeves, C., & Ward, J. (2010). E-science and data support services: A study of ARL member institutions. *Association of Research Libraries*. Retrieved from http://www.arl.org/storage/documents/publications/escience-report-2010.pdf

Steinhart, G. (2007). DataStaR: An institutional approach to research data curation. *IASSIST Quarterly, 31*(34), 34–39.

Tarver, H., & Phillips, M. (2012). Integrating image-based research datasets into an existing digital repository infrastructure. *Cataloging & Classification Quarterly, 51*(1–3), 238–250. http://dx.doi.org/10.1080/01639374.2012.732203

Tenopir, C., Allard, S., Douglass, K., Aydinoglu, A. U., Wu, L., Read, E., . . . Frame, M. (2011). Data sharing by scientists: Practices and perceptions. *PloS One, 6*(6), e21101. http://dx.doi.org/10.1371/journal.pone.0021101

Tenopir, C., Birch, B., & Allard, S. (2012). *Academic libraries and research data services: Currentpractices and plans for the future; an ACRL white paper*. Association of College and Research Libraries, a division of the American Library Association. Retrieved from http://www.ala.org/acrl/sites/ala.org.acrl/files/content/publications/whitepapers/Tenopir_Birch_Allard.pdf

Treloar, A., Groenewegen, D., & Harboe-Ree, C. (2007). The data curation continuum: Managing data objects in institutional repositories. *D-Lib Magazine, 13*(9/10). http://dx.doi.org/10.1045/september2007-treloar

Wright, S. J., Kozlowski, W. A., Dietrich, D., Khan, H. J., Steinhart, G. S., & McIntosh, L. (2013). Using data curation profiles to design the datastar dataset registry. *D-Lib Magazine, 19*(7/8). http://dx.doi.org/10.1045/july2013-wright