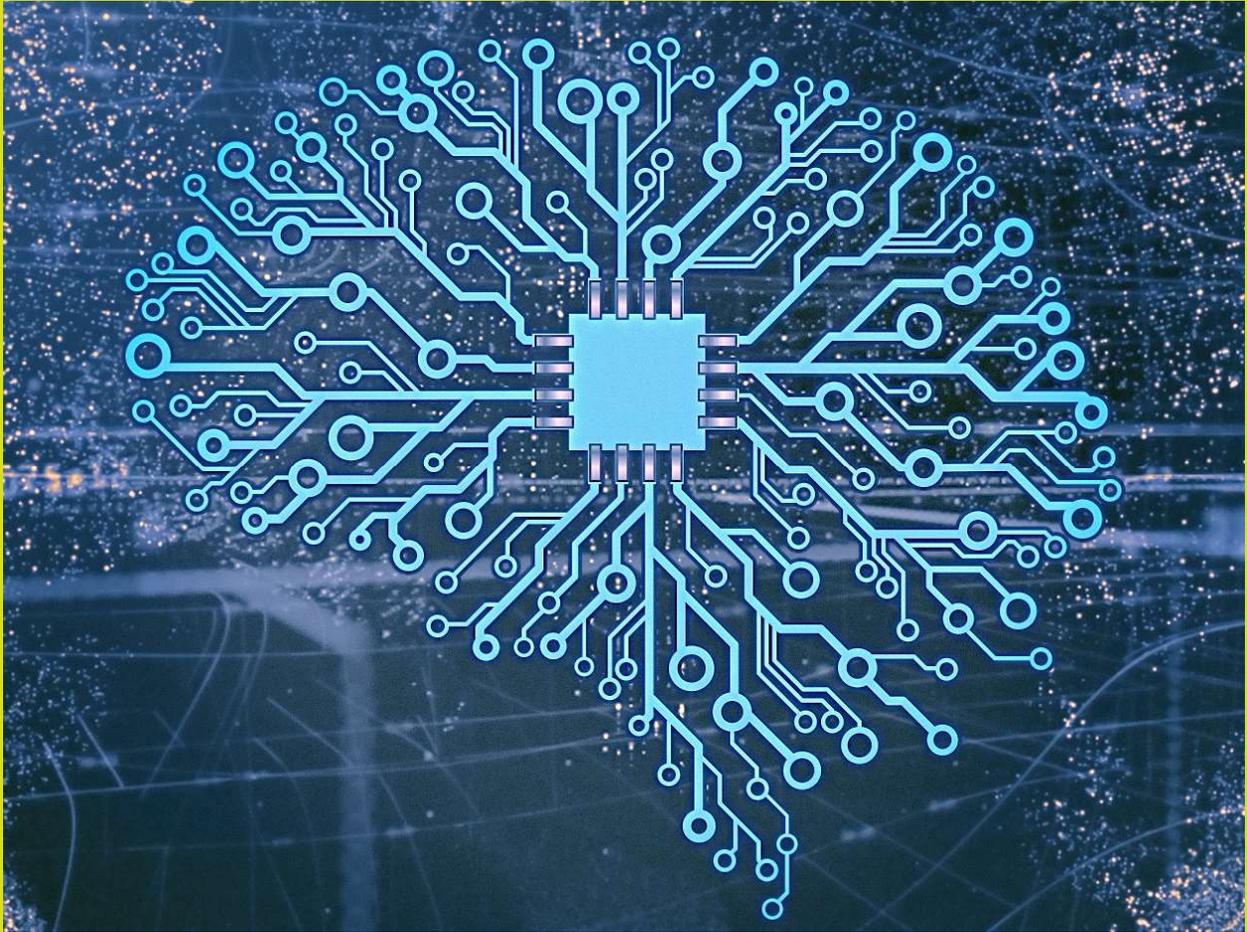


Research Data Curation and Management Bibliography



Charles W. Bailey Jr.

Research Data Curation and Management Bibliography

Copyright © 2021 by Charles W. Bailey, Jr.

This work is licensed under a Creative Commons Attribution 4.0 International License, <https://creativecommons.org/licenses/by/4.0/>.

To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/> or send a letter to Creative Commons, PO Box 1866, Mountain View, CA 94042, USA.

Original cover image before modification by Liam Huang. That image is under a Creative Commons Attribution 2.0 Generic License (<https://creativecommons.org/licenses/by/2.0/>).

Digital Scholarship, Houston, TX.

<http://www.digital-scholarship.org/>

The author makes no warranty of any kind, either express or implied, for information in the *Research Data Curation and Management Bibliography*, which is provided on an "as is" basis. The author does not assume and hereby disclaims any liability to any party for any loss or damage resulting from the use of information in the *Research Data Curation and Management Bibliography*.

Cite as: Bailey, Charles W., Jr. *Research Data Curation and Management Bibliography*. Houston: Digital Scholarship, 2021. <http://www.digital-scholarship.org/rdcmb/rdcmb.htm>.

Preface

The *Research Data Curation and Management Bibliography* includes over 800 selected English-language articles and books that are useful in understanding the curation of digital research data in academic and other research institutions.

The "digital curation" concept is still evolving. In "Digital Curation and Trusted Repositories: Steps toward Success," Christopher A. Lee and Helen R. Tibbo define digital curation as follows:

Digital curation involves selection and appraisal by creators and archivists; evolving provision of intellectual access; redundant storage; data transformations; and, for some materials, a commitment to long-term preservation. Digital curation is stewardship that provides for the reproducibility and re-use of authentic digital data and other digital assets. Development of trustworthy and durable digital repositories; principles of sound metadata creation and capture; use of open standards for file formats and data encoding; and the promotion of information management literacy are all essential to the longevity of digital resources and the success of curation efforts.¹

The *Research Data Curation and Management Bibliography* covers topics such as research data creation, acquisition, metadata, provenance, repositories, management, policies, support services, funding agency requirements, open access, peer review, publication, citation, sharing, reuse, and preservation. It is highly selective in its coverage.

The bibliography does not cover conference proceedings, digital media works (such as MP3 files), editorials, e-mail messages, interviews, letters to the editor, presentation slides or transcripts, technical reports, unpublished e-prints, or weblog postings.

Most sources have been published from January 2009 through December 2019; however, a limited number of earlier key sources are also included. The bibliography has links to included works. URLs may alter without warning (or automatic forwarding) or they may disappear altogether. Where possible, this bibliography uses Digital Object Identifier System (DOI) URLs. DOIs are not rechecked after initial validation. Publisher systems may have temporary DOI

resolution problems. Should a link be dead, try entering it in the Internet Archive [Wayback Machine](#).

Abstracts are included in this bibliography if a work is under a Creative Commons Attribution License (BY and national/international variations), a Creative Commons public domain dedication (CC0), or a Creative Commons Public Domain Mark and this is clearly indicated in the publisher's *current webpage* for the article. Note that a publisher may have changed the licenses for all articles on a journal's website but not have made corresponding license changes in journal's PDF files. The license on the current webpage is deemed to be the correct one. Since publishers can change licenses in the future, the license indicated for a work in this bibliography may not be the one you find upon retrieval of the work.

Unless otherwise noted, article abstracts in this bibliography are under a Creative Commons Attribution 4.0 International License, <https://creativecommons.org/licenses/by/4.0/>. Abstracts are reproduced as written in the source material.

See the Creative Commons' *Frequently Asked Questions* for a discussion of how documents under different Creative Commons licenses can be combined.

¹ Christopher A. Lee and Helen R. Tibbo, "Digital Curation and Trusted Repositories: Steps Toward Success," *Journal of Digital Information* 8, no. 2 (2007). <https://journals.tdl.org/jodi/index.php/jodi/article/view/229>

Bibliography

Aalbersberg, IJsbrand Jan, Sophia Atzeni, Hylke Koers, Beate Specker, and Elena Zudilova-Seinstra. "Bringing Digital Science Deep inside the Scientific Article: The Elsevier Article of the Future Project." *LIBER Quarterly* 23, no. 4 (2014): 275-299. <http://doi.org/10.18352/lq.8446>

In 2009, Elsevier introduced the "Article of the Future" project to define an optimal way for the dissemination of science in the digital age, and in this paper we discuss three of its key dimensions. First we discuss interlinking scientific articles and research data stored with domain-specific data repositories—such interlinking is essential to interpret both article and data efficiently and correctly. We then present easy-to-use 3D visualization tools embedded in online articles: a key example of how the digital article format adds value to scientific communication and helps readers to better understand research results. The last topic covered in this paper is automatic enrichment of journal articles through text-mining or other methods. Here we share insights from a recent survey on the question: how can we find a balance between creating valuable contextual links, without sacrificing the high-quality, peer-reviewed status of published articles?

Aalbersberg, IJsbrand, Judson Dunham, and Hylke Koers. "Connecting Scientific Articles with Research Data: New Directions in Online Scholarly Publishing." *Data Science Journal* 12 (2013): WDS235-WDS242. <https://doi.org/10.2481/dsj.wds-043>

Researchers across disciplines are increasingly utilizing electronic tools to collect, analyze, and organize data. However, when it comes to publishing their work, there are no common, well-established standards on how to make that data available to other researchers. Consequently, data are often not stored in a consistent manner, making it hard or impossible to find data sets associated with an article—even though such data might be essential to reproduce results or to perform further analysis. Data repositories can play an important role in improving this situation, offering increased visibility, domain-specific coordination, and expert knowledge on data management. As a leading STM publisher, Elsevier is actively pursuing opportunities to establish links between the online scholarly article and data repositories. This helps to increase usage and visibility for both articles and data sets and also adds valuable context to the data. These data-linking efforts tie in with other

initiatives at Elsevier to enhance the online article in order to connect with current researchers' workflows and to provide an optimal platform for the communication of science in the digital era.

Abrams, Stephen, Patricia Cruse, Carly Strasser, Perry Willet, Geoffrey Boushey, Julia Kochi, Megan Laurance, and Angela Rizk-Jackson. "DataShare: Empowering Researcher Data Curation." *International Journal of Digital Curation* 9, no. 1 (2014): 110-118. <https://doi.org/10.2218/ijdc.v9i1.305>

Researchers are increasingly being asked to ensure that all products of research activity—not just traditional publications—are preserved and made widely available for study and reuse as a precondition for publication or grant funding, or to conform to disciplinary best practices. In order to conform to these requirements, scholars need effective, easy-to-use tools and services for the long-term curation of their research data. The DataShare service, developed at the University of California, is being used by researchers to: (1) prepare for curation by reviewing best practice recommendations for the acquisition or creation of digital research data; (2) select datasets using intuitive file browsing and drag-and-drop interfaces; (3) describe their data for enhanced discoverability in terms of the DataCite metadata schema; (4) preserve their data by uploading to a public access collection in the UC3 Merritt curation repository; (5) cite their data in terms of persistent and globally-resolvable DOI identifiers; (6) expose their data through registration with well-known abstracting and indexing services and major internet search engines; (7) control the dissemination of their data through enforceable data use agreements; and (8) discover and retrieve datasets of interest through a faceted search and browse environment. Since the widespread adoption of effective data management practices is highly dependent on ease of use and integration into existing individual, institutional, and disciplinary workflows, the emphasis throughout the design and implementation of DataShare is to provide the highest level of curation service with the lowest possible technical barriers to entry by individual researchers. By enabling intuitive, self-service access to data curation functions, DataShare helps to contribute to more widespread adoption of good data curation practices that are critical to open scientific inquiry, discourse, and advancement.

Abrams, Stephen, John Kratz, Stephanie Simms, Marisa Strong, and Perry Willett. "Dash: Data Sharing Made Easy at the University of California." *International Journal of Digital Curation* 11, no. 1 (2016): 118-127. <https://doi.org/10.2218/ijdc.v11i1.408>

Scholars at the ten campuses of the University of California system, like their academic peers elsewhere, increasingly are being asked to ensure that data resulting from their research and teaching activities are subject to effective long-term management, public discovery, and retrieval. The new academic imperative for research data management (RDM) stems from mandates from public and private funding agencies, pre-publication requirements, institutional policies, and evolving norms of scholarly discourse. In order to meet these new obligations, scholars need access to appropriate disciplinary and institutional tools, services, and guidance. When providing help in these areas, it is important that service providers recognize the disparity in scholarly familiarity with data curation concepts and practices. While the UC Curation Center (UC3) at the California Digital Library supports a growing roster of innovative curation services for University use, most were intended originally to meet the needs of institutional information professionals, such as librarians, archivists, and curators. In order to address the new curation concerns of individual scholars, UC3 realized that it needed to deploy new systems and services optimized for stakeholders with widely divergent experiences, expertise, and expectations. This led to the development of Dash, an online data publication service making campus data sharing easy. While Dash gives the appearance of being a full-fledged repository, in actuality it is only a lightweight overlay layer that sits on top of standards-compliant repositories, such as UC3's existing Merritt curation repository. The Dash service offers intuitive, easy-to-use interfaces for dataset submission, description, publication, and discovery. By imposing minimal prescriptive eligibility and submission requirements; automating and hiding the mechanical details of DOI assignment, data packaging, and repository deposit; and featuring a streamlined, self-service user experience that can be integrated easily into scholarly workflows, Dash is an important new service offering with which UC scholars can meet their RDM obligations.

Adamick, Jessica, Rebecca C. Reznik-Zellen, and Matt Sheridan. "Data Management Training for Graduate Students at a Large Research University." *Journal of eScience Librarianship* 1, no. 3 (2013): e1022. <http://dx.doi.org/10.7191/jeslib.2012.1022>

Adams, Sam, and Peter Murray-Rust. "Chempound—A Web 2.0-Inspired Repository for Physical Science Data." *Journal of Digital Information* 13, no. 1 (2012). <http://journals.tdl.org/jodi/index.php/jodi/article/view/5873>

Addison, Aaron, and Jennifer Moore. "Teaching Users to Work with Research Data: Case Studies in Architecture, History and Social Work." *IASSIST Quarterly* 39, no. 4 (2015): 39-43. <https://doi.org/10.29173/iq905>

Addison, Aaron, Jennifer Moore, and Cynthia Hudson-Vitale. "Forging Partnerships: Foundations of Geospatial Data Stewardship." *Journal of Map & Geography Libraries* 11, no. 3 (2015): 359-375. <http://dx.doi.org/10.1080/15420353.2015.1054544>

Adu, Kofi Koranteng, Dube Luyande, and Adjei Emmanuel. "Digital Preservation: The Conduit through which Open Data, Electronic Government and the Right to Information Are Implemented." *Library Hi Tech* 34, no. 4 (2016): 733-747. <https://doi.org/10.1108/LHT-07-2016-0078>

Akers, Katherine G. "Going Beyond Data Management Planning: Comprehensive Research Data Services." *College & Research Libraries News* 75, no. 8 (2014): 435-436. <https://doi.org/10.5860/crln.75.8.9176>

———. "Looking Out for the Little Guy: Small Data Curation." *Bulletin of the American Society for Information Science and Technology* 39, no. 3 (2013): 58-59. <https://doi.org/10.1002/bult.2013.1720390317>

Akers, Katherine G., and Jennifer Doty. "Differences among Faculty Ranks in Views on Research Data Management." *IASSIST Quarterly* 36, no. 2 (2012): 16-20. <https://doi.org/10.29173/iq771>

———. "Disciplinary Differences in Faculty Research Data Management Practices and Perspectives." *International Journal of Digital Curation* 8, no. 2 (2013): 5-26. <https://doi.org/10.2218/ijdc.v8i2.263>

Academic librarians are increasingly engaging in data curation by providing infrastructure (e.g., institutional repositories) and offering services (e.g., data management plan consultations) to support the management of research data on their campuses. Efforts to develop these resources may benefit from a greater understanding of disciplinary differences in research data management needs. After conducting a survey of data management practices and perspectives at our research university, we categorized faculty members into four research domains—arts and humanities, social sciences, medical sciences, and basic sciences—and analyzed variations in their patterns of survey responses. We found statistically significant differences among the four research domains for nearly every survey item, revealing important

disciplinary distinctions in data management actions, attitudes, and interest in support services. Serious consideration of both the similarities and dissimilarities among disciplines will help guide academic librarians and other data curation professionals in developing a range of data-management services that can be tailored to the unique needs of different scholarly researchers.

Akers, Katherine G., and Jennifer A. Green. "Towards a Symbiotic Relationship between Academic Libraries and Disciplinary Data Repositories: A Dryad and University of Michigan Case Study." *International Journal of Digital Curation* 9, no. 1 (2014): 119-131. <https://doi.org/10.2218/ijdc.v9i1.306>

In addition to encouraging the deposit of research data into institutional data repositories, academic librarians can further support research data sharing by facilitating the deposit of data into external disciplinary data repositories.

In this paper, we focus on the University of Michigan Library and Dryad, a repository for scientific and medical data, as a case study to explore possible forms of partnership between academic libraries and disciplinary data repositories. We found that although few University of Michigan researchers have submitted data to Dryad, many have recently published articles in Dryad-integrated journals, suggesting significant opportunities for Dryad use on our campus. We suggest that academic libraries could promote the sharing and preservation of science and medical data by becoming Dryad members, purchasing vouchers to cover researchers' data submission costs, and hosting local curators who could directly work with campus researchers to improve the accuracy and completeness of data packages and thereby increase their potential for re-use.

By enabling the use of both institutional and disciplinary data repositories, we argue that academic librarians can achieve greater success in capturing the vast amounts of data that presently fail to depart researchers' hands and making that data visible to relevant communities of interest.

Akers, Katherine G., Fe C. Sferdean, Natsuko H. Nicholls, and Jennifer A. Green. "Building Support for Research Data Management: Biographies of Eight Research Universities." *International Journal of Digital Curation* 9, no. 2 (2014): 171-191. <https://doi.org/10.2218/ijdc.v9i2.327>

Academic research libraries are quickly developing support for research data management (RDM), including both new services and infrastructure. Here,

we tell the stories of how eight different universities have developed programs of RDM support, focusing on the prominent role of the library in educating and assisting researchers with managing their data throughout the research lifecycle. Based on these stories, we construct timelines for each university depicting key steps in building support for RDM, and we discuss similarities and dissimilarities among universities in motivation to provide RDM support, collaborations among campus units, assessment of needs and services, and changes in staffing.

Akmon, Dharma, and Susan Jekielek. "Restricting Data's Use: A Spectrum of Concerns in Need of Flexible Approaches." *IASSIST Quarterly*, 43, no. 3 (2019): 1–7. <https://doi.org/10.29173/iq941>

Akmon, Dharma, Ann Zimmerman, Morgan Daniels, and Margaret Hedstrom. "The Application of Archival Concepts to a Data-Intensive Environment: Working with Scientists to Understand Data Management and Preservation Needs." *Archival Science* 11, no. 3/4 (2011): 329-348. <https://doi.org/10.1007/s10502-011-9151-4>

Albani, Sergio, and David Giaretta. "Long-Term Preservation of Earth Observation Data and Knowledge in ESA through CASPAR." *International Journal of Digital Curation* 4, no. 3 (2009): 4-16. <https://doi.org/10.2218/ijdc.v4i3.127>

Aleixandre-Benaven, Rafael, Luz María Moreno-Solano, Antonia Ferrer Sapena, and Enrique Alfonso Sánchez Pérez. "Correlation between Impact Factor and Public Availability of Published Research Data in Information Science and Library Science Journals." *Scientometrics* 107, no. 1 (2016): 1-13. <https://doi.org/10.1007/s11192-016-1868-7>

Aleixandre-Benavent, Rafael, Antonia Ferrer Sapena, Silvia Coronado Ferrer, Fernanda Peset, and Alicia García. "Policies Regarding Public Availability of Published Research Data in Pediatrics Journals." *Scientometrics* 118 (2019): 439–451. <https://doi.org/10.1007/s11192-018-2978-1>

Allard, Suzie. "DataONE: Facilitating eScience through Collaboration." *Journal of eScience Librarianship* 1, no. 1 (2012): e1004. <https://doi.org/10.7191/jeslib.2012.1004>

Alma, Bridget. "Perseids: Experimenting with Infrastructure for Creating and Sharing Research Data in the Digital Humanities." *Data Science Journal* 16, no. 19 (2017). <http://doi.org/10.5334/dsj-2017-019>

The Perseids project provides a platform for creating, publishing, and sharing research data, in the form of textual transcriptions, annotations and analyses. An offshoot and collaborator of the Perseus Digital Library (PDL), Perseids is also an experiment in reusing and extending existing infrastructure, tools, and services. This paper discusses infrastructure in the domain of digital humanities (DH). It outlines some general approaches to facilitating data sharing in this domain, and the specific choices we made in developing Perseids to serve that goal. It concludes by identifying lessons we have learned about sustainability in the process of building Perseids, noting some critical gaps in infrastructure for the digital humanities, and suggesting some implications for the wider community.

Alqasab, Mariam, Suzanne M. Embury, and Sandra de F. Mendes Sampaio. "Amplifying Data Curation Efforts to Improve the Quality of Life Science Data." *International Journal of Digital Curation* 12, no. 1 (2017): 1-12.
<http://dx.doi.org/10.2218/ijdc.v12i1.495>

In the era of data science, datasets are shared widely and used for many purposes unforeseen by the original creators of the data. In this context, defects in datasets can have far reaching consequences, spreading from dataset to dataset, and affecting the consumers of data in ways that are hard to predict or quantify. Some form of waste is often the result. For example, scientists using defective data to propose hypotheses for experimentation may waste their limited wet lab resources chasing the wrong experimental targets. Scarce drug trial resources may be used to test drugs that actually have little chance of giving a cure.

Because of the potential real world costs, database owners care about providing high quality data. Automated curation tools can be used to an extent to discover and correct some forms of defect. However, in some areas human curation, performed by highly-trained domain experts, is needed to ensure that the data represents our current interpretation of reality accurately. Human curators are expensive, and there is far more curation work to be done than there are curators available to perform it. Tools and techniques are needed to enable the full value to be obtained from the curation effort currently available.

In this paper, we explore one possible approach to maximising the value obtained from human curators, by automatically extracting information about data defects and corrections from the work that the curators do. This

information is packaged in a source independent form, to allow it to be used by the owners of other databases (for which human curation effort is not available or is insufficient). This amplifies the efforts of the human curators, allowing their work to be applied to other sources, without requiring any additional effort or change in their processes or tool sets. We show that this approach can discover significant numbers of defects, which can also be found in other sources.

Alsheikh-Ali, Alawi A., Waqas Qureshi, Mouaz H. Al-Mallah, and John P. A. Ioannidis. "Public Availability of Published Research Data in High-Impact Journals." *PLoS ONE* 6, no.9 (2011): e24357.
<https://doi.org/10.1371/journal.pone.0024357>

The observations of scientists as coded in their primary data constitute a central commodity in the scientific enterprise [1]. Reproduction of research findings and further exploration of related hypotheses require access to these primary data, and their public availability has been a concern for all stakeholders of the scientific process, including regulatory and funding agencies, journal editors, individual researchers, and patients [2], [3], [4], [5], [6], [7]. Recently, efforts have converged to encourage making data, protocols, and analytical codes available, as part of the growing movement of reproducible research [8], [9], [10]. The benefits and challenges of public data availability and data sharing have long been hotly discussed in the scientific community [11], [12], [13]. Recent analyses have empirically highlighted deficiencies in the practice of making primary data and protocols available in peer-reviewed publications [14], [15], [16]. These analyses, however, have focused on either a particular discipline or area of research or were limited to a single journal. To date, there has not been an empiric evaluation of public availability of primary data and related material and protocols across diverse scientific fields or journals. We aimed to assess the current status of these practices in the most highly-cited journals across the scientific literature.

Altman, Micah, Margaret O. Adams, Jonathan Crabtree, Darrell Donakowski, Marc Maynard, Amy Pienta, and Copeland H. Young. "Digital Preservation through Archival Collaboration: The Data Preservation Alliance for the Social Sciences." *American Archivist* 72, no. 1 (2009): 170-184.
<https://doi.org/10.17723/aarc.72.1.eu7252lhnrp7h188>

Altman, Micah, Christine Borgman, Mercè Crosas, and Maryann Matone. "An Introduction to the Joint Principles for Data Citation." *Bulletin of the Association*

for Information Science and Technology 41, no. 3 (2015): 43-45.
<https://doi.org/10.1002/bult.2015.1720410313>

Altman, Micah, Eleni Castro, Mercè Crosas, Philip Durbin, Alex Garnett, and Jen Whitney. "Open Journal Systems and Dataverse Integration—Helping Journals to Upgrade Data Publication for Reusable Research." *Code4Lib Journal*, no. 30 (2015). <http://journal.code4lib.org/articles/10989>

This article describes the novel open source tools for open data publication in open access journal workflows. This comprises a plugin for Open Journal Systems that supports a data submission, citation, review, and publication workflow; and an extension to the Dataverse system that provides a standard deposit API. We describe the function and design of these tools, provide examples of their use, and summarize their initial reception. We conclude by discussing future plans and potential impact.

This work is licensed under a Creative Commons Attribution 3.0 United States License, <https://creativecommons.org/licenses/by/3.0/us/>.

Altman, Micah, and Gary King. "A Proposed Standard for the Scholarly Citation of Quantitative Data." *D-Lib Magazine* 13, no. 3/4 (2007).
<http://www.dlib.org/dlib/march07/altman/03altman.html>

Amorim, Ricardo, João Castro, João Rocha da Silva, and Cristina Ribeiro. "A Comparison of Research Data Management Platforms: Architecture, Flexible Metadata and Interoperability." *Universal Access in the Information Society* 16, no. 4 (2017): 851-862. <https://doi.org/10.1007/s10209-016-0475-y>

Anastasiadis, Stergios V., Syam Gadde, and Jeffrey S. Chase. "Scale and Performance in Semantic Storage Management of Data Grids." *International Journal on Digital Libraries* 5, no. 2 (2005): 84-98.
<https://doi.org/10.1007/s00799-004-0086-8>

Anderson, W. L. "Some Challenges and Issues in Managing, and Preserving Access to, Long Lived Collections of Digital Scientific and Technical Data." *Data Science Journal* 3 (2004): 191-201.
<https://datascience.codata.org/articles/abstract/301/>

One goal of the Committee on Data for Science and Technology is to solicit information about, promote discussion of, and support action on the many issues related to scientific and technical data preservation, archiving, and

access. This brief paper describes four broad categories of issues that help to organize discussion, learning, and action regarding the work needed to support the long-term preservation of, and access to, scientific and technical data. In each category, some specific issues and areas of concern are described.

Andreoli-Versbach, Patrick, and Frank Mueller-Langer. "Open Access to Data: An Ideal Professed but Not Practised." *Research Policy* 43, no. 9 (2014): 1621-1633. <https://doi.org/10.1016/j.respol.2014.04.008>

Androulakis, Steve, Ashley M. Buckle, Ian Atkinson, David Groenewegen, Nick Nicholas, Andrew Treloar, and Anthony Beitz. "ARCHER—e-Research Tools for Research Data Management." *International Journal of Digital Curation* 4, no. 1 (2009): 22-33. <https://doi.org/10.2218/ijdc.v4i1.75>

Angevaare, Inge. "Taking Care of Digital Collections and Data: 'Curation' and Organisational Choices for Research Libraries." *LIBER Quarterly: The Journal of European Research Libraries* 19, no. 1 (2009): 1-12. <http://doi.org/10.18352/lq.7948>

This article explores the types of digital information research libraries typically deal with and what factors might influence libraries' decisions to take on the work of data curation themselves, to take on the responsibility for data but market out the actual work, or to leave the responsibility to other organisations. The article introduces the issues dealt with in the LIBER Workshop 'Curating Research' to be held in The Hague on 17 April 2009 (<http://www.kb.nl/curatingresearch>) and this corresponding issue of *LIBER Quarterly*.

Aquino, Janine, John Allison, Robert Rilling, Don Stott, Kathryn Young, and Michael Daniels. "Motivation and Strategies for Implementing Digital Object Identifiers (DOIs) at NCAR's Earth Observing Laboratory—Past Progress and Future Collaborations." *Data Science Journal* 16, no. 7 (2017). <http://doi.org/10.5334/dsj-2017-007>

In an effort to lead our community in following modern data citation practices by formally citing data used in published research and implementing standards to facilitate reproducible research results and data, while also producing meaningful metrics that help assess the impact of our services, the National Center for Atmospheric Research (NCAR) Earth Observing Laboratory (EOL) has implemented the use of Digital Object Identifiers

(DOIs) (DataCite 2017) for both physical objects (e.g., research platforms and instruments) and datasets. We discuss why this work is important and timely, and review the development of guidelines for the use of DOIs at EOL by focusing on how decisions were made. We discuss progress in assigning DOIs to physical objects and datasets, summarize plans to cite software, describe a current collaboration to develop community tools to display citations on websites, and touch on future plans to cite workflows that document dataset processing and quality control. Finally, we will review the status of efforts to engage our scientific community in the process of using DOIs in their research publications.

Arora, Ritu, Maria Esteva, and Jessica Trelogan. "Leveraging High Performance Computing for Managing Large and Evolving Data Collections." *International Journal of Digital Curation* 9, no. 2 (2014): 17-27.
<https://doi.org/10.2218/ijdc.v9i2.331>

The process of developing a digital collection in the context of a research project often involves a pipeline pattern during which data growth, data types, and data authenticity need to be assessed iteratively in relation to the different research steps and in the interest of archiving. Throughout a project's lifecycle curators organize newly generated data while cleaning and integrating legacy data when it exists, and deciding what data will be preserved for the long term. Although these actions should be part of a well-oiled data management workflow, there are practical challenges in doing so if the collection is very large and heterogeneous, or is accessed by several researchers contemporaneously. There is a need for data management solutions that can help curators with efficient and on-demand analyses of their collection so that they remain well-informed about its evolving characteristics. In this paper, we describe our efforts towards developing a workflow to leverage open science High Performance Computing (HPC) resources for routinely and efficiently conducting data management tasks on large collections. We demonstrate that HPC resources and techniques can significantly reduce the time for accomplishing critical data management tasks, and enable a dynamic archiving throughout the research process. We use a large archaeological data collection with a long and complex formation history as our test case. We share our experiences in adopting open science HPC resources for large-scale data management, which entails understanding usage of the open source HPC environment and training users. These experiences can be generalized to meet the needs of other data curators working with large collections.

Aschenbrenner, Andreas, Harry Enke, Thomas Fischer, and Jens Ludwig. "Diversity and Interoperability of Repositories in a Grid Curation Environment." *Journal of Digital Information* 12, no. 2 (2011). <http://journals.tdl.org/jodi/index.php/jodi/article/view/1896>

Asher, Andrew, and Lori M. Jahnke. "Curating the Ethnographic Moment." *Archive Journal*, no. 3 (2013). <http://www.archivejournal.net/issue/3/archives-remixed/curating-the-ethnographic-moment/>

Ashley, Kevin. "Data Quality and Curation." *Data Science Journal* 12 (2013): GRDI65-GRDI68. <https://doi.org/10.2481/dsj.grdi-011>

Data quality is an issue that touches on every aspect of the research data landscape and is therefore appropriate to examine in the context of planning for future research data infrastructures. As producers, researchers want to believe that they produce high quality data; as consumers, they want to obtain data of the highest quality. Data centres typically have stringent controls to ensure that they only acquire and disseminate data of the highest quality. Data managers will usually say that they improve the quality of the data they are responsible for. Much of the infrastructure that will emit, transform, integrate, visualise, manage, analyse, and disseminate data during its life will have dependencies, explicit or implicit, on the quality of the data it is dealing with.

——— "Research Data and Libraries: Who Does What." *Insights: The UKSG Journal* 25, no. 2 (2012): 155-157. <https://doi.org/10.1629/2048-7754.25.2.155>

A range of external pressures are causing research data management (RDM) to be an increasing concern at senior level in universities and other research institutions. But as well as external pressures, there are also good reasons for establishing effective research data management services within institutions which can bring benefits to researchers, their institutions and those who publish their research. In this article some of these motivating factors, both positive and negative, are described. Ways in which libraries can play a role—or even lead—in the development of RDM services that work within the institution and as part of a national and international research data infrastructure are also set out.

Assante, Massimiliano, Leonardo Candela, Donatella Castelli, and Alice Tani. "Are Scientific Data Repositories Coping with Research Data Publishing?" *Data Science Journal* 15, no. 6 (2016): 1-24. <http://doi.org/10.5334/dsj-2016-006>

Research data publishing is intended as the release of research data to make it possible for practitioners to (re)use them according to "open science" dynamics. There are three main actors called to deal with research data publishing practices: researchers, publishers, and data repositories. This study analyses the solutions offered by generalist scientific data repositories, i.e., repositories supporting the deposition of any type of research data. These repositories cannot make any assumption on the application domain. They are actually called to face with the almost open ended typologies of data used in science. The current practices promoted by such repositories are analysed with respect to eight key aspects of data publishing, i.e., dataset formatting, documentation, licensing, publication costs, validation, availability, discovery and access, and citation. From this analysis it emerges that these repositories implement well consolidated practices and pragmatic solutions for literature repositories. These practices and solutions can not totally meet the needs of management and use of datasets resources, especially in a context where rapid technological changes continuously open new exploitation prospects.

Atwood, Thea P., Patricia B. Condon, Julie Goldman, Tom Hohenstein, Carolyn V. Mills, and Zachary W. Painter. "Grassroots Professional Development via the New England Research Data Management Roundtables." *Journal of eScience Librarianship* 6, no. 2 (2017): e1111. <https://doi.org/10.7191/jeslib.2017.1111>

Objectives: To meet the changing needs of our campuses, librarians responsible for research data services are often tasked with starting new endeavors with new populations without much support. This paper reports on a collaborative effort to build a community of practice of librarians tasked with addressing the research data needs of their campuses, describes how this effort was evaluated, and presents future opportunities.

Methods: In March of 2015, three librarians found themselves in a situation of serendipitous professional development: one was seeking to provide a new method of mentorship, and two more were working on an event, hoping to broadcast it to a wider community. From these two disparate goals, the Research Data Management (RDM) Roundtables were created. The RDM Roundtables planning committee developed a low-cost professional development day divided into two parts: a morning session that detailed an idea or solution relevant to our practice, and an afternoon roundtable discussion on practical aspects of research data services. Evaluations from these events were coded in NVivo and we report on the common themes.

Results: Participants returned sixty-one evaluations from four events. Five themes emerged from the evaluations: learning, sharing, format, networking, and empathy.

Conclusion: The events provide a valuable professional development experience for attendees, and the authors hope that by providing a description of the events' development, others will establish their own local communities of practice.

Austin, Claire C., Theodora Bloom, Sünje Dallmeier-Tiessen, Varsha K. Khodiyar, Fiona Murphy, Amy Nurnberger, Lisa Raymond, Martina Stockhause, Jonathan Tedds, Mary Vardigan, and Angus Whyte. "Key Components of Data Publishing: Using Current Best Practices to Develop a Reference Model for Data Publishing." *International Journal on Digital Libraries* 18, no. 2 (2017): 77-92. <https://doi.org/10.1007/s00799-016-0178-2>

Austin, Claire C., Susan Brown, Nancy Fong, Chuck Humphrey, Amber Leahey, and Peter Webster. "Research Data Repositories: Review of Current Features, Gap Analysis, and Recommendations for Minimum Requirements." *IASSIST Quarterly* 39, no. 4 (2015): 24-38. <https://doi.org/10.29173/iq904>

Aydinoglu, Arsev Umur, Dogan Guleda, and Taskin Zehra. "Research Data Management in Turkey: Perceptions and Practices." *Library Hi Tech* 35, no. 2 (2017): 271-289. <https://doi.org/10.1108/LHT-11-2016-0134>

Bache, Richard, Simon Miles, Bolaji Coker, and Adel Taweel. "Informative Provenance for Repurposed Data: A Case Study using Clinical Research Data." *International Journal of Digital Curation* 8, no. 2 (2013): 27-46. <https://doi.org/10.2218/ijdc.v8i2.262>

The task repurposing of heterogeneous, distributed data for originally unintended research objectives is a non-trivial problem because the mappings required may not be precise. A particular case is clinical data collected for patient care being used for medical research. The fact that research repositories will record data differently means that assumptions must be made as how to transform of this data. Records of provenance that document how this process has taken place will enable users of the data warehouse to utilise the data appropriately and ensure that future data added from another source is transformed using comparable assumptions. For a provenance-based approach to be reusable and supportable with software tools, the provenance records must use a well-defined model of the transformation process. In this

paper, we propose such a model, including a classification of the individual 'sub-functions' that make up the overall transformation. This model enables meaningful provenance data to be generated automatically. A case study is used to illustrate this approach and an initial classification of transformations that alter the information is created.

Baker, Karen S., Ruth E. Duerr, and Mark A. Parsons. "Scientific Knowledge Mobilization: Co-evolution of Data Products and Designated Communities." *International Journal of Digital Curation* 10, no. 2 (2015): 110-135. <https://doi.org/10.2218/ijdc.v10i2.346>

Digital data are accumulating rapidly, yet issues relating to data production remain unexamined. Data sharing efforts in particular are nascent, disunited and incomplete. We investigate the development of data products tailored for diverse communities with differing knowledge bases. We explore not the technical aspects of how, why, or where data are made available, but rather the socio-scientific aspects influencing what data products are created and made available for use. These products differ from compact data summaries often published in journals. We report on development by a national data center of two data collections describing the changing polar environment. One collection characterizes sea ice products derived from satellite remote sensing data and development unfolds over three decades. The second collection characterizes the Greenland Ice Sheet melt where development of an initial collection of data products over a period of several months was informed by insights gained from earlier experience. In documenting the generation of these two collections, a data product development cycle supported by a data product team is identified as key to mobilizing scientific knowledge. The collections reveal a co-evolution of data products and designated communities where community interest may be triggered by events such as environmental disturbance and new modes of communication. These examples of data product development in practice illustrate knowledge mobilization in the earth sciences; the collections create a bridge between data producers and a growing number of audiences interested in making evidence-based decisions.

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>

Balkestein, Marjan, and Heiko Tjalsma. "The ADA Approach: Retro-archiving Data in an Academic Environment." *Archival Science* 7, no. 1 (2007): 89-105.
<https://doi.org/10.1007/s10502-007-9053-7>

Ball, Alexander, Kevin Ashley, Patrick McCann, Laura Molloy, and Veerle Van den Eynden. "Show Me The Data: The Pilot UK Research Data Registry." *International Journal of Digital Curation* 9, no. 1 (2014): 132-141.
<https://doi.org/10.2218/ijdc.v9i1.307>

The UK Research Data (Metadata) Registry (UKRDR) pilot project is implementing a prototype registry for the UK's research data assets, enabling the holdings of subject-based data centres and institutional data repositories alike to be searched from a single location. The purpose of the prototype is to prove the concept of the registry, and uncover challenges that will need to be addressed if and when the registry is developed into a sustainable service. The prototype is being tested using metadata records harvested from nine UK data centres and the data repositories of nine UK universities.

Ball, Alexander, Sean Chen, Jane Greenberg, Cristina Perez, Keith Jeffery, and Rebecca Koskela. "Building a Disciplinary Metadata Standards Directory." *International Journal of Digital Curation* 9, no. 1 (2014): 142-151.
<https://doi.org/10.2218/ijdc.v9i1.308>

The Research Data Alliance (RDA) Metadata Standards Directory Working Group (MSDWG) is building a directory of descriptive, discipline-specific metadata standards. The purpose of the directory is to promote the discovery, access and use of such standards, thereby improving the state of research data interoperability and reducing duplicative standards development work.

This work builds upon the UK Digital Curation Centre's Disciplinary Metadata Catalogue, a resource created with much the same aim in mind. The first stage of the MSDWG's work was to update and extend the information contained in the catalogue. In the current, second stage, a new platform is being developed in order to extend the functionality of the directory beyond that of the catalogue, and to make it easier to maintain and sustain. Future work will include making the directory more amenable to use by automated tools.

Ball, Alexander, Mansur Darlington, Thomas Howard, Chris McMahon, and Steve Culley. "Visualizing Research Data Records for Their Better Management."

Journal of Digital Information 13, no. 1 (2012).
<http://journals.tdl.org/jodi/index.php/jodi/article/view/5917>

Ball, Alexander, Kellie Snow, Peter Obee, and Greg Simpson. "Choose Your Own Research Data Management Guidance." *International Journal of Digital Curation* 12, no. 1 (2017): 13-21. <https://doi.org/10.2218/ijdc.v12i1.494>

The GW4 Research Data Services Group has developed a Research Data Management Triage Tool to help researchers find answers quickly to the more common research data queries, and direct them to appropriate guidance and sources of advice for more complex queries. The tool takes the form of an interactive web page that asks users questions and updates itself in response. The conversational and dynamic way the tool progresses is similar to the behaviour of text adventures, which are a genre of interactive fiction; this is one of the oldest forms of computer game and was also popular in print form in, for example, the Choose Your Own Adventure and Fighting Fantasy series of books. In fact, the tool was written using interactive fiction software. It was tested with staff and students at the four UK universities within the GW4 collaboration.

Ball, Joanna. "Research Data Management for Libraries: Getting Started." *Insights: The UKSG Journal* 26, no. 3 (2013): 256-260. <https://doi.org/10.1629/2048-7754.70>

Many libraries are keen to take on new roles in providing support for effective research data management (RDM), but lack the necessary skills and resources to do so. This article explores the approach used by the University of Sussex to engage with academic departments about their RDM practices and requirements in order to develop relevant library support services. It describes a project undertaken with three Academic Schools to inform a list of recommendations for senior management, to include areas which should be taken forward by the Library, IT and Research Office in order to create a sustainable RDM service. The article is unflinchingly honest in sharing the differing reactions to the project and the lessons learnt along the way.

Bangani, Siviwe, and Mathew Moyo. "Data Sharing Practices among Researchers at South African Universities." *Data Science Journal*, 18, no. 1 (2019): 28.
<http://doi.org/10.5334/dsj-2019-028>.

Research data management practices have gained momentum the world over. This is due to increased demands by governments and other funding agencies

to have research data archived and shared as widely as possible. This paper sought to establish the data sharing practices of researchers in South Africa. The study further sought to establish the level of collaboration among researchers in sharing research data at the university level. The outcomes of the survey will help the researchers to develop appropriate data literacy awareness programmes meant to stimulate growth in data sharing practices for the benefit of research, not only in South Africa, but the world at large. A survey research method was used to gather data from willing public universities in South Africa. A similar study was conducted in other countries such as the United Kingdom, France and Turkey but the Researchers believe that circumstances in the developed world may differ with the South African research environment, hence the current study. The major finding of this study was that most researchers preferred to use data produced by others but less keen on sharing their own data. This study is the first of its kind in South Africa which investigates data sharing practices of researchers from multi-disciplinary fields at the university level and will contribute immensely to the growing body of literature in the area of research data management.

Barateiro, José, Gonçalo Antunes, Manuel Cabral, José Borbinha, and Rodrigo Rodrigues. "Digital Preservation of Scientific Data." *Lecture Notes in Computer Science* 5173 (2008): 388-391. https://doi.org/10.1007/978-3-540-87599-4_41

———. "Using a Grid for Digital Preservation." *Lecture Notes in Computer Science* 5362 (2008): 225-235. https://doi.org/10.1007/978-3-540-89533-6_23

Barbrow, Sarah, Denise Brush, and Julie Goldman. "Research Data Management and Services: Resources for Novice Data Librarians." *College & Research Libraries News* 78, no. 5 (2017): 274-278. <https://doi.org/10.5860/crln.78.5.274>

Bardi, Alessia, and Paolo Manghi. "Enhanced Publications: Data Models and Information Systems." *LIBER Quarterly* 23, no. 4 (2014): 240-273. <http://doi.org/10.18352/lq.8445>

"Enhanced publications" are commonly intended as digital publications that consist of a mandatory narrative part (the description of the research conducted) plus related "parts", such as datasets, other publications, images, tables, workflows, devices. The state-of-the-art on information systems for enhanced publications has today reached the point where some kind of common understanding is required, in order to provide the methodology and language for scientists to compare, analyse, or simply discuss the multitude of

solutions in the field. In this paper, we thoroughly examined the literature with a two-fold aim: firstly, introducing the terminology required to describe and compare structural and semantic features of existing enhanced publication data models; secondly, proposing a classification of enhanced publication information systems based on their main functional goals.

Bardyn, Tania P., Taryn Resnick, and Susan K. Camina. "Translational Researchers' Perceptions of Data Management Practices and Data Curation Needs: Findings from a Focus Group in an Academic Health Sciences Library." *Journal of Web Librarianship* 6, no. 4 (2012): 274-287.

<https://doi.org/10.1080/19322909.2012.730375>

Barton, Amy, Paul J. Bracke, and Ann Marie Clark. "Digitization, Data Curation, and Human Rights Documents: Case Study of a Library-Researcher-Practitioner Collaboration." *IASSIST Quarterly* 40, no.1 (2016): 27.

<https://doi.org/10.29173/iq625>

Baru, Chaitanya. "Sharing and Caring of eScience Data." *International Journal on Digital Libraries* 7, no. 1/2 (2007): 113-116. <https://doi.org/10.1007/s00799-007-0029-2>

Battista, Andrew, Karen Majewicz, Stephen Balogh, and Darren Hardy. "Consortial Geospatial Data Collection: Toward Standards and Processes for Shared GeoBlacklight Metadata." *Journal of Library Metadata* 17, no. 3-4 (2017): 183-200. <https://doi.org/10.1080/19386389.2018.1443414>

Baum, Benjamin, R. Bauer Christian, Thomas Franke, Harald Kusch, Marcel Parciak, Thorsten Rottmann, Nadine Umbach, and Ulrich Sax. "Opinion Paper: Data Provenance Challenges in Biomedical Research." *it—Information Technology* 59, no. 4 (2017): 191-196. <https://doi.org/10.1515/itit-2016-0031>

Baykoucheva, Svetla. *Managing Scientific Information and Research Data*. Elsevier: Waltham, MA, 2015. <http://www.worldcat.org/oclc/986633735>

Beagrie, Neil, Robert Beagrie, and Ian Rowlands. "Research Data Preservation and Access: The Views of Researchers." *Ariadne*, no. 60 (2009).

<http://www.ariadne.ac.uk/issue60/beagrie-et-al>

Beaujardière, Jeff De La. "NOAA Environmental Data Management." *Journal of Map & Geography Libraries* 12, no. 1 (2016): 5-27.

<https://doi.org/10.1080/15420353.2015.1087446>

Beckett, Mark G., Chris R. Allton, Christine T. H. Davies, Ian Davis, Jonathan M. Flynn, Eilidh J. Grant, Russell S. Hamilton, Alan C. Irving, R. D. Kenway, Radoslaw H. Ostrowski, James T. Perry, Jason R. Swedlow, and Arthur Trew. "Building a Scientific Data Grid with DiGS." *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 367, no. 1897 (2009): 2471-2481. <https://doi.org/10.1098/rsta.2009.0050>

Beckles, Zosia, Stephen Gray, Debra Hiom, Kirsty Merrett, Kellie Snow, and Damian Steer. "Disciplinary Data Publication Guides." *International Journal of Digital Curation* 13, no. 1 (2018): 150-160. <https://doi.org/10.2218/ijdc.v13i1.603>

Many academic disciplines have very comprehensive standard for data publication and clear guidance from funding bodies and academic publishers. In other cases, whilst much good-quality general guidance exists, there is a lack of information available to researchers to help them decide which specific data elements should be shared. This is a particular issue for disciplines with very varied data types, such as engineering, and presents an unnecessary barrier to researchers wishing to meet funder expectations on data sharing. This article outlines a project to provide simple, visual, discipline-specific guidance on data publication, undertaken at the University of Bristol at the request of the Faculty of Engineering.

Belter, Christopher W. "Measuring the Value of Research Data: A Citation Analysis of Oceanographic Data Sets." *PLoS ONE* 9, no. 3 (2014): e92590. <http://dx.doi.org/10.1371/journal.pone.0092590>

Evaluation of scientific research is becoming increasingly reliant on publication-based bibliometric indicators, which may result in the devaluation of other scientific activities—such as data curation—that do not necessarily result in the production of scientific publications. This issue may undermine the movement to openly share and cite data sets in scientific publications because researchers are unlikely to devote the effort necessary to curate their research data if they are unlikely to receive credit for doing so. This analysis attempts to demonstrate the bibliometric impact of properly curated and openly accessible data sets by attempting to generate citation counts for three data sets archived at the National Oceanographic Data Center. My findings suggest that all three data sets are highly cited, with estimated citation counts in most cases higher than 99% of all the journal articles published in *Oceanography* during the same years. I also find that methods of citing and referring to these data sets in scientific publications are highly inconsistent,

despite the fact that a formal citation format is suggested for each data set. These findings have important implications for developing a data citation format, encouraging researchers to properly curate their research data, and evaluating the bibliometric impact of individuals and institutions.

This work is licensed under a Creative Commons CC0 1.0 Universal (CC0 1.0) Public Domain Dedication,
<https://creativecommons.org/publicdomain/zero/1.0/>.

Bender, Stefam, and Jorg Heining. "The Research-Data-Centre in Research-Data-Centre Approach: A First Step towards Decentralised International Data Sharing." *IASSIST Quarterly* 35, no. 3 (2011): 10-16. <https://doi.org/10.29173/iq119>

Berman, Elizabeth A. "An Exploratory Sequential Mixed Methods Approach to Understanding Researchers' Data Management Practices at UVM: Integrated Findings to Develop Research Data Services." *Journal of eScience Librarianship* 5, no. 1 (2017): e1104. <https://doi.org/10.7191/jeslib.2017.1104>

Berman, Francine. "Got Data? A Guide to Data Preservation in the Information Age." *Communications of the ACM* 51, no. 12 (2008): 50-56.
<https://doi.org/10.1145/1409360.1409376>

Bethune, Alec, Butch Lazorchak, and Zsolt Nagy. "GeoMAPP: A Geospatial Multistate Archive and Preservation Partnership." *Journal of Map & Geography Libraries* 6, no. 1 (2009): 45-56. <https://doi.org/10.1080/15420350903432630>

Bhardwaj, Raj Kumar, and Paul Banks, eds. *Research Data Access and Management in Modern Libraries*. Hershey PA: IGI Global, 2019.
<https://melvyl.on.worldcat.org/oclc/1109802207>

Bird, Colin, Simon Coles, Iris Garrelfs, Tom Griffin, Magnus Hagdorn, Graham Klyne, Mike Mineter, and Cerys Willoughby. "Using Metadata Actively." *International Journal of Digital Curation* 11, no. 1 (2016): 76-85.
<https://doi.org/10.2218/ijdc.v11i1.412>

Almost all researchers collect and preserve metadata, although doing so is often seen as a burden. However, when that metadata can be, and is, used actively during an investigation or creative process, the benefits become apparent instantly. Active use can arise in various ways, several of which are being investigated by the Collaboration for Research Enhancement by Active use of Metadata (CREAM) project, which was funded by Jisc as part of their

Research Data Spring initiative. The CREAM project is exploring the concept through understanding the active use of metadata by the partners in the collaboration. This paper explains what it means to use metadata actively and describes how the CREAM project characterises active use by developing use cases that involve documenting the key decision points during a process. Well-documented processes are accordingly more transparent, reproducible, and reusable.

Bird, Colin L., Cerys Willoughby, Simon J. Coles, and Jeremy G. Frey. "Data Curation Issues in the Chemical Sciences." *Information Standards Quarterly* 25, no. 3 (2013): 4-12. <https://www.niso.org/niso-io/2013/09/data-curation-issues-chemical-sciences>

Bishoff, Carolyn, and Lisa Johnston. "Approaches to Data Sharing: An Analysis of NSF Data Management Plans from a Large Research University." *Journal of Librarianship and Scholarly Communication* 3, no. 2 (2015): eP1231. <http://doi.org/10.7710/2162-3309.1231>

INTRODUCTION Sharing digital research data is increasingly common, propelled by funding requirements, journal publishers, local campus policies, or community-driven expectations of more collaborative and interdisciplinary research environments. However, it is not well understood how researchers are addressing these expectations and whether they are transitioning from individualized practices to more thoughtful and potentially public approaches to data sharing that will enable reuse of their data. **METHODS** The University of Minnesota Libraries conducted a local opt-in study of data management plans (DMPs) included in funded National Science Foundation (NSF) grant proposals from January 2011 through June 2014. In order to understand the current data management and sharing practices of campus researchers, we solicited, coded, and analyzed 182 DMPs, accounting for 41% of the total number of plans available. **RESULTS** DMPs from seven colleges and academic units were included. The College of Science of Engineering accounted for 70% of the plans in our review. While 96% of DMPs mentioned data sharing, we found a variety of approaches for how PIs shared their data, where data was shared, the intended audiences for sharing, and practices for ensuring long-term reuse. **CONCLUSION** DMPs are useful tools to investigate researchers' current plans and philosophies for how research outputs might be shared. Plans and strategies for data sharing are inconsistent across this sample, and researchers need to better understand what kind of sharing constitutes public access. More intervention is needed to ensure that

researchers implement the sharing provisions in their plans to the fullest extent possible. These findings will help academic libraries develop practical, targeted data services for researchers that aim to increase the impact of institutional research.

Bishop, Bradley Wade, Tony H. Grubestic, and Sonya Prasertong. "Digital Curation and the GeoWeb: An Emerging Role for Geographic Information Librarians." *Journal of Map & Geography Libraries: Advances in Geospatial Information, Collections & Archives* 9, no. 3 (2013): 296-312.
<https://doi.org/10.1080/15420353.2013.817367>

Bishop, Bradley Wade, and Carolyn Hank. "Measuring FAIR Principles to Inform Fitness for Use." *International Journal of Digital Curation* 13, no. 1 (2018): 35-46.
<https://doi.org/10.2218/ijdc.v13i1.630>

For open science to flourish, data and any related digital outputs should be discoverable and re-usable by a variety of potential consumers. The recent FAIR Data Principles produced by the Future of Research Communication and e-Scholarship (FORCE11) collective provide a compilation of considerations for making data findable, accessible, interoperable, and re-usable. The principles serve as guideposts to 'good' data management and stewardship for data and/or metadata. On a conceptual level, the principles codify best practices that managers and stewards would find agreement with, exist in other data quality metrics, and already implement. This paper reports on a secondary purpose of the principles: to inform assessment of data's FAIR-ness or, put another way, data's fitness for use. Assessment of FAIR-ness likely requires more stratification across data types and among various consumer communities, as how data are found, accessed, interoperated, and re-used differs depending on types and purposes. This paper's purpose is to present a method for qualitatively measuring the FAIR Data Principles through operationalizing findability, accessibility, interoperability, and re-usability from a re-user's perspective. The findings may inform assessments that could also be used to develop situationally-relevant fitness for use frameworks.

Bishop, Libby, and Arja Kuula-Luumi. "Revisiting Qualitative Data Reuse." *SAGE Open* 7, no. 1 (2017): 2158244016685136.
<https://doi.org/10.1177/2158244016685136>

Secondary analysis of qualitative data entails reusing data created from previous research projects for new purposes. Reuse provides an opportunity to study the raw materials of past research projects to gain methodological and substantive insights. In the past decade, use of the approach has grown rapidly in the United Kingdom to become sufficiently accepted that it must now be regarded as mainstream. Several factors explain this growth: the open data movement, research funders' and publishers' policies supporting data sharing, and researchers seeing benefits from sharing resources, including data. Another factor enabling qualitative data reuse has been improved services and infrastructure that facilitate access to thousands of data collections. The UK Data Service is an example of a well-established facility; more recent has been the proliferation of repositories being established within universities. This article will provide evidence of the growth of data reuse in the United Kingdom and in Finland by presenting both data and case studies of reuse that illustrate the breadth and diversity of this maturing research method. We use two distinct data sources that quantify the scale, types, and trends of reuse of qualitative data: (a) downloads of archived data collections held at data repositories and (b) publication citations. Although the focus of this article is on the United Kingdom, some discussion of the international environment is provided, together with data and examples of reuse at the Finnish Social Science Data Archive. The conclusion summarizes the major findings, including some conjectures regarding what makes qualitative data attractive for reuse and sharing.

This work is licensed under a Creative Commons Attribution 3.0 Unported License, <https://creativecommons.org/licenses/by/3.0/>.

Bonifacio, Flavio. "Differences in Data Sharing Attitudes and Behaviours." *IASSIST Quarterly* 42 No 3 (2018). <https://doi.org/10.29173/iq912>

Borgman, Christine L. "The Conundrum of Sharing Research Data." *Journal of the American Society for Information Science and Technology* 63, no. 6 (2012): 1059-1078. <https://doi.org/10.1002/asi.22634>

———. "The Lives and After Lives of Data." *Harvard Data Science Review* 1, no. 1 (2019). <https://doi.org/10.1162/99608f92.9a36bdb6>

The most elusive term in data science is 'data.' While often treated as objects to be computed upon, data is a theory-laden concept with a long history. Data exist within knowledge infrastructures that govern how they are created,

managed, and interpreted. By comparing models of data life cycles, implicit assumptions about data become apparent. In linear models, data pass through stages from beginning to end of life, which suggest that data can be recreated as needed. Cyclical models, in which data flow in a virtuous circle of uses and reuses, are better suited for irreplaceable observational data that may retain value indefinitely. In astronomy, for example, observations from one generation of telescopes may become calibration and modeling data for the next generation, whether digital sky surveys or glass plates. The value and reusability of data can be enhanced through investments in knowledge infrastructures, especially digital curation and preservation. Determining what data to keep, why, how, and for how long, is the challenge of our day.

Borgman, Christine L., Milena S. Golshan, Ashley E. Sands, Jillian C. Wallis, Rebekah L. Cummings, Peter T. Darch, and Bernadette M. Randles. "Data Management in the Long Tail: Science, Software, and Service." *International Journal of Digital Curation* 11, no. 1 (2016): 128-149.
<http://dx.doi.org/10.2218/ijdc.v11i1.428>

Scientists in all fields face challenges in managing and sustaining access to their research data. The larger and longer term the research project, the more likely that scientists are to have resources and dedicated staff to manage their technology and data, leaving those scientists whose work is based on smaller and shorter term projects at a disadvantage. The volume and variety of data to be managed varies by many factors, only two of which are the number of collaborators and length of the project. As part of an NSF project to conceptualize the Institute for Empowering Long Tail Research, we explored opportunities offered by Software as a Service (SaaS). These cloud-based services are popular in business because they reduce costs and labor for technology management, and are gaining ground in scientific environments for similar reasons. We studied three settings where scientists conduct research in small and medium-sized laboratories. Two were NSF Science and Technology Centers (CENS and C-DEBI) and the third was a workshop of natural reserve scientists and managers. These laboratories have highly diverse data and practices, make minimal use of standards for data or metadata, and lack resources for data management or sustaining access to their data, despite recognizing the need. We found that SaaS could address technical needs for basic document creation, analysis, and storage, but did not support the diverse and rapidly changing needs for sophisticated domain-specific tools and services. These are much more challenging knowledge

infrastructure requirements that require long-term investments by multiple stakeholders.

Borgman, Christine L., Andrea Scharnhorst, and Milena S. Golshan. "Digital Data Archives as Knowledge Infrastructures: Mediating Data Sharing and Reuse." *Journal of the Association for Information Science and Technology* 70, no. 8 (2019): 888-904. <https://doi.org/10.1002/asi.24172>

Borgman, Christine L., Jillian C. Wallis, and Noel Enyedy. "Little Science Confronts the Data Deluge: Habitat Ecology, Embedded Sensor Networks, and Digital Libraries." *International Journal on Digital Libraries* 7, no. 1 (2007): 17-30. <https://doi.org/10.1007/s00799-007-0022-9>

Borgman, Christine L., Jillian C. Wallis, and Matthew S. Mayernik. "Who's Got the Data? Interdependencies in Science and Technology Collaborations." *Computer Supported Cooperative Work* 21, no. 6 (2012): 485-523. <https://doi.org/10.1007/s10606-012-9169-z>

Bracke, Marianne Stowell. "Emerging Data Curation Roles for Librarians: A Case Study of Agricultural Data." *Journal of Agricultural & Food Information* 12, no. 1 (2011): 65-74. <https://doi.org/10.1080/10496505.2011.539158>

Brandt, D. Scott, and Eugenia Kim. "Data Curation Profiles as a Means to Explore Managing, Sharing, Disseminating or Preserving Digital Outcomes." *International Journal of Performance Arts and Digital Media* 10, no. 1 (2014): 21-34. <https://doi.org/10.1080/14794713.2014.912498>

Bresnahan, Megan M., and Andrew M. Johnson. "Assessing Scholarly Communication and Research Data Training Needs." *Reference Services Review* 41, no. 3 (2013): 413-433. <https://doi.org/10.1108/rsr-01-2013-0003>

Brewerton, Gary. "Research Data Management: A Case Study." *Ariadne*, no. 74 (2015). <http://www.ariadne.ac.uk/issue74/brewerton>

Briney, Kristin. *Data Management for Researchers: Organize, Maintain and Share Your Data for Research Success* Pelagic Publishing, 2015. <http://www.worldcat.org/oclc/927940305>

———. "The Problem with Dates: Applying ISO 8601 to Research Data Management." *Journal of eScience Librarianship* 7, no. 2 (2018): e1147. <https://doi.org/10.7191/jeslib.2018.1147>

Dates appear regularly in research data and metadata but are a problematic data type to normalize due to a variety of potential formats. This suggests an opportunity for data librarians to assist with formatting dates, yet there are frequent examples of data librarians using diverse strategies for this purpose. Instead, data librarians should adopt the international date standard ISO 8601. This standard provides needed consistency in date formatting, allows for inclusion of several types of date-time information, and can sort dates chronologically. As regular advocates for standardization in research data, data librarians must adopt ISO 8601 and push for its use as a data management best practice.

Briney, Kristin, Abigail Goben, and Lisa Zilinski. "Do You Have an Institutional Data Policy? A Review of the Current Landscape of Library Data Services and Institutional Data Policies." *Journal of Librarianship and Scholarly Communication* 3, no. 2 (2015): eP1232. <http://doi.org/10.7710/2162-3309.1232>

INTRODUCTION Many research institutions have developed research data services in their libraries, often in anticipation of or in response to funder policy. However, policies at the institution level are either not well known or nonexistent. **METHODS** This study reviewed library data services efforts and institutional data policies of 206 American universities, drawn from the July 2014 Carnegie list of universities with "Very High" or "High" research activity designation. Twenty-four different characteristics relating to university type, library data services, policy type, and policy contents were examined. **RESULTS** The study has uncovered findings surrounding library data services, institutional data policies, and content within the policies. **DISCUSSION** Overall, there is a general trend toward the development and implementation of data services within the university libraries. Interestingly, just under half of the universities examined had a policy of some sort that either specified or mentioned research data. Many of these were standalone data policies, while others were intellectual property policies that included research data. When data policies were discoverable, not behind a log in, they focused on the definition of research data, data ownership, data retention, and terms surrounding the separation of a researcher from the institution. **CONCLUSION** By becoming well versed on research data policies, librarians can provide support for researchers by navigating the policies at their institutions, facilitating the activities needed to comply with the requirements of research funders and publishers. This puts academic libraries in a unique position to provide insight and guidance in the development and revisions of institutional data policies.

Brochu, Lauren, and Jane Burns. "Librarians and Research Data Management—A Literature Review: Commentary from a Senior Professional and a New Professional Librarian." *New Review of Academic Librarianship* 25, no. 1 (2019): 49-58. <https://doi.org/10.1080/13614533.2018.1501715>

Broeder, Daan, and Laurence Lannom. "Data Type Registries: A Research Data Alliance Working Group." *D-Lib Magazine* 20, no. 1/2 (2014). <http://www.dlib.org/dlib/january14/broeder/01broeder.html>

Brown, Rebecca A., Malcolm Wolski, and Joanna Richardson. "Developing New Skills for Research Support Librarians." *The Australian Library Journal* 64, no. 3 (2015): 224-234. <http://dx.doi.org/10.1080/00049670.2015.1041215>

Brownlee, Rowan. "Research Data and Repository Metadata: Policy and Technical Issues at the University of Sydney Library." *Cataloging & Classification Quarterly* 47, no. 3/4 (2009): 370-379. <https://doi.org/10.1080/01639370802714182>

Burgess, Lucie, Neil Jefferies, Sally Rumsey, John Southall, David Tomkins, and James A. J. Wilson. "From Compliance to Curation: ORA-Data at the University of Oxford." *Alexandria* 26, no. 2 (2016): 107-135. <https://doi.org/10.1177/0955749016657482>

Burgi, Pierre-Yves, Eliane Blumer, and Basma Makhlouf-Shabou. "Research Data Management in Switzerland." *IFLA Journal* 43, no. 1 (2017): 5-21. <https://doi.org/10.1177/0340035216678238>

Burnette, Margaret H., Sarah C. Williams, and Heidi J. Imker. "From Plan to Action: Successful Data Management Plan Implementation in a Multidisciplinary Project." *Journal of eScience Librarianship* 5, no. 1 (2016): e1101. <https://doi.org/10.7191/jeslib.2016.1101>

Burton, A., D. Groenewegen, C. Love, A. Treloar, and R. Wilkinson. "Making Research Data Available in Australia." *Intelligent Systems* 27, no. 3 (2012): 40-43. <http://doi.ieeecomputersociety.org/10.1109/MIS.2012.57>

Burton, Adrian, Koers Hylke, Manghi Paolo, Bruzzo Sandro La, Aryani Amir, Diepenbroek Michael, and Schindler Uwe. "The Data-Literature Interlinking Service: Towards a Common Infrastructure for Sharing Data-Article Links." *Program* 51, no. 1 (2017): 75-100. <https://doi.org/10.1108/PROG-06-2016-0048>

Burton, Adrian, and Andrew Treloar. "Designing for Discovery and Re-use: The 'ANDS Data Sharing Verbs' Approach to Service Decomposition." *International Journal of Digital Curation* 4, no. 3 (2009): 44-56.

<https://doi.org/10.2218/ijdc.v4i3.124>

Australian National Data Services (ANDS) is designing systems to support data sharing and Re-use. The paper commences with an overview of the setting for ANDS, before introducing ANDS itself. The paper then structures its discussion of ANDS services for Re-use in terms of the ANDS Data Sharing Verbs: Create, Store, Describe, Identify, Register, Discover, Access and Exploit. For each of the data verbs, a rationale for its importance is provided together with a description of how it is being implemented by ANDS. The paper concludes by arguing for the data verbs approach as a useful way to design and structure flexible services in a heterogeneous environment.

Buys, Cunera M., and Pamela L. Shaw. "Data Management Practices Across an Institution: Survey and Report." *Journal of Librarianship and Scholarly Communication* 3, no. 2 (2015): eP1225. <http://doi.org/10.7710/2162-3309.1225>

INTRODUCTION Data management is becoming increasingly important to researchers in all fields. The E-Science Working Group designed a survey to investigate how researchers at Northwestern University currently manage data and to help determine their future needs regarding data management.

METHODS A 21-question survey was distributed to approximately 12,940 faculty, graduate students, postdoctoral candidates, and selected research-affiliated staff at Northwestern's Evanston and Chicago Campuses. Survey questions solicited information regarding types and size of data, current and future needs for data storage, data retention and data sharing, what researchers are doing (or not doing) regarding data management planning, and types of training or assistance needed. There were 831 responses and 788 respondents completed the survey, for a response rate of approximately 6.4%. **RESULTS** Survey results indicate investigators need both short and long term storage and preservation solutions. However, 31% of respondents did not know how much storage they will require. This means that establishing a correctly sized research storage service will be difficult. Additionally, research data is stored on local hard drives, departmental servers or equipment hard drives. These types of storage solutions limit data sharing and long term preservation. Data sharing tends to occur within a research group or with collaborators prior to publication, expanding to more public availability after publication. Survey

responses also indicate a need to provide increased consulting and support services, most notably for data management planning, awareness of regulatory requirements, and use of research software.

Callaghan, Sarah. "Preserving the Integrity of the Scientific Record: Data Citation and Linking." *Learned Publishing* 27, no. 5 (2014): 15-24.
<https://doi.org/10.1087/20140504>

Callaghan, Sarah, Steve Donegan, Sam Pepler, Mark Thorley, Nathan Cunningham, Peter Kirsch, Linda Ault, Patrick Bell, Rod Bowie, Adam Leadbetter, Roy Lowry, Gwen Moncoiffé, Kate Harrison, Ben Smith-Haddon, Anita Weatherby, and Dan Wright. "Making Data a First Class Scientific Output: Data Citation and Publication by NERC's Environmental Data Centres." *International Journal of Digital Curation* 7, no. 1 (2012): 107-113.
<https://doi.org/10.2218/ijdc.v7i1.218>

The NERC Science Information Strategy Data Citation and Publication project aims to develop and formalise a method for formally citing and publishing the datasets stored in its environmental data centres. It is believed that this will act as an incentive for scientists, who often invest a great deal of effort in creating datasets, to submit their data to a suitable data repository where it can properly be archived and curated. Data citation and publication will also provide a mechanism for data producers to receive credit for their work, thereby encouraging them to share their data more freely.

Callaghan, Sarah, Jonathan Tedds, John Kunze, Varsha Khodiyar, Rebecca Lawrence, Matthew S. Mayernik, Fiona Murphy, Timothy Roberts, and Angus Whyte. "Guidelines on Recommending Data Repositories as Partners in Publishing Research Data." *International Journal of Digital Curation* 9, no. 1 (2014): 152-163. <https://doi.org/10.2218/ijdc.v9i1.309>

This document summarises guidelines produced by the UK Jisc-funded PREPARDE data publication project on the key issues of repository accreditation. It aims to lay out the principles and the requirements for data repositories intent on providing a dataset as part of the research record and as part of a research publication. The data publication requirements that repository accreditation may support are rapidly changing, hence this paper is intended as a provocation for further discussion and development in the future.

Candela, Leonardo, Donatella Castelli, Paolo Manghi, and Sarah Callaghan. "On Research Data Publishing." *International Journal on Digital Libraries* 18, no. 2 (2017): 73-75. <https://doi.org/10.1007/s00799-017-0213-y>

Candela, Leonardo, Donatella Castelli, Paolo Manghi, and Alice Tani. "Data Journals: A Survey." *Journal of the Association for Information Science and Technology* 66, no. 9 (2015): 1747-1762. <http://dx.doi.org/10.1002/asi.23358>

Capó-Lugo, Carmen E., Abel N. Kho, Linda C. O'Dwyer, and Marc B. Rosenman. "Data Sharing and Data Registries in Physical Medicine and Rehabilitation." *PM&R* 9, no. 5 (2017): S59-S74. <http://dx.doi.org/10.1016/j.pmrj.2017.04.003>

Carlhed, Carina, and Iris Alfredsson. "Swedish National Data Service's Strategy for Sharing and Mediating Data." *IASSIST Quarterly* 32, no. 1-4 (2010). 30. <https://doi.org/10.29173/iq654>

Carlson, Jake. "Demystifying the Data Interview: Developing a Foundation for Reference Librarians to Talk with Researchers about Their Data." *Reference Services Review* 40, no. 1 (2012): 7-23. <https://doi.org/10.1108/00907321211203603>

——— "Opportunities and Barriers for Librarians in Exploring Data: Observations from the Data Curation Profile Workshops." *Journal of eScience Librarianship* 2, no. 2 (2013): 17-33. <http://dx.doi.org/10.7191/jeslib.2013.1042>

Carlson, Jake, Megan Sapp Nelson, Lisa R. Johnston, and Amy Koshoffer. "Developing Data Literacy Programs: Working with Faculty, Graduate Students and Undergraduates." *Bulletin of the Association for Information Science and Technology* 41, no. 6 (2015): 14-17. <https://doi.org/10.1002/bult.2015.1720410608>

Carlson, Jake, and Marianne Stowell-Bracke. "Data Management and Sharing from the Perspective of Graduate Students: An Examination of the Culture and Practice at the Water Quality Field Station." *portal: Libraries and the Academy* 13, no. 4 (2013): 343-361. <https://doi.org/10.1353/pla.2013.0034>

Carroll, Michael W. "Sharing Research Data and Intellectual Property Law: A Primer." *PLOS Biology* 13, no. 8 (2015): e1002235. <https://doi.org/10.1371/journal.pbio.1002235>

Sharing research data by depositing it in connection with a published article or otherwise making data publicly available sometimes raises intellectual

property questions in the minds of depositing researchers, their employers, their funders, and other researchers who seek to reuse research data. In this context or in the drafting of data management plans, common questions are (1) what are the legal rights in data; (2) who has these rights; and (3) how does one with these rights use them to share data in a way that permits or encourages productive downstream uses? Leaving to the side privacy and national security laws that regulate sharing certain types of data, this Perspective explains how to work through the general intellectual property and contractual issues for all research data.

Castle, Clair. "Getting the Central RDM Message Across: A Case Study of Central versus Discipline-Specific Research Data Services (RDS) at the University of Cambridge." *Libri* 69, no. 2 (2019): 105-116. <https://doi.org/10.1515/libri-2018-0064>

Castro, Eleni, Mercè Crosas, Alex Garnett, Kasey Sheridan, and Micah Altman. "Evaluating and Promoting Open Data Practices in Open Access Journals." *Journal of Scholarly Publishing* 49, no. 1 (2017): 66-88. <https://doi.org/10.3138/jsp.49.1.66>

Castro, Eleni, and Alex Garnett. "Building a Bridge Between Journal Articles and Research Data: The PKP-Dataverse Integration Project." *International Journal of Digital Curation* 9, no. 1 (2014): 176-184. <https://doi.org/10.2218/ijdc.v9i1.311>

A growing number of funding agencies and international scholarly organizations are requesting that research data be made more openly available to help validate and advance scientific research. Thus, this is an opportune moment for research data repositories to partner with journal editors and publishers in order to simplify and improve data curation and publishing practices. One practical example of this type of cooperation is currently being facilitated by a two year (2012-2014) one million dollar Sloan Foundation grant, integrating two well-established open source systems: the Public Knowledge Project's (PKP) Open Journal Systems (OJS), developed by Stanford University and Simon Fraser University; and Harvard University's Dataverse Network web application, developed by the Institute for Quantitative Social Science (IQSS). To help make this interoperability possible, an OJS Dataverse plugin and Data Deposit API are being developed, which together will allow authors to submit their articles and datasets through an existing journal management interface, while the underlying data are seamlessly deposited into a research data repository, such as the Harvard

Dataverse. This practice paper will provide an overview of the project, and a brief exploration of some of the specific challenges to and advantages of this integration.

Chad, Ken, and Suzanne Enright. "The Research Cycle and Research Data Management (RDM): Innovating Approaches at the University of Westminster." *Insights: The UKSG Journal* 27, no. 2 (2014): 147-153.
<https://doi.org/10.1629/2048-7754.152>

This article presents a case study based on experience of delivering a more joined-up approach to supporting institutional research activity and processes, research data management (RDM) and open access (OA). The result of this small study, undertaken at the University of Westminster in 2013, indicates that a more holistic approach should be adopted, embedding RDM more fully into the wider research management landscape and taking researchers' priorities into consideration. Rapid development of an innovative pilot system followed closely on from a positive engagement with researchers, and today a purpose-built, integrated and fully working set of tools are functioning within the virtual research environment (VRE). This provides a coherent 'thread' to support researchers, doctoral students and professional support staff throughout the research cycle. The article describes the work entailed in more detail, together with the impact achieved so far and what future work is planned.

Chao, Tiffany C., Melissa H. Cragin, and Carole L. Palmer. "Data Practices and Curation Vocabulary (DPCVocab): An Empirically Derived Framework of Scientific Data Practices and Curatorial Processes." *Journal of the Association for Information Science and Technology* 66, no. 3 (2015): 616-633.
<https://doi.org/10.1002/asi.23184>

Chapple, Michael J. "Speaking the Same Language: Building a Data Governance Program for Institutional Impact." *EDUCAUSE Review* 48, no. 6 (2013): 14-27.
<https://er.educause.edu/articles/2013/12/speaking-the-same-language-building-a-data-governance-program-for-institutional-impact>

Charbonneau, Deborah H. "Strategies for Data Management Engagement." *Medical Reference Services Quarterly* 32, no. 3 (2013): 365-374.
<https://doi.org/10.1080/02763869.2013.807089>

Charbonneau, Deborah H., and Joan E. Beaudoin. "State of Data Guidance in Journal Policies: A Case Study in Oncology." *International Journal of Digital Curation* 10, no. 2 (2015): 136-156. <https://doi.org/10.2218/ijdc.v10i2.375>

This article reports the results of a study examining the state of data guidance provided to authors by 50 oncology journals. The purpose of the study was the identification of data practices addressed in the journals' policies. While a number of studies have examined data sharing practices among researchers, little is known about how journals address data sharing. Thus, what was discovered through this study has practical implications for journal publishers, editors, and researchers. The findings indicate that journal publishers should provide more meaningful and comprehensive data guidance to prospective authors. More specifically, journal policies requiring data sharing, should direct researchers to relevant data repositories, and offer better metadata consultation to strengthen existing journal policies. By providing adequate guidance for authors, and helping investigators to meet data sharing mandates, scholarly journal publishers can play a vital role in advancing access to research data.

Chawinga, Winner Dominic, and Sandy Zinn. "Global Perspectives of Research Data Sharing: A Systematic Literature Review." *Library & Information Science Research* 41, no. 2 (2019): 109-122. <https://doi.org/10.1016/j.lisr.2019.04.004>

Chen, Xiujuan, and Ming Wu. "Survey on the Needs for Chemistry Research Data Management and Sharing." *The Journal of Academic Librarianship* 43, no. 4 (2017): 346-353. <https://doi.org/10.1016/j.acalib.2017.06.006>

Chervenaka, Ann, Ian Foster, Carl Kesselman, Charles Salisbury, and Steven Tuecke. "The Data Grid: Towards an Architecture for the Distributed Management and Analysis of Large Scientific Datasets." *Journal of Network and Computer Applications* 23, no. 3 (2000): 187-200. <https://doi.org/10.1006/jnca.2000.0110>

Childs, Sue, Julie McLeod, Elizabeth Lomas, and Glenda Cook. "Opening Research Data: Issues and Opportunities." *Records Management Journal* 24, no. 2 (2014): 14-162. <http://dx.doi.org/10.1108/RMJ-01-2014-0005>

Chiwere, Elisha R.T., and Zanele Mathe. "Academic Libraries' Role in Research Data Management Services: A South African Perspective." *South African Journal of Libraries and Information Science* 81, No 2 (2015). <http://sajlis.journals.ac.za/pub/article/view/1563>

Chou, Chiu-chuang Lu. "50 Years of Social Science Data Services: A Case Study from the University of Wisconsin-Madison." *International Journal of Librarianship* 2, no. 1 (2017): 42-52. <https://doi.org/10.23974/ijol.2017.vol2.1.23>

The Data and Information Services Center (DISC), formerly known as the Data and Program Library Services (DPLS) has provided learning, teaching and research support to students, staff and faculty in social sciences at the University of Wisconsin-Madison for 50 years. What changes have our organization, collections, and services experienced? How has DISC evolved with the advancement of technology? What role does DISC play in the current and future landscape of social science data services on our campus and beyond? This paper gives answers to these questions and recommends a few simple steps in adding social science data services in academic libraries.

Choudhury, G. Sayeed. "Case Study in Data Curation at Johns Hopkins University." *Library Trends* 57, no. 2 (2008): 211-220. <https://doi.org/10.1353/lib.0.0028>

——— "Data Curation: An Ecological Perspective." *College & Research Libraries News* 71, no. 4 (2010): 194-196. <https://doi.org/10.5860/crln.71.4.8354>

Choudhury, Sayeed, Tim DiLauro, Alex Szalay, Ethan Vishniac, Robert J. Hanisch, Julie Steffen, Robert Milkey, Teresa Ehling, and Ray Plante. "Digital Data Preservation for Scholarly Publications in Astronomy." *International Journal of Digital Curation* 2, no. 2 (2007): 20-30. <https://doi.org/10.2218/ijdc.v2i2.26>

Christen, Peter. "Data Linkage: The Big Picture." *Harvard Data Science Review* 1, no. 2 (2019): <https://doi.org/10.1162/99608f92.84deb5c4>

Claibourn, Michele P. "Bigger on the Inside: Building Research Data Services at the University of Virginia." *Insights: The UKSG journal* 28, no. 2 (2015): 100-106. <https://doi.org/10.1629/uksg.239>

Every story has a beginning, where the narrator chooses to start, though this is rarely the genesis. This story begins with the launch of the University of Virginia Library's new Research Data Services unit in October 2013. Born from the conjoining of a data management team and a data analysis team, Research Data Services expanded to encompass data discovery and acquisitions, research software support, and new expertise in the use of restricted data. Our purpose is to respond to the challenges created by the growing ubiquity and scale of data by helping researchers acquire, analyze,

manage, and archive these resources. We have made serious strides toward becoming 'the face of data services at U.Va.' This article tells a bit of our story so far, relays some early challenges and how we've responded to them, outlines several initial successes, and summarizes a few lessons going forward.

Clare, Connie, Maria Cruz, Elli Papadopoulou, James Savage, Marta Teperek, Yan Wang, Iza Witkowska, and Joanne Yeomans. *Engaging Researchers with Data Management: The Cookbook*. Open Book Publishers, 2019.
<https://doi.org/10.11647/OBP.0185>.

Clement, Ryan, Amy Blau, Parvaneh Abbaspour, and Eli Gandour-Rood. "Team-based Data Management Instruction at Small Liberal Arts Colleges." *IFLA Journal* 43, no. 1 (2017): 105-118. <https://doi.org/10.1177/0340035216678239>

Clements, Anna. "Research Information Meets Research Data Management. . . in the Library?" *Insights: The UKSG journal* 26, no. 3 (2013): 298-304.
<https://doi.org/10.1629/2048-7754.99>

Research data management (RDM) is a major priority for many institutions as they struggle to cope with the plethora of pronouncements including funder policies, a G8 statement, REF2020 consultations, all stressing the importance of open data in driving everything from global innovation through to more accountable governance; not to mention the more direct possibility that non-compliance could result in grant income drying up. So, at the coalface, how do we become part of this global movement?

In this article the author explains the approach being taken at the University of St Andrews, building on the research information management infrastructure (data, systems and people) that has evolved since 2006. Continuing to navigate through the rapidly evolving research policy and cultural landscape, they aim to establish services to support their research community as it moves to this 'open by default' requirement of funders and governments.

Coates, Heather L. "Building Data Services From the Ground Up: Strategies and Resources." *Journal of eScience Librarianship* 3, no. 1 (2014): e1063.
<http://dx.doi.org/10.7191/jeslib.2014.1063>

Colavizza, Giovanni, Iain Hrynaszkiewicz, Isla Staden, Kirstie Whitaker, and Barbara McGillivray. "The Citation Advantage of Linking Publications to

Research Data." *PLoS ONE* 15, no. 4 (2019): e0230416.
<https://doi.org/10.1371/journal.pone.0230416>

Efforts to make research results open and reproducible are increasingly reflected by journal policies encouraging or mandating authors to provide data availability statements. As a consequence of this, there has been a strong uptake of data availability statements in recent literature. Nevertheless, it is still unclear what proportion of these statements actually contain well-formed links to data, for example via a URL or permanent identifier, and if there is an added value in providing such links. We consider 531, 889 journal articles published by PLOS and BMC, develop an automatic system for labelling their data availability statements according to four categories based on their content and the type of data availability they display, and finally analyze the citation advantage of different statement categories via regression. We find that, following mandated publisher policies, data availability statements become very common. In 2018 93.7% of 21,793 PLOS articles and 88.2% of 31,956 BMC articles had data availability statements. Data availability statements containing a link to data in a repository—rather than being available on request or included as supporting information files—are a fraction of the total. In 2017 and 2018, 20.8% of PLOS publications and 12.2% of BMC publications provided DAS containing a link to data in a repository. We also find an association between articles that include statements that link to data in a repository and up to 25.36% ($\pm 1.07\%$) higher citation impact on average, using a citation prediction model. We discuss the potential implications of these results for authors (researchers) and journal publishers who make the effort of sharing their data in repositories. All our data and code are made available in order to reproduce and extend our results.

Cole, Gareth John. "Establishing a Research Data Management Service at Loughborough University." *International Journal of Digital Curation* 11, no. 1 (2016): 68-75. <https://doi.org/10.2218/ijdc.v11i1.407>

In common with most UK universities Loughborough University needed to be compliant with the EPSRC Data Expectations by May 2015. This paper explains the process the University went through to meet these expectations. The paper also demonstrates how University senior management took the opportunity to look beyond compliance with EPSRC requirements. Project staff were challenged to identify a solution which would help to increase the University's research visibility and reach. The solution to all of these challenges is an innovative and ground-breaking relationship between the

University and three external partners. Investment has also been made in professional services staff to help manage and oversee the service. This paper explores the ways in which each element of Loughborough's research data service helps to reduce the burden on researchers, how much of the infrastructure is invisible to the research community, and how the service is being embedded in existing infrastructure and workflows.

Collie, W. Aaron, and Michael Witt. "A Practice and Value Proposal for Doctoral Dissertation Data Curation." *International Journal of Digital Curation* 6, no. 2 (2011): 165-175. <https://doi.org/10.2218/ijdc.v6i2.194>

Collins, Ellen. "Use and Impact of UK Research Data Centres." *International Journal of Digital Curation* 6, no. 1 (2011): 20-31. <https://doi.org/10.2218/ijdc.v6i1.169>

Conrad, Anders Sparre, Rasmus Handberg, and Michael Svendsen. "Reuse for Research: Curating Astrophysical Datasets for Future Researchers." *International Journal of Digital Curation* 12, no. 2 (2017): 37-46. <https://doi.org/10.2218/ijdc.v12i2.516>

"Our data are going to be valuable for science for the next 50 years, so please make sure you preserve them and keep them accessible for active research for at least that period."

These were approximately the words used by the principal investigator of the Kepler Asteroseismic Science Consortium (KASC) when he presented our task to us. The data in question consists of data products produced by KASC researchers and working groups as part of their research, as well as underlying data imported from the NASA archives.

The overall requirements for 50 years of preservation while, at the same time, enabling reuse of the data for active research presented a number of specific challenges, closely intertwining data handling and data infrastructure with scientific issues. This paper reports our work to deliver the best possible solution, performed in close cooperation between the research team and library personnel.

Conway, Esther, David Giaretta, Simon Lambert, and Brian Matthews. "Curating Scientific Research Data for the Long Term: A Preservation Analysis Method in Context." *International Journal of Digital Curation* 6, no. 2 (2011): 38-52. <https://doi.org/10.2218/ijdc.v6i2.204>

Conway, Esther, Brian Matthews, David Giaretta, Simon Lambert, Michael Wilson, and Nick Draper. "Managing Risks in the Preservation of Research Data with Preservation Networks." *International Journal of Digital Curation* 7, no. 1 (2012): 3-15. <https://doi.org/10.2218/ijdc.v7i1.210>

Network modelling provides a framework for the systematic analysis of needs and options for preservation. A number of general strategies can be identified, characterised and applied to many situations; these strategies may be combined to produce robust preservation solutions tailored to the needs of the community and responsive to their environment. This paper provides an overview of this approach. We describe the components of a Preservation Network Model and go on to show how it may be used to plan preservation actions according to the requirements of the particular situation using illustrative examples from scientific archives.

Conway, Esther, Sam Pepler, Wendy Garland, David Hooper, Fulvio Marelli, Luca Liberti, Emanuela Piervitali, Katrin Molch, Helen Graves, and Lucio Badiali. "Ensuring the Long Term Impact of Earth Science Data through Data Curation and Preservation." *Information Standards Quarterly* 25, no. 3 (2013): 28-36. <https://www.niso.org/niso-io/2013/09/ensuring-long-term-impact-earth-science-data-through-data-curation-and-preservation>

Cope, Jez, and James Baker. "Library Carpentry: Software Skills Training for Library Professionals." *International Journal of Digital Curation* 12, no. 2 (2017): 266-273. <https://doi.org/10.2218/ijdc.v12i2.576>

Much time and energy is now being devoted to developing the skills of researchers in the related areas of data analysis and data management. However, less attention is currently paid to developing the data skills of librarians themselves: these skills are often brought in by recruitment in niche areas rather than considered as a wider development need for the library workforce, and are not widely recognised as important to the professional career development of librarians. We believe that building computational and data science capacity within academic libraries will have direct benefits for both librarians and the users we serve.

Library Carpentry is a global effort to provide training to librarians in technical areas that have traditionally been seen as the preserve of researchers, IT support and systems librarians. Established non-profit volunteer organisations, such as Software Carpentry and Data Carpentry,

offer introductory research software skills training with a focus on the needs and requirements of research scientists. Library Carpentry is a comparable introductory software skills training programme with a focus on the needs and requirements of library and information professionals. This paper describes how the material was developed and delivered, and reports on challenges faced, lessons learned and future plans.

Corrall, Sheila, Mary Anne Kennan, and Waseem Afzal. "Bibliometrics and Research Data Management Services: Emerging Trends in Library Support for Research." *Library Trends* 61, no. 3 (2013): 636-674.
<https://doi.org/10.1353/lib.2013.0005>

Corti, Louise, Veerle Van den Eynden, Libby Bishop, and Matthew Woollard. *Managing and Sharing Research Data: A Guide to Good Practice*. Los Angeles: SAGE, 2014. <http://www.worldcat.org/oclc/900727807>

Costello, Mark J., and John Wieczorek. "Best Practice for Biodiversity Data Management and Publication." *Biological Conservation* 173, no. 1 (2014): 68-73.
<http://dx.doi.org/10.1016/j.biocon.2013.10.018>

Council on Library and Information Resources, ed. *Research Data Management: Principles, Practices, and Prospects*. Washington, DC: Council on Library and Information Resources, 2013. <https://www.clir.org/pubs/reports/pub160/>

Cousijn, Helena, Patricia Feeney, Daniella Lowenberg, Eleonora Presani, and Natasha Simons. "Bringing Citations and Usage Metrics Together to Make Data Count." *Data Science Journal*, 18, no. 1 (2019): 9. <http://doi.org/10.5334/dsj-2019-009>.

Over the last years, many organizations have been working on infrastructure to facilitate sharing and reuse of research data. This means that researchers now have ways of making their data available, but not necessarily incentives to do so. Several Research Data Alliance (RDA) working groups have been working on ways to start measuring activities around research data to provide input for new Data Level Metrics (DLMs). These DLMs are a critical step towards providing researchers with credit for their work. In this paper, we describe the outcomes of the work of the Scholarly Link Exchange (Scholix) working group and the Data Usage Metrics working group. The Scholix working group developed a framework that allows organizations to expose and discover links between articles and datasets, thereby providing an indication of data citations. The Data Usage Metrics group works on a

standard for the measurement and display of Data Usage Metrics. Here we explain how publishers and data repositories can contribute to and benefit from these initiatives. Together, these contributions feed into several hubs that enable data repositories to start displaying DLMs. Once these DLMs are available, researchers are in a better position to make their data count and be rewarded for their work.

Covey, Denise Troll. "ORCID @ CMU: Successes and Failures." *Journal of eScience Librarianship* 4, no. 2 (2015): e1083.
<https://doi.org/10.7191/jeslib.2015.1083>

Cox, Andrew M., Mary Anne Kennan, Liz Lyon, and Stephen Pinfield. "Developments in Research Data Management in Academic Libraries: Towards an Understanding of Research Data Service Maturity." *Journal of the Association for Information Science and Technology* 68, no. 9 (2017): 2182-2200.
<http://dx.doi.org/10.1002/asi.23781>

This article reports an international study of research data management (RDM) activities, services, and capabilities in higher education libraries. It presents the results of a survey covering higher education libraries in Australia, Canada, Germany, Ireland, the Netherlands, New Zealand, and the UK. The results indicate that libraries have provided leadership in RDM, particularly in advocacy and policy development. Service development is still limited, focused especially on advisory and consultancy services (such as data management planning support and data-related training), rather than technical services (such as provision of a data catalog, and curation of active data). Data curation skills development is underway in libraries, but skills and capabilities are not consistently in place and remain a concern. Other major challenges include resourcing, working with other support services, and achieving "buy in" from researchers and senior managers. Results are compared with previous studies in order to assess trends and relative maturity levels. The range of RDM activities explored in this study are positioned on a "landscape maturity model," which reflects current and planned research data services and practice in academic libraries, representing a "snapshot" of current developments and a baseline for future research.

Cox, Andrew M., Mary Anne Kennan, Liz Lyon, Stephen Pinfield, and Laura Sbaffi. "Progress in Research Data Services." *International Journal of Digital Curation* 14 no. 1 (2019): 126-135. <https://doi.org/10.2218/ijdc.v14i1.595>

University libraries have played an important role in constructing an infrastructure of support for Research Data Management at an institutional level. This paper presents a comparative analysis of two international surveys of libraries about their involvement in Research Data Services conducted in 2014 and 2018. The aim was to explore how services had developed over this time period, and to explore the drivers and barriers to change. In particular, there was an interest in how far the FAIR data principles had been adopted.

Services in nearly every area were more developed in 2018 than before, but technical services remained less developed than advisory. Progress on institutional policy was also evident. However, priorities did not seem to have shifted significantly. Open ended answers suggested that funder policy, rather than researcher demand, remained the main driver of service development and that resources and skills gaps remained issues. While widely understood as an important reference point and standard, because of their relatively recent publication date, FAIR principles had not been widely adopted explicitly in policy.

Cox, Andrew M., and Stephen Pinfield. "Research Data Management and Libraries: Current Activities and Future Priorities." *Journal of Librarianship and Information Science* 46 no. 4 (2014): 299-316.
<https://doi.org/10.1177/0961000613492542>

Cox, Andrew M., Stephen Pinfield, and Jennifer Smith. "Moving a Brick Building: UK Libraries Coping with Research Data Management as a 'Wicked' Problem " *Journal of Librarianship and Information Science* 48 no. 1 (2016): 3-17.
<https://doi.org/10.1177/0961000614533717>

The purpose of this paper is to explore the value to librarians of seeing research data management as a 'wicked' problem. Wicked problems are unique, complex problems which are defined differently by different stakeholders making them particularly intractable. Data from 26 semi-structured in-depth telephone interviews with librarians was analysed to see how far their perceptions of research data management aligned with the 16 features of a wicked problem identified from the literature. To a large extent research data management is perceived to be wicked, though over time good practices may emerge to help to 'tame' the problem. How interviewees thought research data management should be approached reflected this realisation. The generic value of the concept of wicked problems is

considered and some first thoughts about how the curriculum for new entrants to the profession can prepare them for such problems are presented.

This work is licensed under a Creative Commons Attribution 3.0 Unported License. <https://creativecommons.org/licenses/by/3.0/> .

Cox, Andrew M., and Eddy Verbaan. *Exploring Research Data Management*. London: Facet Publishing, 2018. <https://melvyl.on.worldcat.org/oclc/1037950491>

Cox, Andrew M., Eddy Verbaan, and Barbara Sen. "A Spider, an Octopus, or an Animal Just Coming into Existence? Designing a Curriculum for Librarians to Support Research Data Management." *Journal of eScience Librarianship* 3, no. 1 (2014): e1055. <http://dx.doi.org/10.7191/jeslib.2014.1055>

———. "Upskilling Liaison Librarians for Research Data Management." *Ariadne*, no. 70 (2012). <http://www.ariadne.ac.uk/issue70/cox-et-al>

In this context, JISC have funded the White Rose consortium of academic libraries at Leeds, Sheffield and York, working closely with the Sheffield Information School, in the RDMRose Project (link is external), to develop learning materials that will help librarians grasp the opportunity that RDM offers. The learning materials will be used in the Information School's Masters courses, and are also to be made available to other information sector training providers on a share-alike licence. A version will also be made available (from January 2013) as an Open Educational Resource for use by information professionals who want to update their competencies as part of their continuing professional development (CPD). The learning materials are being developed specifically for liaison librarians, to upskill existing professionals and to expand the knowledge base for new entrants to librarianship. It is hoped to accommodate the perspectives of any information professional, but the scope is not intended to encompass a syllabus for a data management specialist role (following the distinction made by Corral [1]).

This article summarises current thinking developed within the project about the scope and level of such learning materials. This thinking is based on a number of sources: the literature and existing curricula and also the project vision and data collected during the project in focus groups with staff at the participating libraries.

This work is licensed under a Creative Commons Attribution 3.0 Unported License. <https://creativecommons.org/licenses/by/3.0/>.

Cox, Andrew Martin, and Winnie Wan Ting Tam. "A Critical Analysis of Lifecycle Models of the Research Process and Research Data Management." *Aslib Journal of Information Management* 70, no. 2 (2018): 142-157. <https://doi.org/10.1108/AJIM-11-2017-0251>

Cragin, Melissa H., Carole L. Palmer, Jacob R. Carlson, and Michael Witt. "Data Sharing, Small Science and Institutional Repositories." *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 368, no. 1926 (2010): 4023-4038. <https://doi.org/10.1098/rsta.2010.0165>

Creamer, Andrew. "Current Issues and Approaches to Curating Student Research Data." *Bulletin of the Association for Information Science and Technology* 41, no. 6 (2015): 22-25. <https://doi.org/10.1002/bult.2015.1720410610>

Creamer, Andrew, Myrna E. Morales, Javier Crespo, Donna Kafel, and Elaine R. Martin. "An Assessment of Needed Competencies to Promote the Data Curation and Management Librarianship of Health Sciences and Science and Technology Librarians in New England." *Journal of eScience Librarianship* 1, no. 1 (2012): e1006. <http://dx.doi.org/10.7191/jeslib.2012.1006>

Creamer, Andrew T., Myrna E. Morales, Donna Kafel, Javier Crespo, and Elaine R. Martin. "A Sample of Research Data Curation and Management Courses." *Journal of eScience Librarianship* 1, no. 2 (2012). <http://dx.doi.org/10.7191/jeslib.2012.1016>

Crosas, Mercè. "A Data Sharing Story." *Journal of eScience Librarianship* 1, no. 3 (2012): e1020. <http://dx.doi.org/10.7191/jeslib.2012.1020>

———. "The Dataverse Network: An Open-Source Application for Sharing, Discovering and Preserving Data." *D-Lib Magazine* 17, no. 1/2 (2011). <http://www.dlib.org/dlib/january11/crosas/01crosas.html>

———. "The Evolution of Data Citation: From Principles to Implementation." *IASSIST Quarterly* 37, no. 1-4 (2013): 62-70. <https://doi.org/10.29173/iq504>

Crosas, Mercè, Gary King, James Honaker, and Latanya Sweeney. "Automating Open Science for Big Data." *The Annals of the American Academy of Political and Social Science* 659, no. 1 (2014): 260-273. <http://gking.harvard.edu/publications/automating-open-science-big-data>

Crowston, Kevin. "'Personas' to Support Development of Cyberinfrastructure for Scientific Data Sharing." *Journal of eScience Librarianship* 4, no. 2 (2015): e1082. <https://doi.org/10.7191/jeslib.2015.1082>

Cuevas-Vicentín, Víctor, Parisa Kianmajd, Bertram Ludäscher, Paolo Missier, Fernando Chirigati, Yaxing Wei, David Koop, and Saumen Dey. "The PBase Scientific Workflow Provenance Repository." *International Journal of Digital Curation* 9, no. 2 (2014): 28-38. <https://doi.org/10.2218/ijdc.v9i2.332>

Scientific workflows and their supporting systems are becoming increasingly popular for compute-intensive and data-intensive scientific experiments. The advantages scientific workflows offer include rapid and easy workflow design, software and data reuse, scalable execution, sharing and collaboration, and other advantages that altogether facilitate "reproducible science". In this context, provenance—information about the origin, context, derivation, ownership, or history of some artifact—plays a key role, since scientists are interested in examining and auditing the results of scientific experiments.

However, in order to perform such analyses on scientific results as part of extended research collaborations, an adequate environment and tools are required. Concretely, the need arises for a repository that will facilitate the sharing of scientific workflows and their associated execution traces in an interoperable manner, also enabling querying and visualization. Furthermore, such functionality should be supported while taking performance and scalability into account.

With this purpose in mind, we introduce PBase: a scientific workflow provenance repository implementing the ProvONE proposed standard, which extends the emerging W3C PROV standard for provenance data with workflow specific concepts. PBase is built on the Neo4j graph database, thus offering capabilities such as declarative and efficient querying. Our experiences demonstrate the power gained by supporting various types of queries for provenance data. In addition, PBase is equipped with a user friendly interface tailored for the visualization of scientific workflow provenance data, making the specification of queries and the interpretation of their results easier and more effective.

Curdt, Constanze, and Dirk Hoffmeister. "Research Data Management Services for a Multidisciplinary, Collaborative Research Project: Design and Implementation of

the TR32DB project Database." *Program* 49, no. 4 (2015): 494-512.
<http://dx.doi.org/10.1108/PROG-02-2015-0016>

Curdt, Constanze, Dirk Hoffmeister, Guido Waldhoff, Christian Jekel, and Georg Bareth. "Scientific Research Data Management for Soil-Vegetation-Atmosphere Data—The TR32DB." *International Journal of Digital Curation* 7, no. 2 (2012): 68-80. <https://doi.org/10.2218/ijdc.v7i2.208>

The implementation of a scientific research data management system is an important task within long-term, interdisciplinary research projects. Besides sustainable storage of data, including accurate descriptions with metadata, easy and secure exchange and provision of data is necessary, as well as backup and visualisation. The design of such a system poses challenges and problems that need to be solved.

This paper describes the practical experiences gained by the implementation of a scientific research data management system, established in a large, interdisciplinary research project with focus on Soil-Vegetation-Atmosphere Data.

Curty, Renata Gonçalves. "Factors Influencing Research Data Reuse in the Social Sciences: An Exploratory Study." *International Journal of Digital Curation* 11, no. 1 (2016): 96-117. <https://doi.org/10.2218/ijdc.v11i1.401>

The development of e-Research infrastructure has enabled data to be shared and accessed more openly. Policy mandates for data sharing have contributed to the increasing availability of research data through data repositories, which create favourable conditions for the re-use of data for purposes not always anticipated by original collectors. Despite the current efforts to promote transparency and reproducibility in science, data re-use cannot be assumed, nor merely considered a 'thrifting' activity where scientists shop around in data repositories considering only the ease of access to data. The lack of an integrated view of individual, social and technological influential factors to intentional and actual data re-use behaviour was the key motivator for this study. Interviews with 13 social scientists produced 25 factors that were found to influence their perceptions and experiences, including both their unsuccessful and successful attempts to re-use data. These factors were grouped into six theoretical variables: perceived benefits, perceived risks, perceived effort, social influence, facilitating conditions, and perceived re-usability. These research findings provide an in-depth understanding about

the re-use of research data in the context of open science, which can be valuable in terms of theory and practice to help leverage data re-use and make publicly available data more actionable.

Curty, Renata Gonçalves, Kevin Crowston, Alison Specht, Bruce W. Grant, and Elizabeth D. Dalton. "Attitudes and Norms Affecting Scientists' Data Reuse." *PLOS ONE* 12, no. 12 (2017): e0189288.
<https://doi.org/10.1371/journal.pone.0189288>

The value of sharing scientific research data is widely appreciated, but factors that hinder or prompt the reuse of data remain poorly understood. Using the Theory of Reasoned Action, we test the relationship between the beliefs and attitudes of scientists towards data reuse, and their self-reported data reuse behaviour. To do so, we used existing responses to selected questions from a worldwide survey of scientists developed and administered by the DataONE Usability and Assessment Working Group (thus practicing data reuse ourselves). Results show that the perceived efficacy and efficiency of data reuse are strong predictors of reuse behaviour, and that the perceived importance of data reuse corresponds to greater reuse. Expressed lack of trust in existing data and perceived norms against data reuse were not found to be major impediments for reuse contrary to our expectations. We found that reported use of models and remotely-sensed data was associated with greater reuse. The results suggest that data reuse would be encouraged and normalized by demonstration of its value. We offer some theoretical and practical suggestions that could help to legitimize investment and policies in favor of data sharing.

Dai, Yun. "How Many Ways Can We Teach Data Literacy?" *IASSIST Quarterly* 43, no. 4 (2019): 1–11. <https://doi.org/10.29173/iq963>

Dallmeier-Tiessen, Suenje, Mariella Guercio, Robert Darby, Kathrin Gitmans, Simon Lambert, Brian Matthews, Jari Suhonen Salvatore Mele, and Michael Wilson. "Enabling Sharing and Reuse of Scientific Data." *New Review of Information Networking* 19, no. 1 (2014).
<https://doi.org/10.1080/13614576.2014.883936>

Dallmeier-Tiessen, Suenje, Varsha Khodiyar, Fiona Murphy, Amy Nurnberger, Lisa Raymond, and Angus Whyte. "Connecting Data Publication to the Research Workflow: A Preliminary Analysis." *International Journal of Digital Curation* 12, no. 1 (2017): 88-105. <https://doi.org/10.2218/ijdc.v12i1.533>

The data curation community has long encouraged researchers to document collected research data during active stages of the research workflow, to provide robust metadata earlier, and support research data publication and preservation. Data documentation with robust metadata is one of a number of steps in effective data publication. Data publication is the process of making digital research objects 'FAIR', i.e. findable, accessible, interoperable, and reusable; attributes increasingly expected by research communities, funders and society. Research data publishing workflows are the means to that end. Currently, however, much published research data remains inconsistently and inadequately documented by researchers. Documentation of data closer in time to data collection would help mitigate the high cost that repositories associate with the ingest process. More effective data publication and sharing should in principle result from early interactions between researchers and their selected data repository. This paper describes a short study undertaken by members of the Research Data Alliance (RDA) and World Data System (WDS) working group on Publishing Data Workflows. We present a collection of recent examples of data publication workflows that connect data repositories and publishing platforms with research activity 'upstream' of the ingest process. We re-articulate previous recommendations of the working group, to account for the varied upstream service components and platforms that support the flow of contextual and provenance information downstream. These workflows should be open and loosely coupled to support interoperability, including with preservation and publication environments. Our recommendations aim to stimulate further work on researchers' views of data publishing and the extent to which available services and infrastructure facilitate the publication of FAIR data. We also aim to stimulate further dialogue about, and definition of, the roles and responsibilities of research data services and platform providers for the 'FAIRness' of research data publication workflows themselves.

Darch, Peter T., and Emily J. M. Knox. "Ethical Perspectives on Data and Software Sharing in the Sciences: A Research Agenda." *Library & Information Science Research* 39, no. 4 (2017): 295-302. <https://doi.org/10.1016/j.lisr.2017.11.008>

Darlington, Jeffrey. "A National Archive of Datasets." *Ariadne*, no. 39 (2004). <http://www.ariadne.ac.uk/issue39/ndad>

Davis, Hilary M., and William M. Cross. "Using a Data Management Plan Review Service as a Training Ground for Librarians." *Journal of Librarianship and*

Scholarly Communication 3, no. 2 (2015): eP1243. <http://doi.org/10.7710/2162-3309.1243>

INTRODUCTION Research Data Management (RDM) offers opportunities and challenges at the interface of library support and researcher needs. Libraries are in a position of balancing the capacity to provide support at the point of need while also implementing training for subject liaison librarians grounded in the practical issues and realities facing researchers and their institutions. **DESCRIPTION OF PROGRAM/SERVICE** The North Carolina State University (NCSU) Libraries has deployed a Data Management Plan (DMP) Review service managed by a committee of librarians with diverse experience in data management and domain expertise. By rotating librarians through membership on the committee and by inviting subject liaisons librarians to participate in the DMP Review process, our training ground model aims to develop needed competencies and support researchers through relevant services and partnerships. **AUDIT OF PROGRAM/SERVICE** This article presents an audit of the DMP Review service as a training ground to develop and enhance competencies as identified by the Joint Task Force on Librarians' Competencies in Support of E-Research and Scholarly Communication. **NEXT STEPS AND CONCLUSIONS** The DMP Review service creates opportunities for librarians to learn valuable skills while simultaneously providing a time-sensitive service to researchers. The process of auditing competencies developed by participating in the DMP Review service highlights gaps needed to more fully support RDM and reinforces the capacity of the DMP Review service as a training ground to sustain and iterate learning opportunities for librarians engaged in research support and partnerships.

De La Beaujardière, Jeff. "NOAA Environmental Data Management." *Journal of Map & Geography Libraries* 12, no. 1 (2016): 5-27. <http://dx.doi.org/10.1080/15420353.2015.1087446>

De Waard, Anita. "Research Data Management at Elsevier: Supporting Networks of Data and Workflows." *Information Services & Use* 36, no. 1-2 (2016): 49-55. <https://doi.org/10.3233/isu-160805>

Dearborn, Carly C., Amy J. Barto, and Neal A. Harmeyer. "The Purdue University Research Repository: HUBzero Customization for Dataset Publication and Digital Preservation." *OCLC Systems & Services: International Digital Library Perspectives* 30, no. 1 (2014): 15-27. <https://doi.org/10.1108/oclc-07-2013-0022>

Dearborn, Dylanne, Steve Marks, and Leanne Trimble. "The Changing Influence of Journal Data Sharing Policies on Local RDM Practices." *International Journal of Digital Curation* 12, no. 2 (2017): 376-389.
<https://doi.org/10.2218/ijdc.v12i2.583>

The purpose of this study was to examine changes in research data deposit policies of highly ranked journals in the physical and applied sciences between 2014 and 2016, as well as to develop an approach to examining the institutional impact of deposit requirements. Policies from the top ten journals (ranked by impact factor from the Journal Citation Reports) were examined in 2014 and again in 2016 in order to determine if data deposits were required or recommended, and which methods of deposit were listed as options. For all 2016 journals with a required data deposit policy, publication information (2009-2015) for the University of Toronto was pulled from Scopus and departmental affiliation was determined for each article. The results showed that the number of high-impact journals in the physical and applied sciences requiring data deposit is growing. In 2014, 71.2% of journals had no policy, 14.7% had a recommended policy, and 13.9% had a required policy (n=836). In contrast, in 2016, there were 58.5% with no policy, 19.4% with a recommended policy, and 22.0% with a required policy (n=880). It was also evident that U of T chemistry researchers are by far the most heavily affected by these journal data deposit requirements, having published 543 publications, representing 32.7% of all publications in the titles requiring data deposit in 2016. The Python scripts used to retrieve institutional publications based on a list of ISSNs have been released on GitHub so that other institutions can conduct similar research.

Dehnhard, I., E. Weichselgartner, and G. Krampen. "Researcher's Willingness to Submit Data for Data Sharing: A Case Study on a Data Archive for Psychology." *Data Science Journal* 12 (2013): 172-180. <http://dx.doi.org/10.2481/dsj.12-037>

Data sharing has gained importance in scientific communities because scientific associations and funding organizations require long term preservation and dissemination of data. To support psychology researchers in data archiving and data sharing, the Leibniz Institute for Psychology Information developed an archiving facility for psychological research data in Germany: PsychData. In this paper we report different types of data requests that were sent to researchers with the aim of building up a sustainable data archive. Resulting response rates were rather low, however, comparable to

those published by other authors. Possible reasons for the reluctance of researchers to submit data are discussed.

This work is licensed under a Creative Commons Attribution 3.0 Unported License, <https://creativecommons.org/licenses/by/3.0/>.

Delasalle, Jenny. "Research Data Management at the University of Warwick: Recent Steps towards a Joined-up Approach at a UK University." *LIBREAS. Library Ideas*, no. 23 (2013): 97-105. <http://libreas.eu/ausgabe23/10delasalle/>

This paper charts the steps taken and possible ways forward for the University of Warwick in its approach to research data management, providing a typical example of a UK research university's approach in two strands: requirements and support. The UK government approach and funding landscape in relation to research data management provided drivers for the University of Warwick to set requirements and provide support, and examples of good practice at other institutions, support from a central national body (the UK Digital Curation Centre) and learning from other universities' experiences all proved valuable to the University of Warwick. Through interviews with researchers at Warwick, various issues and challenges are revealed: perhaps the biggest immediate challenges for Warwick going forward are overcoming scepticism amongst researchers, overcoming costs, and understanding the implications of involving third party companies in research data management. Building technical infrastructure could sit alongside and beyond those immediate steps and beyond the challenges that face one University are those that affect academia as a whole. Researchers and university administrators need to work together to address the broader challenges, such as the accessibility of data for future use and the reward for researchers who practice data management in exemplary ways, and indeed it may be that a wider, national or international but disciplinary technical infrastructure affects what an individual university needs to achieve. As we take these steps, universities and institutions are all learning from each other.

Delserone, Leslie M. "At the Watershed: Preparing for Research Data Management and Stewardship at the University of Minnesota Libraries." *Library Trends* 57, no. 2 (2008): 202-210. <https://doi.org/10.1353/lib.0.0032>

Descoteaux, Danielle, Chiara Farinelli, Marina Soares e Silva, and Anita de Waard. "Playing Well on the Data FAIRground: Initiatives and Infrastructure in

Research Data Management." *Data Intelligence* 1, no. 4 (2019): 350–367. https://doi.org/10.1162/dint_a_00020

Dietrich, Dianne. "Metadata Management in a Data Staging Repository." *Journal of Library Metadata* 10, no. 2/3 (2010): 79-98. <https://doi.org/10.1080/19386389.2010.506376>

Dietrich, Dianne, Trisha Adamus, Alison Miner, and Gail Steinhart. "De-Mystifying the Data Management Requirements of Research Funders." *Issues in Science & Technology Librarianship*, no. 70 (2012). <http://www.istl.org/12-summer/refereed1.html>

Dillo, Ingrid, and Peter Doorn. "The Front Office-Back Office Model: Supporting Research Data Management in the Netherlands." *International Journal of Digital Curation* 9, no. 2 (2014): 39-46. <https://doi.org/10.2218/ijdc.v9i2.333>

High quality and timely data management and secure storage of data, both during and after completion of research, are an essential prerequisite for sharing that data. It is therefore crucial that universities and research institutions themselves formulate a clear policy on data management within their organization. For the implementation of this data management policy, high quality support for researchers and an adequate technical infrastructure are indispensable.

This practice paper will present an overview of the merging federated data infrastructure in the Netherlands with its front office-back office model, as a use case of an efficient and effective national support infrastructure for researchers.

We will elaborate on the stakeholders involved, on the services they offer each other, and on the benefits of this model not only for the front and back offices themselves, but also for the researchers. We will also pay attention to a number of challenges that we are facing, like the implementation of a technical infrastructure for automatic data ingest and integrating access to research data.

Donaldson, Devan Ray, Ingrid Dillo, Robert Downs, and Sarah Ramdeen. "The Perceived Value of Acquiring Data Seals of Approval." *International Journal of Digital Curation* 12, no. 1 (2017): 130-151. <https://doi.org/10.2218/ijdc.v12i1.481>

The Data Seal of Approval (DSA) is one of the most widely used standards for Trusted Digital Repositories to date. Those who developed this standard have articulated seven main benefits of acquiring DSAs: 1) stakeholder confidence, 2) improvements in communication, 3) improvement in processes, 4) transparency, 5) differentiation from others, 6) awareness raising about digital preservation, and 7) less labor- and time-intensive. Little research has focused on if and how those who have acquired DSAs actually perceive these benefits. Consequently, this study examines the benefits of acquiring DSAs from the point of view of those who have them. In a series of 15 semi-structured interviews with representatives from 16 different organizations, participants described the benefits of having DSAs in their own words. Our findings suggest that participants experience all of the seven benefits that those who developed the standard promised. Additionally, our findings reflect the greater importance of some of those benefits compared to others. For example, participants mentioned the benefits of stakeholder confidence, transparency, improvement in processes and awareness raising about digital preservation more frequently than they discussed less labor- and time-intensive (e.g. it being less labor- and time-intensive to acquire DSAs than becoming certified by other standards), improvements in communication, and differentiation from others. Participants also mentioned two additional benefits of acquiring DSAs that are not explicitly listed on the DSA website that were very important to them: 1) the impact of acquiring the DSA on documentation of their workflows, and 2) assurance that they were following best practice. Implications and future directions for research are discussed.

Donaldson, Devan Ray, Shawn Martin, and Thomas Proffen. "Understanding Perspectives on Sharing Neutron Data at Oak Ridge National Laboratory." *Data Science Journal* 16 (2017). <https://doi.org/10.5334/dsj-2017-035>

Even though the importance of sharing data is frequently discussed, data sharing appears to be limited to a few fields, and practices within those fields are not well understood. This study examines perspectives on sharing neutron data collected at Oak Ridge National Laboratory's neutron sources. Operation at user facilities has traditionally focused on making data accessible to those who create them. The recent emphasis on open data is shifting the focus to ensure that the data produced are reusable by others. This mixed methods research study included a series of surveys and focus group interviews in which 13 data consumers, data managers, and data producers answered questions about their perspectives on sharing neutron data. Data consumers

reported interest in reusing neutron data for comparison/verification of results against their own measurements and testing new theories using existing data. They also stressed the importance of establishing context for data, including how data are produced, how samples are prepared, units of measurement, and how temperatures are determined. Data managers expressed reservations about reusing others' data because they were not always sure if they could trust whether the people responsible for interpreting data did so correctly. Data producers described concerns about their data being misused, competing with other users, and over-reliance on data producers to understand data. We present the Consumers Managers Producers (CMP) Model for understanding the interplay of each group regarding data sharing. We conclude with policy and system recommendations and discuss directions for future research.

Donnelly, Martin. "The DCC's Institutional Engagements: Raising Research Data Management Capacity in UK Higher Education." *Bulletin of the American Society for Information Science and Technology* 39, no. 6 (2013): 37-40. <https://doi.org/10.1002/bult.2013.1720390612>

Donnelly, Martin, Sarah Jones, and John W. Pattenden-Fail. "DMP Online: A Demonstration of the Digital Curation Centre's Web-based Tool for Creating, Maintaining and Exporting Data Management Plans." *Lecture Notes in Computer Science* 6273 (2010): 530-533. <https://doi.org/10.2218/ijdc.v5i1.152>

Donnelly, Martin, and Robin North. "The Milieu and the MESSAGE: Talking to Researchers about Data Curation Issues in a Large and Diverse E-science Project." *International Journal of Digital Curation* 6, no. 1 (2011): 32-44. <https://doi.org/10.2218/ijdc.v6i1.170>

Doty, Jennifer, Joel Herndon, Jared Lyle, and Libbie Stephenson. "Learning to Curate." *Bulletin of the American Society for Information Science and Technology* 40, no. 6 (2014): 31-34. <https://doi.org/10.1002/bult.2014.1720400610>

Doty, Jennifer, Melanie T. Kowalski, Bethany C. Nash, and Simon F. O'Riordan. "Making Student Research Data Discoverable: A Pilot Program Using Dataverse." *Journal of Librarianship and Scholarly Communication* 3, no. 2 (2015): eP1234. <http://doi.org/10.7710/2162-3309.1234>

INTRODUCTION The support and curation of research data underlying theses and dissertations are an opportunity for institutions to enhance their ETD collections. This article describes a pilot data archiving service that leverages Emory University's existing Electronic Theses and Dissertations

(ETDs) program. DESCRIPTION OF PROGRAM This pilot service tested the appropriateness of Dataverse, a data repository, as a data archiving and access solution for Emory University using research data identified in Emory University's ETD repository, developed the legal documents necessary for a full implementation of Dataverse on campus, and expanded outreach efforts to meet the research data needs of graduate students. This article also situates the pilot service within the context of Emory Libraries and explains how it relates to other library efforts currently underway. NEXT STEPS The pilot project team plans to seek permission from alumni whose data were included in the pilot to make them available publicly in Dataverse, and the team will revise the ETD license agreement to allow this type of use. The team will also automate the ingest of supplemental ETD research data into the data repository where possible and create a workshop series for students who are creating research data as part of their theses or dissertations.

Douglass, Kimberly, Suzie Allard, Carol Tenopir, Lei Wu, and Mike Frame. "Managing Scientific Data as Public Assets: Data Sharing Practices and Policies among Full-Time Government Employees." *Journal of the Association for Information Science and Technology* 65, no. 2 (2014): 251-262. <http://dx.doi.org/10.1002/asi.22988>

Downs, Robert R., and Robert S. Chen. "Designing Submission and Workflow Services for Preserving Interdisciplinary Scientific Data." *Earth Science Informatics* 3, no. 1/2 (2010): 101-110. <https://doi.org/10.1007/s12145-010-0051-6>

——— "Organizational Needs for Managing and Preserving Geospatial Data and Related Electronic Records." *Data Science Journal* 4 (2005). <https://doi.org/10.2481/dsj.4.255>

Government agencies and other organizations are required to manage and preserve records that they create and use to facilitate future access and reuse. The increasing use of geospatial data and related electronic records presents new challenges for these organizations, which have relied on traditional practices for managing and preserving records in printed form. This article reports on an investigation of current and future needs for managing and preserving geospatial electronic records on the part of local and state-level organizations in the New York City metropolitan region. It introduces the study and describes organizational needs observed, including needs for organizational coordination and interorganizational cooperation throughout the entire data lifecycle.

Downs, Robert R., Ruth Duerr, and Denise J. Hills. "Data Stewardship in the Earth Sciences." *D-Lib Magazine* 21, no. 7/8 (2015).

<http://www.dlib.org/dlib/july15/downs/07downs.html>

Drachen, Thea Marie, Ole Ellegaard, Asger Væring Larsen, and Søren Bertil Fabricius Dorch. "Sharing Data Increases Citations." *LIBER Quarterly* 26, no. 2 (2016): 67-82. <https://doi.org/10.18352/lq.10149>

Dressel, Willow F. "Research Data Management Instruction for Digital Humanities." *Journal of eScience Librarianship* 6, no 2 (2017): e1115.

<https://doi.org/10.7191/jeslib.2017.1115>

eScience related library services at Princeton University started in response to the National Science Foundation's (NSF) data management plan requirements, and grew to encompass a range of services including data management plan consultation, assistance with depositing into a disciplinary or institutional repository, and research data management instruction. These services were initially directed at science and engineering disciplines on campus, but the eScience Librarian soon realized the relevance of research data management instruction for humanities disciplines with digital approaches. Applicability to the digital humanities was initially recognized by discovery of related efforts from the history department's Information Technology (IT) manager in the form of a graduate-student workshop on file and digital-asset management concepts. Seeing the common ground these activities shared with research data management, a collaboration was formed between the history department's IT Manager and the eScience Librarian to provide a research data management overview to the entire campus community. The eScience Librarian was then invited to participate in the history department's graduate student file and digital asset management workshop to provide an overview of other research data management concepts. Based on the success of the collaboration with the history department IT, the eScience Librarian offered to develop a workshop for the newly formed Center for Digital Humanities at Princeton. To develop the workshop, background research on digital humanities curation was performed revealing similarities and differences between digital humanities curation and research data management in the sciences. These similarities and differences, workshop results, and areas of further study are discussed.

Dressler, Virginia A., Kristin Yeager, and Elizabeth Richardson. "Developing a Data Management Consultation Service for Faculty Researchers: A Case Study

from a Large Midwestern Public University." *International Journal of Digital Curation* 14 no. 1 (2019): 1–23. <https://doi.org/10.2218/ijdc.v14i1.590>

To inform the development of data management services, a library research team at Kent State University conducted a survey of all tenured, tenure-track, and non-tenure track faculty about their data management practices and perceptions. The methodology and results will be presented in the article, as well as how this information was used to inform future work in the library's internal working group. Recommendations will be presented that other academic libraries could model in order to develop similar services at their institutions. Personal anecdotes are included that help ascertain current practices and sentiments around research data from the perspective of the researcher. The article addresses the particular needs of a large Midwestern U.S. academic campus, which are not currently reflected in literature on the topic.

Durantea, Kim, and Darren Hardya. "Discovery, Management, and Preservation of Geospatial Data Using Hydra." *Journal of Map & Geography Libraries: Advances in Geospatial Information, Collections & Archives* 11, no. 2 (2015). <https://doi.org/10.1080/15420353.2015.1041630>

Duranti, L. "The Long-Term Preservation of Accurate and Authentic Digital Data: The INTERPARES Project." *Data Science Journal* 4 (2006). <https://doi.org/10.2481/dsj.4.106>

This article presents the InterPARES Project, a multidisciplinary international research initiative aimed at developing the theoretical and methodological knowledge necessary for the long-term preservation of digital entities produced in the course of business or research activity so that their authenticity can be presumed or verified. The methodology, research activities, preliminary findings and projected products are discussed in the context of the issues that the project attempts to address.

Dürr, Eugène, Kees van der Meer, Wim Luxemburg, and Ronald Dekker. "Dataset Preservation for the Long Term: Results of the DareLux Project." *International Journal of Digital Curation* 3, no. 1 (2008): 29-43. <https://doi.org/10.2218/ijdc.v3i1.40>

Dürr, Eugène, Kees van der Meer, Wim Luxemburg, Maria Heijne, and Ronald Dekker. "Long-Time Preservation of Data Sets, Results of the DareLux Project."

Information Services and Use 28, no. 3/4 (2008): 281-294.
<https://doi.org/10.3233/isu-2008-0571>

Dyke, Kevin R., Ryan Mattke, Len Kne, and Shawn Rounds. "Placing Data in the Land of 10,000 Lakes: Navigating the History and Future of Geospatial Data Production, Stewardship, and Archiving in Minnesota." *Journal of Map & Geography Libraries* 12, no. 1 (2016): 52-72.
<https://doi.org/10.1080/15420353.2015.1073655>

Eaker, Christopher. "Planning Data Management Education Initiatives: Process, Feedback, and Future Directions." *Journal of eScience Librarianship* 3, no. 1 (2014): e1054. <http://dx.doi.org/10.7191/jeslib.2014.1054>

Eclevia, Marian Ramos, John Christopher La Torre Fredeluces, Carlos Lagrosas Eclevia, Jr., and Roselle Saguibo Maestro. "What Makes a Data Librarian?." *Qualitative and Quantitative Methods in Libraries*, 8, no. 3, (2019): 273-290.
<http://qqml-journal.net/index.php/qqml/article/view/541>

Edmunds, Scott C., Peter Li, Christopher I. Hunter, Si Zhe Xiao, Robert L. Davidson, Nicole Nogoy, and Laurie Goodman. "Experiences in Integrated Data and Research Object Publishing Using GigaDB." *International Journal on Digital Libraries* (2016): 1-13. <http://dx.doi.org/10.1007/s00799-016-0174-6>

Erwin, Tracey, and Julie Sweetkind-Singer. "The National Geospatial Digital Archive: A Collaborative Project to Archive Geospatial Data." *Journal of Map & Geography Libraries* 6, no. 1 (2010): 6-25.
<https://doi.org/10.1080/15420350903432440>

Erwin, Tracey, Julie Sweetkind-Singer, and Mary Lynette Larsgaard. "The National Geospatial Digital Archives—Collection Development: Lessons Learned." *Library Trends* 57, no. 3 (2009): 490-515.
<https://doi.org/10.1353/lib.0.0049>

Eschenfelder, Kristin R., and Andrew Johnson. "Managing the Data Commons: Controlled Sharing of Scholarly Data." *Journal of the Association for Information Science and Technology* 65, no. 9 (2014): 1757-1774,
<http://dx.doi.org/10.1002/asi.23086>

Eschenfelder, Kristin R., and Kalpana Shankar. "Organizational Resilience in Data Archives: Three Case Studies in Social Science Data Archives." *Data Science Journal* 16, no. 12 (2017). <http://doi.org/10.5334/dsj-2017-012>

As public investment in archiving research data grows, there has been increasing attention to the longevity or sustainability of the data repositories that curate such data. While there have been many conceptual frameworks developed and case reports of individual archives and digital repositories, there have been few empirical studies of how such archives persist over time. In this paper, we draw upon organizational studies theories to approach the issue of sustainability from an organizational perspective, focusing specifically on the organizational histories of three social science data archives (SSDA): ICPSR, UKDA, and LIS. Using a framework of organizational resilience to understand how archives perceive crisis, respond to it, and learn from experience, this article reports on an empirical study of sustainability in these long-lived SSDAs. The study draws from archival documents and interviews to examine how sustainability can and should be conceptualized as on-going processes over time and not as a quality at a single moment. Implications for research and practice in data archive sustainability are discussed.

Eynden, Veerle, and Louise Corti. "Advancing Research Data Publishing Practices for the Social Sciences: From Archive Activity to Empowering Researchers." *International Journal on Digital Libraries* 18, no. 2 (2017): 113-121.
<https://doi.org/10.1007/s00799-016-0177-3>

Fallow, Colleen, Elise Dunham, Elizabeth Wickes, Dena Strong, Ayla Stein, Qian Zhang, Kyle Rimkus, Bill Ingram, and Heidi J. Imker. "Overly Honest Data Repository Development." *Code4Lib Journal*, no. 34 (2016).
<http://journal.code4lib.org/articles/11980>

After a year of development, the library at the University of Illinois at Urbana-Champaign has launched a repository, called the Illinois Data Bank (<https://databank.illinois.edu/>), to provide Illinois researchers with a free, self-serve publishing platform that centralizes, preserves, and provides persistent and reliable access to Illinois research data. This article presents a holistic view of development by discussing our overarching technical, policy, and interface strategies. By openly presenting our design decisions, the rationales behind those decisions, and associated challenges this paper aims to contribute to the library community's work to develop repository services that meet growing data preservation and sharing needs.

This work is licensed under a Creative Commons Attribution 3.0 United States License, <https://creativecommons.org/licenses/by/3.0/us/>.

Fan, Zhenjia. "Context-Based Roles and Competencies of Data Curators in Supporting Research Data Lifecycle Management: Multi-Case Study in China." *Libri* 69, no. 2 (2019): 127-137. <https://doi.org/10.1515/libri-2018-0065>

Faniel, Ixchel M., and Ann Zimmerman. "Beyond the Data Deluge: A Research Agenda for Large-Scale Data Sharing and Reuse." *International Journal of Digital Curation* 6, no. 1 (2011): 58-69. <https://doi.org/10.2218/ijdc.v6i1.172>

Farrell, Shannon L., and Megan Kocher. "Examining the Research Practices of Agricultural Scholars at the University of Minnesota Twin Cities." *Journal of Agricultural & Food Information* 18, no. 3-4 (2017): 357-372. <http://dx.doi.org/10.1080/10496505.2017.1318072>

Fary, Michael, and Kim Owen. *Developing an Institutional Research Data Management Plan Service*. Louisville, CO: EDUCAUSE, 2013. <https://library.educause.edu/resources/2013/1/developing-an-institutional-research-data-management-plan-service>

Faundeen, John L. "The Challenge of Archiving and Preserving Remotely Sensed Data." *Data Science Journal* 2 (2003): 159-163. <https://doi.org/10.2481/dsj.2.159>

Few would question the need to archive the scientific and technical (S&T) data generated by researchers. At a minimum, the data are needed for change analysis. Likewise, most people would value efforts to ensure the preservation of the archived S&T data. Future generations will use analysis techniques not even considered today. Until recently, archiving and preserving these data were usually accomplished within existing infrastructures and budgets. As the volume of archived data increases, however, organizations charged with archiving S&T data will be increasingly challenged (U.S. General Accounting Office, 2002). The U.S. Geological Survey has had experience in this area and has developed strategies to deal with the mountain of land remote sensing data currently being managed and the tidal wave of expected new data. The Agency has dealt with archiving issues, such as selection criteria, purging, advisory panels, and data access, and has met with preservation challenges involving photographic and digital media. That experience has allowed the USGS to develop management approaches, which this paper outlines.

Faundeen, John L., and Vivian B. Hutchison. "The Evolution, Approval and Implementation of the U.S. Geological Survey Science Data Lifecycle Model." *Journal of eScience Librarianship* 6, no. 2 (2017): e1117. <https://doi.org/10.7191/jeslib.2017.1117>

This paper details how the U.S. Geological Survey (USGS) Community for Data Integration (CDI) Data Management Working Group developed a Science Data Lifecycle Model, and the role the Model plays in shaping agency-wide policies and data management applications. Starting with an extensive literature review of existing data lifecycle models, representatives from various backgrounds in USGS attended a two-day meeting where the basic elements for the Science Data Lifecycle Model were determined. Refinements and reviews spanned two years, leading to finalization of the model and documentation in a formal agency publication¹.

The Model serves as a critical framework for data management policy, instructional resources, and tools. The Model helps the USGS address both the Office of Science and Technology Policy (OSTP)² for increased public access to federally funded research, and the Office of Management and Budget (OMB)³ 2013 Open Data directives, as the foundation for a series of agency policies related to data management planning, metadata development, data release procedures, and the long-term preservation of data. Additionally, the agency website devoted to data management instruction and best practices (www2.usgs.gov/datamanagement) is designed around the Model's structure and concepts. This paper also illustrates how the Model is being used to develop tools for supporting USGS research and data management processes.

Fear, Kathleen. "Building Outreach on Assessment: Researcher Compliance with Journal Policies for Data Sharing." *Bulletin of the Association for Information Science and Technology* 41, no. 6 (2015): 18-21.

<https://doi.org/10.1002/bult.2015.1720410609>

———. "'You Made It, You Take Care of It': Data Management as Personal Information Management." *International Journal of Digital Curation* 6, no. 2 (2011): 53-77. <https://doi.org/10.2218/ijdc.v6i2.190>

This study explores how researchers at a major Midwestern university are managing their data, as well as the factors that have shaped their practices and those that motivate or inhibit changes to that practice. A combination of survey (n=363) and interview data (n=15) yielded both qualitative and quantitative results bearing on my central research question: In what types of data management activities do researchers at this institution engage? Corollary to that, I also explored the following questions: What do researchers feel could be improved about their data management practices?

Which services might be of interest to them? How do they feel those services could most effectively be implemented?

In this paper, I situate researchers' data management practices within a theory of personal information management. I present a view of data management and preservation needs from researchers' perspectives across a range of domains. Additionally, I discuss the implications that understanding research data management as personal information management has for introducing services to support and improve data management practice.

Fear, Kathleen, and Devan Ray Donaldson. "Provenance and Credibility in Scientific Data Repositories." *Archival Science* 12, no. 3 (2012): 319-339. <https://doi.org/10.1007/s10502-012-9172-7>

Fearon, David, Jr., Betsy Gunia, Barbara E. Pralle, Sherry Lake, and Andrew L. Sallans. *SPEC Kit 334: Research Data Management Services*. Washington, DC: ARL, 2013. <http://www.worldcat.org/oclc/855211766>

Fecher, Benedikt, Sascha Friesike, and Marcel Hebing. "What Drives Academic Data Sharing?" *PLoS ONE* 10, no. 2 (2015): e0118053. <http://doi.org/10.1371/journal.pone.0118053>

Despite widespread support from policy makers, funding agencies, and scientific journals, academic researchers rarely make their research data available to others. At the same time, data sharing in research is attributed a vast potential for scientific progress. It allows the reproducibility of study results and the reuse of old data for new research questions. Based on a systematic review of 98 scholarly papers and an empirical survey among 603 secondary data users, we develop a conceptual framework that explains the process of data sharing from the primary researcher's point of view. We show that this process can be divided into six descriptive categories: Data donor, research organization, research community, norms, data infrastructure, and data recipients. Drawing from our findings, we discuss theoretical implications regarding knowledge creation and dissemination as well as research policy measures to foster academic collaboration. We conclude that research data cannot be regarded as knowledge commons, but research policies that better incentivise data sharing are needed to improve the quality of research results and foster scientific progress.

Fecher, Benedikt, Sascha Friesike, Marcel Hebing, and Stephanie Linek. "A Reputation Economy: How Individual Reward Considerations Trump Systemic

Arguments for Open Access to Data." *Palgrave Communications* 3 (2017): 17051.
<http://dx.doi.org/10.1057/palcomms.2017.51>

Federer, Lisa. "The Librarian as Research Informationist: A Case Study." *Journal of the Medical Library Association* 101, no. 4 (2013): 298-302.
<https://doi.org/10.3163/1536-5050.101.4.011>

———. Research Data Management in the Age of Big Data: Roles and Opportunities for Librarians. *Information Services & Use* 36, no. 1-2 (2016):35-43.
<https://doi.org/10.3233/isu-160797>

Ferguson, Jen. "Description and Annotation of Biomedical Data Sets." *Journal of eScience Librarianship* 1, no. 1 (2012): e1000).
<http://dx.doi.org/10.7191/jeslib.2012.1000>

Ferreira, Filipe, Miguel E. Coimbra, Raquel Bairrão, Ricardo Viera, Ana T. Freitas, Luís M. S. Russo, and José Borbinha. "Data Management in Metagenomics: A Risk Management Approach." *International Journal of Digital Curation* 9, no. 1 (2014): 41-56. <https://doi.org/10.2218/ijdc.v9i1.299>

In eScience, where vast data collections are processed in scientific workflows, new risks and challenges are emerging. Those challenges are changing the eScience paradigm, mainly regarding digital preservation and scientific workflows. To address specific concerns with data management in these scenarios, the concept of the Data Management Plan was established, serving as a tool for enabling digital preservation in eScience research projects. We claim risk management can be jointly used with a Data Management Plan, so new risks and challenges can be easily tackled. Therefore, we propose an analysis process for eScience projects using a Data Management Plan and ISO 31000 in order to create a Risk Management Plan that can complement the Data Management Plan. The motivation, requirements and validation of this proposal are explored in the MetaGen-FRAME project, focused in Metagenomics.

Figueiredo, Ana Sofia. "Data Sharing: Convert Challenges into Opportunities. " *Frontiers in Public Health* 5, no. 327 (2017).
<https://doi.org/10.3389/fpubh.2017.00327>.

Initiatives for sharing research data are opportunities to increase the pace of knowledge discovery and scientific progress. The reuse of research data has the potential to avoid the duplication of data sets and to bring new views from

multiple analysis of the same data set. For example, the study of genomic variations associated with cancer profits from the universal collection of such data and helps in selecting the most appropriate therapy for a specific patient. However, data sharing poses challenges to the scientific community. These challenges are of ethical, cultural, legal, financial, or technical nature. This article reviews the impact that data sharing has in science and society and presents guidelines to improve the efficient sharing of research data.

Finney, K. "Managing Antarctic Data—A Practical Use Case." *Data Science Journal* 13 (2014): PDA8-PDA14. <https://doi.org/10.2481/dsj.ifpda-02>

Scientific data management is performed to ensure that data are curated in a manner that supports their qualified reuse. Curation usually involves actions that must be performed by those who capture or generate data and by a facility with the capability to sustainably archive and publish data beyond an individual project's lifecycle. The Australian Antarctic Data Centre is such a facility. How this centre is approaching the administration of Antarctic science data is described in the following paper and serves to demonstrate key facets necessary for undertaking polar data management in an increasingly connected global data environment.

Fitt, Alistai, Rowena Rouse, and Sarah Taylor. "RDM: An Approach from a Modern University with a Growing Research Portfolio." *Journal of Digital Media Management* 3, no. 4 (2015): 320-328. <https://doi.org/10.2218/ijdc.v10i1.353>

Florance, Patrick, Marc McGee, Christopher Barnett, Stephen McDonald. "The Open Geoportal Federation." *Journal of Map & Geography Libraries* 11, no. 3 (2015): 376-394. <http://dx.doi.org/10.1080/15420353.2015.1054543>

Fong, Bonnie L., and Minglu Wang. "Required Data Management Training for Graduate Students in an Earth and Environmental Sciences Department." *Journal of eScience Librarianship* 4, no. 1 (2015): e1067. <https://doi.org/10.7191/jeslib.2015.1067>

Fox, Robert. "The Art and Science of Data Curation." *OCLC Systems & Services: International Digital Library Perspectives* 29, no. 4 (2013): 195-199. <https://doi.org/10.1108/OCLC-07-2013-0021>

Fowler, Dan, Jo Barratt, and Paul Walsh. "Frictionless Data: Making Research Data Quality Visible." *International Journal of Digital Curation*, 12, no. 2 (2017): 274-285. <https://doi.org/10.2218/ijdc.v12i2.577>

There is significant friction in the acquisition, sharing, and reuse of research data. It is estimated that eighty percent of data analysis is invested in the cleaning and mapping of data (Dasu and Johnson, 2003). This friction hampers researchers not well versed in data preparation techniques from reusing an ever-increasing amount of data available within research data repositories. Frictionless Data is an ongoing project at Open Knowledge International focused on removing this friction. We are doing this by developing a set of tools, specifications, and best practices for describing, publishing, and validating data. The heart of this project is the "Data Package", a containerization format for data based on existing practices for publishing open source software. This paper will report on current progress toward that goal.

Frank, Rebecca D., Elizabeth Yake, and Ixchel M. Faniel.

"Destruction/Reconstruction: Preservation of Archaeological and Zoological Research Data." *Archival Science* 15, no. 2 (2015): 141-167.

<https://doi.org/10.1007/s10502-014-9238-9>

Frerichs, Wendy. "Data Management for Researchers: Organize, Maintain and Share Your Data for Research Success." *Australian Academic & Research Libraries* 47, no. 4 (2016): 325-326.

<https://doi.org/10.1080/00048623.2016.1262736>

Frey, Jeremy. "Curation of Laboratory Experimental Data as Part of the Overall Data Lifecycle." *International Journal of Digital Curation* 3, no. 1 (2008): 44-62.

<https://doi.org/10.2218/ijdc.v3i1.41>

Frey, Jeremy, Simon J. Coles, Colin Bird, and Cerys Willoughby. "Collection, Curation, Citation at Source: Publication@Source 10 Years On." *International Journal of Digital Curation* 10, no. 2 (2015): 1-11.

<https://doi.org/10.2218/ijdc.v10i2.377>

The Southampton chemical information group had its genesis in 2001, when we began an e-Science pilot project to investigate structure-property mapping, combinatorial chemistry, and the Grid. CombeChem instigated a range of activities that have since been underway for more than ten years, in many ways matching the expansion of interest in using the Web as a vehicle for collection, curation, dissemination, reuse, and exploitation of scientific data and information. Chemistry has frequently provided the exemplar case studies, notably for the series of projects—funded by Jisc and EPSRC—that

investigated the issues associated with the long-term preservation of data to support the scholarly knowledge cycle, such as the eBank UK project.

Rapid developments in Internet access and mobile technology have significantly influenced the way researchers view connectivity, data standards, and the increasing importance and power of semantics and the Semantic Web. These technical advances interact strongly with the social dimension and have led to a reconsideration of the responsibilities of researchers for the quality of their research and for satisfying the requirements of modern stakeholders. Such obligations have given rise to discussions about Open Access and Open Data, creating a range of alternatives that are now technically feasible but need to be socially acceptable. Business plans are changing too, but in a strange contradiction, desire can run ahead of what is possible, sensible, and affordable, while lagging behind in imagination of what would be technically possible and potentially game-changing!

Taking the chemical sciences as our example and focusing on the curation of research data, we explore from our perspective, ten years back and ten years forward, how far we have been able to re-imagine the data/information value pathway from bench to publication. We assess not only the major advances and changes that have been achieved, but also where we have been less successful than we might have hoped. We explore the directions for the future, based on what is clearly already possible and on what we can envisage becoming feasible in the near future.

Friddell, J., E. LeDrew, and W. Vincent. "The Polar Data Catalogue: Best Practices for Sharing and Archiving Canada's Polar Data." *Data Science Journal* 13 (2014): PDA1-PDA7. <https://doi.org/10.2481/dsj.ifpda-01>.

The Polar Data Catalogue (PDC) is a growing Canadian archive and public access portal for Arctic and Antarctic research and monitoring data. In partnership with a variety of Canadian and international multi-sector research programs, the PDC encompasses the natural, social, and health sciences. From its inception, the PDC has adopted international standards and best practices to provide a robust infrastructure for reliable security, storage, discoverability, and access to Canada's polar data and metadata. Current efforts focus on developing new partnerships and incentives for data archiving and sharing and on expanding connections to other data centres through metadata interoperability protocols.

Frisz, Chris, Geoffrey Brown, and Samuel Waggoner. "Assessing Migration Risk for Scientific Data Formats." *International Journal of Digital Curation* 7, no. 1 (2012): 27-38. <https://doi.org/10.2218/ijdc.v7i1.212>

The majority of information about science, culture, society, economy and the environment is born digital, yet the underlying technology is subject to rapid obsolescence. One solution to this obsolescence, format migration, is widely practiced and supported by many software packages, yet migration has well known risks. For example, newer formats—even where similar in function—do not generally support all of the features of their predecessors, and, where similar features exist, there may be significant differences of interpretation.

There appears to be a conflict between the wide use of migration and its known risks. In this paper we explore a simple hypothesis—that, where migration paths exist, the majority of data files can be safely migrated leaving only a few that must be handled more carefully—in the context of several scientific data formats that are or were widely used. Our approach is to gather information about potential migration mismatches and, using custom tools, evaluate a large collection of data files for the incidence of these risks. Our results support our initial hypothesis, though with some caveats. Further, we found that writing a tool to identify "risky" format features is considerably easier than writing a migration tool.

Fuhr, Justin. "How Do I Do That? A Literature Review of Research Data Management Skill Gaps of Canadian Health Sciences Information Professionals." *Journal of the Canadian Health Libraries Association* 40, no. 2 (2019): 51–60. <https://doi.org/10.29173/jchla29371>

Gabridge, Tracy. "The Last Mile: Liaison Roles in Curating Science and Engineering Research Data." *Research Library Issues: A Bimonthly Report from ARL, CNI, AND SPARC*, no. 265 (2009): 15-21. <https://doi.org/10.29242/rli.265.4>

Ganley, Emma. "PLOS Data Policy: Catalyst for a Better Research Process." *College & Research Libraries News* 75, no. 6 (2014): 305-308. <https://doi.org/10.5860/crln.75.6.9138>

Garnett, Alex, Amber Leahey, Dany Savard, Barbara Towell, and Lee Wilson. "Open Metadata for Research Data Discovery in Canada." *Journal of Library Metadata* 17, no. 3-4 (2017): 201-217. <https://doi.org/10.1080/19386389.2018.1443698>

Garrett, Leigh, Marie-Therese Gramstadt, and Carlos Silva. "Here, KAPTUR This! Identifying and Selecting the Infrastructure Required to Support the Curation and Preservation of Visual Arts Research Data." *International Journal of Digital Curation* 8, no. 2 (2013): 68-88. <https://doi.org/10.2218/ijdc.v8i2.273>

Research data is increasingly perceived as a valuable resource and, with appropriate curation and preservation, it has much to offer learning, teaching, research, knowledge transfer and consultancy activities in the visual arts. However, very little is known about the curation and preservation of this data: none of the specialist arts institutions have research data management policies or infrastructure and anecdotal evidence suggests that practice is ad hoc, left to individual researchers and teams with little support or guidance. In addition, the curation and preservation of such diverse and complex digital resources as found in the visual arts is, in itself, challenging. Led by the Visual Arts Data Service, a research centre of the University for the Creative Arts, in collaboration with the Glasgow School of Art; Goldsmiths College, University of London; and University of the Arts London, and funded by JISC, the KAPTUR project (2011-2013) seeks to address the lack of awareness and explore the potential of research data management systems in the arts by discovering the nature of research data in the visual arts, investigating the current state of research data management, developing a model of best practice applicable to both specialist arts institutions and arts departments in multidisciplinary institutions, and by applying, testing and piloting the model with the four institutional partners. Utilising the findings of the KAPTUR user requirement and technical review, this paper will outline the method and selection of an appropriate research data management system for the visual arts and the issues the team encountered along the way.

Garrett, Leigh, Marie-Therese Gramstadt, Carlos Silva, and Anne Spalding. "KAPTUR the Highlights: Exploring Research Data Management in the Visual Arts." *Ariadne*, no. 71 (2013). <http://www.ariadne.ac.uk/issue71/garrett-et-al>

Gaudette, Glenn R., and Donna Kafel. "A Case Study: Data Management in Biomedical Engineering." *Journal of eScience Librarianship* 1, no. 3 (2012): e1027. <http://dx.doi.org/10.7191/jeslib.2012.1027>

Gelernter, Judith, and Michael Lesk. "Use of Ontologies for Data Integration and Curation." *International Journal of Digital Curation* 6, no. 1 (2011). <https://doi.org/10.2218/ijdc.v6i1.173>

Geraghty, Ruth. "Curation after the Fact: Practical and Ethical Challenges of Archiving Legacy Evaluation Data." *International Journal of Digital Curation* 12, no. 1 (2017): 152-161. <https://doi.org/10.2218/ijdc.v12i1.550>

Over a 12-year period, the Atlantic Philanthropies invested more than €127m in agencies and community groups, running 52 prevention and early intervention (PEI) programmes and services in the children and youth sector throughout Ireland. As a condition of this funding, each PEI programme was evaluated by a university-based research team, resulting in a substantial collection of metric and qualitative information about ways to improve the lives of vulnerable Irish families. In 2016, the Atlantic Philanthropies funded the Prevention and Early Intervention Research Initiative at the Children's Research Network of Ireland and Northern Ireland (hereafter, the Initiative) to gather, prepare and share this evaluation data through the public data archives.

The Initiative faces several challenges in its objective to archive this extensive collection of legacy data, and this paper will present two of the more salient challenges: how to share this data so that it is both (1) meaningful and (2) ethical. The paper pays particular attention to the challenges of safely sharing evaluation data through anonymisation and restricted access conditions; and also, the practical and ethical challenges of retroactively preparing these datasets for the archive.

A series of publicly available documents that guide each stage of the Initiative are in development, and are emerging as a key output. This paper will describe two pivotal documents, namely the CRN-PEI Guiding Principles, and the CRN-PEI Protocols for preparing and archiving evaluation data. The CRN-PEI Guiding Principles outline the key legal and ethical obligations of archiving this legacy evaluation data, and act as moral compass to steer our progress through these uncharted waters. The CRN-PEI Protocols define the standards for how data included in the Initiative is prepared for deposition in the public data archives, so they are easily located, interpretable and comparable in the long term. This protocol is based upon best practice documentation from a number of international sources and our primary aim is to generate 'safe, useful data' (Elliot et al., 2016).

Getler, Magdalena, Diana Sisu, Sarah Jones, and Kerry Miller. "DMPonline Version 4.0: User-Led Innovation." *International Journal of Digital Curation* 9, no. 1 (2014): 193-219. <https://doi.org/10.2218/ijdc.v9i1.312>

DMPonline is a web-based tool to help researchers and research support staff produce data management and sharing plans. Between October and December 2012, we examined DMPonline in unprecedented detail. The results of this evaluation led to some major changes. We have shortened the DCC Checklist for a Data Management Plan and revised how this is used in the tool. We have also amended the data model for DMPonline, improved workflows and redesigned the user interface.

This paper reports on the evaluation, outlining the methods used, the results gathered and how they have been acted upon. We conducted usability testing on v.3 of DMPonline and the v.4 beta prior to release. The results from these two rounds of usability testing are compared to validate the changes made. We also put forward future plans for a more iterative development approach and greater community input.

Giarlo, Michael J. "Academic Libraries as Data Quality Hubs." *Journal of Librarianship and Scholarly Communication* 1, no. 3 (2013): eP1059.
<http://dx.doi.org/10.7710/2162-3309.1059>

Academic libraries have a critical role to play as data quality hubs on campus. There is an increased need to ensure data quality within 'e-science'. Given academic libraries' curation and preservation expertise, libraries are well suited to support the data quality process. Data quality measurements are discussed, including the fundamental elements of trust, authenticity, understandability, usability and integrity, and are applied to the Digital Curation Lifecycle model to demonstrate how these measures can be used to understand and evaluate data quality within the curatorial process. Opportunities for improvement and challenges are identified as areas that are fruitful for future research and exploration.

Giofrè, David, Geoff Cumming, Luca Fresc, Ingrid Boedker, and Patrizio Tressoldi. "The Influence of Journal Submission Guidelines On Authors' Reporting of Statistics and Use of Open Research Practices." *PLOS ONE* 12, no. 4 (2017): e0175583. <https://doi.org/10.1371/journal.pone.0175583>

From January 2014, *Psychological Science* introduced new submission guidelines that encouraged the use of effect sizes, estimation, and meta-analysis (the "new statistics"), required extra detail of methods, and offered badges for use of open science practices. We investigated the use of these practices in empirical articles published by Psychological Science and, for

comparison, by the *Journal of Experimental Psychology: General*, during the period of January 2013 to December 2015. The use of null hypothesis significance testing (NHST) was extremely high at all times and in both journals. In *Psychological Science*, the use of confidence intervals increased markedly overall, from 28% of articles in 2013 to 70% in 2015, as did the availability of open data (3 to 39%) and open materials (7 to 31%). The other journal showed smaller or much smaller changes. Our findings suggest that journal-specific submission guidelines may encourage desirable changes in authors' practices.

Goben, Abigail, and Megan R. Sapp Nelson. "The Data Engagement Opportunities Scaffold: Development and Implementation." *Journal of eScience Librarianship* 7, no. 2 (2018): e1128. <https://doi.org/10.7191/jeslib.2018.1128>

While interest in research data management (RDM) services have grown, clarifying the path between traditional library responsibilities and RDM remains a challenge. While the literature has provided ideas about services and student-/researcher-focused data information literacy (DIL) competencies, nothing has yet brought these skill sets together to provide a pathway for librarians engaging in RDM. The Data Engagement Opportunities scaffold was developed to provide a strategic trajectory relating information science skills, the DIL competencies, the stages of the data life cycle, three levels of RDM engagement activities, and potential measurable outcomes. This scaffold provides direction for librarians looking to identify their current abilities and explore new opportunities.

Goben, Abigail, and Rebecca Raszewski. "Research Data Management Self-Education for Librarians: A Webliography." *Issues in Science and Technology Librarianship*, no. 82 (2015). <http://istl.org/15-fall/internet2.html>

As data as a scholarly object continues to grow in importance in the research community, librarians are undertaking increasing responsibilities regarding data management and curation. New library initiatives include assisting researchers in finding data sets for reuse; locating and hosting repositories for required archiving; consultations on workflow, data management plans, and best practices; responding to changing funder policies (Whitmire, et al. 2015) and development of department or institutional policies. Librarians looking to provide services or expand into these areas will need both foundational resources and information about engaging the network of librarians exploring data. This webliography is intended for librarians seeking to enhance their

own knowledge and assist peers in improving their data management awareness.

Goben, Abigail, and Dorothea Salo. "Federal Research Data Requirements Set to Change." *College & Research Libraries News* 74, no. 8 (2013): 421-425. <https://doi.org/10.5860/crln.74.8.8994>

Goldman, Julie, Donna Kafel, and Elaine R. Martin. "Assessment of Data Management Services at New England Region Resource Libraries." *Journal of eScience Librarianship* 4, no. 1 (2015): e1068. <https://doi.org/10.7191/jeslib.2015.1068>

Goldstein, Justin C., Matthew S. Mayernik, and Hampapuram K. Ramapriyan. "Identifiers for Earth Science Data Sets: Where We Have Been and Where We Need to Go." *Data Science Journal* 16, no. 23 (2017). <http://doi.org/10.5334/dsj-2017-023>

Considerable attention has been devoted to the use of persistent identifiers for assets of interest to scientific and other communities alike over the last two decades. Among persistent identifiers, Digital Object Identifiers (DOIs) stand out quite prominently, with approximately 133 million DOIs assigned to various objects as of February 2017. While the assignment of DOIs to objects such as scientific publications has been in place for many years, their assignment to Earth science data sets is more recent. Applying persistent identifiers to data sets enables improved tracking of their use and reuse, facilitates the crediting of data producers, and aids reproducibility through associating research with the exact data set(s) used. Maintaining provenance—i.e., tracing back lineage of significant scientific conclusions to the entities (data sets, algorithms, instruments, satellites, etc.) that lead to the conclusions, would be prohibitive without persistent identifiers. This paper provides a brief background on the use of persistent identifiers in general within the US, and DOIs more specifically. We examine their recent use for Earth science data sets, and outline successes and some remaining challenges. Among the challenges, for example, is the ability to conveniently and consistently obtain data citation statistics using the DOIs assigned by organizations that manage data sets.

Goodison, Crystal, Alexis Guillaume Thomas, and Sam Palmer. "The Florida Geographic Data Library: Lessons Learned and Workflows for Geospatial Data

Management." *Journal of Map & Geography Libraries* 12, no. 1 (2016): 73-99. <http://dx.doi.org/10.1080/15420353.2015.1038861>

Goodman, Alyssa, Alberto Pepe, Alexander W. Blocker, Christine L. Borgman, Kyle Cranmer, Merce Crosas, Rosanne Di Stefano, Yolanda Gil, Paul Groth, Margaret Hedstrom, David W. Hogg, Vinay Kashyap, Ashish Mahabal, Aneta Siemiginowska, and Aleksandra Slavkovic. "Ten Simple Rules for the Care and Feeding of Scientific Data." *PLoS Computational Biology* 10, no. 4 (2014): e1003542. <http://dx.doi.org/10.1371/journal.pcbi.1003542>

This article offers a short guide to the steps scientists can take to ensure that their data and associated analyses continue to be of value and to be recognized. In just the past few years, hundreds of scholarly papers and reports have been written on questions of data sharing, data provenance, research reproducibility, licensing, attribution, privacy, and more—but our goal here is not to review that literature. Instead, we present a short guide intended for researchers who want to know why it is important to "care for and feed" data, with some practical advice on how to do that. The final section at the close of this work (Links to Useful Resources) offers links to the types of services referred to throughout the text.

Gordon, Andrew S., David S. Millman, Lisa Steiger, Karen E. Adolph, and Rick O. Gilmore. "Researcher-Library Collaborations: Data Repositories as a Service for Researchers." *Journal of Librarianship and Scholarly Communication* 3, no. 2 (2015): eP1238. <http://doi.org/10.7710/2162-3309.1238>

INTRODUCTION New interest has arisen in organizing, preserving, and sharing the raw materials—the data and metadata—that undergird the published products of research. Library and information scientists have valuable expertise to bring to bear in the effort to create larger, more diverse, and more widely used data repositories. However, for libraries to be maximally successful in providing the research data management and preservation services required of a successful data repository, librarians must work closely with researchers and learn about their data management workflows.

DESCRIPTION OF SERVICES Databrary is a data repository that is closely linked to the needs of a specific scholarly community—researchers who use video as a main source of data to study child development and learning. The project's success to date is a result of its focus on community outreach and providing services for scholarly communication, engaging institutional partners, offering services for data curation with the guidance of closely

involved information professionals, and the creation of a strong technical infrastructure. NEXT STEPS Databrary plans to improve its curation tools that allow researchers to deposit their own data, enhance the user-facing feature set, increase integration with library systems, and implement strategies for long-term sustainability.

Grabus, Sam, and Jane Greenberg. "The Landscape of Rights and Licensing Initiatives for Data Sharing." *Data Science Journal*, 18, no. 1 (2019): 29. <http://doi.org/10.5334/dsj-2019-029>

Over the last twenty years, a wide variety of resources have been developed to address the rights and licensing problems inherent with contemporary data sharing practices. The landscape of developments in this area is increasingly confusing and difficult to navigate, due to the complexity of intellectual property and ethics issues associated with sharing sensitive data. This paper seeks to address this challenge, examining the landscape and presenting a Version 1.0 directory of resources. A multi-method study was pursued, with an environmental scan examining 20 resources, resulting in three high-level categories: standards, tools, and community initiatives; and a content analysis revealing the subcategories of rights, licensing, metadata & ontologies. A timeline confirms a shift in licensing standardization priorities from open data to more nuanced and technologically robust solutions, over time, to accommodate for more sensitive data types. This paper reports on the research undertaking, and comments on the potential for using license-specific metadata supplements and developing data-centric rights and licensing ontologies.

Grant, Rebecca. "Identifying HSS Research Data for Preservation: A Snapshot of Current Policy and Guidelines." *New Review of Information Networking* 20, no. 1-2 (2015): 97-103. <http://dx.doi.org/10.1080/13614576.2015.1110400>

Grant, Rebecca, and Iain Hrynaszkiewicz. "The Impact on Authors and Editors of Introducing Data Availability Statements at Nature Journals." *International Journal of Digital Curation* 13, no. 1 (2018): 195-203. <https://doi.org/10.2218/ijdc.v13i1.614>

This article describes the adoption of a standard policy for the inclusion of data availability statements in all research articles published at the Nature family of journals, and the subsequent research which assessed the impacts that these policies had on authors, editors, and the availability of datasets. The

key findings of this research project include the determination of average and median times required to add a data availability statement to an article; and a correlation between the way researchers make their data available, and the time required to add a data availability statement.

Green, Katie, Kieron Niven, and Georgina Field. "Migrating 2 and 3D Datasets: Preserving AutoCAD at the Archaeology Data Service." *ISPRS International Journal of Geo-Information* 5, no. 4 (2016): 44. <http://dx.doi.org/10.3390/ijgi5040044>

Greenberg, Jane. "Big Metadata, Smart Metadata, and Metadata Capital: Toward Greater Synergy Between Data Science and Metadata." *Journal of Data and Information Science* 2, no. 3 (2017): 19-36. <https://doi.org/10.1515/jdis-2017-0012>

———. "Metadata for Scientific Data: Historical Considerations, Current Practice, and Prospects." *Journal of Library Metadata* 10, no. 2/3 (2010): 75-78. <https://doi.org/10.1080/19386389.2010.520262>

Greenberg, Jane, Hollie C. White, Sarah Carrier, and Ryan Scherle. "A Metadata Best Practice for a Scientific Data Repository." *Journal of Library Metadata* 9, no. 3/4 (2009): 194-212. <https://doi.org/10.1080/19386380903405090>

Greene, Gretchen, Raymond Plante, and Robert Hanisch. "Building Open Access to Research (OAR) Data Infrastructure at NIST." *Data Science Journal*, 18, no. 1 (2019): 30. <http://doi.org/10.5334/dsj-2019-030>.

As a National Metrology Institute (NMI), the USA National Institute of Standards and Technology (NIST) scientists, engineers and technology experts conduct research across a full spectrum of physical science domains. NIST is a non-regulatory agency within the U.S. Department of Commerce with a mission to promote U.S. innovation and industrial competitiveness by advancing measurement science, standards, and technology in ways that enhance economic security and improve our quality of life. NIST research results in the production and distribution of standard reference materials, calibration services, and datasets. These are generated from a wide range of complex laboratory instrumentation, expert analyses, and calibration processes. In response to a government open data policy, and in collaboration with the broader research community, NIST has developed a federated Open Access to Research (OAR) scientific data infrastructure aligned with FAIR (Findable, Accessible, Interoperable, Reusable) data principles. Through the OAR initiatives, NIST's Material Measurement Laboratory Office of Data and

Informatics (ODI) recently released a new scientific data discovery portal and public data repository. These science-oriented applications provide dissemination and public access for data from across the broad spectrum of NIST research disciplines, including chemistry, biology, materials science (such as crystallography, nanomaterials, etc.), physics, disaster resilience, cyberinfrastructure, communications, forensics, and others. NIST's public data consist of carefully curated Standard Reference Data, legacy high valued data, and new research data publications. The repository is thus evolving both in content and features as the nature of research progresses. Implementation of the OAR infrastructure is key to NIST's role in sharing high integrity reproducible research for measurement science in a rapidly changing world.

Griffin, Philippa C., Jyoti Khadake, Kate S. LeMay, Suzanna E. Lewis, Sandra Orchard, Andrew Pask, Bernard Pope, Ute Roessner, Keith Russell, Torsten Seemann, Andrew Treloar, Sonika Tyagi, Jeffrey H. Christiansen, Saravanan Dayalan, Simon Gladman, Sandra B. Hangartner, Helen L. Hayden, William W. H. Ho, Gabriel Keeble-Gagnère, Pasi K. Korhonen, Peter Neish, Priscilla R. Prestes, Mark F. Richardson, Nathan S. Watson-Haigh, Kelly L. Wyres, Neil D. Young, and Maria Victoria Schneider. "Best Practice Data Life Cycle Approaches for the Life Sciences." *F1000Research* 6 (2017): 1618-1618.
<https://doi.org/10.12688/f1000research.12344.2>

Throughout history, the life sciences have been revolutionised by technological advances; in our era this is manifested by advances in instrumentation for data generation, and consequently researchers now routinely handle large amounts of heterogeneous data in digital formats. The simultaneous transitions towards biology as a data science and towards a 'life cycle' view of research data pose new challenges. Researchers face a bewildering landscape of data management requirements, recommendations and regulations, without necessarily being able to access data management training or possessing a clear understanding of practical approaches that can assist in data management in their particular research domain. Here we provide an overview of best practice data life cycle approaches for researchers in the life sciences/bioinformatics space with a particular focus on 'omics' datasets and computer-based data processing and analysis. We discuss the different stages of the data life cycle and provide practical suggestions for useful tools and resources to improve data management practices.

Griffiths, Aaron. "The Publication of Research Data: Researcher Attitudes and Behaviour." *International Journal of Digital Curation* 4, no. 1 (2009): 46-56. <https://doi.org/10.2218/ijdc.v4i1.77>

Groenewegen, David, and Andrew Treloar. "Adding Value by Taking a National and Institutional Approach to Research Data: The ANDS Experience." *International Journal of Digital Curation* 8, no. 2 (2013): 89-98. <https://doi.org/10.2218/ijdc.v8i2.274>

The Australian National Data Service (ANDS) has been working to add value to Australia's research data environment since 2009. This paper looks at the changes that have occurred over this time, ANDS' role in those changes and the current state of the Australian research sector at this time, using case studies of selected institutions.

Grootveld, Marjan, and Jeff van Egmond. "Peer-Reviewed Open Research Data: Results of a Pilot." *International Journal of Digital Curation* 7, no. 2 (2012): 81-91. <https://doi.org/10.2218/ijdc.v7i2.231>

Peer review of publications is at the core of science and primarily seen as instrument for ensuring research quality. However, it is less common to independently value the quality of the underlying data as well. In the light of the 'data deluge' it makes sense to extend peer review to the data itself and this way evaluate the degree to which the data are fit for re-use. This paper describes a pilot study at EASY—the electronic archive for (open) research data at our institution. In EASY, researchers can archive their data and add metadata themselves. Devoted to open access and data sharing, at the archive we are interested in further enriching these metadata with peer reviews.

As a pilot, we established a workflow where researchers who have downloaded data sets from the archive were asked to review the downloaded data set. This paper describes the details of the pilot including the findings, both quantitative and qualitative. Finally, we discuss issues that need to be solved when such a pilot is turned into a structural peer review functionality for the archiving system.

Grynock, Tess. "Implementing Research Data Management Services in a Canadian Context." *Dalhousie Journal of Interdisciplinary Management* 12, no. 1 (2016). <https://ojs.library.dal.ca/djim/article/view/6458>

Gunjal, Bhojaraju, and Panorea Gaitano." A Proposed Framework to Boost Research in Higher Educational Institutes." *IASSIST Quarterly* 41 no. 1-4 (2017): 12. <https://doi.org/10.29173/iq12>

Gutmann, M., K. Schürer, D. Donakowski, and Hilary Beedham. "The Selection, Appraisal, and Retention of Social Science Data." *Data Science Journal* 3 (2004): 209-221. <http://doi.org/10.2481/dsj.3.209>

The number of data collections produced in the social sciences prohibits the archiving of every scientific study. It is therefore necessary to make decisions regarding what can be preserved and why it should be preserved. This paper reviews the processes used by two data archives, one from the United States and one from the United Kingdom, to illustrate how data are selected for archiving, how they are appraised, and what steps are required to retain the usefulness of the data for future use. It also presents new initiatives that seek to encourage an increase in the long-term preservation of digital resources.

Gutmann, Myron P., Mark Abrahamson, Margaret O. Adams, Micah Altman, Caroline Arms, Kenneth Bollen, Michael Carlson, Jonathan Crabtree, Darrell Donakowski, Gary King, Jared Lyle, Marc Maynard, Amy Pienta, Richard Rockwell, Lois Timms-Ferrara, and Copeland H. Young. "From Preserving the Past to Preserving the Future: The Data-PASS Project and the Challenges of Preserving Digital Social Science Data." *Library Trends* 57, no. 3 (2009): 315-337. <https://doi.org/10.1353/lib.0.0039>

Guy, Marieke, Martin Donnelly, and Laura Molloy. "Pinning It Down: Towards a Practical Definition of 'Research Data' for Creative Arts Institutions." *International Journal of Digital Curation* 8, no. 2 (2013): 99-110. <https://doi.org/10.2218/ijdc.v8i2.275>

There is a widespread understanding among scientific researchers about what is meant by 'research data'; however this does not readily translate into a creative context. As part of its engagement with the University of the Arts London (UAL) and via its support for the JISC Managing Research Data Programme, the Digital Curation Centre (DCC) and partners have worked towards an acceptable and practical definition of research data for creative arts institutions. This paper describes the activities carried out to help pin down such a definition, including a literature review, short and extended interviews with researchers, interactions with an academic arts research

practitioner, and distillation of the results from a one-day workshop which took place in London in September 2012.

Hadley, Martin John, and Howard Noble. "Promoting Interactive Visualisation at University of Oxford: The Live Data Network." *International Journal of Digital Curation* 11, no. 1 (2016): 172-182. <https://doi.org/10.2218/ijdc.v11i1.418>

This article introduces the Live Data project funded by the Research IT Board of the University of Oxford's IT Services department. The primary aim of the project is to support academics in creating interactive visualisations using a variety of cloud-based visualisation services, which the academic can freely embed within academic journals, blogs and personal websites through the use of iframes. To achieve this the project has been funded from October 2015 to March 2017 to recruit visualisation case studies from across the University and to develop software agnostic workflows for the creation of interactive visualisations.

Within this report we present interactive visualisations as a vital component of the academic's toolkit for engaging potential collaborators and the general public with their research data—thereby bridging the so-called 'data gap' between data, publication and researcher.

Halbert, Martin. "The Problematic Future of Research Data Management: Challenges, Opportunities and Emerging Patterns Identified by the DataRes Project." *International Journal of Digital Curation* 8, no. 2 (2013): 111-122. <https://doi.org/10.2218/ijdc.v8i2.276>

This paper describes findings and projections from a project that has examined emerging policies and practices in the United States regarding the long-term institutional management of research data. The DataRes project at the University of North Texas (UNT) studied institutional transitions taking place during 2011-2012 in response to new mandates from U.S. governmental funding agencies requiring research data management plans to be submitted with grant proposals. Additional synergistic findings from another UNT project, termed iCAMP, will also be reported briefly.

This paper will build on these data analysis activities to discuss conclusions and prospects for likely developments within coming years based on the trends surfaced in this work. Several of these conclusions and prospects are surprising, representing both opportunities and troubling challenges, for not only the library profession but the academic research community as a whole.

Hansen, Nadja Hansen, Kruse, Filip, and Thestrup, Jesper Boserup. "Managing Data in Cross-Institutional Projects." *IASSIST Quarterly* 43, no. 3 (2019): 1–10. <https://doi.org/10.29173/iq950>

Hanson, Karen L., Theodora A. Bakker, Mario A. Svirsky, Arlene C. Neuman, and Neil Rambo. "Informationist Role: Clinical Data Management in Auditory Research." *Journal of eScience Librarianship* 2, no. 1 (2013): e1030. <http://dx.doi.org/10.7191/jeslib.2013.1030>

Hardwicke, Tom E., Maya B. Mathur, Kyle MacDonald, Gustav Nilsson, George C. Banks, Mallory C. Kidwell, Alicia Hofelich Mohr, Elizabeth Clayton, Erica J. Yoon, Michael Henry Tessler, Richie L. Lenne, Sara Altman, Bria Long, and Michael C. Frank. "Data Availability, Reusability, and Analytic Reproducibility: Evaluating the Impact of a Mandatory Open Data Policy at the Journal Cognition." *Royal Society Open Science* 5 (2018): 180448180448. <http://doi.org/10.1098/rsos.180448>

Access to data is a critical feature of an efficient, progressive and ultimately self-correcting scientific ecosystem. But the extent to which in-principle benefits of data sharing are realized in practice is unclear. Crucially, it is largely unknown whether published findings can be reproduced by repeating reported analyses upon shared data ('analytic reproducibility'). To investigate this, we conducted an observational evaluation of a mandatory open data policy introduced at the journal Cognition. Interrupted time-series analyses indicated a substantial post-policy increase in data available statements (104/417, 25% pre-policy to 136/174, 78% post-policy), although not all data appeared reusable (23/104, 22% pre-policy to 85/136, 62%, post-policy). For 35 of the articles determined to have reusable data, we attempted to reproduce 1324 target values. Ultimately, 64 values could not be reproduced within a 10% margin of error. For 22 articles all target values were reproduced, but 11 of these required author assistance. For 13 articles at least one value could not be reproduced despite author assistance. Importantly, there were no clear indications that original conclusions were seriously impacted. Mandatory open data policies can increase the frequency and quality of data sharing. However, suboptimal data curation, unclear analysis specification and reporting errors can impede analytic reproducibility, undermining the utility of data sharing and the credibility of scientific findings.

Harp, Matthew R., and Matt Ogborn. "Collaborating Externally and Training Internally to Support Research Data Services." *Journal of eScience Librarianship* 8, no. 2 (2019): e1165. <https://doi.org/10.7191/jeslib.2019.1165>.

The ASU Library is actively building relationships around and increasing its expertise in research data services. We have established a collaboration with our university's research administration in order to coordinate our distinct areas of expertise in research data services so that both entities can better support researchers all the way through the research data lifecycle. The Library embedded itself into research administration's learning management system and works with their research advancement officers to engage with researchers and staff we have not traditionally reached. Forging this new collaboration increased expectations that the Library will expand existing research data services to more investigators, so we have grown Library professionals' internal competencies by providing research data management training opportunities to meet these demands. In addition, the Library's Research Services Working Group established data competencies, workflows, and trainings so more librarians gain skills necessary to answer and assist patrons with data needs. Greater expertise throughout the Library enables us to authentically and confidently scale our research data services and form new collaborations.

Harris-Pierce, Rebecca L., and Yan Quan Liu. "Is Data Curation Education at Library and Information Science Schools in North America Adequate?" *New Library World* 113, no. 11/12, (2012): 598-613. <http://dx.doi.org/10.1108/03074801211282957>

Harvey, Matthew J., Andrew McLean, and Henry S. Rzepa. "A Metadata-Driven Approach to Data Repository Design." *Journal of Cheminformatics* 9, no. 1 (2017): 4. <https://doi.org/10.1186/s13321-017-0190-6>

He, Lin, and Vinita Nahar. "Reuse of Scientific Data in Academic Publications." *Aslib Journal of Information Management* 68, no. 4 (2016): 478-494. <https://doi.org/10.1108/ajim-01-2016-0008>

He, Lin, and Han Zhengbiao. "Do Usage Counts of Scientific Data Make Sense? An Investigation of the Dryad Repository." *Library Hi Tech* 35, no. 2 (2017): 332-342. <https://doi.org/10.1108/LHT-12-2016-0158>

Hedges, Mark, Tobias Blanke, Stella Fabiane, Gareth Knight, and Eric Liao. "Sheer Curation of Experiments: Data, Process, Provenance." *Journal of Digital*

Information 13, no. 1 (2012).

<http://journals.tdl.org/jodi/index.php/jodi/article/view/5883>

Hedges, Mark, Tobias Blanke, and Adil Hasan. "Rule-Based Curation and Preservation of Data: A Data Grid Approach using iRODS." *Future Generation Computer Systems* 25, no. 4 (2009): 446-452.

<https://doi.org/10.1016/j.future.2008.10.003>

Hedges, Mark, Mike Haft, and Gareth Knight. "FISHNet: Encouraging Data Sharing and Reuse in the Freshwater Science Community " *Journal of Digital Information* 13, no. 1 (2012).

<http://journals.tdl.org/jodi/index.php/jodi/article/view/5884>

Heidorn, P. Bryan. "The Emerging Role of Libraries in Data Curation and E-Science." *Journal of Library Administration* 51, no. 7/8 (2011): 662-672.

<https://doi.org/10.1080/01930826.2011.601269>

——— "Shedding Light on the Dark Data in the Long Tail of Science." *Library Trends* 57, no. 2 (2008): 280-29. <https://doi.org/10.1353/lib.0.0036>

Helbig, Kerstin. "Research Data Management Training for Geographers: First Impressions." *ISPRS International Journal of Geo-Information* 5, no. 4 (2016): 40.

<http://dx.doi.org/10.3390/ijgi5040040>

Helbig, Kerstin, Brigitte Hausstein, and Ralf Toepfer. "Supporting Data Citation: Experiences and Best Practices of a DOI Allocation Agency for Social Sciences." *Journal of Librarianship and Scholarly Communication* 3, no. 2 (2015): eP1220.

<http://doi.org/10.7710/2162-3309.1220>

INTRODUCTION As more and more research data becomes better and more easily available, data citation gains in importance. The management of research data has been high on the agenda in academia for more than five years. Nevertheless, not all data policies include data citation, and problems like versioning and granularity remain. **SERVICE DESCRIPTION** da|ra operates as an allocation agency for DataCite and offers the registration service for social and economic research data in Germany. The service is jointly run by GESIS and ZBW, thereby merging experiences on the fields of Social Sciences and Economics. The authors answer questions pertaining to the most frequent aspects of research data registration like versioning and granularity as well as recommend the use of persistent identifiers linked with enriched metadata at the landing page. **NEXT STEPS** The promotion of data

sharing and the development of a citation culture among the scientific community are future challenges. Interoperability becomes increasingly important for publishers and infrastructure providers. The already existent heterogeneity of services demands solutions for better user guidance. Building information competence is an asset of libraries, which can and should be expanded to research data.

Henderson, Margaret. *Data Management: A Practical Guide for Librarians*. Lanham: Rowman & Littlefield, 2017. <http://www.worldcat.org/oclc/962081716>

Henderson, Margaret, Yasmeen Shorish, and Steve Van Tuyl. "Research Data Management on a Shoestring Budget." *Bulletin of the American Society for Information Science and Technology* 40, no. 6 (2014): 14-17. <https://doi.org/10.1002/bult.2014.1720400606>

Henderson, Margaret E., and Teresa L. Knott. "Starting a Research Data Management Program Based in a University Library." *Medical Reference Services Quarterly* 34, no. 1 (2015): 387-403. <https://dx.doi.org/10.1080%2F02763869.2015.986783>

Hendren, Christine Ogilvie, Christina M. Powers, Mark D, Hoover, and Stacey L. Harper. "The Nanomaterial Data Curation Initiative: A Collaborative Approach to Assessing, Evaluating, and Advancing the State of the Field." *Beilstein Journal of Nanotechnology* 6 (2015): 1752-1762. <http://dx.doi.org/10.3762/bjnano.6.179>

The Nanomaterial Data Curation Initiative (NDCI), a project of the National Cancer Informatics Program Nanotechnology Working Group (NCIP NanoWG), explores the critical aspect of data curation within the development of informatics approaches to understanding nanomaterial behavior. Data repositories and tools for integrating and interrogating complex nanomaterial datasets are gaining widespread interest, with multiple projects now appearing in the US and the EU. Even in these early stages of development, a single common aspect shared across all nanoinformatics resources is that data must be curated into them. Through exploration of sub-topics related to all activities necessary to enable, execute, and improve the curation process, the NDCI will provide a substantive analysis of nanomaterial data curation itself, as well as a platform for multiple other important discussions to advance the field of nanoinformatics. This article outlines the NDCI project and lays the foundation for a series of papers on nanomaterial data curation. The NDCI purpose is to: 1) present and evaluate

the current state of nanomaterial data curation across the field on multiple specific data curation topics, 2) propose ways to leverage and advance progress for both individual efforts and the nanomaterial data community as a whole, and 3) provide opportunities for similar publication series on the details of the interactive needs and workflows of data customers, data creators, and data analysts. Initial responses from stakeholder liaisons throughout the nanoinformatics community reveal a shared view that it will be critical to focus on integration of datasets with specific orientation toward the purposes for which the individual resources were created, as well as the purpose for integrating multiple resources. Early acknowledgement and undertaking of complex topics such as uncertainty, reproducibility, and interoperability is proposed as an important path to addressing key challenges within the nanomaterial community, such as reducing collateral negative impacts and decreasing the time from development to market for this new class of technologies.

This work is licensed under a Creative Commons Attribution 2.0 Generic Licensee, <https://creativecommons.org/licenses/by/2.0/>.

Hense, Andreas, and Florian Quadt. "Acquiring High Quality Research Data." *D-Lib Magazine* 17, no. 1/2 (2011).
<http://www.dlib.org/dlib/january11/hense/01hense.html>

Herold, Philip. "Data Sharing Among Ecology, Evolution, and Natural Resources Scientists: An Analysis of Selected Publications." *Journal of Librarianship and Scholarly Communication* 3, no. 2 (2015): eP1244. <http://doi.org/10.7710/2162-3309.1244>

INTRODUCTION Understanding the differing data management practices among academic disciplines is an important way to inform existing and emerging library research support and services. This paper reports findings from a study of data sharing practices among ecology, evolution, and natural resources scientists at the University of Minnesota. It examines data sharing rates, methods, and disciplinary differences and discusses the characteristics of researchers, data, methods, and aspects of data sharing across this group of disciplines. **METHODS** Data sharing practices are investigated by reviewing the two most recently published research articles (n=155) for each faculty member (n=78) in three departments at a single large research university. All mentions of data sharing in each publication were pursued in order to locate, analyze, and characterize shared data. **RESULTS** Seventy-two of 155 (46%)

articles indicated that related research data was publicly shared by some method. The most prevalent method for data sharing was via journal websites, with 91% of data sharing articles using this method. Ecology, evolution, and behavior scientists shared data at the highest rate (70% of their articles), contrasting with fisheries, wildlife, and conservation biologists (18%), and forest resources (16%). **DISCUSSION** Differences between data sharing practices may be attributable to a range of influences: funder, journal, and institutional policies; disciplinary norms; and perceived or real rewards or incentives, as well as contrasting concerns, cost, or other barriers to sharing data. **CONCLUSION** Study results suggest differential approaches to data services outreach based on discipline and research type and support the need for education and influence on both scientist and journal practices.

Herterich, Patricia, and Sünje Dallmeier-Tiessen. "Data Citation Services in the High-Energy Physics Community." *D-Lib Magazine* 22, no. 1/2 (2016).
<http://www.dlib.org/dlib/january16/herterich/01herterich.html>

Hickson, Susan, Kylie Ann Poulton, Maria Connor, Joanna Richardson, and Malcolm Wolski. "Modifying Researchers' Data Management Practices: A Behavioural Framework for Library Practitioners." *IFLA Journal* 42, no. 4 (2016): 253–265. <https://doi.org/10.1177/0340035216673856>

Higman, Rosie, Daniel Bangert, and Sarah Jones. "Three Camps, One Destination: The Intersections of Research Data Management, FAIR and Open." *Insights* 32, no. 1 (2019): 18. <http://doi.org/10.1629/uksg.468>

Open data, FAIR (findable, accessible, interoperable and reusable) and research data management (RDM) are three overlapping but distinct concepts, each emphasizing different aspects of handling and sharing research data. They have different strengths in terms of informing and influencing how research data is treated, and there is much scope for enrichment of data if they are applied collectively. This paper explores the boundaries of each concept and where they intersect and overlap. As well as providing greater definitional clarity, this will help researchers to manage and share their data, and those supporting researchers, such as librarians and data stewards, to understand how these concepts can best be used in an advocacy setting. FAIR and open both focus on data sharing, ensuring content is made available in ways that promote access and reuse. Data management by contrast is about the stewardship of data from the point of conception onwards. It makes no assumptions about access, but is essential if data are to be meaningful to

others. The concepts of FAIR and open are more noble aspirations and are, this paper argues, a useful way to engage researchers and encourage good data practices from the outset.

Higman, Rosie, and Stephen Pinfield. "Research Data Management and Openness: The Role of Data Sharing in Developing Institutional Policies and Practices." *Program* 49, no. 4 (2015): 364-381. <https://doi.org/10.1108/prog-01-2015-0005>

Hiom, Debra, Dom Fripp, Stephen Gray, Kellie Snow, and Damian Steer. "Research Data Management at the University of Bristol: Charting a Course from Project to Service." *Program* 49, no. 4 (2015): 475-493. <http://dx.doi.org/10.1108/PROG-02-2015-0019>

Hiom, Debra, Stephen Gray, Damian Steer, Kirsty Merrett, Kellie Snow, and Zosia Beckles. "Introducing Safe Access to Sensitive Data at the University of Bristol." *International Journal of Digital Curation* 12, no. 2 (2017): 26-36. <https://doi.org/10.2218/ijdc.v12i2.506>

The economic and societal benefits of making research data available for reuse and verification are now widely understood and accepted. However, there are some research studies, particularly those involving human participants, which face particular challenges in making their data openly available due to the sensitivities of the data. Despite its potential value to society this material is invariably kept locked away due to concerns over its inappropriate disclosure. The University of Bristol's Research Data Service has developed the institutional infrastructure, including policies and procedures, required to safely grant access to sensitive research data in a way that is transparent, secure, sustainable and crucially, replicable by other institutions.

This paper looks at the background and challenges faced by the institution in dealing with sensitive data, outlines the approach taken and some of the outstanding issues to be tackled.

This paper looks at the background and challenges faced by the institution in dealing with sensitive data, outlines the approach taken and some of the outstanding issues to be tackled.

Horstmann, Wolfram, and Jan Brase. "Libraries and Data—Paradigm Shifts and Challenges." *Bibliothek Forschung und Praxis* 40, no. 2 (2016): 273–277. <https://doi.org/10.1515/bfp-2016-0034>

Hou, Chung-Yi, and Matthew Mayernik. "Formalizing an Attribution Framework for Scientific Data/Software Products and Collections." *International Journal of Digital Curation* 11, No. 2 (2016): 87-103. <https://doi.org/10.2218/ijdc.v11i2.404>

As scientific research and development become more collaborative, the diversity of skills and expertise involved in producing scientific data are expanding as well. Since recognition of contribution has significant academic and professional impact for participants in scientific projects, it is important to integrate attribution and acknowledgement of scientific contributions into the research and data lifecycle. However, defining and clarifying contributions and the relationship of specific individuals and organizations can be challenging, especially when balancing the needs and interests of diverse partners. Designing an implementation method for attributing scientific contributions within complex projects that can allow ease of use and integration with existing documentation formats is another crucial consideration.

To provide a versatile mechanism for organizing, documenting, and storing contributions to different types of scientific projects and their related products, an attribution and acknowledgement matrix and XML schema have been created as part of the Attribution and Acknowledgement Content Framework (AACF). Leveraging the taxonomies of contribution roles and types that have been developed and published previously, the authors consolidated 16 contribution types that could be considered and used when accrediting team member's contributions. Using these contribution types, specific information regarding the contributing organizations and individuals can be documented using the AACF.

This paper provides the background and motivations for creating the current version of the AACF Matrix and Schema, followed by demonstrations of the process and the results of using the Matrix and the Schema to record the contribution information of different sample datasets. The paper concludes by highlighting the key feedback and features to be examined in order to improve the next revisions of the Matrix and the Schema.

———. "Recognizing the Diversity of Contributions: A Case Study for Framing Attribution and Acknowledgement for Scientific Data." *International Journal of Digital Curation* 11, no. 1 (2016): 33-52. <https://doi.org/10.2218/ijdc.v11i1.357>

As scientific data volumes, format types, and sources increase rapidly with the invention and improvement of scientific capabilities, the resulting datasets are becoming more complex to manage as well. One of the significant management challenges is pulling apart the individual contributions of specific people and organizations within large, complex projects. This is important for two aspects: 1) assigning responsibility and accountability for scientific work, and 2) giving professional credit to individuals (e.g. hiring, promotion, and tenure) who work within such large projects. This paper aims to review the extant practice of data attribution and how it may be improved. Through a case study of creating a detailed attribution record for a climate model dataset, the paper evaluates the strengths and weaknesses of the current data attribution method and proposes an alternative attribution framework accordingly. The paper concludes by demonstrating that, analogous to acknowledging the different roles and responsibilities shown in movie credits, the methodology developed in the study could be used in general to identify and map out the relationships among the organizations and individuals who had contributed to a dataset. As a result, the framework could be applied to create data attribution for other dataset types beyond climate model datasets.

Hou, Chung-Yi, Heather Soyka, Vivian Hutchison, Isis Sema, Chris Allen, and Amber Budden. "Evaluating the Effectiveness of Data Management Training: DataONE's Survey Instrument." *International Journal of Digital Curation* 12, no. 2 (2017): 47-60. <https://doi.org/10.2218/ijdc.v12i2.508>

Effective management is a key component for preparing data to be retained for future long term access, use, and reuse by a broader community. Developing the skills to plan and perform data management tasks is important for individuals and institutions. Teaching data literacy skills may also help to mitigate the impact of data deluge and other effects of being overexposed to and overwhelmed by data.

The process of learning how to manage data effectively for the entire research data lifecycle can be complex. There are often multiple stages involved within a lifecycle for managing data, and each stage may require specific knowledge, expertise, and resources. Additionally, although a range of organizations offers data management education and training resources, it can often be difficult to assess how effective the resources are for educating users to meet their data management requirements.

In the case of Data Observation Network for Earth (DataONE), DataONE's extensive collaboration with individuals and organizations has informed the development of multiple educational resources. Through these interactions, DataONE understands that the process of creating and maintaining educational materials that remain responsive to community needs is reliant on careful evaluations. Therefore, the impetus for a comprehensive, customizable Education EVAluation instrument (EEVA) is grounded in the need for tools to assess and improve current and future training and educational resources for research data management.

In this paper, the authors outline and provide context for the background and motivations that led to creating EEVA for evaluating the effectiveness of data management educational resources. The paper details the process and results of the current version of EEVA. Finally, the paper highlights the key features, potential uses, and the next steps in order to improve future extensions and revisions of EEVA.

Hrynaszkiewicz, Iain, Aliaksandr Birukou, Mathias Astell, Sowmya Swaminathan, Amye Kenall, and Varsha Khodiyar. "Standardising and Harmonising Research Data Policy in Scholarly Publishing." *International Journal of Digital Curation* 12, no. 1 (2017): 65-71. <https://doi.org/10.2218/ijdc.v12i1.531>

To address the complexities researchers face during publication, and the potential community-wide benefits of wider adoption of clear data policies, the publisher Springer Nature has developed a standardised, common framework for the research data policies of all its journals. An expert working group was convened to audit and identify common features of research data policies of the journals published by Springer Nature, where policies were present. The group then consulted with approximately 30 editors, covering all research disciplines within the organisation. The group also consulted with academic editors, librarians and funders, which informed development of the framework and the creation of supporting resources. Four types of data policy were defined in recognition that some journals and research communities are more ready than others to adopt strong data policies. As of January 2017 more than 700 journals have adopted a standard policy and this number is growing weekly. To potentially enable standardisation and harmonisation of data policy across funders, institutions, repositories, societies and other publishers, the policy framework was made available under a Creative Commons license. However, the framework requires wider debate with these stakeholders and an

Interest Group within the Research Data Alliance (RDA) has been formed to initiate this process.

Hswe, Patricia, and Ann Holt. "Joining in the Enterprise of Response in the Wake of the NSF Data Management Planning Requirement." *Research Library Issues*, no. 274 (2011): 11-17. <https://doi.org/10.29242/rli.274.2>

Huang, Hong, Corinne Jörgensen, Besiki Stvilia. "Genomics Data Curation Roles, Skills and Perception of Data Quality." *Library & Information Science Research* 37, no. 1 (2015): 10-20. <http://dx.doi.org/10.1016/j.lisr.2014.08.003>

Hudson-Vitale, Cynthia, Heidi Imker, Jake Carlson, Lisa R. Johnston, Robert Olendorf, Wendy Kozlowski, and Claire Stewart. *SPEC Kit 354: Data Curation*. Washington, DC: ARL, 2017. <https://doi.org/10.29242/spec.354>

Humphrey, Chuck, Kathleen Shearer, and Martha Whitehead. "Towards a Collaborative National Research Data Management Network." *International Journal of Digital Curation* 11, no. 1 (2016): 195-207. <https://doi.org/10.2218/ijdc.v11i1.411>

This paper describes the plans and strategies to develop Portage, a national network of sustainable, shared services for research data management (RDM) in Canada. A description of the RDM context in Canada is provided. This environment has heightened expectations around the Government of Canada's Open Science plans and includes deliverables aimed at improving access to publications and data resulting from federally funded scientific activities. At the same time, a recent environmental scan published by Canada's three federal research granting councils reveals significant gaps in services, infrastructure, and funding mechanisms to support RDM. In addition, Canada's RDM environment consists of stakeholders from a variety of communities with minimal ongoing coordination or cooperation.

The Portage network was conceived as a collaborative network model based on libraries' strong connections with researchers across the disciplines, an ethos of curation and preservation, and experience with systems for managing data in all its forms. A pilot project provided Portage with a vision and set of principles, and identified several objectives as the small wins that would build the trust and shared understanding required for a successful network. Current services and activities of Portage, including a data management planning tool and an infrastructure project, are described in this paper.

Portage now faces the challenge of moving from project to operational network, and the challenge of establishing a sustainable governance model. CARL appointed a Steering Committee that will be proposing a full governance model at the conclusion of this transition period. Using a framework of factors identified in the literature, several relevant collaborative and network governance models are being explored.

This paper outlines experience to date with Portage and matters under consideration for long-term sustainability, with a goal of engaging international colleagues in discussion and furthering the concepts for the benefit of RDM networks everywhere.

Hunter, Jane. "Scientific Publication Packages—A Selective Approach to the Communication and Archival of Scientific Output." *International Journal of Digital Curation* 1, no. 1 (2006): 33-52. <https://doi.org/10.2218/ijdc.v1i1.4>

The use of digital technologies within research has led to a proliferation of data, many new forms of research output and new modes of presentation and analysis. Many scientific communities are struggling with the challenge of how to manage the terabytes of data and new forms of output, they are producing. They are also under increasing pressure from funding organizations to publish their raw data, in addition to their traditional publications, in open archives. In this paper I describe an approach that involves the selective encapsulation of raw data, derived products, algorithms, software and textual publications within "scientific publication packages". Such packages provide an ideal method for: encapsulating expert knowledge; for publishing and sharing scientific process and results; for teaching complex scientific concepts; and for the selective archival, curation and preservation of scientific data and output. They also provide a bridge between technological advances in the Digital Libraries and eScience domains. In particular, I describe the RDF-based architecture that we are adopting to enable scientists to construct, publish and manage "scientific publication packages"—compound digital objects that encapsulate and relate the raw data to its derived products, publications and the associated contextual, provenance and administrative metadata.

Ikeshoji-Orlati, Veronica A., and Clifford B. Anderson. "Developing Data Curation Protocols for Digital Projects at Vanderbilt: Une Micro-Histoire." *International Journal of Digital Curation*, 12, no. 2 (2017): 246-254. <https://doi.org/10.2218/ijdc.v12i2.574>

This paper examines the intersection of legacy digital humanities projects and the ongoing development of research data management services at Vanderbilt University's Jean and Alexander Heard Library. Future directions for data management and curation protocols are explored through the lens of a case study: the (re)curation of data from an early 2000s e-edition of Raymond Poggenburg's *Charles Baudelaire: Une Micro-histoire*. The vagaries of applying the Library of Congress Metadata Object Description Schema (MODS) to the data and metadata of the *Micro-histoire* will be addressed. In addition, the balance between curating data and metadata for preservation vs. curating it for (re)use by future researchers is considered in order to suggest future avenues for holistic research data management services at Vanderbilt.

Ishida, Mayu. "The New England Collaborative Data Management Curriculum Pilot at the University of Manitoba: A Canadian Experience." *Journal of eScience Librarianship* 4, no. 2 (2015): e1061. <https://doi.org/10.7191/jeslib.2014.1061>

Canada's federal funding agencies are following the directions of funding agencies in the United States and United Kingdom, and will soon require a data management plan in grant applications. The University of Manitoba Libraries in Canada has started planning and implementing research data services, and education is seen as a key component. In June 2014, the New England Collaborative Data Management Curriculum (NECDMC) (Lamar Soutter Library, University of Massachusetts Medical School 2014) was piloted and used to provide data management training for a group of subject librarians at the University of Manitoba Libraries, in combination with information about data-related policies of the Canadian funding agencies and the University of Manitoba. The seven NECDMC modules were delivered in a seminar style, with emphasis on group discussions and Canadian content. The benefits of NECDMC—adaptability and flexible framework—should be weighed against the challenges experienced in the pilot, mainly the significant amount of time needed to create local content and complement the existing curriculum. Overall, the pilot showed that NECDMC is a good, thorough introduction to data management, and that it is possible to adapt NECDMC to the local and Canadian settings in an effective way.

Jacobs, Clifford A., and Steven J. Worley. "Data Curation in Climate and Weather: Transforming Our Ability to Improve Predictions through Global Knowledge Sharing." *International Journal of Digital Curation* 4, no. 2 (2009): 68-79. <https://doi.org/10.2218/ijdc.v4i2.94>

Jahnke, Lori, Andrew Asher, and Spencer D. C. Keralis. *The Problem of Data*. Washington, DC: Council on Library and Information Resources, 2012.
<http://www.clir.org/pubs/reports/pub154>

Jeng, Wei, He Daqing, and Chi Yu. "Social Science Data Repositories in Data Deluge: A Case Study of ICPSR's Workflow and Practices." *The Electronic Library* 35, no. 4 (2017): 626-649. <https://doi.org/10.1108/EL-11-2016-0243>

Jeng, Wei, and Liz Lyon. "A Report of Data-Intensive Capability, Institutional Support, and Data Management Practices in Social Sciences." *International Journal of Digital Curation* 11, no. 1 (2016): 156-171.
<https://doi.org/10.2218/ijdc.v11i1.398>

We report on a case study which examines the social science community's capability and institutional support for data management. Fourteen researchers were invited for an in-depth qualitative survey between June 2014 and October 2015. We modify and adopt the Community Capability Model Framework (CCMF) profile tool to ask these scholars to self-assess their current data practices and whether their academic environment provides enough supportive infrastructure for data related activities. The exemplar disciplines in this report include anthropology, political sciences, and library and information science.

Our findings deepen our understanding of social disciplines and identify capabilities that are well developed and those that are poorly developed. The participants reported that their institutions have made relatively slow progress on economic supports and data science training courses, but acknowledged that they are well informed and trained for participants' privacy protection. The result confirms a prior observation from previous literature that social scientists are concerned with ethical perspectives but lack technical training and support. The results also demonstrate intra- and inter-disciplinary commonalities and differences in researcher perceptions of data-intensive capability, and highlight potential opportunities for the development and delivery of new and impactful research data management support services to social sciences researchers and faculty.

John, Kaye, Bruce Rachel, and Fripp Dom. "Establishing a Shared Research Data Service for UK Universities." *Insights: The UKSG Journal* 30, no. 1 (2017): 59-70.
<https://doi.org/10.1629/uksg.346>

Johnson, Andrew. "The DataQ Story." *Bulletin of the Association for Information Science and Technology* 42, no. 5 (2016): 38-40.
<https://doi.org/10.1002/bul2.2016.1720420509>

Johnson, Andrew M., and Shelley Knuth. "Data Management Plan Requirements for Campus Grant Competitions: Opportunities for Research Data Services Assessment and Outreach." *Journal of eScience Librarianship* 5, no. 1 (2016): e1089. <https://doi.org/10.7191/jeslib.2016.1089>

Objective: To examine the effects of research data services (RDS) on the quality of data management plans (DMPs) required for a campus-level faculty grant competition, as well as to explore opportunities that the local DMP requirement presented for RDS outreach.

Methods: Nine reviewers each scored a randomly assigned portion of DMPs from 82 competition proposals. Each DMP was scored by three reviewers, and the three scores were averaged together to obtain the final score. Interrater reliability was measured using intraclass correlation. Unpaired t-tests were used to compare mean DMP scores for faculty who utilized RDS services with those who did not. Unpaired t-tests were also used to compare mean DMP scores for proposals that were funded with proposals that were not funded. One-way ANOVA was used to compare mean DMP scores among proposals from six broad disciplinary categories.

Results: Analyses showed that RDS consultations had a statistically significant effect on DMP scores. Differences between DMP scores for funded versus unfunded proposals and among disciplinary categories were not significant. The DMP requirement also provided a number of both expected and unexpected outreach opportunities for RDS services.

Conclusions: Requiring DMPs for campus grant competitions can provide important assessment and outreach opportunities for research data services. While these results might not be generalizable to DMP review processes at federal funding agencies, they do suggest the importance, at any level, of developing a shared understanding of what constitutes a high quality DMP among grant applicants, grant reviewers, and RDS providers.

Johnson, Andrew W., and Megan M. Bresnahan. "DataDay!: Designing and Assessing a Research Data Workshop for Subject Librarians." *Journal of Librarianship and Scholarly Communication* 3, no. 2 (2015): eP1229.
<http://doi.org/10.7710/2162-3309.1229>

BACKGROUND Many libraries have launched or adapted services to address the research data needs of campus faculty and students. At the University of Colorado Boulder (CU-Boulder), local demand for research data training emerged from a broader assessment of training needs for subject librarians. The findings from this assessment led to the development of a day-long workshop called DataDay! that aimed to expand and translate the skills of subject librarians into the context of research data support. **DESCRIPTION OF PROGRAM** The DataDay! workshop incorporated hands-on exercises with expert presentations, informal discussions, and print handouts. The workshop allowed participants to gain experience with activities like working with real data sets and developing materials for outreach about research data services. Several instruments were used to assess the workshop learning outcomes, which included changes in knowledge and comfort levels related to engaging in research data support. Assessment activities also measured how well participants applied concepts taught in the workshop to novel situations. **NEXT STEPS** Future research data training efforts for CU-Boulder librarians will be informed by the DataDay! workshop assessment results, and this workshop may provide a model for other institutions to use to train subject librarians to adapt to new roles in support of research data. There is also a need for the lessons learned from local training efforts like DataDay! to inform the development of resources to support the broader subject librarian community as their institutions launch and grow research data services.

Johnston, Lisa R., ed. *Curating Research Data*. Chicago: Association of College and Research Libraries, 2017. <http://www.worldcat.org/oclc/971359114>

Johnston, Lisa R., Jake R. Carlson, Patricia Hswe, Cynthia Hudson-Vitale, Heidi Imker, Wendy Kozlowski, Robert K. Olendorf, and Claire Stewart,. "Data Curation Network: How Do We Compare? A Snapshot of Six Academic Library Institutions' Data Repository and Curation Services." *Journal of eScience Librarianship* 6, no. 1 (2017): e1102. <https://doi.org/10.7191/jeslib.2017.1102>

Johnston, Lisa R., Jake Carlson, Cynthia Hudson-Vitale, Heidi Imker, Wendy Kozlowski, Robert Olendorf, Claire Stewart, Mara Blake, Joel Herndon, Timothy M. McGeary, and Elizabeth Hull. "Data Curation Network: A Cross-Institutional Staffing Model for Curating Research Data." *International Journal of Digital Curation* 13, no. 1 (2018):125-140. <https://doi.org/10.2218/ijdc.v13i1.616>.

Funders increasingly require that data sets arising from sponsored research must be preserved and shared, and many publishers either require or

encourage that data sets accompanying articles are made available through a publicly accessible repository. Additionally, many researchers wish to make their data available regardless of funder requirements both to enhance their impact and also to propel the concept of open science. However, the data curation activities that support these preservation and sharing activities are costly, requiring advanced curation practices, training, specific technical competencies, and relevant subject expertise. Few colleges or universities will be able to hire and sustain all of the data curation expertise locally that its researchers will require, and even those with the means to do more will benefit from a collective approach that will allow them to supplement at peak times, access specialized capacity when infrequently-curated types arise, and stabilize service levels to account for local staff transition, such as during turn-over periods. The Data Curation Network (DCN) provides a solution for partners of all sizes to develop or to supplement local curation expertise with the expertise of a resilient, distributed network, and creates a funding stream to both sustain central services and support expansion of distributed expertise over time. This paper presents our next steps for piloting the DCN, scheduled to launch in the spring of 2018 across nine partner institutions. Our implementation plan is based on planning phase research performed from 2016-2017 that monitored the types, disciplines, frequency, and curation needs of data sets passing through the curation services at the six planning phase institutions. Our DCN implementation plan includes a well-coordinated and tiered staffing model, a technology-agnostic submission workflow, standardized curation procedures, and a sustainability approach that will allow the DCN to prevail beyond the grant-supported implementation phase as a curation-as-service model.

Johnston, Lisa R., Meghan Lafferty, and Beth Petsan. "Training Researchers on Data Management: A Scalable, Cross-Disciplinary Approach." *Journal of eScience Librarianship* 1, no. 2 (2012): e1012. <http://dx.doi.org/10.7191/jeslib.2012.1012>

Johnston, Wayne. "Digital Preservation Initiatives in Ontario: Trusted Digital Repositories and Research Data Repositories." *Partnership: The Canadian Journal of Library and Information Practice and Research* 7, no. 2 (2012). <https://doi.org/10.21083/partnership.v7i2.2014>

Joint, Nicholas. "Data Preservation, the New Science and the Practitioner Librarian." *Library Review* 56, no. 6 (2007): 451-455. <https://doi.org/10.1108/00242530710760337>

Jones, Sarah. "Developments in Research Funder Data Policy." *International Journal of Digital Curation* 7, no. 1 (2012): 114-125.
<https://doi.org/10.2218/ijdc.v7i1.219>

This paper reviews developments in funders' data management and sharing policies, and explores the extent to which they have affected practice. The Digital Curation Centre has been monitoring UK research funders' data policies since 2008. There have been significant developments in subsequent years, most notably the joint Research Councils UK's Common Principles on Data Policy and the Engineering and Physical Sciences Research Council's Policy Framework on Research Data. This paper charts these changes and highlights shifting emphasises in the policies. Institutional data policies and infrastructure are increasingly being developed as a result of these changes. While action is clearly being taken, questions remain about whether the changes are affecting practice on the ground.

Jones, Sarah, Alexander Ball, and Çuna Ekmekcioglu. "The Data Audit Framework: A First Step in the Data Management Challenge." *International Journal of Digital Curation* 3, no. 2 (2008): 112-120.
<https://doi.org/10.2218/ijdc.v3i2.62>

Joo, Soohyung, Kim Sujin, and Youngseek Kim. "An Exploratory Study of Health Scientists' Data Reuse Behaviors: Examining Attitudinal, Social, and Resource Factors." *Aslib Journal of Information Management* 69, no. 4 (2017): 389-407.
<https://doi.org/10.1108/AJIM-12-2016-0201>

Joo, Yeon Kyoung, and Youngseek Kim. "Engineering Researchers' Data Reuse Behaviours: a Structural Equation Modelling Approach." *The Electronic Library* 35, no. 6 (2017): 1141-1161. <https://doi.org/10.1108/el-08-2016-0163>

Jünger, Stefan, Kerrin Borschewski, and Wolfgang Zenk-Möltgen. "Documenting Georeferenced Social Science Survey Data: Limits of Metadata Standards and Possible Solutions." *Journal of Map and Geography Libraries*, 15, no. 1 (2019): 68-95. <https://doi.org/10.1080/15420353.2019.1659903>

Kaboré, Esther Dzalé Yeumo. "Opening and Linking Agricultural Research Data." *D-Lib Magazine* 20, no. 1/2 (2014).
<http://www.dlib.org/dlib/january14/kabore/01kabore.html>

Kafel, Donna, Andrew T. Creamer, and Elaine R. Martin. "Building the New England Collaborative Data Management Curriculum." *Journal of eScience Librarianship* 3, no. 1 (2014): e1066. <http://dx.doi.org/10.7191/jeslib.2014.1066>

Kanao, M., M. Okada, J. Friddell, and A. Kadokura. "Science Metadata Management, Interoperability and Data Citations of the National Institute of Polar Research, Japan." *Data Science Journal* 17, no. 1.(2018). <http://doi.org/10.5334/dsj-2018-001>

The Polar Data Centre (PDC) of the National Institute of Polar Research (NIPR) has a responsibility to manage polar science data as part of the National Antarctic Data Centre and the Science Committee on Antarctic Research. During the International Polar Year (IPY 2007-2008), a remarkable number of data/metadata involving multi-disciplinary science activities were compiled. Although the long-term stewardship of the accumulation of metadata falls to the data center of NIPR, the work has been in collaboration with the Global Change Master Directory, the Polar Information Commons, the World Data System and other data science bodies/communities under the International Council for Science. In addition, links with other data centers, such as the Data Integration and Analysis System Program of the Global Earth Observation System of Systems and the Polar Data Catalogue of Canada were initiated in 2014 using the Open Archives Initiative Protocol for Metadata Harvesting. The metadata compiled by the PDC were recently modified using an automatic attributing system and DataCite through the Japan Link Center.

Kansa, Eric C., Sarah Whitcher Kansa, and Benjamin Arbuckle. "Publishing and Pushing: Mixing Models for Communicating Research Data in Archaeology." *International Journal of Digital Curation* 9, no. 1 (2014): 57-70. <https://doi.org/10.2218/ijdc.v9i1.301>

We present a case study of data integration and reuse involving 12 researchers who published datasets in Open Context, an online data publishing platform, as part of collaborative archaeological research on early domesticated animals in Anatolia. Our discussion reports on how different editorial and collaborative review processes improved data documentation and quality, and created ontology annotations needed for comparative analyses by domain specialists. To prepare data for shared analysis, this project adapted editor-supervised review and revision processes familiar to conventional publishing, as well as more novel models of revision adapted from open source software

development of public version control. Preparing the datasets for publication and analysis required significant investment of effort and expertise, including archaeological domain knowledge and familiarity with key ontologies. To organize this work effectively, we emphasized these different models of collaboration at various stages of this data publication and analysis project. Collaboration first centered on data editors working with data contributors, then widened to include other researchers who provided additional peer-review feedback, and finally the widest research community, whose collaboration is facilitated by GitHub's version control system. We demonstrate that the "publish" and "push" models of data dissemination need not be mutually exclusive; on the contrary, they can play complementary roles in sharing high quality data in support of research. This work highlights the value of combining multiple models in different stages of data dissemination.

Kaplan, Diane E. "The Stanley Milgram Papers: A Case Study on Appraisal of and Access to Confidential Data Files." *American Archivist* 59, no. 3 (2009): 288-297. <https://doi.org/10.17723/aarc.59.3.k3245057x1902078>

Karasti, Helena, and Karen S. Baker. "Digital Data Practices and the Long Term Ecological Research Program Growing Global." *International Journal of Digital Curation* 3, no. 2 (2008): 42-58. <https://doi.org/10.2218/ijdc.v3i2.57>

Karasti, Helena, Karen S. Baker, and Eija Halkola. "Enriching the Notion of Data Curation in E-science: Data Managing and Information Infrastructuring in the Long Term Ecological Research (LTER) Network." *Computer Supported Cooperative Work* 15, no. 4 (2006): 321-358. <https://doi.org/10.1007/s10606-006-9023-2>

Karcher, Sebastian, Dessislava Kirilova, and Nicholas Weber. "Beyond the Matrix: Repository Services for Qualitative Data." *IFLA Journal* 42, no. 4 (2016): 292–302. <https://doi.org/10.1177/0340035216672870>

Keil, Deborah E. "Research Data Needs from Academic Libraries: The Perspective of a Faculty Researcher." *Journal of Library Administration* 54, no. 3 (2014): 233-240. <https://doi.org/10.1080/01930826.2014.915168>

Kellam, Lynda, and Kristi Thompson, eds. *Databrarianship: The Academic Data Librarian in Theory and Practice* Chicago: ALA, 2016. <http://www.worldcat.org/oclc/1026783184>

Kennan, Mary Anne, Sheila Corral, and Waseem Afzal. "'Making Space' in Practice and Education: Research Support Services in Academic Libraries." *Library Management* 35, no. 8/9 (2014): 666-683. <https://doi.org/10.1108/lm-03-2014-0037>

Keenan, Teressa, and Wendy Walker. "Considerations and Challenges for Describing Historical Research Data: A Case Study." *Journal of Library Metadata* 17, no. 3-4 (2017): 241-252. <https://doi.org/10.1080/19386389.2018.1440919>

Kenyon, Jeremy, Bruce Godfrey, and Gail Z. Eckwright. "Geospatial Data Curation at the University of Idaho." *Journal of Web Librarianship* 6, no. 4 (2012): 251-262. <https://doi.org/10.1080/19322909.2012.729983>

Kenyon, Jeremy, Nancy Sprague, and Edward Flathers. "The Journal Article as a Means to Share Data: a Content Analysis of Supplementary Materials from Two Disciplines." *Journal of Librarianship and Scholarly Communication* 4 (2016): p.eP2112. <http://doi.org/10.7710/2162-3309.2112>

INTRODUCTION The practice of publishing supplementary materials with journal articles is becoming increasingly prevalent across the sciences. We sought to understand better the content of these materials by investigating the differences between the supplementary materials published by authors in the geosciences and plant sciences. **METHODS** We conducted a random stratified sampling of four articles from each of 30 journals published in 2013. In total, we examined 297 supplementary data files for a range of different factors. **RESULTS** We identified many similarities between the practices of authors in the two fields, including the formats used (Word documents, Excel spreadsheets, PDFs) and the small size of the files. There were differences identified in the content of the supplementary materials: the geology materials contained more maps and machine-readable data; the plant science materials included much more tabular data and multimedia content. **DISCUSSION** Our results suggest that the data shared through supplementary files in these fields may not lend itself to reuse. Code and related scripts are not often shared, nor is much 'raw' data. Instead, the files often contain summary data, modified for human reading and use. **CONCLUSION** Given these and other differences, our results suggest implications for publishers, librarians, and authors, and may require shifts in behavior if effective data sharing is to be realized.

Kerby, Erin E. "Research Data Practices in Veterinary Medicine: A Case Study." *Journal of eScience Librarianship* 4, no. 1 (2015): e1073.
<https://doi.org/10.7191/jeslib.2015.1073>

Kervin, Karina E., William K. Michener, and Robert B. Cook. "Common Errors in Ecological Data Sharing." *Journal of eScience Librarianship* 2, no. 2 (2013): e1024. <http://dx.doi.org/10.7191/jeslib.2013.1024>

Ketchum, Andrea M. "The Research Life Cycle and the Health Sciences Librarian: Responding to Change in Scholarly Communication." *Journal of the Medical Library Association* 105, no. 1 (2017): 80-83.
<https://doi.org/10.5195/jmla.2017.110>

Kethers, Stefanie, Andrew Treloar, and Mingfang Wu. "Building Tools to Facilitate Data Reuse." *International Journal of Digital Curation* 11, no. 2 (2017): 1-12. <https://doi.org/10.2218/ijdc.v11i2.409>

The Australian National Data Service (ANDS) has been funded by the Australian Government since 2009, with a goal to increase the value of data to researchers, research institutions and the nation. To achieve this goal, ANDS has funded more than 200 projects under seven programs. This paper provides an overview of one of these programs, the Applications Program, which focused on funding software infrastructure to enable data reuse to demonstrate the value of making data available to researchers. The paper also presents some representative projects, a summary of what the program has achieved, and lessons learned.

Khan, Huda, Brian Caruso, Jon Corson-Rikert, Dianne Dietrich, Brian Lowe, and Gail Steinhart. "DataStaR: Using the Semantic Web approach for Data Curation." *International Journal of Digital Curation* 62, no. 2 (2011).
<https://doi.org/10.2218/ijdc.v6i2.197>

In disciplines as varied as medicine, social sciences, and economics, data and their analyses are essential parts of researchers' contributions to their respective fields. While sharing research data for review and analysis presents new opportunities for furthering research, capturing these data in digital forms and providing the digital infrastructure for sharing data and metadata pose several challenges. This paper reviews the motivations behind and design of the Data Staging Repository (DataStaR) platform that targets specific portions of the research data curation lifecycle: data and metadata capture and sharing prior to publication, and publication to permanent archival repositories. The

goal of DataStaR is to support both the sharing and publishing of data while at the same time enabling metadata creation without imposing additional overheads for researchers and librarians. Furthermore, DataStaR is intended to provide cross-disciplinary support by being able to integrate different domain-specific metadata schemas according to researchers' needs. DataStaR's strategy of a usable interface coupled with metadata flexibility allows for a more scalable solution for data sharing, publication, and metadata reuse.

Khayat, Mohammad, and Steven J. Kempler. "Life Cycle Management Considerations of Remotely Sensed Geospatial Data and Documentation for Long Term Preservation." *Journal of Map & Geography Libraries: Advances in Geospatial Information, Collections & Archives* 11, no. 3 (2015): 271-288. <https://doi.org/10.1080/15420353.2015.1072122>

Kim, Jeonghyun. "Data Sharing and Its Implications for Academic Libraries." *New Library World* 114, no. 11/12 (2013): 494-506. <https://doi.org/10.1108/nlw-06-2013-0051>

Kim, Jihyun. "Data Sharing from the Perspective of Faculty in Korea." *Libri* 67, no. 3 (2017): 179-192. <https://doi.org/10.1515/libri-2016-0116>

Kim, Suntae, and Wongoo Lee. "Global Data Repository Status and Analysis: Based on Korea, China and Japan." *Library Hi Tech* 32, no. 4 (2014): 706-722. <https://doi.org/10.1108/lht-06-2014-0064>

Kim, Youngseek. "Fostering Scientists' Data Sharing Behaviors via Data Repositories, Journal Supplements, and Personal Communication Methods." *Information Processing & Management* 53, no. 4 (2017): 871-885. <https://doi.org/10.1016/j.ipm.2017.03.003>

Kim, Youngseek, Benjamin K. Addom, and Jeffrey M. Stanton. "Education for eScience Professionals: Integrating Data Curation and Cyberinfrastructure." *International Journal of Digital Curation* 6, no. 1 (2011): 125-138. <https://doi.org/10.2218/ijdc.v6i1.177>

Kim, Youngseek, and Melissa Adler. "Social Scientists' Data Sharing Behaviors: Investigating the Roles of Individual Motivations, Institutional Pressures, and Data Repositories." *International Journal of Information Management* 35, no. 4 (August 2015): 408-418. <https://doi.org/10.1016/j.ijinfomgt.2015.04.007>

Kim, Youngseek, and C. Sean Burns. "Norms of Data Sharing in Biological Sciences: The Roles of Metadata, Data Repository, and Journal and Funding Requirements." *Journal of Information Science* 42, no. 2 (July 9, 2015): 230-245. <https://doi.org/10.1177/0165551515592098>

Kim, Youngseek, and Sujin Kim. "Institutional, Motivational, and Resource Factors Influencing Health Scientists' Data-Sharing Behaviours." *Journal of Scholarly Publishing* 46, no. 4 (2015): 366-389. <https://doi.org/10.3138/jsp.46.4.05>

Kim, Youngseek, and Nah Seungahn. "Internet Researchers' Data Sharing Behaviors: An Integration of Data Reuse Experience, Attitudinal Beliefs, Social Norms, and Resource Factors." *Online Information Review* 42, no. 1 (2017): 124-142. <https://doi.org/10.1108/OIR-10-2016-0313>

Kim, Youngseek, and Jeffrey M. Stanton. "Institutional and Individual Factors Affecting Scientists' Data Sharing Behaviors: A Multilevel Analysis." *Journal of the Association for Information Science and Technology* 67 (2016): 776-799. <https://doi.org/10.1002/asi.23424>

———. "Institutional and Individual Influences on Scientists' Data Sharing Practices." *The Journal of Computational Science Education* 3, no. 1 (2012): 47-56. <https://doi.org/10.22369/issn.2153-4136/3/1/6>

Kim, Youngseek, and Ayoung Yoon. "Scientists' Data Reuse Behaviors: A Multilevel Analysis." *Journal of the Association for Information Science and Technology* 68, no. 12 (2017): 2709-2719. <https://doi.org/10.1002/asi.23892>

Kim, Youngseek, and Ping Zhang. "Understanding Data Sharing Behaviors of STEM Researchers: The Roles of Attitudes, Norms, and Data Repositories." *Library & Information Science Research* 37, no. 3 (2015): 189-200. <https://doi.org/10.1016/j.lisr.2015.04.006>

Kindling, Maxi, Heinz Pampel, Stephanie van de Sandt, Jessika Rücknagel, Paul Vierkant, Gabriele Kloska, Michael Witt, Peter Schirmbacher, Roland Bertelmann, and Frank Scholze. "The Landscape of Research Data Repositories in 2015: A re3data Analysis." *D-Lib Magazine* 23, no. 3/4 (2017). <https://doi.org/10.1045/march2017-kindling>

Kirilova, Dessi, and Sebastian Karcher. "Rethinking Data Sharing and Human Participant Protection in Social Science Research: Applications from the

Qualitative Realm." *Data Science Journal* 16 (2017). <https://doi.org/10.5334/dsj-2017-043>

While data sharing is becoming increasingly common in quantitative social inquiry, qualitative data are rarely shared. One factor inhibiting data sharing is a concern about human participant protections and privacy. Protecting the confidentiality and safety of research participants is a concern for both quantitative and qualitative researchers, but it raises specific concerns within the epistemic context of qualitative research. Thus, the applicability of emerging protection models from the quantitative realm must be carefully evaluated for application to the qualitative realm. At the same time, qualitative scholars already employ a variety of strategies for human-participant protection implicitly or informally during the research process. In this practice paper, we assess available strategies for protecting human participants and how they can be deployed. We describe a spectrum of possible data management options, such as de-identification and applying access controls, including some already employed by the Qualitative Data Repository (QDR) in tandem with its pilot depositors. Throughout the discussion, we consider the tension between modifying data or restricting access to them, and retaining their analytic value. We argue that developing explicit guidelines for sharing qualitative data generated through interaction with humans will allow scholars to address privacy concerns and increase the secondary use of their data.

Kitchin, John R., Ana E. Van Gulick, and Lisa D. Zilinski. "Automating Data Sharing through Authoring Tools." *International Journal on Digital Libraries* 18, no. 2 (2016): 93-98. <https://doi.org/10.1007/s00799-016-0173-7>

Klump, Jens, Roland Bertelmann, Jan Brase, Michael Diepenbroek, Hannes Grobe, Heinke Höck, Michael Lautenschlager, Uwe Schindler, Irina Sens, and Joachim Wächter. "Data Publication in the Open Access Initiative." *Data Science Journal*. 5 (2006): 79-83. <http://doi.org/10.2481/dsj.5.79>

The 'Berlin Declaration' was published in 2003 as a guideline to policy makers to promote the Internet as a functional instrument for a global scientific knowledge base. Because knowledge is derived from data, the principles of the 'Berlin Declaration' should apply to data as well. Today, access to scientific data is hampered by structural deficits in the publication process. Data publication needs to offer authors an incentive to publish data through long-term repositories. Data publication also requires an adequate

licence model that protects the intellectual property rights of the author while allowing further use of the data by the scientific community.

Knight, Gareth. "Building a Research Data Management Service for the London School of Hygiene & Tropical Medicine." *Program* 49, no. 4 (2015): 424-439. <http://dx.doi.org/10.1108/PROG-01-2015-0011>

———. "A Digital Curate's Egg: A Risk Management Approach to Enhancing Data Management Practices." *Journal of Web Librarianship* 6, no. 4 (2012): 228-250. <https://doi.org/10.1080/19322909.2012.729992>

Knight, Gareth, and Maureen Pennock. "Data without Meaning: Establishing the Significant Properties of Digital Research." *International Journal of Digital Curation* 4, no. 1 (2009): 159-174. <https://doi.org/10.2218/ijdc.v4i1.86>

Knuth, Shelley L., Andrew M. Johnson, and Thomas Hauser. "Research Data Services at the University of Colorado Boulder." *Bulletin of the Association for Information Science and Technology* 41, no. 6 (2015): 35-38. <https://doi.org/10.1002/bult.2015.1720410614>

Knuth, Shelley L., Andrew Johnson, Thea Lindquist, Debra Weiss, Deborah Hamrick, Thomas Hauser, and Leslie Reynolds. "The Center for Research Data and Digital Scholarship at the University of Colorado-Boulder." *Bulletin of the Association for Information Science and Technology* 43, no. 2 (2017): 46-48. <https://doi.org/10.1002/bul2.2017.1720430215>

Kolb, Tracy L., E. Agnes Blukacz-Richards, Andrew M. Muir, Randall M. Claramunt, Marten A. Koops, William W. Taylor, Trent M. Sutton, Michael T. Arts, and Ed Bissel. "How to Manage Data to Enhance Their Potential for Synthesis, Preservation, Sharing, and Reuse—A Great Lakes Case Study." *Fisheries* 38, no. 2 (2013): 52-64. <https://doi.org/10.1080/03632415.2013.757975>

Koltay, Tibor. "Data Governance, Data Literacy and the Management of Data Quality." *IFLA Journal* 42, no. 4 (2016): 303–312. <https://doi.org/10.1177/0340035216672238>

———. "Data Literacy for Researchers and Data Librarians." *Journal of Librarianship and Information Science* 49, no. 1 (2016): 3-14. <https://doi.org/10.1177/0961000615616450>

———. "Data Literacy: In Search of a Name and Identity." *Journal of Documentation* 71, no. 2 (2015): 401-415. <http://dx.doi.org/10.1108/JD-02-2014-0026>

———. "Research 2.0 and Research Data Services in Academic and Research Libraries: Priority Issues." *Library Management* 38, no. 6/7 (2017): 345-353. <https://doi.org/10.1108/LM-11-2016-0082>

Kong, Nicole Ningning. "Exploring Best Management Practices for Geospatial Data in Academic Libraries." *Journal of Map & Geography Libraries* 11, no. 2 (2015): 207-225. <http://dx.doi.org/10.1080/15420353.2015.1043170>

Konkiel, Stacy. "Tracking Citations and Altmetrics for Research Data: Challenges and Opportunities." *Bulletin of the American Society for Information Science and Technology* 39, no. 6 (2013): 27-32. <https://doi.org/10.1002/bult.2013.1720390610>

Koshoffer, Amy, Amy E. Neeser, Linda Newman, and Lisa R. Johnston. "Giving Datasets Context: A Comparison Study of Institutional Repositories That Apply Varying Degrees of Curation." *International Journal of Digital Curation* 13, no. 1 (2018): 15-34. <https://doi.org/10.2218/ijdc.v13i1.632>

This research study compared four academic libraries' approaches to curating the metadata of dataset submissions in their institutional repositories and classified them in one of four categories: no curation, pre-ingest curation, selective curation, and post-ingest curation. The goal is to understand the impact that curation may have on the quality of user-submitted metadata. The findings were 1) the metadata elements varied greatly between institutions, 2) repositories with more options for authors to contribute metadata did not result in more metadata contributed, 3) pre- or post-ingest curation process could have a measurable impact on the metadata but are difficult to separate from other factors, and 4) datasets submitted to a repository with pre- or post-ingest curation more often included documentation.

Kouper, Inna. "CLIR/DLF Digital Curation Postdoctoral Fellowship—The Hybrid Role of Data Curator." *Bulletin of the American Society for Information Science and Technology* 39, no. 2 (2013): 46-47. <https://doi.org/10.1002/bult.2013.1720390213>

Kowalczyk, Stacy, and Kalpana Shankar. "Data Sharing in the Sciences." *Annual Review of Information Science and Technology* 45, no. 1 (2011): 247-294. <https://doi.org/10.1002/aris.2011.1440450113>

Kowalczyk, Stacy T. "Modelling the Research Data Lifecycle." *International Journal of Digital Curation* 12, no. 2 (2017): 331-361.
<https://doi.org/10.2218/ijdc.v12i2.429>

This paper develops and tests a lifecycle model for the preservation of research data by investigating the research practices of scientists. This research is based on a mixed-method approach. An initial study was conducted using case study analytical techniques; insights from these case studies were combined with grounded theory in order to develop a novel model of the Digital Research Data Lifecycle. A broad-based quantitative survey was then constructed to test and extend the components of the model. The major contribution of these research initiatives are the creation of the Digital Research Data Lifecycle, a data lifecycle that provides a generalized model of the research process to better describe and explain both the antecedents and barriers to preservation. The antecedents and barriers to preservation are data management, contextual metadata, file formats, and preservation technologies. The availability of data management support and preservation technologies, the ability to create and manage contextual metadata, and the choices of file formats all significantly effect the preservability of research data.

Kozlowski, Wendy. "Funding Agency Responses to Federal Requirements for Public Access to Research Results." *Bulletin of the American Society for Information Science and Technology* 40, no. 6 (2014): 26-30.
<https://doi.org/10.1002/bult.2014.1720400609>

Kraft, Angelina, Matthias Razum, Jan Potthoff, Andrea Porzel, Thomas Engel, Frank Lange, Karina van den Broek, and Filipe Furtado. "The RADAR Project—A Service for Research Data Archival and Publication." *ISPRS International Journal of Geo-Information* 5, no. 3 (2016): 28. <http://dx.doi.org/10.3390/ijgi5030028>

Kratz, John, and Carly Strasser. "Data Publication Consensus and Controversies." *F1000Research*, no. 3 (2014): 94. <https://doi.org/10.12688/f1000research.3979.3>

The movement to bring datasets into the scholarly record as first class research products (validated, preserved, cited, and credited) has been inching forward for some time, but now the pace is quickening. As data publication venues proliferate, significant debate continues over formats, processes, and terminology. Here, we present an overview of data publication initiatives underway and the current conversation, highlighting points of consensus and

issues still in contention. Data publication implementations differ in a variety of factors, including the kind of documentation, the location of the documentation relative to the data, and how the data is validated. Publishers may present data as supplemental material to a journal article, with a descriptive "data paper," or independently. Complicating the situation, different initiatives and communities use the same terms to refer to distinct but overlapping concepts. For instance, the term published means that the data is publicly available and citable to virtually everyone, but it may or may not imply that the data has been peer-reviewed. In turn, what is meant by data peer review is far from defined; standards and processes encompass the full range employed in reviewing the literature, plus some novel variations. Basic data citation is a point of consensus, but the general agreement on the core elements of a dataset citation frays if the data is dynamic or part of a larger set. Even as data publication is being defined, some are looking past publication to other metaphors, notably "data as software," for solutions to the more stubborn problems.

Krier, Laura, and Carly A. Strasser. *Data Management for Libraries: A LITA Guide*. Chicago: ALA, 2014. <http://www.worldcat.org/oclc/941189062>

Kriesberg, Adam, Kerry Huller, Ricardo Punzalan, and Cynthia Parr. "An Analysis of Federal Policy on Public Access to Scientific Research Data." *Data Science Journal* 16, no. 27 (2017). <http://doi.org/10.5334/dsj-2017-027>

The 2013 Office of Science and Technology Policy (OSTP) Memo on federally-funded research directed agencies with research and development budgets above \$100 million to develop and release plans to increase and broaden access to research results, both published literature and data. The agency responses have generated discussion and interest but are yet to be analyzed and compared. In this paper, we examine how 19 federal agencies responded to the memo, written by John Holdren, on issues of scientific data and the extent of their compliance to the directives outlined in the memo. We present a varied picture of the readiness of federal science agencies to comply with the memo through a comparative analysis and close reading of the contents of these responses. While some agencies, particularly those with a long history of supporting and conducting science, scored well, other responses indicate that some agencies have only taken a few steps towards implementing policies that comply with the memo. These results are of interest to the data curation community as they reveal how different agencies across the federal government approach their responsibilities for research data

management, and how new policies and requirements might continue to affect scientists and research communities.

Kristof, Cindy. "Data and Copyright." *Bulletin of the Association for Information Science and Technology* 42, no. 6 (2016): 20-22.
<https://doi.org/10.1002/bul2.2016.1720420607>

Kruse, Filip, and Jesper Boserup Thestrup. "Research Libraries' New Role in Research Data Management, Current Trends and Visions in Denmark." *LIBER Quarterly* 23, no. 4 (2014): 310-333. <http://doi.org/10.18352/lq.9173>

The amount of research data is growing constantly, due to new technology with new potentials for collecting and analysing both digital data and research objects. This growth creates a demand for a coherent IT-infrastructure. Such an infrastructure must be able to provide facilities for storage, preservation and a more open access to data in order to fulfil the demands from the researchers themselves, the research councils and research foundations.

This paper presents the findings of a research project carried out under the auspices of DEFF (Danmarks Elektroniske Fag-og Forskningsbibliotek—Denmark's Electronic research Library)[i] to analyse how the Danish universities store, preserve and provide access to research data. It shows that they do not have a common IT-infrastructure for research data management. This paper describes the various paths chosen by individual universities and research institutions, and the background for their strategies of research data management. Among the main reasons for the uneven practices are the lack of a national policy in this field, the different scientific traditions and cultures and the differences in the use and organization of IT-services.

This development contains several perspectives that are of particular relevance to research libraries. As they already curate digital collections and are active in establishing web archives, the research libraries become involved in research and dissemination of knowledge in new ways. This paper gives examples of how The State and University Library's services facilitate research data management with special regard to digitization of research objects, storage, preservation and sharing of research data. This paper concludes that the experience and skills of research libraries make the libraries important partners in a research data management infrastructure.

Kruse, Filip, and Jesper Boserup Thestrup, eds. *Research Data Management: A European Perspective*. Berlin; Boston: De Gruyter Saur, 2018.
<https://melvyl.on.worldcat.org/oclc/1017095732>

Krzton, Ali. "Supporting the Proliferation of Data-Sharing Scholars in the Research Ecosystem." *Journal of eScience Librarianship* 7, no. 2 (2018): e1145.
<https://doi.org/10.7191/jeslib.2018.1145>

Librarians champion the value of openness in scholarship and have been powerful advocates for the sharing of research data. College and university administrators have recently joined in the push for data sharing due to funding mandates. However, the researchers who create and control the data usually determine whether and how data is shared, so it is worthwhile to look at what they are incentivized to do. The current scholarly publishing landscape plus the promotion and tenure process create a "prisoner's dilemma" for researchers as they decide whether or not to share data, consistent with the observation that researchers in general are eager for others to share data but reluctant to do so themselves. If librarians encourage researchers to share data and promote openness without simultaneously addressing the academic incentive structure, those who are intrinsically motivated to share data will be selected against via the promotion and tenure process. This will cause those who are hostile to sharing to be disproportionately recruited into the senior ranks of academia. To mitigate the risk of this unintended consequence, librarians must advocate for a change in incentives alongside the call for greater openness. Highly-cited datasets must be given similar weight to highly-cited articles in promotion and tenure decisions in order for researchers to reap the rewards of their sharing. Librarians can help by facilitating data citation to track the impact of datasets and working to persuade higher administration of the value of rewarding data sharing in tenure and promotion.

Kubas, Alicia, and McBurney, Jenny. "Frustrations and Roadblocks in Data Reference Librarianship." *IASSIST Quarterly*, 43, no. 1 (2019): 1–18.
<https://doi.org/10.29173/iq939>

Kugler, Tracy A., David C. Van Riper, Steven M. Manson, David A. Haynes II, Joshua Donato, and Katie Stinebaugh. "Terra Populus: Workflows for Integrating and Harmonizing Geospatial Population and Environmental Data." *Journal of Map & Geography Libraries* 11, no. 2 (2015): 180-206.
<http://dx.doi.org/10.1080/15420353.2015.1036484>

Kurata, Keiko, Mamiko Matsubayashi, and Shinji Mine. "Identifying the Complex Position of Research Data and Data Sharing Among Researchers in Natural Science." *SAGE Open* (July 2017). <https://doi.org/10.1177/2158244017717301>

This article aims to provide an overview of researchers' practices and perceptions on data use and sharing. Semistructured interviews were conducted with 23 Japanese researchers in the natural sciences to identify their research practices and data use, including data sharing. We divided the interview scripts into meaningful phrases as a unit of analysis. Next, we focused on 406 statements on research data and reanalyzed them based on four aspects: stance on research data, practices and perceptions of data use, range of data sharing, and data type. A cluster analysis identified 14 clusters, which were divided into five groups: open access for data, restricted access for data, data interpretation, data processing and preservation, and data infrastructure. Our results reveal the complexity and diversity of the relationship between data and research practices. That is, the practice of research data sharing is heterogeneous, with no "one size fits all" between and among researchers.

Kutay, Stephen. "Advancing Digital Repository Services for Faculty Primary Research Assets: An Exploratory Study." *The Journal of Academic Librarianship* 40, no. 6 (2014): 642-649. <https://doi.org/10.1016/j.acalib.2014.08.006>

Lafia, Sara, and Werner Kuhn. "Spatial Discovery of Linked Research Datasets and Documents at a Spatially Enabled Research Library." *Journal of Map & Geography Libraries* 14, no. 1 (2018): 21-39. <https://doi.org/10.1080/15420353.2018.1478923>

Lage, Kathryn, Barbara Losoff, and Jack Maness. "Receptivity to Library Involvement in Scientific Data Curation: A Case Study at the University of Colorado Boulder." *portal: Libraries & the Academy* 11, no. 4 (2011): 915-937. <https://doi.org/10.1353/pla.2011.0049>

Lagoze, Carl. "eBird: Curating Citizen Science Data for Use by Diverse Communities." *International Journal of Digital Curation* 9, no. 1 (2014): 71-82. <https://doi.org/10.2218/ijdc.v9i1.302>

In this paper we describe eBird, a highly successful citizen science project. With over 150,000 participants worldwide and an accumulation of over 140,000,000 bird observations globally in the last decade, eBird has evolved into a major tool for scientific investigations in diverse fields such as

ornithology, computer science, statistics, ecology and climate change. eBird's impact in scientific research is grounded in careful data curation practices that pay attention to all stages of the data lifecycle, and attend to the needs of stakeholders engaged in that data lifecycle. We describe the important aspects of eBird, paying particular attention to the mechanisms to improve data quality; describe the data products that are available to the global community; investigate some aspects of the downloading community; and demonstrate significant results that derive from the use of openly-available eBird data.

Lamb, Ian, and Catherine Larson. "Shining a Light on Scientific Data: Building a Data Catalog to Foster Data Sharing and Reuse." *Code4Lib Journal*, no. 32 (2016). <http://journal.code4lib.org/articles/11421>

The scientific community's growing eagerness to make research data available to the public provides libraries—with our expertise in metadata and discovery—an interesting new opportunity. This paper details the in-house creation of a "data catalog" which describes datasets ranging from population-level studies like the US Census to small, specialized datasets created by researchers at our own institution. Based on Symphony2 and Solr, the data catalog provides a powerful search interface to help researchers locate the data that can help them, and an administrative interface so librarians can add, edit, and manage metadata elements at will. This paper will outline the successes, failures, and total redos that culminated in the current manifestation of our data catalog.

This work is licensed under a Creative Commons Attribution 3.0 United States License, <https://creativecommons.org/licenses/by/3.0/us/>.

Lambert, Paul, Vernon Gayle, Larry Tan, Ken Turner, Richard Sinnott, and Ken Prandy. "Data Curation Standards and Social Science Occupational Information Resources." *International Journal of Digital Curation 2*, no. 1 (2007): 73-91. <https://doi.org/10.2218/ijdc.v2i1.15>

Lassi, Monica, Maria Johnsson, Koraljka Golub. "Research Data Services: An Exploration of Requirements at Two Swedish Universities." *IFLA Journal 42*, no. 4 (2016): 266-277. <https://doi.org/10.1177/0340035216671963>

Latham, Bethany. "Research Data Management: Defining Roles, Prioritizing Services, and Enumerating Challenges." *The Journal of Academic Librarianship 43*, no. 3 (2017): 263-265. <https://doi.org/10.1016/j.acalib.2017.04.004>

Latham, Bethany, and Jodi Welch Poe. "The Library as Partner in University Data Curation: A Case Study in Collaboration." *Journal of Web Librarianship* 6, no. 4 (2012): 288-304. <https://doi.org/10.1080/19322909.2012.729429>

Laughton, P., and T. du Plessis. "Data Curation in the World Data System: Proposed Framework." *Data Science Journal* 12 (2013): 56-70. <https://doi.org/10.2481/dsj.13-029>

The value of data in society is increasing rapidly. Organisations that work with data should have standard practices in place to ensure successful curation of data. The World Data System (WDS) consists of a number of data centres responsible for curating research data sets for the scientific community. The WDS has no formal data curation framework or model in place to act as a guideline for member data centres. The objective of this research was to develop a framework for the curation of data in the WDS. A multiple-case case study was conducted. Interviews were used to gather qualitative data and analysis of the data, which led to the development of this framework. The proposed framework is largely based on the Open Archival Information System (OAIS) functional model and caters for the curation of both analogue and digital data.

Laure, Erwin, and Dejan Vitlacil. "Data Storage and Management for Global Research Data Infrastructures—Status and Perspectives." *Data Science Journal* 12 (2013): GRDI37-GRDI42. <https://doi.org/10.2481/dsj.grdi-007>

In the vision of Global Research Data Infrastructures (GRDIs), data storage and management plays a crucial role. A successful GRDI will require a common globally interoperable distributed data system, formed out of data centres, that incorporates emerging technologies and new scientific data activities. The main challenge is to define common certification and auditing frameworks that will allow storage providers and data communities to build a viable partnership based on trust. To achieve this, it is necessary to find a long-term commitment model that will give financial, legal, and organisational guarantees of digital information preservation. In this article we discuss the state of the art in data storage and management for GRDIs and point out future research directions that need to be tackled to implement GRDIs.

Lawrence, Bryan, Catherine Jones, Brian Matthews, Sam Pepler, and Sarah Callaghan. "Citation and Peer Review of Data: Moving towards Formal Data

Publication." *International Journal of Digital Curation* 6, no. 2 (2011): 4-37.
<https://doi.org/10.2218/ijdc.v6i2.205>

This paper discusses many of the issues associated with formally publishing data in academia, focusing primarily on the structures that need to be put in place for peer review and formal citation of datasets. Data publication is becoming increasingly important to the scientific community, as it will provide a mechanism for those who create data to receive academic credit for their work and will allow the conclusions arising from an analysis to be more readily verifiable, thus promoting transparency in the scientific process. Peer review of data will also provide a mechanism for ensuring the quality of datasets, and we provide suggestions on the types of activities one expects to see in the peer review of data. A simple taxonomy of data publication methodologies is presented and evaluated, and the paper concludes with a discussion of dataset granularity, transience and semantics, along with a recommended human-readable citation syntax.

Layne, R., A. Capel, N. Coe, and M. Wheatley. "Long Term Preservation of Scientific Data: Lessons from Jet and Other Domains." *Fusion Engineering and Design* 87, no. 12 (2012): 2209-2212.
<https://doi.org/10.1016/j.fusengdes.2012.07.004>

Leadbetter, A., L. Raymond, C. Chandler, L. Pikula, P. Pissierssens, and E. Urban. *Ocean Data Publication Cookbook*. Oostende, Belgium: UNESCO, 2013.
http://www.iode.org/index.php?option=com_oe&task=viewDocumentRecord&docID=10574

Lee, Christopher, Suzie Allard, Nancy McGovern, and Alice Bishop. "Open Data Meets Digital Curation: An Investigation of Practices and Needs." *International Journal of Digital Curation* 11, no. 2 (2017): 115-125.
<https://doi.org/10.2218/ijdc.v11i2.403>

In the United States, research funded by the government produces a significant portion of data. US law mandates that these data should be freely available to the public through 'public access', which is defined as fully discoverable and usable by the public. The U.S. government executive branch supported the public access requirements by issuing an Executive Directive titled 'Increasing Access to the Results of Federally Funded Scientific Research' that required federal agencies with annual research and development expenditures of more than \$100 million to create public access

plans by 22 August 2013. The directive applied to 19 federal agencies, some with multiple divisions. Additional direction for this initiative was provided by the Executive Order 'Making Open and Machine Readable the New Default for Government Information' which was accompanied by a memorandum with specific guidelines for information management and instructions to find ways to reduce compliance costs through interagency cooperation.

In late 2013, the Institute of Museum and Library Services (IMLS) funded the Council on Library and Information Resources (CLIR) to conduct a project to help IMLS and its constituents understand the implications of the US federal public access mandate and how needs and gaps in digital curation can best be addressed. Our project has three research components: (1) a structured content analysis of federal agency plans supporting public access to data and publications, identifying both commonalities and differences among plans; (2) case studies (interviews and analysis of project deliverables) of seven projects previously funded by IMLS to identify lessons about skills, capabilities and institutional arrangements that can facilitate data curation activities; and (3) a gap analysis of continuing education and readiness assessment of the workforce. Research and cultural institutions urgently need to rethink the professional identities of those responsible for collecting, organizing, and preserving data for future use. This paper reports on a project to help inform further investments.

Lee, Dong Joon, and Besiki Stvilia. "Developing a Data Identifier Taxonomy." *Cataloging & Classification Quarterly* 52, no. 3 (2014): 303-336.
<https://doi.org/10.1080/01639374.2014.880166>

———. "Practices of Research Data Curation in Institutional Repositories: A Qualitative View from Repository Staff." *PLOS ONE* 12, no. 3 (2017): e0173987.
<https://doi.org/10.1371/journal.pone.0173987>

The importance of managing research data has been emphasized by the government, funding agencies, and scholarly communities. Increased access to research data increases the impact and efficiency of scientific activities and funding. Thus, many research institutions have established or plan to establish research data curation services as part of their Institutional Repositories (IRs). However, in order to design effective research data curation services in IRs, and to build active research data providers and user communities around those IRs, it is essential to study current data curation practices and provide rich

descriptions of the sociotechnical factors and relationships shaping those practices. Based on 13 interviews with 15 IR staff members from 13 large research universities in the United States, this paper provides a rich, qualitative description of research data curation and use practices in IRs. In particular, the paper identifies data curation and use activities in IRs, as well as their structures, roles played, skills needed, contradictions and problems present, solutions sought, and workarounds applied. The paper can inform the development of best practice guides, infrastructure and service templates, as well as education in research data curation in Library and Information Science (LIS) schools.

Leonelli, Sabina. "Data Governance is Key to Interpretation: Reconceptualizing Data in Data Science." *Harvard Data Science Review* 1, no. 1 (2019).

<https://doi.org/10.1162/99608f92.17405bb6>

I provide a philosophical perspective on the characteristics of data-centric research and the conceptualization of data that underpins it. The transformative features of contemporary data science derive not only from the availability of Big Data and powerful computing, but also from a fundamental shift in the conceptualization of data as research materials and sources of evidence. A relational view of data is proposed, within which the meaning assigned to data depends on the motivations and instruments used to analyze them and to defend specific interpretations. The presentation of data, the way they are identified, selected and included (or excluded) in databases and the information provided to users to re-contextualize them are fundamental to producing knowledge—and significantly influence its content. Concerns around interpreting data and assessing their quality can be tackled by cultivating governance strategies around how data are collected, managed and processed.

Lewis, Stuart, Lorraine Beard, Mary McDerby, Robin Taylor, Thomas Higgins, and Claire Knowles. "Developing a Data Vault." *International Journal of Digital Curation* 11, no. 1 (2016): 86-95. <https://doi.org/10.2218/ijdc.v11i1.406>

Research data is being generated at an ever-increasing rate. This brings challenges in how to store, analyse, and care for the data. A component of this problem is the stewardship of data and associated files that need a safe and secure home for the medium to long-term.

As part of typical suites of Research Data Management services, researchers are provided with large allocations of 'active data storage'. This is often stored on expensive and fast disks to enable efficient transfer and working with large amounts of data. However, over time this active data store fills up, and researchers need a facility to move older but still valuable data to cheaper storage for long-term care. In addition, research funders are increasingly requiring data to be stored in forms that allow it to be described and retrieved in the future. For data that can't be shared publicly in an open repository, a closed solution is required that can make use of offline or near-line storage for cost efficiency.

This paper describes a solution to these requirements, called the Data Vault.

Li, Si, Xiaozhe Zhuang, Wenming Xing, and Weining Guo. "The Cultivation of Scientific Data Specialists: Development of LIS Education Oriented to E-science Service Requirements." *Library Hi Tech* 31, no. 4 (2013): 700-724.
<https://doi.org/10.1108/lht-06-2013-0070>

Linde, Peter, Merel Noorman, Bridgette A. Wessels, and Thordis Sveinsdottir. "How Can Libraries and Other Academic Stakeholders Engage in Making Data Open?" *Information Services and Use* 34, no. 3-4 (2014): 211-219.
<https://doi.org/10.3233/isu-140741>

Linek, Stephanie B., Benedikt Fecher, Sascha Friesike, and Marcel Hebing. "Data Sharing as Social Dilemma: Influence of the Researcher's Personality." *PLOS ONE* 12, no. 8 (2017): e0183216. <https://doi.org/10.1371/journal.pone.0183216>

It is widely acknowledged that data sharing has great potential for scientific progress. However, so far making data available has little impact on a researcher's reputation. Thus, data sharing can be conceptualized as a social dilemma. In the presented study we investigated the influence of the researcher's personality within the social dilemma of data sharing. The theoretical background was the appropriateness framework. We conducted a survey among 1564 researchers about data sharing, which also included standardized questions on selected personality factors, namely the so-called Big Five, Machiavellianism and social desirability. Using regression analysis, we investigated how these personality domains relate to four groups of dependent variables: attitudes towards data sharing, the importance of factors that might foster or hinder data sharing, the willingness to share data, and actual data sharing. Our analyses showed the predictive value of personality

for all four groups of dependent variables. However, there was not a global consistent pattern of influence, but rather different compositions of effects. Our results indicate that the implications of data sharing are dependent on age, gender, and personality. In order to foster data sharing, it seems advantageous to provide more personal incentives and to address the researchers' individual responsibility.

Linne, Monika, and Wolfgang Zenk-Möltgen. "Strengthening Institutional Data Management and Promoting Data Sharing in the Social and Economic Sciences." *LIBER Quarterly* 27, no. 1 (2017): 8-72. <http://doi.org/10.18352/lq.10195>

In the German social and economic sciences there is a growing awareness of flexible data distribution and research data reuse, especially as increasing numbers of research funders recommend publishing research data as the basis for scientific insight. However, a data-sharing mentality has not yet been established in Germany attributable to researchers' strong reservations about publishing their data. This attitude is exacerbated by the fact that, at present, there is no trusted national data sharing repository that covers the particular requirements of institutions regarding research data. This article discusses how this objective can be achieved with the project initiative SowiDataNet. The development of a community-driven data repository is a logically consistent and important step towards an attitude shift concerning data sharing in the social and economic sciences.

Littauer, Richard, Karthik Ram, Bertram Ludäscher, William Michener, and Rebecca Koskela. "Trends in Use of Scientific Workflows: Insights from a Public Repository and Recommendations for Best Practice." *International Journal of Digital Curation* 7, no. 2 (2012): 92-100. <https://doi.org/10.2218/ijdc.v7i2.232>

Scientific workflows are typically used to automate the processing, analysis and management of scientific data. Most scientific workflow programs provide a user-friendly graphical user interface that enables scientists to more easily create and visualize complex workflows that may be comprised of dozens of processing and analytical steps. Furthermore, many workflows provide mechanisms for tracing provenance and methodologies that foster reproducible science. Despite their potential for enabling science, few studies have examined how the process of creating, executing, and sharing workflows can be improved. In order to promote open discourse and access to scientific methods as well as data, we analyzed a wide variety of workflow systems and publicly available workflows on the public repository myExperiment. It is

hoped that understanding the usage of workflows and developing a set of recommended best practices will lead to increased contribution of workflows to the public domain.

Liu, Xia, and Ning Ding. "Research Data Management in Universities of Central China." *Electronic Library* 34, no. 5 (2016): 808-822.
<http://dx.doi.org/10.1108/EL-04-2015-0063>

Llebot, Clara, and Steven Van Tuyl. "Peer Review of Research Data Submissions to ScholarsArchive@OSU: How Can We Improve the Curation of Research Datasets to Enhance Reusability?" *Journal of eScience Librarianship* 8, no. 2 (2019): e1166. <https://doi.org/10.7191/jeslib.2019.1166>.

Objective: Best practices such as the FAIR Principles (Findability, Accessibility, Interoperability, Reusability) were developed to ensure that published datasets are reusable. While we employ best practices in the curation of datasets, we want to learn how domain experts view the reusability of datasets in our institutional repository, ScholarsArchive@OSU. Curation workflows are designed by data curators based on their own recommendations, but research data is extremely specialized, and such workflows are rarely evaluated by researchers. In this project we used peer-review by domain experts to evaluate the reusability of the datasets in our institutional repository, with the goal of informing our curation methods and ensure that the limited resources of our library are maximizing the reusability of research data.

Methods: We asked all researchers who have datasets submitted in Oregon State University's repository to refer us to domain experts who could review the reusability of their data sets. Two data curators who are non-experts also reviewed the same datasets. We gave both groups review guidelines based on the guidelines of several journals. Eleven domain experts and two data curators reviewed eight datasets. The review included the quality of the repository record, the quality of the documentation, and the quality of the data. We then compared the comments given by the two groups.

Results: Domain experts and non-expert data curators largely converged on similar scores for reviewed datasets, but the focus of critique by domain experts was somewhat divergent. A few broad issues common across reviews were: insufficient documentation, the use of links to journal articles in the place of documentation, and concerns about duplication of effort in creating

documentation and metadata. Reviews also reflected the background and skills of the reviewer. Domain experts expressed a lack of expertise in data curation practices and data curators expressed their lack of expertise in the research domain.

Conclusions: The results of this investigation could help guide future research data curation activities and align domain expert and data curator expectations for reusability of datasets. We recommend further exploration of these common issues and additional domain expert peer-review project to further refine and align expectations for research data reusability.

Locher, Anita E. "Starting Points for Lowering the Barrier to Spatial Data Preservation." *Journal of Map & Geography Libraries* 12, no. 1 (2016): 28-51. <http://dx.doi.org/10.1080/15420353.2015.1080781>

Losee, Robert M. "Informational Facts and the Metainformation Inherent in IFacts: The Soul of Data Sciences." *Journal of Library Metadata* 13, no. 1 (2013): 59-74. <https://doi.org/10.1080/19386389.2013.778732>

Lötter, Lucia, and Christa van Zyl. "A Reflection on a Data Curation Journey." *Journal of Empirical Research on Human Research Ethics* 10, no. 3 (2015): 338-343. <http://dx.doi.org/10.1177/1556264615592846>

This commentary is a reflection on experience of data preservation and sharing (i.e., data curation) practices developed in a South African research organization. The lessons learned from this journey have echoes in the findings and recommendations emerging from the present study in Low and Middle-Income Countries (LMIC) and may usefully contribute to more general reflection on the management of change in data practice.

This work is licensed under a Creative Commons Attribution 3.0 Unported License, <https://creativecommons.org/licenses/by/3.0/>.

Lubell, Josh, Sudarsan Rachuri, Mahesh Mani, and Eswaran Subrahmanian. "Sustaining Engineering Informatics: Toward Methods and Metrics for Digital Curation." *International Journal of Digital Curation* 3, no. 2 (2008). <https://doi.org/10.2218/ijdc.v3i2.58>

Lynch, Clifford. "The Need for Research Data Inventories and the Vision for SHARE." *Information Standards Quarterly* 26, no. 2 (2014): 29-31.

<https://www.niso.org/niso-io/2014/06/need-research-data-inventories-and-vision-share>

Lyon, Liz. "The Informatics Transform: Re-engineering Libraries for the Data Decade." *International Journal of Digital Curation* 7, no. 1 (2012): 126-138. <https://doi.org/10.2218/ijdc.v7i1.220>

———. "Librarians in the Lab: Toward Radically Re-Engineering Data Curation Services at the Research Coalface." *New Review of Academic Librarianship* 22, no. 4 (2016): 391-409. <http://dx.doi.org/10.1080/13614533.2016.1159969>

———. "Transparency: The Emerging Third Dimension of Open Science and Open Data." *LIBER Quarterly* 25, no. 4 (2016):153-171. <http://doi.org/10.18352/lq.10113>

This paper presents an exploration of the concept of research transparency. The policy context is described and situated within the broader arena of open science. This is followed by commentary on transparency within the research process, which includes a brief overview of the related concept of reproducibility and the associated elements of research integrity, fraud and retractions. A two-dimensional model or continuum of open science is considered and the paper builds on this foundation by presenting a three-dimensional model, which includes the additional axis of 'transparency'. The concept is further unpacked and preliminary definitions of key terms are introduced: transparency, transparency action, transparency agent and transparency tool. An important linkage is made to the research lifecycle as a setting for potential transparency interventions by libraries. Four areas are highlighted as foci for enhanced engagement with transparency goals: Leadership and Policy, Advocacy and Training, Research Infrastructures and Workforce Development.

Lyon, Liz, Wei Jeng, and Eleanor Mattern. "Research Transparency: A Preliminary Study of Disciplinary Conceptualisation, Drivers, Tools and Support Services." *International Journal of Digital Curation* 12, no. 1 (2017): 46-64. <https://doi.org/10.2218/ijdc.v12i1.530>

This paper describes a preliminary study of research transparency, which draws on the findings from four focus group sessions with faculty in chemistry, law, urban and social studies, and civil and environmental engineering. The multi-faceted nature of transparency is highlighted by the broad ways in which the faculty conceptualised the concept (data sharing,

ethics, replicability) and the vocabulary they used with common core terms identified (data, methods, full disclosure). The associated concepts of reproducibility and trust are noted. The research lifecycle stages are used as a foundation to identify the action verbs and software tools associated with transparency. A range of transparency drivers and motivations are listed. The role of libraries and data scientists is discussed in the context of the provision of transparency services for researchers.

Macdonald, Stuart, Ingrid Dillo, Sophia Lafferty-Hess, Lynn Woolfrey, and Mary Vardigan. "Demonstrating Repository Trustworthiness through the Data Seal of Approval." *IASSIST Quarterly*, 40, no. 3 (2017): 6. <https://doi.org/10.29173/iq396>.

Macdonald, Stuart, and Luis Martinez-Urbe. "Collaboration to Data Curation: Harnessing Institutional Expertise." *New Review of Academic Librarianship* 16, no. supplement 1 (2010): 4-16. <https://doi.org/10.1080/13614533.2010.505823>

MacMillan, Don. "Data Sharing and Discovery: What Librarians Need to Know." *The Journal of Academic Librarianship* 40, no. 5 (2014): 541-549. <https://doi.org/10.1016/j.acalib.2014.06.011>

———. "Developing Data Literacy Competencies to Enhance Faculty Collaborations." *LIBER Quarterly* 24, no. 3 (2015): 140-160. <http://doi.org/10.18352/lq.9868>

In order to align information literacy instruction with changing faculty and student needs, librarians must expand their skills and competencies beyond traditional information sources. In the sciences, this increasingly means integrating the data resources used by researchers into instruction for undergraduate students. Open access data repositories allow students to work with more primary data than ever before, but only if they know how and where to look. This paper will describe the development of two information literacy workshops designed to scaffold student learning in the biological sciences across two second-year courses, detailing the long-term collaboration between a librarian and an instructor that now serves over 500 students per semester. In each workshop, students are guided through the discovery and analysis of life sciences data from multiple sites, encouraged to integrate text and data sources, and supported in completing research assignments.

Macy, Katharine V., and Heather L. Coates. "Data Information Literacy Instruction in Business and Public Health: Comparative Case Studies." *IFLA Journal* 42, no. 4 (2016): 313–327. <https://doi.org/10.1177/0340035216673382>

Maday, Charlotte, and Magalie Moysan. "Records Management for Scientific Data." *Archives and Manuscripts* 42, no. 2 (2014): 190-192.
<https://doi.org/10.1108/eb027016>

Makani, Joyline. "Knowledge Management, Research Data Management, and University Scholarship: Towards an Integrated Institutional Research Data Management Support-System Framework." *VINE* 45, no. 3 (2015): 344-359.
<http://dx.doi.org/10.1108/VINE-07-2014-0047>

Mallery, Mary. "DMPtool: Guidance and Resources for Your Data Management Plan." *Technical Services Quarterly* 31, no. 2 (2014): 197-199.
<http://dx.doi.org/10.1080/07317131.2014.875394>

Malone, James, Andy Brown, Allyson L. Lister, Jon Ison, Duncan Hull, Helen Parkinson, and Robert Stevens. "The Software Ontology (SWO): A Resource for Reproducibility in Biomedical Data Analysis, Curation and Digital Preservation." *Journal of Biomedical Semantics* 5, no. 1 (2014): 25.
<http://dx.doi.org/10.1186/2041-1480-5-25>

Manghi, Paolo, Lukasz Bolikowski, Natalia Manold, Jochen Schirrwagen, and Tim Smith. "OpenAIREplus: the European Scholarly Communication Data Infrastructure." *D-Lib Magazine* 18, no. 9/10 (2012).
<http://dx.doi.org/10.1045/september2012-manghi>

Mannheimer, Sara. "Toward a Better Data Management Plan: The Impact of DMPs on Grant Funded Research Practices." *Journal of eScience Librarianship* 7, no. 3 (2018): e1155. <https://doi.org/10.7191/jeslib.2018.1155>

Data Management Plans (DMPs) are often required for grant applications. But do strong DMPs lead to better data management and sharing practices? Several recent research projects in the Library and Information Science field have investigated data management planning and practice through DMP content analysis and data-management-related interviews. However, research hasn't yet shown how DMPs ultimately affect data management and data sharing practices during grant-funded research. The research described in this article contributes to the existing literature by examining the impact of DMPs on grant awards and on Principal Investigators' (PIs) data management and sharing practices. The results of this research suggest the following key takeaways: (1) Most PIs practice internal data management in order to prevent data loss, to facilitate sharing within the research team, and to seamlessly continue their research during personnel turnover; (2) PIs still have room to

grow in understanding specialized concepts such as metadata and policies for use and reuse; (3) PIs may need guidance on practices that facilitate FAIR data, such as using metadata standards, assigning licenses to their data, and publishing in data repositories. Ultimately, the results of this research can inform academic library services and support stronger, more actionable DMPs.

Mannheimer, Sara, Leila Belle Sterman, and Susan Borda. "Discovery and Reuse of Open Datasets: An Exploratory Study." *Journal of eScience Librarianship* 5, no. 1 (2016): e1091. <http://dx.doi.org/10.7191/jeslib.2016.1091>.

Mannheimer, Sara, Ayoung Yoon, Jane Greenberg, Elena Feinstein, and Ryan Scherle. "A Balancing Act: The Ideal and the Realistic in Developing Dryad's Preservation Policy." *First Monday* 19, no. 8 (2014). <http://dx.doi.org/10.5210/fm.v19i8.5415>

Manninen, Lauren. "Describing Data: A Review of Metadata for Datasets in the Digital Commons Institutional Repository Platform: Problems and Recommendations." *Journal of Library Metadata* 18, no. 1 (2018): 1-11. <https://doi.org/10.1080/19386389.2018.1454379>

Objective: This article analyzes twenty cited or downloaded datasets and the repositories that house them, in order to produce insights that can be used by academic libraries to encourage discovery and reuse of research data in institutional repositories.

Methods: Using Thomson Reuters' Data Citation Index and repository download statistics, we identified twenty cited/downloaded datasets. We documented the characteristics of the cited/downloaded datasets and their corresponding repositories in a self-designed rubric. The rubric includes six major categories: basic information; funding agency and journal information; linking and sharing; factors to encourage reuse; repository characteristics; and data description.

Results: Our small-scale study suggests that cited/downloaded datasets generally comply with basic recommendations for facilitating reuse: data are documented well; formatted for use with a variety of software; and shared in established, open access repositories. Three significant factors also appear to contribute to dataset discovery: publishing in discipline-specific repositories; indexing in more than one location on the web; and using persistent identifiers. The cited/downloaded datasets in our analysis came from a few

specific disciplines, and tended to be funded by agencies with data publication mandates.

Conclusions: The results of this exploratory research provide insights that can inform academic librarians as they work to encourage discovery and reuse of institutional datasets. Our analysis also suggests areas in which academic librarians can target open data advocacy in their communities in order to begin to build open data success stories that will fuel future advocacy efforts.

Marcial, Laura Haak, and Bradley M. Hemminger. "Scientific Data Repositories on the Web: An Initial Survey." *Journal of the American Society for Information Science and Technology* 61, no. 10 (2010): 2029-2048.
<http://dx.doi.org/10.1002/asi.21339>

Marshall, Brianna, Katherine O'Bryan, Na Qin, and Rebecca Vernon. "Organizing, Contextualizing, and Storing Legacy Research Data: A Case Study of Data Management for Librarians." *Issues in Science and Technology Librarianship*, no. 74 (2013). <http://www.istl.org/13-fall/article1.html>

Martin, Caroline, Colette Cadiou, and Emmanuelle Jannès-Ober. "Data Management: New Tools, New Organization, and New Skills in a French Research Institute." *LIBER Quarterly* 27, no. 1 (2017): 73–88.
<http://doi.org/10.18352/lq.10196>

In the context of E-science and open access, visibility and impact of scientific results and data have become important aspects for spreading information to users and to the society in general. The objective of this general trend of the economy is to feed the innovation process and create economic value. In our institute, the French National Research Institute of Science and Technology for Environment and Agriculture, Irstea, the department in charge of scientific and technical information, with the help of other professionals (Scientists, IT professionals, ethics advisors...), has recently developed suitable services for the researchers and for their needs concerning the data management in order to answer European recommendations for open data. This situation has demanded to review the different workflows between databases, to question the organizational aspects between skills, occupations, and departments in the institute. In fact, the data management involves all professionals and researchers to asset their working ways together.

Martin, Erika G., Jennie Law, Weijia Ran, Natalie Helbig, and Guthrie S. Birkhead. "Evaluating the Quality and Usability of Open Data for Public Health

Research: A Systematic Review of Data Offerings on 3 Open Data Platforms." *Journal of Public Health Management and Practice* 23, no. 4 (2017): e5-e13. <https://doi.org/10.1097/phh.0000000000000388>

Martinez-Uribe, Luis, and Stuart Macdonald. "User Engagement in Research Data Curation." *Lecture Notes in Computer Science* 5714 (2009): 309-314. https://doi.org/10.1007/978-3-642-04346-8_30

Martinsen, David. "Primary Research Data and Scholarly Communication." *Chemistry International* 39, no. 3 (2017): 35-38. <https://doi.org/10.1515/ci-2017-0309>

Mathiak, Brigitte, and Katarina Boland. "Challenges in Matching Dataset Citation Strings to Datasets in Social Science." *D-Lib Magazine* 21, no. 1/2 (2015). <https://doi.org/10.1045/january2015-mathiak>

Matlatse, Refiloe, Heila Pienaar, and Martie van Deventer. "Mobilising a Nation: RDM Training and Education in South Africa." *International Journal of Digital Curation* 12, no. 2 (2017): 299-310. <https://doi.org/10.2218/ijdc.v12i2.579>

The South African Network of Data and Information Curation Communities (NeDICC) was formed to promote the development and use of standards and best practices among South African data stewards and data librarians (NeDICC, 2015). The steering committee has members from various South African HEIs and research councils. As part of their service offerings NeDICC arranges seminars, workshops and conferences to promote awareness regarding digital curation. NeDICC has contributed to the increase in awareness, and growth of knowledge, on the subject of digital and data curation in South Africa (Kahn et al.,2014).NeDICC members are involved in the UP M.IT and Continued Professional Development training, and serve as external examiners for the UCT M.Phil in Digital Curation degree. NeDICC is responsible for the Research Data Management track at the annual e-Research conference in SA and develops an annual training-focussed programme to provide workshop opportunities with both SA and foreign trainers. This paper specifically addresses the efforts by this community to mobilise and upskill South African librarians so that they would be willing and able to provide the necessary RDM services that would strengthen the national data effort.

Mattern, Eleanor, Aaron Brenner, and Liz Lyon. "Learning by Teaching about RDM: An Active Learning Model for Internal Library Education." *International*

Journal of Digital Curation 11, no. 2 (2017): 27-38.
<https://doi.org/10.2218/ijdc.v11i2.414>

This paper reports on the design, delivery and assessment of a model for internal library education around research data management (RDM). Conducted at the University of Pittsburgh Library System (ULS), the exercise and resultant instructional session employed an active learning approach, in which a group of librarians and archivists explored data issues and conventions in a discipline of their own selection and presented their findings to an audience of library colleagues. In this paper, we put forth an adaptable active learning model for internal RDM education and offer guidance for its implementation by peer libraries that are similarly building internal capacity for the design and delivery of RDM services that are responsive to disciplinary needs.

Mattern, Eleanor, Wei Jeng, Daqing He, Liz Lyon, and Aaron Brenner. "Using Participatory Design and Visual Narrative Inquiry to Investigate Researchers—Data Challenges and Recommendations for Library Research Data Services." *Program* 49, no. 4 (2015): 408-423. <https://doi.org/10.1108/prog-01-2015-0012>

Matthews, Brian, Shoaib Sufi, Damian Flannery, Laurent Lerusse, Tom Griffin, Michael Gleaves, and Kerstin Kleese. "Using a Core Scientific Metadata Model in Large-Scale Facilities." *International Journal of Digital Curation* 5, no. 1 (2010): 106-118. <https://doi.org/10.2218/ijdc.v5i1.146>

In this paper, we present the Core Scientific Metadata Model (CSMD), a model for the representation of scientific study metadata developed within the Science & Technology Facilities Council (STFC) to represent the data generated from scientific facilities. The model has been developed to allow management of and access to the data resources of the facilities in a uniform way, although we believe that the model has wider application, especially in areas of "structural science" such as chemistry, materials science and earth sciences. We give some motivations behind the development of the model, and an overview of its major structural elements, centred on the notion of a scientific study formed by a collection of specific investigations. We give some details of the model, with the description of each investigation associated with a particular experiment on a sample generating data, and the associated data holdings are then mapped to the investigation with the appropriate parameters. We then go on to discuss the instantiation of the metadata model within a production quality data management infrastructure,

the Information CATalogue (ICAT), which has been developed within STFC for use in large-scale photon and neutron sources. Finally, we give an overview of the relationship between CSMD, and other initiatives, and give some directions for future developments.

Mattmann, C., Crichton, D. J., A. F. Hart, S. C. Kelly, and J. S. Hughes. "Experiments with Storage and Preservation of NASA's Planetary Data via the Cloud." *IT Professional* 12, no. 5 (2010): 28-35.
<https://doi.org/10.1109/mitp.2010.97>

Mauthner, Natasha Susan, and Odette Parry. "Open Access Digital Data Sharing: Principles, Policies and Practices." *Social Epistemology: A Journal of Knowledge, Culture and Policy* 27, no. 1 (2013): 47-67.
<https://doi.org/10.1080/02691728.2012.760663>

Mayernik, Matthew S. "Data Citation Initiatives and Issues." *Bulletin of the American Society for Information Science and Technology* 38, no. 5 (2012): 23-28.
<https://doi.org/10.1002/bult.2012.1720380508>

———. "Research Data and Metadata Curation as Institutional Issues." *Journal of the Association for Information Science and Technology* 67, no. 4 (2015): 973-993.
<https://doi.org/10.1002/asi.23425>

Mayernik, Matthew S., Sarah Callaghan, Roland Leighm, Jonathan Tedds, and Steven Worley. "Peer Review of Datasets: When, Why, and How." *Bulletin of the American Meteorological Society* 96 (2015): 191-201.
<http://dx.doi.org/10.1175/BAMS-D-13-00083.1>

Mayernik, Matthew S., G. Sayeed Choudhury, Tim DiLauro, Elliot Metsger, Barbara Pralle, Mike Rippin, and Ruth Duerr. "The Data Conservancy Instance: Infrastructure and Organizational Services for Research Data Curation." *D-Lib Magazine* 18, no. 9/10 (2012). <https://doi.org/10.1045/september2012-mayernik>

Mayernik, Matthew S., Tim DiLauro, Ruth Duerr, Elliot Metsger, Anne E. Thessen, and G. Sayeed Choudhury. "Data Conservancy Provenance, Context, and Lineage Services: Key Components for Data Preservation and Curation." *Data Science Journal* 12 (2013): 158-171. <https://doi.org/10.2481/dsj.12-039>

Among the key services that institutional data management infrastructures must provide are provenance and lineage tracking and the ability to associate data with contextual information needed for understanding and use. These

functionalities are critical for addressing a number of key issues faced by data collectors and users, including trust in data, results traceability, data transparency, and data citation support. In this paper, we describe the support for these services within the Data Conservancy Service (DCS) software. The DCS provenance, context, and lineage services cross the four layers in the DCS data curation stack model: storage, archiving, preservation, and curation.

Mayernik, Matthew S., Jennifer Phillips, and Eric Nienhouse. "Linking Publications and Data: Challenges, Trends, and Opportunities." *D-Lib Magazine* 22, no. 5/6 (2016). <https://doi.org/10.1045/may2016-mayernik>

Mayo, Christine, Todd J. Vision, and Elizabeth A. Hull. "The Location of the Citation: Changing Practices in How Publications Cite Original Data in the Dryad Digital Repository." *International Journal of Digital Curation* 11, no. 1 (2016): 150-155. <https://doi.org/10.2218/ijdc.v11i1.400>

While stakeholders in scholarly communication generally agree on the importance of data citation, there is not consensus on where those citations should be placed within the publication—particularly when the publication is citing original data. Recently, CrossRef and the Digital Curation Center (DCC) have recommended as a best practice that original data citations appear in the works cited sections of the article. In some fields, such as the life sciences, this contrasts with the common practice of only listing data identifier(s) within the article body (intratextually). We inquired whether data citation practice has been changing in light of the guidance from CrossRef and the DCC. We examined data citation practices from 2011 to 2014 in a corpus of 1,125 articles associated with original data in the Dryad Digital Repository. The percentage of articles that include no reference to the original data has declined each year, from 31% in 2011 to 15% in 2014. The percentage of articles that include data identifiers intratextually has grown from 69% to 83%, while the percentage that cite data in the works cited section has grown from 5% to 8%. If the proportions continue to grow at the current rate of 19-20% annually, the proportion of articles with data citations in the works cited section will not exceed 90% until 2030.

McEwen, Leah, and Ye Li. "Academic Librarians at Play in the Field of Cheminformatics: Building the Case for Chemistry Research Data Management." *Journal of Computer-Aided Molecular Design* 28, no. 10 (2014): 975-988. <http://dx.doi.org/10.1007/s10822-014-9777-4>

McGarva, Guy, Steve Morris, and Greg Janée. *Preserving Geospatial Data*. York, UK: Digital Preservation Coalition, 2009.

http://www.dpconline.org/component/docman/doc_download/363-preserving-geospatial-data-by-guy-mcgarva-steve-morris-and-gred-greg-janee

McKinney, Bill, Peter A. Meyer, Mercè Crosas, and Piotr Sliz. "Extension of Research Data Repository System to Support Direct Compute Access to Biomedical Datasets: Enhancing Dataverse to Support Large Datasets." *Annals of the New York Academy of Sciences* 13871, no. 1 (2017): 95-104.

<https://doi.org/10.1111/nyas.13272>

McLure, Merinda, Allison V. Level, Catherine L. Cranston, Beth Oehlert, and Mike Culbertson. "Data Curation: A Study of Researcher Practices and Needs." *portal: Libraries and the Academy* 14, no. 2 (2014).

<https://doi.org/10.1353/pla.2014.0009>

McNally, Ruth, Adrian Mackenzie, Allison Hui, and Jennifer Tomomitsu. "Understanding the 'Intensive' in 'Data Intensive Research': Data Flows in Next Generation Sequencing and Environmental Networked Sensors." *International Journal of Digital Curation* 7, no. 1 (2012): 81-94.

<https://doi.org/10.2218/ijdc.v7i1.216>

Genomic and environmental sciences represent two poles of scientific data. In the first, highly parallel sequencing facilities generate large quantities of sequence data. In the latter, loosely networked remote and field sensors produce intermittent streams of different data types. Yet both genomic and environmental sciences are said to be moving to data intensive research. This paper explores and contrasts data flow in these two domains in order to better understand how data intensive research is being done. Our case studies are next generation sequencing for genomics and environmental networked sensors.

Our objective was to enrich understanding of the 'intensive' processes and properties of data intensive research through a 'sociology' of data using methods that capture the relational properties of data flows. Our key methodological innovation was the staging of events for practitioners with different kinds of expertise in data intensive research to participate in the collective annotation of visual forms. Through such events we built a substantial digital data archive of our own that we then analysed in terms of three traits of data flow: durability, replicability and metrology.

Our findings are that analysing data flow with respect to these three traits provides better insight into how doing data intensive research involves people, infrastructures, practices, things, knowledge and institutions. Collectively, these elements shape the topography of data and condition how it flows. We argue that although much attention is given to phenomena such as the scale, volume and speed of data in data intensive research, these are measures of what we call 'extensive' properties rather than intensive ones. Our thesis is that extensive changes, that is to say those that result in non-linear changes in metrics, can be seen to result from intensive changes that bring multiple, disparate flows into confluence.

If extensive shifts in the modalities of data flow do indeed come from the alignment of disparate things, as we suggest, then we advocate the staging of workshops and other events with the purpose of developing the 'missing' metrics of data flow.

Medeiros, Norm. "A Public Trust: Libraries and Data Curation " *OCLC Systems & Services: International Digital Library Perspectives* 29, no. 4 (2013): 192-194.
<https://doi.org/10.1108/oclc-06-2013-0018>

Mehnert, Andrew James, Andrew Janke, Marco Gruwel, Wojtek James Goscinski, Thomas Close, Dean Taylor, Aswin Narayanan, George Vidalis, Graham Galloway, and Andrew Treloar. "Putting the Trust into Trusted Data Repositories: A Federated Solution for the Australian National Imaging Facility." *International Journal Of Digital Curation* 14 no. 1 (2019): 102-113.
<https://doi.org/10.2218/ijdc.v14i1.594>

The National Imaging Facility (NIF) provides Australian researchers with state-of-the-art instrumentation—including magnetic resonance imaging (MRI), positron emission tomography (PET), X-ray computed tomography (CT) and multispectral imaging – and expertise for the characterisation of animals, plants and materials.

To maximise research outcomes, as well as to facilitate collaboration and sharing, it is essential not only that the data acquired using these instruments be managed, curated and archived in a trusted data repository service, but also that the data itself be of verifiable quality. In 2017, several NIF nodes collaborated on a national project to define the requirements and best practices necessary to achieve this, and to establish exemplar services for both preclinical MRI data and clinical ataxia MRI data.

In this paper we describe the project, its key outcomes, challenges and lessons learned, and future developments, including extension to other characterisation facilities and instruments/modalities.

Meghini, Carlo. "Data Preservation." *Data Science Journal* 12 (2013): GRDI51-GRDI57. <https://doi.org/10.2481/dsj.grdi-009>

Digital information is a vital resource in our knowledge economy, valuable for research and education, science and the humanities, creative and cultural activities, and public policy (The Blue Ribbon Task Force on Sustainable Digital Preservation and Access, 2010). New high-throughput instruments, telescopes, satellites, accelerators, supercomputers, sensor networks, and running simulations are generating massive amounts of data (Thanos, 2011). These data are used by decision makers for improving the quality of life of citizens. Moreover, researchers are employing sophisticated technologies to analyse these data to address questions that were unapproachable just a few years ago (Helbing & Balietti, 2011). Digital technologies have fostered a new world of research characterized by immense datasets, unprecedented levels of openness among researchers, and new connections among researchers, policy makers, and the public (The National Academy of Sciences, 2009).

Michener, William K. "Ecological Data Sharing." *Ecological Informatics* 29, part 1 (2015): 33-44. <http://dx.doi.org/10.1016/j.ecoinf.2015.06.010>

Data sharing is the practice of making data available for use by others. Ecologists are increasingly generating and sharing an immense volume of data. Such data may serve to augment existing data collections and can be used for synthesis efforts such as meta-analysis, for parameterizing models, and for verifying research results (i.e., study reproducibility). Large volumes of ecological data may be readily available through institutions or data repositories that are the most comprehensive available and can serve as the core of ecological analysis. Ecological data are also employed outside the research context and are used for decision-making, natural resource management, education, and other purposes. Data sharing has a long history in many domains such as oceanography and the biodiversity sciences (e.g., taxonomic data and museum specimens), but has emerged relatively recently in the ecological sciences.

A review of several of the large international and national ecological research programs that have emerged since the mid-1900s highlights the initial failures and more recent successes as well as the underlying causes—from a near absence of effective policies to the emergence of community and data sharing policies coupled with the development and adoption of data and metadata standards and enabling tools. Sociocultural change and the move towards more open science have evolved more rapidly over the past two decades in response to new requirements set forth by governmental organizations, publishers and professional societies. As the scientific culture has changed so has the cyberinfrastructure landscape. The introduction of community-based data repositories, data and metadata standards, software tools, persistent identifiers, and federated search and discovery have all helped promulgate data sharing. Nevertheless, there are many challenges and opportunities especially as we move towards more open science. Cyberinfrastructure challenges include a paucity of easy-to-use metadata management systems, significant difficulties in assessing data quality and provenance, and an absence of analytical and visualization approaches that facilitate data integration and harmonization. Challenges and opportunities abound in the sociocultural arena where funders, researchers, and publishers all have a stake in clarifying policies, roles and responsibilities, as well as in incentivizing data sharing. A set of best practices and examples of software tools are presented that can enable research transparency, reproducibility and new knowledge by facilitating idea generation, research planning, data management and the dissemination of data and results.

Michener, William K., Suzie Allard, Amber Budden, Robert B. Cook, Kimberly Douglass, Mike Frame, Steve Kelling, Rebecca Koskela, Carol Tenopir, and David A. Vieglais. "Participatory Design of DataONE—Enabling Cyberinfrastructure for the Biological and Environmental Sciences." *Ecological Informatics* 11 (2012): 5-15. <http://dx.doi.org/10.1016/j.ecoinf.2011.08.007>

Michener, William K, and Matthew B. Jones. "Ecoinformatics: Supporting Ecology as a Data-Intensive Science." *Trends in Ecology & Evolution* 27, no. 2 (2012): 85-93. <http://dx.doi.org/10.1016/j.tree.2011.11.016>

Michener, William K., Todd Vision, Patricia Cruse, Dave Vieglais, John Kunze, and Greg Janée. "DataONE: Data Observation Network for Earth—Preserving Data and Enabling Innovation in the Biological and Environmental Sciences." *D-Lib Magazine* 17, no. 1/2 (2011). <https://doi.org/10.1045/january2011-michener>

Miksa, Tomasz, Andreas Rauber, Roman Ganguly, and Paolo Budroni.
"Information Integration for Machine Actionable Data Management Plans."
International Journal of Digital Curation 12, no. 1 (2017): 22-35.
<https://doi.org/10.2218/ijdc.v12i1.529>

Data management plans are free-form text documents describing the data used and produced in scientific experiments. The complexity of data-driven experiments requires precise descriptions of tools and datasets used in computations to enable their reproducibility and reuse. Data management plans fall short of these requirements. In this paper, we propose machine-actionable data management plans that cover the same themes as standard data management plans, but particular sections are filled with information obtained from existing tools. We present mapping of tools from the domains of digital preservation, reproducible research, open science, and data repositories to data management plan sections. Thus, we identify the requirements for a good solution and identify its limitations. We also propose a machine-actionable data model that enables information integration. The model uses ontologies and is based on existing standards.

Miksa, Tomasz, Stephan Strodl, and Andreas Rauber. "Process Management Plans." *International Journal of Digital Curation* 9, no. 1 (2014): 83-97.
<https://doi.org/10.2218/ijdc.v9i1.303>

In the era of research infrastructures and big data, sophisticated data management practices are becoming essential building blocks of successful science. Most practices follow a data-centric approach, which does not take into account the processes that created, analysed and presented the data. This fact limits the possibilities for reliable verification of results. Furthermore, it does not guarantee the reuse of research, which is one of the key aspects of credible data-driven science. For that reason, we propose the introduction of the new concept of Process Management Plans, which focus on the identification, description, sharing and preservation of the entire scientific processes. They enable verification and later reuse of result data and processes of scientific experiments. In this paper we describe the structure and explain the novelty of Process Management Plans by showing in what way they complement existing Data Management Plans. We also highlight key differences, major advantages, as well as references to tools and solutions that can facilitate the introduction of Process Management Plans.

Minor, David, Matt Critchlow, Arwen Hutt, Declan Fleming, Mary Linn Bergstrom, and Don Sutton. "Research Data Curation Pilots: Lessons Learned." *International Journal of Digital Curation* 9, no. 1 (2014): 220-230. <https://doi.org/10.2218/ijdc.v9i1.313>

In the spring of 2011, the UC San Diego Research Cyberinfrastructure (RCI) Implementation Team invited researchers and research teams to participate in a research curation and data management pilot program. This invitation took the form of a campus-wide solicitation. More than two dozen applications were received and, after due deliberation, the RCI Oversight Committee selected five curation-intensive projects. These projects were chosen based on a number of criteria, including how they represented campus research, varieties of topics, researcher engagement, and the various services required. The pilot process began in September 2011, and will be completed in early 2014. Extensive lessons learned from the pilots are being compiled and are being used in the on-going design and implementation of the permanent Research Data Curation Program in the UC San Diego Library.

In this paper, we present specific implementation details of these various services, as well as lessons learned. The program focused on many aspects of contemporary scholarship, including data creation and storage, description and metadata creation, citation and publication, and long term preservation and access. Based on the lessons learned in our processes, the Research Data Curation Program will provide a suite of services from which campus users can pick and choose, as necessary. The program will provide support for the data management requirements from national funding agencies.

Minor, David, Don Sutton, Ardys Kozbial, Brad Westbrook, Michael Burek, and Michael Smorul. "Chronopolis Digital Preservation Network." *International Journal of Digital Curation* 5, no. 1 (2010). <https://doi.org/10.2218/ijdc.v5i1.147>

Mischo, William H., Mary C. Schlembach, and Megan N. O'Donnell. "An Analysis of Data Management Plans in University of Illinois National Science Foundation Grant Proposals." *Journal of eScience Librarianship* 3, no. 1 (2014): e1060. <http://dx.doi.org/10.7191/jeslib.2014.1060>

Missier, Paolo. "Data Trajectories: Tracking Reuse of Published Data for Transitive Credit Attribution." *International Journal of Digital Curation* 11, no. 1 (2016): 1-16. <https://doi.org/10.2218/ijdc.v11i1.425>

The ability to measure the use and impact of published data sets is key to the success of the open data/open science paradigm. A direct measure of impact would require tracking data (re)use in the wild, which is difficult to achieve. This is therefore commonly replaced by simpler metrics based on data download and citation counts. In this paper we describe a scenario where it is possible to track the trajectory of a dataset after its publication, and show how this enables the design of accurate models for ascribing credit to data originators. A Data Trajectory (DT) is a graph that encodes knowledge of how, by whom, and in which context data has been re-used, possibly after several generations. We provide a theoretical model of DTs that is grounded in the W3C PROV data model for provenance, and we show how DTs can be used to automatically propagate a fraction of the credit associated with transitively derived datasets, back to original data contributors. We also show this model of transitive credit in action by means of a Data Reuse Simulator. In the longer term, our ultimate hope is that credit models based on direct measures of data reuse will provide further incentives to data publication. We conclude by outlining a research agenda to address the hard questions of creating, collecting, and using DTs systematically across a large number of data reuse instances in the wild.

Missier, Paolo, Bertram Ludäscher, Saumen Dey, Michael Wang, Tim McPhillips, Shawn Bowers, Michael Agun, and Ilkay Altintas. "Golden Trail: Retrieving the Data History That Matters from a Comprehensive Provenance Repository." *International Journal of Digital Curation* 7, no. 1 (2012): 139-150.
<https://doi.org/10.2218/ijdc.v7i1.221>

Experimental science can be thought of as the exploration of a large research space, in search of a few valuable results. While it is this "Golden Data" that gets published, the history of the exploration is often as valuable to the scientists as some of its outcomes. We envision an e-research infrastructure that is capable of systematically and automatically recording such history—an assumption that holds today for a number of workflow management systems routinely used in e-science. In keeping with our gold rush metaphor, the provenance of a valuable result is a "Golden Trail". Logically, this represents a detailed account of how the Golden Data was arrived at, and technically it is a sub-graph in the much larger graph of provenance traces that collectively tell the story of the entire research (or of some of it).

In this paper we describe a model and architecture for a repository dedicated to storing provenance traces and selectively retrieving Golden Trails from it.

As traces from multiple experiments over long periods of time are accommodated, the trails may be sub-graphs of one trace, or they may be the logical representation of a virtual experiment obtained by joining together traces that share common data.

The project has been carried out within the Provenance Working Group of the Data Observation Network for Earth (DataONE) NSF project. Ultimately, our longer-term plan is to integrate the provenance repository into the data preservation architecture currently being developed by DataONE.

Mitchell, Erik T. "Research Support: The New Mission for Libraries." *Journal of Web Librarianship* 7, no. 1 (2013): 109-113.

<https://doi.org/10.1080/19322909.2013.757930>

Mohr, Alicia Hofelich, Josh Bishoff, Carolyn Bishoff, Steven Braun, Christine Storino, and Lisa R. Johnston. "When Data Is a Dirty Word: A Survey to Understand Data Management Needs Across Diverse Research Disciplines." *Bulletin of the Association for Information Science and Technology* 42, no. 1 (2015): 51-53.

<https://onlinelibrary.wiley.com/doi/full/10.1002/bul2.2015.1720420114>

Molloy, Laura. "Digital Curation Skills in the Performing Arts—An Investigation of Practitioner Awareness and Knowledge of Digital Object Management and Preservation." *International Journal of Performance Arts and Digital Media* 10, no. 1 (2014): 7-20. <https://doi.org/10.1080/14794713.2014.912496>

Molloy, Laura, Simon Hodson, Meik Poschen, and Jonathan Tedds. "Gathering Evidence of Benefits: A Structured Approach from the JISC Managing Research Data Programme." *International Journal of Digital Curation* 8, no. 2 (2013): 123-133. <https://doi.org/10.2218/ijdc.v8i2.277>

The work of the Jisc Managing Research Data programme is—along with the rest of the UK higher education sector—taking place in an environment of increasing pressure on research funding. In order to justify the investment made by Jisc in this activity—and to help make the case more widely for the value of investing time and money in research data management—individual projects and the programme as a whole must be able to clearly express the resultant benefits to the host institutions and to the broader sector. This paper describes a structured approach to the measurement and description of benefits provided by the work of these projects for the benefit of funders, institutions and researchers. We outline the context of the programme and its

work; discuss the drivers and challenges of gathering evidence of benefits; specify benefits as distinct from aims and outputs; present emerging findings and the types of metrics and other evidence which projects have provided; explain the value of gathering evidence in a structured way to demonstrate benefits generated by work in this field; and share lessons learned from progress to date.

Molloy, Laura, and Kellie Snow. "The Data Management Skills Support Initiative: Synthesising Postgraduate Training in Research Data Management." *International Journal of Digital Curation* 7, no. 2 (2012): 101-109.

<https://doi.org/10.2218/ijdc.v7i2.233>

This paper will describe the efforts and findings of the JISC Data Management Skills Support Initiative ('DaMSSI'). DaMSSI was co-funded by the JISC Managing Research Data programme and the Research Information Network (RIN), in partnership with the Digital Curation Centre, to review, synthesise and augment the training offerings of the JISC Research Data Management Training Materials ('RDMTrain') projects.

DaMSSI tested the effectiveness of the Society of College, National and University Libraries' Seven Pillars of Information Literacy model (SCONUL, 2011), and Vitae's Researcher Development Framework ('Vitae RDF') for consistently describing research data management ('RDM') skills and skills development paths in UK HEI postgraduate courses.

With the collaboration of the RDMTrain projects, we mapped individual course modules to these two models and identified basic generic data management skills alongside discipline-specific requirements. A synthesis of the training outputs of the projects was then carried out, which further investigated the generic versus discipline-specific considerations and other successful approaches to training that had been identified as a result of the projects' work. In addition we produced a series of career profiles to help illustrate the fact that data management is an essential component—in obvious and not-so-obvious ways—of a wide range of professions.

We found that both models had potential for consistently and coherently describing data management skills training and embedding this within broader institutional postgraduate curricula. However, we feel that additional discipline-specific references to data management skills could also be beneficial for effective use of these models. Our synthesis work identified that

the majority of core skills were generic across disciplines at the postgraduate level, with the discipline-specific approach showing its value in engaging the audience and providing context for the generic principles.

Findings were fed back to SCOUNL and Vitae to help in the refinement of their respective models, and we are working with a number of other projects, such as the DCC and the EC-funded Digital Curator Vocational Education Europe (DigCurV2) initiative, to investigate ways to take forward the training profiling work we have begun.

Mongeon, Philippe, Robinson-Garcia Nicolas, Jeng Wei, and Costas Rodrigo. "Incorporating Data Sharing to the Reward System of Science: Linking DataCite Records to Authors in the Web of Science." *Aslib Journal of Information Management* 69, no. 5 (2017): 545-556. <https://doi.org/10.1108/AJIM-01-2017-0024>

Mons, Barend. *Data Stewardship for Open Science: Implementing FAIR Principles*. New York: Chapman and Hall/CRC, 2018. <https://doi.org/10.1201/9781315380711>

Moon, Jeff. "Developing a Research Data Management Service—A Case Study." *Partnership: the Canadian Journal of Library and Information Practice and Research* 9, no. 1 (2014). <https://doi.org/10.21083/partnership.v9i1.2988>

Mooney, Hailey, and Mark P. Newton. "The Anatomy of a Data Citation: Discovery, Reuse, and Credit." *Journal of Librarianship and Scholarly Communication* 1, no. 1 (2012). <http://dx.doi.org/10.7710/2162-3309.1035>

INTRODUCTION Data citation should be a necessary corollary of data publication and reuse. Many researchers are reluctant to share their data, yet they are increasingly encouraged to do just that. Reward structures must be in place to encourage data publication, and citation is the appropriate tool for scholarly acknowledgment. Data citation also allows for the identification, retrieval, replication, and verification of data underlying published studies. **METHODS** This study examines author behavior and sources of instruction in disciplinary and cultural norms for writing style and citation via a content analysis of journal articles, author instructions, style manuals, and data publishers. Instances of data citation are benchmarked against a Data Citation Adequacy Index. **RESULTS** Roughly half of journals point toward a style manual that addresses data citation, but the majority of journal articles failed to include an adequate citation to data used in secondary analysis studies.

DISCUSSION Full citation of data is not currently a normative behavior in scholarly writing. Multiplicity of data types and lack of awareness regarding existing standards contribute to the problem. CONCLUSION Citations for data must be promoted as an essential component of data publication, sharing, and reuse. Despite confounding factors, librarians and information professionals are well-positioned and should persist in advancing data citation as a normative practice across domains. Doing so promotes a value proposition for data sharing and secondary research broadly, thereby accelerating the pace of scientific research.

Moore, Reagan W. "Building Preservation Environments with Data Grid Technology." *American Archivist* 69, no. 1 (2007): 139-158. <https://doi.org/10.17723/aarc.69.1.176p5112w5278567>

———. "Geospatial Web Services and Geoarchiving: New Opportunities and Challenges in Geographic Information Service." *Library Trends* 55, no. 2 (2006): 285-303. <https://doi.org/10.1353/lib.2006.0059>

Morris, Chris. "The Life Cycle of Structural Biology Data." *Data Science Journal* 17, no. 26 (2018). <http://doi.org/10.5334/dsj-2018-026>

Research data is acquired, interpreted, published, reused, and sometimes eventually discarded. Understanding this life cycle better will help the development of appropriate infrastructural services, ones which make it easier for researchers to preserve, share, and find data.

Structural biology is a discipline within the life sciences, one that investigates the molecular basis of life by discovering and interpreting the shapes and motions of macromolecules. Structural biology has a strong tradition of data sharing, expressed by the founding of the Protein Data Bank (PDB) in 1971. The culture of structural biology is therefore already in line with the perspective that data from publicly funded research projects are public data.

This review is based on the data life cycle as defined by the UK Data Archive. It identifies six stages: creating data, processing data, analysing data, preserving data, giving access to data, and re-using data. For clarity, 'preserving data' and 'giving access to data' are discussed together. A final stage to the life cycle, 'discarding data', is also discussed.

The review concludes with recommendations for future improvements to the IT infrastructure for structural biology.

Morris, Steven. "The North Carolina Geospatial Data Archiving Project: Challenges and Initial Outcomes." *Journal of Map & Geography Libraries* 6, no. 1 (2009): 26-44. <https://doi.org/10.1080/15420350903432507>

Morris, Steven, James Tuttle, and Jefferson Essic. "A Partnership Framework for Geospatial Data Preservation in North Carolina." *Library Trends* 57, no. 3 (2009): 516-540. <https://doi.org/10.1353/lib.0.0050>

Mothersky, Jessica, Heidi Walsh, Meredith Parsons, Tristan McIntosh, Kari Baldwin, and James M DuBois. "Are We Ready to Share Qualitative Research Data? Knowledge and Preparedness among Qualitative Researchers, IRB Members, and Data Repository Curators." *IASSIST Quarterly* 43, no. 4 (2019) 1–23. <https://doi.org/10.29173/iq952>

Murillo, Angela P. "Data at Risk Initiative: Examining and Facilitating the Scientific Process in Relation to Endangered Data." *Data Science Journal* 12 (2014): 207-219 <https://doi.org/10.2481/dsj.12-048>

Examining the scientific process in relation to endangered data, data reuse, and sharing is crucial in facilitating scientific workflow. Deterioration, format obsolescence, and insufficient metadata for discovery are significant problems leading to loss of scientific data. The research presented in this paper considers these potentially lost data. Four one-hour focus groups and a demographic survey were conducted with 14 scientists to learn about their attitudes toward endangered data, data sharing, data reuse, and their opinions of the DARI inventory. The results indicate that unavailability, lack of context, accessibility issues, and potential endangerment are key concerns to scientists.

Murphy, Fiona. "Data and Scholarly Publishing: The Transforming Landscape." *Learned Publishing* 27, no. 5 (2014): 3-7. <https://doi.org/10.1087/20140502>

———. "An Update on Peer Review and Research Data." *Learned Publishing* 29, no. 1 (2016): 51-53. <https://doi.org/10.1002/leap.1005>

Murray, Matthew, Megan O'Donnell, Mark J. Laufersweiler, John Novak, Betty Rozum, and Santi Thompson. "A Survey of the State of Research Data Services in 35 U.S. Academic Libraries, or 'Wow, What a Sweeping Question!'" *Research Ideas and Outcomes* 5 (2009): e48809. <https://doi.org/10.3897/rio.5.e48809>

Musgrave, Simon. "Improving Access to Recorded Language Data." *D-Lib Magazine* 20, no. 1/2 (2014). <https://doi.org/10.1045/january2014-musgrave>

National Academy of Sciences Committee on Ensuring the Utility and Integrity of Research Data in a Digital Age. *Ensuring the Integrity, Accessibility, and Stewardship of Research Data in the Digital Age*. Washington, DC: National Academies Press, 2009. http://www.nap.edu/catalog.php?record_id=12615

National Research Council Committee on Archiving and Accessing Environmental and Geospatial Data at NOAA. *Environmental Data Management at NOAA: Archiving, Stewardship, and Access*. Washington, DC: National Academies Press, 2007. http://www.nap.edu/catalog.php?record_id=12017

National Science Board. *Long-Lived Digital Data Collections Enabling Research and Education in the 21st Century*. Washington, DC: National Science Foundation, 2005. <http://www.nsf.gov/pubs/2005/nsb0540/>

Naughton, Linda, and David Kernohan. "Making Sense of Journal Research Data Policies." *Insights* 29, no. 1 (2016): 84-89. <http://doi.org/10.1629/uksg.284>

This article gives an overview of the findings from the first phase of the Jisc Journal Research Data Policy Registry pilot (JRDPR), which is currently under way. The project continues from the initial study, 'Journal of Research Data policy bank' (JoRD), carried out by Nottingham University's Centre for Research Communication from 2012 to 2014. The project undertook an analysis of 250 journal research data policies to assess the feasibility of developing a policy registry to assist researchers and support staff to comply with research data publication requirements. The evidence shows that the current research data policy ecosystem is in critical need of standardization and harmonization if such services are to be built and implemented. To this end, the article proposes the next steps for the project with the objective of ultimately moving towards a modern research infrastructure based on machine-readable policies that support a more open scholarly communications environment.

Naum, Alexandra. "Research Data Storage and Management: Library Staff Participation in Showcasing Research Data at the University of Adelaide." *The Australian Library Journal* 63, no. 1 (2014): 35-44. <http://dx.doi.org/10.1080/00049670.2014.890019>

Navale, Vivek, and McAuliffe M. "Long-Term Preservation of Biomedical Research Data." *F1000Research* 7 (2018): 1353
<https://doi.org/10.12688/f1000research.16015.1>

Genomics and molecular imaging, along with clinical and translational research have transformed biomedical science into a data-intensive scientific endeavor. For researchers to benefit from Big Data sets, developing long-term biomedical digital data preservation strategy is very important. In this opinion article, we discuss specific actions that researchers and institutions can take to make research data a continued resource even after research projects have reached the end of their lifecycle. The actions involve utilizing an Open Archival Information System model comprised of six functional entities: Ingest, Access, Data Management, Archival Storage, Administration and Preservation Planning. We believe that involvement of data stewards early in the digital data life-cycle management process can significantly contribute towards long term preservation of biomedical data. Developing data collection strategies consistent with institutional policies, and encouraging the use of common data elements in clinical research, patient registries and other human subject research can be advantageous for data sharing and integration purposes. Specifically, data stewards at the onset of research program should engage with established repositories and curators to develop data sustainability plans for research data. Placing equal importance on the requirements for initial activities (e.g., collection, processing, storage) with subsequent activities (data analysis, sharing) can improve data quality, provide traceability and support reproducibility. Preparing and tracking data provenance, using common data elements and biomedical ontologies are important for standardizing the data description, making the interpretation and reuse of data easier. The Big Data biomedical community requires scalable platform that can support the diversity and complexity of data ingest modes (e.g. machine, software or human entry modes). Secure virtual workspaces to integrate and manipulate data, with shared software programs (e.g., bioinformatics tools), can facilitate the FAIR (Findable, Accessible, Interoperable and Reusable) use of data for near- and long-term research needs.

Ndhlovu, Phillip. "The State of Preparedness for Digital Curation and Preservation: A Case Study of a Developing Country Academic Library." *IASSIST Quarterly* 42 No 3 (2018). <https://doi.org/10.29173/iq929>

Neuroth, Heike, Felix Lohmeier, and Kathleen Marie Smith. "TextGrid—Virtual Research Environment for the Humanities." *International Journal of Digital Curation* 6, no. 2 (2011): 222-231. <https://doi.org/10.2218/ijdc.v6i2.198>

Neylon, Cameron. "Compliance Culture or Culture Change? The Role of Funders in Improving Data Management and Sharing Practice amongst Researchers." *Research Ideas and Outcomes* 3 (2017): e14673. <https://doi.org/10.3897/rio.3.e14673>

Nicholl, Natsuko H., Sara M. Samuel, Leena N. Lalwani, Paul F. Grochowski, and Jennifer A. Green. "Resources to Support Faculty Writing Data Management Plans: Lessons Learned from an Engineering Pilot." *International Journal of Digital Curation* 9, no. 1 (2014): 242-252. <https://doi.org/10.2218/ijdc.v9i1.315>

Recent years have seen a growing emphasis on the need for improved management of research data. Academic libraries have begun to articulate the conceptual foundations, roles, and responsibilities involved in data management planning and implementation. This paper provides an overview of the Engineering data support pilot at the University of Michigan Library as part of developing new data services and infrastructure. Through this pilot project, a team of librarians had an opportunity to identify areas where the library can play a role in assisting researchers with data management, and has put forth proposals for immediate steps that the library can take in this regard. The paper summarizes key findings from a faculty survey and discusses lessons learned from an analysis of data management plans from accepted NSF proposals. A key feature of this Engineering pilot project was to ensure that these study results will provide a foundation for librarians to educate and assist researchers with managing their data throughout the research lifecycle.

Nicholson, Shawn W., and Terrence B. Bennett. "Data Sharing: Academic Libraries and the Scholarly Enterprise." *portal: Libraries and the Academy* 11, no. 1 (2011): 505-516. <http://muse.jhu.edu/article/409890>

———. "The Good News About Bad News: Communicating Data Services to Cognitive Misers." *Journal of Electronic Resources Librarianship* 29, no. 3 (2017): 151-158. <https://doi.org/10.1080/1941126X.2017.1340720>

Nielsen, Hans Jørn, and Birger Hjørland. "Curating Research Data: The Potential Roles of Libraries and Information Professionals." *Journal of Documentation* 70, no. 2 (2014): 221-240. <https://doi.org/10.1108/jd-03-2013-0034>

Nitecki, Danuta A., and Mary Ellen K. Davis. "Expanding Academic Librarians' Roles in the Research Life Cycle." *Libri* 69, no. 2 (2019): 117-125.
<https://doi.org/10.1515/libri-2018-0066>

Niua, Jinfang. "Aggregate Control of Scientific Data." *Archives and Records: The Journal of the Archives and Records Association* 37, no. 1 (2016): 53-64.
<https://doi.org/10.1080/23257962.2016.1145578>

Noonan, Daniel, and Tamar Chute. "Data Curation and the University Archives." *The American Archivist* 77, no. 1 (2014):201-240.
<http://dx.doi.org/10.17723/aarc.77.1.m49r46526847g587>

Norman, Belinda, and Kate Valentine Stanton. "From Project to Strategic Vision: Taking the Lead in Research Data Management Support at the University of Sydney Library." *International Journal of Digital Curation* 9, no. 1 (2014): 253-262. <https://doi.org/10.2218/ijdc.v9i1.316>

This paper explores three stories, each occurring a year apart, illustrating an evolution toward a strategic vision for Library leadership in supporting research data management at the University of Sydney. The three stories describe activities undertaken throughout the Seeding the Commons project and beyond, as the establishment of ongoing roles and responsibilities transition the Library from project partner to strategic leader in the delivery of research data management support. Each story exposes key ingredients that characterise research data management support: researcher engagement; partnerships; and the complementary roles of policy and practice.

Norman, Hazel. "Mandating Data Archiving: Experiences from the Frontline." *Learned Publishing* 27, no. 5 (2014): 35-38. <https://doi.org/10.1087/20140507>

Norris, Timothy B., and Christopher C. Mader. "Data Curation for Big Interdisciplinary Science: The Pulley Ridge Experience." *Journal of eScience Librarianship* 8, no. 2 (2019): e1172. <https://doi.org/10.7191/jeslib.2019.1172>.

The curation and preservation of scientific data has long been recognized as an essential activity for the reproducibility of science and the advancement of knowledge. While investment into data curation for specific disciplines and at individual research institutions has advanced the ability to preserve research data products, data curation for big interdisciplinary science remains relatively unexplored terrain. To fill this lacunae, this article presents a case study of the data curation for the National Centers for Coastal Ocean Science

(NCCOS) funded project “Understanding Coral Ecosystem Connectivity in the Gulf of Mexico-Pulley Ridge to the Florida Keys” undertaken from 2011 to 2018 by more than 30 researchers at several research institutions. The data curation process is described and a discussion of strengths, weaknesses and lessons learned is presented. Major conclusions from this case study include: the reimplementation of data repository infrastructure builds valuable institutional data curation knowledge but may not meet data curation standards and best practices; data from big interdisciplinary science can be considered as a special collection with the implication that metadata takes the form of a finding aid or catalog of datasets within the larger project context; and there are opportunities for data curators and librarians to synthesize and integrate results across disciplines and to create exhibits as stories that emerge from interdisciplinary big science.

Norton, Hannah F., Michele R. Tennant, Cecilia Botero, and Rolando Garcia-Milian. "Assessment of and Response to Data Needs of Clinical and Translational Science Researchers and Beyond." *Journal of eScience Librarianship* 5, no. 1 (2016): e1090. <https://doi.org/10.7191/jeslib.2016.1090>

Objective and Setting: As universities and libraries grapple with data management and "big data," the need for data management solutions across disciplines is particularly relevant in clinical and translational science (CTS) research, which is designed to traverse disciplinary and institutional boundaries. At the University of Florida Health Science Center Library, a team of librarians undertook an assessment of the research data management needs of CTS researchers, including an online assessment and follow-up one-on-one interviews.

Design and Methods: The 20-question online assessment was distributed to all investigators affiliated with UF's Clinical and Translational Science Institute (CTSI) and 59 investigators responded. Follow-up in-depth interviews were conducted with nine faculty and staff members.

Results: Results indicate that UF's CTS researchers have diverse data management needs that are often specific to their discipline or current research project and span the data lifecycle. A common theme in responses was the need for consistent data management training, particularly for graduate students; this led to localized training within the Health Science Center and CTSI, as well as campus-wide training. Another campus-wide outcome was the creation of an action-oriented Data Management/Curation

Task Force, led by the libraries and with participation from Research Computing and the Office of Research.

Conclusions: Initiating conversations with affected stakeholders and campus leadership about best practices in data management and implications for institutional policy shows the library's proactive leadership and furthers our goal to provide concrete guidance to our users in this area.

Ogburn, Joyce L. "The Imperative for Data Curation." *portal: Libraries & the Academy* 10, no. 2 (2010): 241-246. <https://doi.org/10.1353/pla.0.0100>

Olendorf, Robert, and Steve Koch. "Beyond the Low Hanging Fruit: Data Services and Archiving at the University of New Mexico." *Journal of Digital Information* 13, no. 1 (2012). <http://journals.tdl.org/jodi/index.php/jodi/article/view/5878>

O'Malley, Donna L. "Gaining Traction in Research Data Management Support: A Case Study." *Journal of eScience Librarianship* 3, no. 1 (2014): e1059. <http://dx.doi.org/10.7191/jeslib.2014.1059>

Oostdijk, Nelleke, Henk van den Heuvel, and Maaske Treurniet. "The CLARIN-NL Data Curation Service: Bringing Data to the Foreground." *International Journal of Digital Curation* 8, no. 2 (2013): 134-145. <https://doi.org/10.2218/ijdc.v8i2.278>

After decades in which a great deal of effort was spent on the creation of resources, there are currently several initiatives worldwide that aim to create an interoperable, sustainable research infrastructure. An integral part of such an infrastructure constitutes the resources (data and tools) which researchers in the various disciplines employ. Whether the infrastructure will be successful in supporting the needs of the research communities it intends to cater for depends on a number of factors. One factor is that resources that are or could be relevant to the wider research community are made visible through this infrastructure and, to the greatest extent possible, accessible and usable. In practice, the durable availability of resources is often not properly regulated within research projects.

CLARIN-NL is directed at creating an interoperable language resources infrastructure for the humanities in the Netherlands. The Data Curation Service was established in order to salvage language resources in this field that are threatened to be lost. In the CLARIN context, a great deal of attention is given to standards, formats and intellectual property rights. Consequently,

the Data Curation Service (DCS) has a role as mediator in bringing researchers in the field of humanities and existing data centres closer together.

This article consists of two parts: the first part provides the background to the work of the DCS while the second part illustrates the work of the DCS by describing the actual curation of a collection of language learner data.

Palaiologk, Anna S., Anastasios A. Economides, Heiko D. Tjalsma, and Laurents B. Sesink. "An Activity-Based Costing Model for Long-Term Preservation and Dissemination of Digital Research Data: The Case of DANS." *International Journal on Digital Libraries* 12, no. 4 (2012): 195-214.
<https://doi.org/10.1007/s00799-012-0092-1>

Palmer, Carole L., Bryan P. Heidorn, Dan Wright, and Melissa H. Cragin. "Graduate Curriculum for Biological Information Specialists: A Key to Integration of Scale in Biology." *International Journal of Digital Curation* 2, no. 2 (2007): 31-40. <https://doi.org/10.2218/ijdc.v2i2.27>

Scientific data problems do not stand in isolation. They are part of a larger set of challenges associated with the escalation of scientific information and changes in scholarly communication in the digital environment. Biologists in particular are generating enormous sets of data at a high rate, and new discoveries in the biological sciences will increasingly depend on the integration of data across multiple scales. This work will require new kinds of information expertise in key areas. To build this professional capacity we have developed two complementary educational programs: a Biological Information Specialist (BIS) masters degree and a concentration in Data Curation (DC). We believe that BISs will be central in the development of cyberinfrastructure and information services needed to facilitate interdisciplinary and multi-scale science. Here we present three sample cases from our current research projects to illustrate areas in which we expect information specialists to make important contributions to biological research practice.

Palmer, Carole L., Nicholas M. Weber, Trevor Muñoz, and Allen H. Renear. "Foundations of Data Curation: The Pedagogy and Practice of 'Purposeful Work' with Research Data." *Archive Journal*, no. 3 (2013).
<http://www.archivejournal.net/issue/3/archives-remixed/foundations-of-data-curation-the-pedagogy-and-practice-of-purposeful-work-with-research-data/>

Palumbo, Laura B., Ron Jantz, Yu-Hung Lin, Aletia Morgan, Minglu Wang, Krista White, Ryan Womack, Yingting Zhang, and Yini Zhu. "Preparing to Accept Research Data: Creating Guidelines for Librarians." *Journal of eScience Librarianship* 4, no. 2 (2015): e1080. <https://doi.org/10.7191/jeslib.2015.1080>

Parham, Susan Wells, Jon Bodnar, and Sara Fuchs. "Supporting Tomorrow's Research: Assessing Faculty Data Curation Needs at Georgia Tech." *College & Research Libraries News* 73, no. 1 (2012): 10-13. <https://doi.org/10.5860/crln.73.1.8686>

Parham, Susan Wells, Jake Carlson, Patricia Hswe, Brian Westra, and Amanda Whitmire. "Using Data Management Plans to Explore Variability in Research Data Management Practices Across Domains." *International Journal of Digital Curation* 11, no. 1 (2016): 53-67. <https://doi.org/10.2218/ijdc.v11i1.423>

This paper describes an investigation into how researchers in different fields are interpreting and responding to the U.S. National Science Foundation's data management plan (DMP) requirement. As documents written by the researchers themselves, DMPs can provide insight into researchers' understanding of the potential value of their data to others; the environment in which their data are developed and prepared; and their willingness and ability to ensure the data are available to others now and in the long-term. With support from the Institute of Museum and Library Services, the authors conducted a content analysis of DMPs generated at their respective institutions using a shared rubric. By developing and testing a rubric designed to understand and evaluate the content of DMPs, the authors intend to develop a more complete understanding, at a larger scale, of how researchers plan for managing, sharing, and archiving their data.

Parham, Susan Wells, and Chris Doty. "NSF DMP Content Analysis: What Are Researchers Saying?" *Bulletin of the American Society for Information Science and Technology* 39, no. 1 (2012): 37-38. <http://doi.org/10.1002/bult.2012.1720390113>

Park, Hyoungjoo, and Dietmar Wolfram. "An Examination of Research Data Sharing and Re-Use: Implications for Data Citation Practice." *Scientometrics* 111, no. 1 (2017): 443-461. <https://doi.org/10.1007/s11192-017-2240-2>

Parsons, M., and P. Fox. "Is Data Publication the Right Metaphor?" *Data Science Journal* 12 (2013): WDS32-WDS46. <https://doi.org/10.2481/dsj.wds-042>

International attention to scientific data continues to grow. Opportunities emerge to re-visit long-standing approaches to managing data and to critically examine new capabilities. We describe the cognitive importance of metaphor. We describe several metaphors for managing, sharing, and stewarding data and examine their strengths and weaknesses. We particularly question the applicability of a "publication" approach to making data broadly available. Our preliminary conclusions are that no one metaphor satisfies enough key data system attributes and that multiple metaphors need to co-exist in support of a healthy data ecosystem. We close with proposed research questions and a call for continued discussion.

Parsons, Mark A. "Organizational Status of RDA." *D-Lib Magazine* 20, no. 1/2 (2014). <https://doi.org/10.1045/january2014-parsons>

Parsons, Rebecca, and Summers, Scott. "The Research Data Alliance: Implementing the Technology, Practice and Connections of a Data Infrastructure." *Bulletin of the American Society for Information Science and Technology* 39, no. 6 (2013): 33-36. <https://doi.org/10.1002/bult.2013.1720390611>

———. "The Role of Case Studies in Effective Data Sharing, Reuse and Impact." *IASSIST Quarterly* 40 no. 3 (2017) 14. <https://doi.org/10.29173/iq397>

Parsons, Thomas. "Creating a Research Data Management Service." *International Journal of Digital Curation* 8, no. 2 (2013): 146-156. <https://doi.org/10.2218/ijdc.v8i2.279>

This paper provides an overview of the elements required to create a sustainable research data management (RDM) service. The paper summarises key learning and lessons learnt from the University of Nottingham's project to create an RDM service for researchers. Collective experiences and learning from three key areas are covered, including: data management requirements gathering and validation, RDM training, and the creation of an RDM website.

Pascuzzi, Pete E., and Megan R. Sapp Nelson. "Integrating Data Science Tools into a Graduate Level Data Management Course." *Journal of eScience Librarianship* 7, no. 3 (2018): e1152. <https://doi.org/10.7191/jeslib.2018.1152>

Objective: This paper describes a project to revise an existing research data management (RDM) course to include instruction in computer skills with robust data science tools.

Setting: A Carnegie R1 university.

Brief Description: Graduate student researchers need training in the basic concepts of RDM. However, they generally lack experience with robust data science tools to implement these concepts holistically. Two library instructors fundamentally redesigned an existing research RDM course to include instruction with such tools. The course was divided into lecture and lab sections to facilitate the increased instructional burden. Learning objectives and assessments were designed at a higher order to allow students to demonstrate that they not only understood course concepts but could use their computer skills to implement these concepts.

Results: Twelve students completed the first iteration of the course. Feedback from these students was very positive, and they appreciated the combination of theoretical concepts, computer skills and hands-on activities. Based on student feedback, future iterations of the course will include more "flipped" content including video lectures and interactive computer tutorials to maximize active learning time in both lecture and lab.

Pasquetto, Irene V., Christine L. Borgman, and Morgan F. Wofford. "Uses and Reuses of Scientific Data: The Data Creators' Advantage." *Harvard Data Science Review* 1, no.2 (2019). <https://doi.org/10.1162/99608f92.fc14bf2d>

Open access to data, as a core principle of open science, is predicated on assumptions that scientific data can be reused by other researchers. We test those assumptions by asking where scientists find reusable data, how they reuse those data, and how they interpret data they did not collect themselves. By conducting a qualitative meta-analysis of evidence on two long-term, distributed, interdisciplinary consortia, we found that scientists frequently sought data from public collections and from other researchers for comparative purposes such as "ground-truthing" and calibration. When they sought others' data for reanalysis or for combining with their own data, which was relatively rare, most preferred to collaborate with the data creators. We propose a typology of data reuses ranging from comparative to integrative. Comparative data reuse requires interactional expertise, which involves knowing enough about the data to assess their quality and value for a specific comparison such as calibrating an instrument in a lab experiment. Integrative reuse requires contributory expertise, which involves the ability to perform the action, such as reusing data in a new experiment. Data integration requires more specialized scientific knowledge and deeper levels of epistemic trust in

the knowledge products. Metadata, ontologies, and other forms of curation benefit interpretation for any kind of data reuse. Based on these findings, we theorize the data creators' advantage, that those who create data have intimate and tacit knowledge that can be used as barter to form collaborations for mutual advantage. Data reuse is a process that occurs within knowledge infrastructures that evolve over time, encompassing expertise, trust, communities, technologies, policies, resources, and institutions.

Pasquetto, Irene V., Bernadette M. Randles, and Christine L. Borgman. "On the Reuse of Scientific Data." *Data Science Journal* 16, no. 8 (2017).

<http://doi.org/10.5334/dsj-2017-008>

While science policy promotes data sharing and open data, these are not ends in themselves. Arguments for data sharing are to reproduce research, to make public assets available to the public, to leverage investments in research, and to advance research and innovation. To achieve these expected benefits of data sharing, data must actually be reused by others. Data sharing practices, especially motivations and incentives, have received far more study than has data reuse, perhaps because of the array of contested concepts on which reuse rests and the disparate contexts in which it occurs. Here we explicate concepts of data, sharing, and open data as a means to examine data reuse. We explore distinctions between use and reuse of data. Lastly we propose six research questions on data reuse worthy of pursuit by the community: How can uses of data be distinguished from reuses? When is reproducibility an essential goal? When is data integration an essential goal? What are the tradeoffs between collecting new data and reusing existing data? How do motivations for data collection influence the ability to reuse data? How do standards and formats for data release influence reuse opportunities? We conclude by summarizing the implications of these questions for science policy and for investments in data reuse.

Patel, Dimple. "Research Data Management: A Conceptual Framework." *Library Review* 65, no. 4/5 (2016): 226-241. <http://dx.doi.org/10.1108/LR-01-2016-0001>

Patel, Manjula, and Alexander Ball. "Challenges and Issues Relating to the Use of Representation Information for the Digital Curation of Crystallography and Engineering Data." *International Journal of Digital Curation* 3, no. 1 (2008): 76-88. <https://doi.org/10.2218/ijdc.v3i1.43>

Peer, Limor, and Ann Green. "Building an Open Data Repository for a Specialized Research Community: Process, Challenges and Lessons." *International Journal of Digital Curation* 7, no. 1 (2012): 151-162. <https://doi.org/10.2218/ijdc.v7i1.222>

In 2009, the Institution for Social and Policy Studies (ISPS) at Yale University began building an open access digital collection of social science experimental data, metadata, and associated files produced by ISPS researchers. The digital repository was created to support the replication of research findings and to enable further data analysis and instruction. Content is submitted to a rigorous process of quality assessment and normalization, including transformation of statistical code into R, an open source statistical software. Other requirements included: (a) that the repository be integrated with the current database of publications and projects publicly available on the ISPS website; (b) that it offered open access to datasets, documentation, and statistical software program files; (c) that it utilized persistent linking services and redundant storage provided within the Yale Digital Commons infrastructure; and (d) that it operated in accordance with the prevailing standards of the digital preservation community. In partnership with Yale's Office of Digital Assets and Infrastructure (ODAI), the ISPS Data Archive was launched in the fall of 2010. We describe the process of creating the repository, discuss prospects for similar projects in the future, and explain how this specialized repository fits into the larger digital landscape at Yale.

Peer, Limor, Ann Green, and Elizabeth Stephenson. "Committing to Data Quality Review." *International Journal of Digital Curation* 9, no. 1 (2014): 263-291. <https://doi.org/10.2218/ijdc.v9i1.317>

Amid the pressure and enthusiasm for researchers to share data, a rapidly growing number of tools and services have emerged. What do we know about the quality of these data? Why does quality matter? And who should be responsible for data quality? We believe an essential measure of data quality is the ability to engage in informed reuse, which requires that data are independently understandable. In practice, this means that data must undergo quality review, a process whereby data and associated files are assessed and required actions are taken to ensure files are independently understandable for informed reuse. This paper explains what we mean by data quality review, what measures can be applied to it, and how it is practiced in three domain-specific archives. We explore a selection of other data repositories in the research data ecosystem, as well as the roles of researchers, academic libraries, and scholarly journals in regard to their application of data quality

measures in practice. We end with thoughts about the need to commit to data quality and who might be able to take on those tasks.

Peer, Limor, and Stephanie Wykstra. "New Curation Software: Step-by-Step Preparation of Social Science Data and Code for Publication and Preservation." *IASSIST Quarterly* 39, no. 4 (2015): 6-13. <https://doi.org/10.29173/iq902>

Pejša, Stanislav, Shirley J. Dyke, and Thomas J. Hacker. "Building Infrastructure for Preservation and Publication of Earthquake Engineering Research Data." *International Journal of Digital Curation* 9, no. 2 (2014): 83-97. <https://doi.org/10.2218/ijdc.v9i2.335>

The objective of this paper is to showcase the progress of the earthquake engineering community during a decade-long effort supported by the National Science Foundation in the George E. Brown Jr., Network for Earthquake Engineering Simulation (NEES). During the four years that NEES network operations have been headquartered at Purdue University, the NEEScomm management team has facilitated an unprecedented cultural change in the ways research is performed in earthquake engineering. NEES has not only played a major role in advancing the cyberinfrastructure required for transformative engineering research, but NEES research outcomes are making an impact by contributing to safer structures throughout the USA and abroad. This paper reflects on some of the developments and initiatives that helped instil change in the ways that the earthquake engineering and tsunami community share and reuse data and collaborate in general.

Peng, Ge. "The State of Assessing Data Stewardship Maturity—An Overview." *Data Science Journal* 17 (2018): 7. <http://doi.org/10.5334/dsj-2018-007>

Data stewardship encompasses all activities that preserve and improve the information content, accessibility, and usability of data and metadata. Recent regulations, mandates, policies, and guidelines set forth by the U.S. government, federal other, and funding agencies, scientific societies and scholarly publishers, have levied stewardship requirements on digital scientific data. This elevated level of requirements has increased the need for a formal approach to stewardship activities that supports compliance verification and reporting. Meeting or verifying compliance with stewardship requirements requires assessing the current state, identifying gaps, and, if necessary, defining a roadmap for improvement. This, however, touches on standards and best practices in multiple knowledge domains. Therefore, data

stewardship practitioners, especially those at data repositories or data service centers or associated with data stewardship programs, can benefit from knowledge of existing maturity assessment models. This article provides an overview of the current state of assessing stewardship maturity for federally funded digital scientific data. A brief description of existing maturity assessment models and related application(s) is provided. This helps stewardship practitioners to readily obtain basic information about these models. It allows them to evaluate each model's suitability for their unique verification and improvement needs.

Peng, Ge, Anna Milan, Nancy A. Ritchey, Robert P. Partee II, Sonny Zinn, Evan McQuinn, Kenneth S. Casey, Paul Lemieux III, Raisa Ionin, Philip Jones, Arianna Jakositz, and Donald Collins. "Practical Application of a Data Stewardship Maturity Matrix for the NOAA OneStop Project." *Data Science Journal*, 18, no. 1 (2019): 41. <http://doi.org/10.5334/dsj-2019-041>

Assessing the stewardship maturity of individual datasets is an essential part of ensuring and improving the way datasets are documented, preserved, and disseminated to users. It is a critical step towards meeting U.S. federal regulations, organizational requirements, and user needs. However, it is challenging to do so consistently and quantifiably. The Data Stewardship Maturity Matrix (DSMM), developed jointly by NOAA's National Centers for Environmental Information (NCEI) and the Cooperative Institute for Climate and Satellites–North Carolina (CICS-NC), provides a uniform framework for consistently rating stewardship maturity of individual datasets in nine key components: preservability, accessibility, usability, production sustainability, data quality assurance, data quality control/monitoring, data quality assessment, transparency/traceability, and data integrity. So far, the DSMM has been applied to over 800 individual datasets that are archived and/or managed by NCEI, in support of the NOAA's OneStop Data Discovery and Access Framework Project. As a part of the OneStop-ready process, tools, implementation guidance, workflows, and best practices are developed to assist the application of the DSMM and described in this paper. The DSMM ratings are also consistently captured in the ISO standard-based dataset-level quality metadata and citable quality descriptive information documents, which serve as interoperable quality information to both machine and human end-users. These DSMM implementation and integration workflows and best practices could be adopted by other data management and stewardship projects or adapted for applications of other maturity assessment models.

Peng, Ge, Nancy A. Ritchey, Kenneth S. Casey, Edward J. Kearns, Jeffrey L. Privette, Drew Saunders, Philip Jones, Tom Maycock, and Steve Ansari. "Scientific Stewardship in the Open Data and Big Data Era—Roles and Responsibilities of Stewards and Other Major Product Stakeholders." *D-Lib Magazine* 22, no. 5/6 (2016). <https://doi.org/10.1045/may2016-peng>

Pepe, Alberto, Alyssa Goodman, August Muench, Merce Crosas, and Christopher Erdmann. "How Do Astronomers Share Data? Reliability and Persistence of Datasets Linked in AAS Publications and a Qualitative Study of Data Practices among US Astronomers." *PLoS ONE* 9, no. 8 (2014): e104798. <http://dx.doi.org/10.1371/journal.pone.0104798>

We analyze data sharing practices of astronomers over the past fifteen years. An analysis of URL links embedded in papers published by the American Astronomical Society reveals that the total number of links included in the literature rose dramatically from 1997 until 2005, when it leveled off at around 1500 per year. The analysis also shows that the availability of linked material decays with time: in 2011, 44% of links published a decade earlier, in 2001, were broken. A rough analysis of link types reveals that links to data hosted on astronomers' personal websites become unreachable much faster than links to datasets on curated institutional sites. To gauge astronomers' current data sharing practices and preferences further, we performed in-depth interviews with 12 scientists and online surveys with 173 scientists, all at a large astrophysical research institute in the United States: the Harvard-Smithsonian Center for Astrophysics, in Cambridge, MA. Both the in-depth interviews and the online survey indicate that, in principle, there is no philosophical objection to data-sharing among astronomers at this institution. Key reasons that more data are not presently shared more efficiently in astronomy include: the difficulty of sharing large data sets; over reliance on non-robust, non-reproducible mechanisms for sharing data (e.g. emailing it); unfamiliarity with options that make data-sharing easier (faster) and/or more robust; and, lastly, a sense that other researchers would not want the data to be shared. We conclude with a short discussion of a new effort to implement an easy-to-use, robust, system for data sharing in astronomy, at theastrodata.org, and we analyze the uptake of that system to-date.

Pepe, Alberto, Matthew Mayernik, Christine L. Borgman, and Herbert Van de Sompel. "From Artifacts to Aggregations: Modeling Scientific Life Cycles on the Semantic Web." *Journal of the American Society for Information Science and Technology* 61, no. 3 (2010): 567-582. <https://doi.org/10.1002/asi.21263>

Pepler, Sam, and Sarah Callaghan. "Twenty Years of Data Management in the British Atmospheric Data Centre." *International Journal of Digital Curation* 10, no. 2 (2015): 23-32. <https://doi.org/10.2218/ijdc.v10i2.379>

The British Atmospheric Data Centre (BADC) has existed in its present form for 20 years, having been formally created in 1994. It evolved from the GDF (Geophysical Data Facility), a SERC (Science and Engineering Research Council) facility, as a result of research council reform where NERC (Natural Environment Research Council) extended its remit to cover atmospheric data below 10km altitude. With that change the BADC took on data from many other atmospheric sources and started interacting with NERC research programmes.

The BADC has now hit early adulthood. Prompted by this milestone, we examine in this paper whether the data centre is creaking at the seams or is looking forward to the prime of its life, gliding effortlessly into the future. Which parts of it are bullet proof and which parts are held together with double-sided sticky tape? Can we expect to see it in its present form in another twenty years' time?

To answer these questions, we examine the interfaces, technology, processes and organisation used in the provision of data centre services by looking at three snapshots in time, 1994, 2004 and 2014, using metrics and reports from the time to compare and contrast the services using BADC. The repository landscape has changed massively over this period and has moved the focus for technology and development as the broader community followed emerging trends, standards and ways of working. The incorporation of these new ideas has been both a blessing and a curse, providing the data centre staff with plenty of challenges and opportunities.

We also discuss key data centre functions including: data discovery, data access, ingestion, data management planning, preservation plans, agreements/licences and data policy, storage and server technology, organisation and funding, and user management. We conclude that the data centre will probably still exist in some form in 2024 and that it will most likely still be reliant on a file system. However, the technology delivering this service will change and the host organisation and funding routes may vary.

Pergl, Robert, Rob Hooft, Marek Suchánek, Vojtěch Knaisl, and Jan Slifka. "Data Stewardship Wizard": A Tool Bringing Together Researchers, Data Stewards, and

Data Experts around Data Management Planning." *Data Science Journal*, 18, no. 1 (2019): 59. <http://doi.org/10.5334/dsj-2019-059>

The Data Stewardship Wizard is a tool for data management planning that is focused on getting the most value out of data management planning for the project itself rather than on fulfilling obligations. It is based on FAIR Data Stewardship, in which each data-related decision in a project acts to optimize the Findability, Accessibility, Interoperability and/or Reusability of the data. The background to this philosophy is that the first reuser of the data is the researcher themselves. The tool encourages the consulting of expertise and experts, can help researchers avoid risks they did not know they would encounter by confronting them with practical experience from others, and can help them discover helpful technologies they did not know existed.

In this paper, we discuss the context and motivation for the tool, we explain its architecture and we present key functions, such as the knowledge model evolvability and migrations, assembling data management plans, metrics and evaluation of data management plans.

Perrier, Laure, Erik Blondal, A. Patricia Ayala, Dylanne Dearborn, Tim Kenny, David Lightfoot, Roger Reka, Mindy Thuna, Leanne Trimble, and Heather MacDonald. "Research Data Management in Academic Institutions: A Scoping Review." *PLOS ONE* 12, no. 5 (2017): e0178261. <https://doi.org/10.1371/journal.pone.0178261>

Objective

The purpose of this study is to describe the volume, topics, and methodological nature of the existing research literature on research data management in academic institutions.

Materials and methods

We conducted a scoping review by searching forty literature databases encompassing a broad range of disciplines from inception to April 2016. We included all study types and data extracted on study design, discipline, data collection tools, and phase of the research data lifecycle.

Results

We included 301 articles plus 10 companion reports after screening 13,002 titles and abstracts and 654 full-text articles. Most articles (85%) were published from 2010 onwards and conducted within the sciences (86%). More than three-quarters of the articles (78%) reported methods that included interviews, cross-sectional, or case studies. Most articles (68%) included the Giving Access to Data phase of the UK Data Archive Research Data Lifecycle that examines activities such as sharing data. When studies were grouped into five dominant groupings (Stakeholder, Data, Library, Tool/Device, and Publication), data quality emerged as an integral element.

Conclusion

Most studies relied on self-reports (interviews, surveys) or accounts from an observer (case studies) and we found few studies that collected empirical evidence on activities amongst data producers, particularly those examining the impact of research data management interventions. As well, fewer studies examined research data management at the early phases of research projects. The quality of all research outputs needs attention, from the application of best practices in research data management studies, to data producers depositing data in repositories for long-term use.

Perrier, Laure, Erik Blondal, and Heather MacDonald. "Exploring the Experiences of Academic Libraries with Research Data Management: A Meta-Ethnographic Analysis of Qualitative Studies." *Library & Information Science Research* 40, no. 3 (2018): 173-183. <https://doi.org/10.5281/zenodo.1324412>

Peters, Christie, and Anita Riley Dryden. "Assessing the Academic Library's Role in Campus-Wide Research Data Management: A First Step at the University of Houston." *Science & Technology Libraries* 30, no. 4 (2011): 387-403. <https://doi.org/10.1080/0194262x.2011.626340>

Peters, Christie, and Porcia Vaughn. "Initiating Data Management Instruction to Graduate Students at the University of Houston Using the New England Collaborative Data Management Curriculum." *Journal of eScience Librarianship* 3, no. 1 (2014): e1064. <https://doi.org/10.7191/jeslib.2014.1064>

Peters, Isabella, Peter Kraker, Elisabeth Lex, Christian Gumpenberger, and Juan Gorraiz. "Research Data Explored: An Extended Analysis of Citations and Altmetrics." *Scientometrics* 107, no. 2 (2016): 723-744. <https://doi.org/10.1007/s11192-016-1887-4>

Petters, Jonathan L., George C. Brooks, Jennifer A. Smith, and Carola A. Haas. "The Impact of Targeted Data Management Training for Field Research Projects—A Case Study." *Data Science Journal*, 18, no. 1 (2019): 43. <http://doi.org/10.5334/dsj-2019-043>

We present a joint effort at Virginia Tech between a research group in the Department of Fish and Wildlife Conservation and Data Services in the University Libraries to improve data management for long-term ecological field research projects in the Florida Panhandle. Consultative research data management support from Data Services in the University Libraries played an integral role in the development of the training curriculum. Emphasizing the importance of data quality to the field workers at the beginning of this training curriculum was a vital part of its success. Also critical for success was the research group's investment of time and effort to work with field workers and improve data management systems. We compare this case study to three others in the literature to compare and contrast data management processes and procedures. This case study serves as one example of how targeted training and efforts in data and project management for a research project can lead to substantial improvements in research data quality.

Peuch, Benjamin. "Elaborating a Crosswalk Between Data Documentation Initiative (DDI) and Encoded Archival Description (EAD) for an Emerging Data Archive Service Provider." *IASSIST Quarterly* 42, no. 2 (2018): 1–24. <https://doi.org/10.29173/iq924>

Peukert, Hagen. "Curating Humanities Research Data: Managing Workflows for Adjusting a Repository Framework." *International Journal of Digital Curation* 12, no. 2 (2017): 234-245. <https://doi.org/10.2218/ijdc.v12i2.571>

Handling heterogeneous data, subject to minimal costs, can be perceived as a classic management problem. The approach at hand applies established managerial theorizing to the field of data curation. It is argued, however, that data curation cannot merely be treated as a standard case of applying management theory in a traditional sense. Rather, the practice of curating humanities research data, the specifications and adjustments of the model suggested here reveal an intertwined process, in which knowledge of both strategic management and solid information technology have to be considered. Thus, suggestions on the strategic positioning of research data, which can be used as an analytical tool to understand the proposed workflow mechanisms, and the definition of workflow modules, which can be flexibly

used in designing new standard workflows to configure research data repositories, are put forward.

Pienta, Amy M., Dharma Akmon, Justin Noble, Lynette Hoelter, and Susan Jekielek. "A Data-Driven Approach to Appraisal and Selection at a Domain Data Repository." *International Journal of Digital Curation* 12, no. 2 (2017): <https://doi.org/10.2218/ijdc.v12i2.500>.

Social scientists are producing an ever-expanding volume of data, leading to questions about appraisal and selection of content given finite resources to process data for reuse. We analyze users' search activity in an established social science data repository to better understand demand for data and more effectively guide collection development. By applying a data-driven approach, we aim to ensure curation resources are applied to make the most valuable data findable, understandable, accessible, and usable. We analyze data from a domain repository for the social sciences that includes over 500,000 annual searches in 2014 and 2015 to better understand trends in user search behavior. Using a newly created search-to-study ratio technique, we identified gaps in the domain data repository's holdings and leveraged this analysis to inform our collection and curation practices and policies. The evaluative technique we propose in this paper will serve as a baseline for future studies looking at trends in user demand over time at the domain data repository being studied with broader implications for other data repositories.

Pinfield, Stephen, Andrew M. Cox, and Jen Smith. "Research Data Management and Libraries: Relationships, Activities, Drivers and Influences." *PLoS ONE* 9, no. 12 (2014): e114734. <https://doi.org/10.1371/journal.pone.0114734>

The management of research data is now a major challenge for research organisations. Vast quantities of born-digital data are being produced in a wide variety of forms at a rapid rate in universities. This paper analyses the contribution of academic libraries to research data management (RDM) in the wider institutional context. In particular it: examines the roles and relationships involved in RDM, identifies the main components of an RDM programme, evaluates the major drivers for RDM activities, and analyses the key factors influencing the shape of RDM developments. The study is written from the perspective of library professionals, analysing data from 26 semi-structured interviews of library staff from different UK institutions. This is an early qualitative contribution to the topic complementing existing quantitative and case study approaches. Results show that although libraries are playing a

significant role in RDM, there is uncertainty and variation in the relationship with other stakeholders such as IT services and research support offices. Current emphases in RDM programmes are on developments of policies and guidelines, with some early work on technology infrastructures and support services. Drivers for developments include storage, security, quality, compliance, preservation, and sharing with libraries associated most closely with the last three. The paper also highlights a 'jurisdictional' driver in which libraries are claiming a role in this space. A wide range of factors, including governance, resourcing and skills, are identified as influencing ongoing developments. From the analysis, a model is constructed designed to capture the main aspects of an institutional RDM programme. This model helps to clarify the different issues involved in RDM, identifying layers of activity, multiple stakeholders and drivers, and a large number of factors influencing the implementation of any initiative. Institutions may usefully benchmark their activities against the data and model in order to inform ongoing RDM activity.

Pink, Catherine. "Meeting the Data Management Compliance Challenge: Funder Expectations and Institutional Reality." *International Journal of Digital Curation* 8, no. 2 (2013): 157-171. <https://doi.org/10.2218/ijdc.v8i2.280>

In common with many global research funding agencies, in 2011 the UK Engineering and Physical Sciences Research Council (EPSRC) published its Policy Framework on Research Data along with a mandate that institutions be fully compliant with the policy by May 2015. The University of Bath has a strong applied science and engineering research focus and, as such, the EPSRC is a major funder of the university's research. In this paper, the Jisc-funded Research360 project shares its experience in developing the infrastructure required to enable a research-intensive institution to achieve full compliance with a particular funder's policy, in such a way as to support the varied data management needs of both the University of Bath and its external stakeholders. A key feature of the Research360 project was to ensure that after the project's completion in summer 2013 the newly developed data management infrastructure would be maintained up to and beyond the EPSRC's 2015 deadline. Central to these plans was the 'University of Bath Roadmap for EPSRC', which was identified as an exemplar response by the EPSRC. This paper explores how a roadmap designed to meet a single funder's requirements can be compatible with the strategic goals of an institution. Also discussed is how the project worked with Charles Beagrie Ltd to develop a supporting business case, thus ensuring implementation of

these long-term objectives. This paper describes how two new data management roles, the Institutional Data Scientist and Technical Data Coordinator, have contributed to delivery of the Research360 project and the importance of these new types of cross-institutional roles for embedding a new data management infrastructure within an institution. Finally, the experience of developing a new institutional data policy is shared. This policy represents a particular example of the need to reconcile a funder's expectations with the needs of individual researchers and their collaborators.

Pinnick, Jaana. "Exploring Digital Preservation Requirements: A Case Study from the National Geoscience Data Centre (NGDC)." *Records Management Journal* 27, no. 2 (2017): 175-191. <https://doi.org/10.1108/RMJ-04-2017-0009>

Piorun, Mary E., Donna Kafel, Tracey Leger-Hornby, Siamak Najafi, Elaine R. Martin, Paul Colombo, and Nancy R. LaPelle. "Teaching Research Data Management: An Undergraduate/Graduate Curriculum." *Journal of eScience Librarianship* 1, no. 1 (2012): e1003. <http://dx.doi.org/10.7191/jeslib.2012.1003>

Piwowar, Heather A. Who Shares? Who Doesn't? Factors Associated with Openly Archiving Raw Research Data. *PLoS ONE*, 6, no. 7 (2011:) e18657. <http://doi.org/10.1371/journal.pone.0018657>

Many initiatives encourage investigators to share their raw datasets in hopes of increasing research efficiency and quality. Despite these investments of time and money, we do not have a firm grasp of who openly shares raw research data, who doesn't, and which initiatives are correlated with high rates of data sharing. In this analysis I use bibliometric methods to identify patterns in the frequency with which investigators openly archive their raw gene expression microarray datasets after study publication.

Automated methods identified 11,603 articles published between 2000 and 2009 that describe the creation of gene expression microarray data. Associated datasets in best-practice repositories were found for 25% of these articles, increasing from less than 5% in 2001 to 30%-35% in 2007-2009. Accounting for sensitivity of the automated methods, approximately 45% of recent gene expression studies made their data publicly available.

First-order factor analysis on 124 diverse bibliometric attributes of the data creation articles revealed 15 factors describing authorship, funding, institution, publication, and domain environments. In multivariate regression, authors were most likely to share data if they had prior experience sharing or

reusing data, if their study was published in an open access journal or a journal with a relatively strong data sharing policy, or if the study was funded by a large number of NIH grants. Authors of studies on cancer and human subjects were least likely to make their datasets available.

These results suggest research data sharing levels are still low and increasing only slowly, and data is least available in areas where it could make the biggest impact. Let's learn from those with high rates of sharing to embrace the full potential of our research output.

Piwowar, Heather A., and Wendy W. Chapman. "Public Sharing of Research Datasets: A Pilot Study of Associations." *Journal of Informetrics* 4, no. 2 (2010): 148-156. <http://dx.doi.org/10.1016/j.joi.2009.11.010>

Piwowar, Heather A., Roger S. Day, and Douglas B. Fridsma. "Sharing Detailed Research Data Is Associated with Increased Citation Rate." *PLoS ONE* 2, no. (2007): e308. <http://dx.doi.org/10.1371/journal.pone.0000308>

Background

Sharing research data provides benefit to the general scientific community, but the benefit is less obvious for the investigator who makes his or her data available.

Principal Findings

We examined the citation history of 85 cancer microarray clinical trial publications with respect to the availability of their data. The 48% of trials with publicly available microarray data received 85% of the aggregate citations. Publicly available data was significantly ($p=0.006$) associated with a 69% increase in citations, independently of journal impact factor, date of publication, and author country of origin using linear regression.

Significance

This correlation between publicly available data and increased literature impact may further motivate investigators to share their detailed research data.

Plale, Beth. "Synthesis of Working Group and Interest Group Activity One Year into the Research Data Alliance." *D-Lib Magazine* 20, no. 1/2 (2014). <http://www.dlib.org/dlib/january14/plale/01plale.html>

Plale, Beth, Bin Cao, Chathura Herath, and Yiming Sun. "Data Provenance for Preservation of Digital Geoscience Data." *Geological Society of America Special Papers* 482 (2011): 125-137. [http://dx.doi.org/10.1130/2011.2482\(11\)](http://dx.doi.org/10.1130/2011.2482(11))

Plale, Beth, Robert H. McDonald, Kavitha Chandraseka, Inna Kouper, Stacy Konkiel, Margaret L. Hedstrom, James Myers, and Praveen Kumar. "SEAD Virtual Archive: Building a Federation of Institutional Repositories for Long-Term Data Preservation in Sustainability Science." *International Journal of Digital Curation* 8, no. 2 (2014): 172-180. <https://doi.org/10.2218/ijdc.v8i2.281>

Major research universities are grappling with their response to the deluge of scientific data emerging through research by their faculty. Many are looking to their libraries and the institutional repositories for a solution. Scientific data introduces substantial challenges that the document-based institutional repository may not be suited to deal with. The Sustainable Environment-Actionable Data (SEAD) Virtual Archive (VA) specifically addresses the challenges of 'long tail' scientific data. In this paper, we propose requirements, policy and architecture to support not only the preservation of scientific data today using institutional repositories, but also rich access to data and their use into the future.

Ponce de León, Mireia Alcalá, and Lluís Anglada i de Ferrer. "From Open Access to Open Data: Collaborative Work in the University Libraries of Catalonia." *LIBER Quarterly* 28, no. 1 (2018): 1-14. <http://doi.org/10.18352/lq.10253>

In the last years, the scientific community and funding bodies have paid attention to collected, generated or used data throughout different research activities. The dissemination of these data becomes one of the constituent elements of Open Science. For this reason, many funders are requiring or promoting the development of Data Management Plans, and depositing open data following the FAIR principles (Findable, Accessible, Interoperable and Reusable). Libraries and research offices of Catalan universities—which coordinately work within the Open Science Area of CSUC—offer support services to research data management. The different works carried out at the Consortium level will be presented, as well the implementation of the service in each university.

Poole, Alex H. "The Conceptual Landscape of Digital Curation." *Journal of Documentation* 72, no. 5 (2016): 961-986. <https://doi.org/10.1108/JD-10-2015-0123>

———. "How Has Your Science Data Grown? Digital Curation and the Human Factor: A Critical Literature Review." *Archival Science* 15, no. 2, (2015): 101-139. <http://dx.doi.org/10.1007/s10502-014-9236-y>

———. "Now is the Future Now? The Urgency of Digital Curation in the Digital Humanities." *Digital Humanities Quarterly* 7, no. 2 (2013). <http://www.digitalhumanities.org/dhq/vol/7/2/000163/000163.html>

Poole, Alex H., and Deborah A. Garwood. "Digging into Data Management in Public-Funded, International Research in Digital Humanities." *Journal of the Association for Information Science and Technology* 71, no. 1 (2019): 84-97. <https://doi.org/10.1002/asi.24213>

Porcal-Gonzalo, Maria C. "A Strategy for the Management, Preservation, and Reutilization of Geographical Information Based on the Lifecycle of Geospatial Data: An Assessment and a Proposal Based on Experiences from Spain and Europe." *Journal of Map & Geography Libraries* 11, no. 3 (2015) 289-329. <http://dx.doi.org/10.1080/15420353.2015.1064054>

Pouchard, Line. "Revisiting the Data Lifecycle with Big Data Curation." *International Journal of Digital Curation* 10, no. 2 (2015): 176-192. <https://doi.org/10.2218/ijdc.v10i2.342>

As science becomes more data-intensive and collaborative, researchers increasingly use larger and more complex data to answer research questions. The capacity of storage infrastructure, the increased sophistication and deployment of sensors, the ubiquitous availability of computer clusters, the development of new analysis techniques, and larger collaborations allow researchers to address grand societal challenges in a way that is unprecedented. In parallel, research data repositories have been built to host research data in response to the requirements of sponsors that research data be publicly available. Libraries are re-inventing themselves to respond to a growing demand to manage, store, curate and preserve the data produced in the course of publicly funded research. As librarians and data managers are developing the tools and knowledge they need to meet these new expectations, they inevitably encounter conversations around Big Data. This paper explores definitions of Big Data that have coalesced in the last decade

around four commonly mentioned characteristics: volume, variety, velocity, and veracity. We highlight the issues associated with each characteristic, particularly their impact on data management and curation. We use the methodological framework of the data life cycle model, assessing two models developed in the context of Big Data projects and find them lacking. We propose a Big Data life cycle model that includes activities focused on Big Data and more closely integrates curation with the research life cycle. These activities include planning, acquiring, preparing, analyzing, preserving, and discovering, with describing the data and assuring quality being an integral part of each activity. We discuss the relationship between institutional data curation repositories and new long-term data resources associated with high performance computing centers, and reproducibility in computational science. We apply this model by mapping the four characteristics of Big Data outlined above to each of the activities in the model. This mapping produces a set of questions that practitioners should be asking in a Big Data project.

Pouchard, Line, Andrew Woolf, and David Bernholdt. "Data Grid Discovery and Semantic Web Technologies for the Earth Sciences." *International Journal on Digital Libraries* 5, no. 2 (2005): 72-83. <https://doi.org/10.1007/s00799-004-0085-9>

Pronk, Tessa E. "The Time Efficiency Gain in Sharing and Reuse of Research Data." *Data Science Journal*, 18, no. 1 (2019): 10. <http://doi.org/10.5334/dsj-2019-010>

Among the frequently stated benefits of sharing research data are time efficiency or increased productivity. The assumption is that reuse or secondary use of research data saves researchers time in not having to produce data for a publication themselves. This can make science more efficient and productive. However, if there is no reuse, time costs in making data available for reuse will have been made with no return on this investment. In this paper a mathematical model is used to calculate the break-even point for time spent sharing in a scientific community, versus time gain by reuse. This is done for several scenarios; from simple to complex datasets to share and reuse, and at different sharing rates. The results indicate that sharing research data can indeed cause an efficiency revenue for the scientific community. However, this is not a given in all modeled scenarios. The scientific community with the lowest reuse needed to reach a break-even point is one that has few sharing researchers and low time investments for sharing and reuse. This suggests it would be beneficial to have a critical

selection of datasets that are worth the effort to prepare for reuse in other scientific studies. In addition, stimulating reuse of datasets in itself would be beneficial to increase efficiency in scientific communities.

Pronk, Tessa E., Paulien H. Wiersma, Anne van Weerden, and Feike Schieving. "A Game Theoretic Analysis of Research Data Sharing." *PeerJ* 3 (2015): e1242. <https://doi.org/10.7717/peerj.1242>

Prost, Hélène, Cécile Malleret, and Joachim Schöpfel. "Hidden Treasures: Opening Data in PhD Dissertations in Social Sciences and Humanities." *Journal of Librarianship and Scholarly Communication* 3, no. 2 (2015): eP1230. <http://doi.org/10.7710/2162-3309.1230>

PURPOSE The paper provides empirical evidence on research data submitted together with PhD dissertations in social sciences and humanities.

APPROACH We conducted a survey on nearly 300 print and electronic dissertations in social sciences and humanities from the University of Lille 3 (France), submitted between 1987 and 2013. **FINDINGS** After a short overview on open access to electronic dissertations, on small data in dissertations, on data management and curation, and on the challenge for academic libraries, the paper presents the results of the survey. Special attention is paid to the size of the research data in appendices, to their presentation and link to the text, to their sources and typology, and to their potential for further research. Methodological shortfalls of the study are discussed, and barriers to open data (metadata, structure, format) and legal questions (privacy, third-party rights) are addressed. The conclusion provides some recommendations for the assistance and advice to PhD students in managing and depositing their research data. **PRACTICAL IMPLICATIONS** Our survey can be helpful for academic libraries to develop assistance and advice for PhD students in managing their research data in collaboration with the research structures and the graduate schools. **ORIGINALITY** There is a growing body of research papers on data management and curation. Produced along with PhD dissertations, little is known about the characteristics of this material, in particular in social sciences and humanities and the impact on the role of academic libraries.

Pryor, Graham. "Attitudes and Aspirations in a Diverse World: The Project StORe Perspective on Scientific Repositories." *International Journal of Digital Curation* 2, no. 1 (2007): 135-144. <https://doi.org/10.2218/ijdc.v2i1.21>

Project StORe was conceived as an initiative to apply digital library technologies in the creation of new value for published research. Ostensibly a technical project, its primary objective was the design of middleware to enable bi-directional links between source repositories containing research data and output repositories containing research publications. Hence, researchers would be able to navigate directly from within an electronic article to the source or synthesised data from which that article was derived. To achieve a product that directly reflects user needs, a survey of researchers was conducted across seven scientific disciplines. This survey exposed the spectrum of cultural pressures, preferences and prejudices that influence the research process, as well as a range of practices in the production and management of research data. Aspects of the research environment revealed by the survey are considered in this paper in the context of repository use and, more broadly, the requirements, roles and responsibilities necessary to good data management.

———. "A Maturing Process of Engagement: Raising Data Capabilities in UK Higher Education." *International Journal of Digital Curation* 8, no. 2 (2013): 181-193. <https://doi.org/10.2218/ijdc.v8i2.282>

In the spring of 2011, the UK's Digital Curation Centre (DCC) commenced a programme of outreach designed to assist individual universities in their development of aptitude for managing research data. This paper describes the approaches taken, covering the context in which these institutional engagements have been discharged and examining the aims, methodology and processes employed. It also explores what has worked and why, as well as the pitfalls encountered, including example outcomes and identifiable or predicted impact. Observing how the research data landscape is constantly undergoing change, the paper concludes with an indication of the steps being taken to refit the DCC institutional engagement to the evolving needs of higher education.

———. "Multi-Scale Data Sharing in the Life Sciences: Some Lessons for Policy Makers." *International Journal of Digital Curation* 4, no. 3 (2009): 71-82. <https://doi.org/10.2218/ijdc.v4i3.115>

Drawing on the final report on a recent series of case studies in the life sciences at the University of Edinburgh, this paper explores the attitudes and perceptions of researchers towards data sharing and contrasts these with the policies of the major research funders. Notwithstanding economic, technical

and cultural inhibitors, the general ethos in the Life Sciences is one of support to the principle of data sharing. However, this position is subject to a complex range of qualifications, not least the crucial need for sharing through collaboration. The kind of generic vision for data sharing that is currently promoted by national agencies is judged to be neither productive nor effective. Only close engagement with research practitioners in the identification of bottom-up strategies that preserve the exercise of informed choice—a fundamental and persistent element of scientific research—will produce change on a national scale.

Pryor, Graham, ed. *Delivering Research Data Management Services: Fundamentals of Good Practice*. London : Facet Publishing, 2014.
<https://melvyl.on.worldcat.org/oclc/878406164>

Pryor, Graham, ed. *Managing Research Data*. London: Facet Publishing, 2012.
<http://www.worldcat.org/oclc/878838741>

Pryor, Graham, and Martin Donnelly. "Skilling Up to Do Data: Whose Role, Whose Responsibility, Whose Career?" *International Journal of Digital Curation* 4, no. 2 (2009): 158-170. <https://doi.org/10.2218/ijdc.v4i2.105>

Pryor, Graham, Sarah Jones, and Angus Whyte. *Delivering Research Data Management Services*. London: Facet Publishing, 2014.
<http://www.worldcat.org/oclc/889922926>

Punzalan, Ricardo L., and Adam Kriesberg. "Library-Mediated Collaborations: Data Curation at the National Agricultural Library." *Library Trends* 65, no. 3 (2017): 429-447. <https://doi.org/10.1353/lib.2017.0010>

Qin, Jian, Kevin Crowston, and Arden Kirkland. "Pursuing Best Performance in Research Data Management by Using the Capability Maturity Model and Rubrics." *Journal of eScience Librarianship* 6, no. 2 (2017): e1113.
<https://doi.org/10.7191/jeslib.2017.1113>

Objective: To support the assessment and improvement of research data management (RDM) practices to increase its reliability, this paper describes the development of a capability maturity model (CMM) for RDM. Improved RDM is now a critical need, but low awareness of—or lack of—data management is still common among research projects.

Methods: A CMM includes four key elements: key practices, key process areas, maturity levels, and generic processes. These elements were determined for RDM by a review and synthesis of the published literature on and best practices for RDM.

Results: The RDM CMM includes five chapters describing five key process areas for research data management: 1) data management in general; 2) data acquisition, processing, and quality assurance; 3) data description and representation; 4) data dissemination; and 5) repository services and preservation. In each chapter, key data management practices are organized into four groups according to the CMM's generic processes: commitment to perform, ability to perform, tasks performed, and process assessment (combining the original measurement and verification). For each area of practice, the document provides a rubric to help projects or organizations assess their level of maturity in RDM.

Conclusions: By helping organizations identify areas of strength and weakness, the RDM CMM provides guidance on where effort is needed to improve the practice of RDM.

Qin, Jian, and John D'ignazio. "The Central Role of Metadata in a Science Data Literacy Course." *Journal of Library Metadata* 10, no. 2/3 (2010): 188-204.
<https://doi.org/10.1080/19386389.2010.506379>

Qin, Jian, Brian Dobreski, and Duncan Brown. "Metadata and Reproducibility: A Case Study of Gravitational Wave Research Data Management." *International Journal of Digital Curation* 11, no. 1 (2016): 218-231.
<https://doi.org/10.2218/ijdc.v11i1.399>

The complexity of computationally-intensive scientific research poses great challenges for both research data management and research reproducibility. What metadata needs to be captured for tracking, reproducing, and reusing computational results is the starting point in developing metadata models to fulfil these functions of data management. This paper reports the findings from interviews with gravitational wave (GW) researchers, which were designed to gather user requirements to develop a metadata model. Motivations for keeping documentation of data and analysis results include trust, accountability and continuity of work. Research reproducibility relies on metadata that represents code dependencies and versions and has good documentation for verification. Metadata specific to GW data, workflows and

outputs tend to differ from those currently available in metadata standards. The paper also discusses the challenges in representing code dependencies and workflows.

Raboin, Regina, Rebecca C. Reznik-Zellen, and Dorothea Salo. "Forging New Service Paths: Institutional Approaches to Providing Research Data Management Services." *Journal of eScience Librarianship* 1, no. 3 (2012): e1021. <http://dx.doi.org/10.7191/jeslib.2012.1021>

Radio, Erik, Fernando Rios, Jeffrey C. Oliver, Benjamin Hickson, and Niamh Wallace. "Manifestations of Metadata Structures in Research Datasets and Their Ontic Implications." *Journal of Library Metadata* 17, no. 3-4 (2017): 161-182. <https://doi.org/10.1080/19386389.2018.1439278>

Ragon, Bart. "The Political Economy of Federally Sponsored Data." *Journal of eScience Librarianship* 2, no. 2 (2013): e1050. <http://dx.doi.org/10.7191/jeslib.2013.1050>

Rajasekar, Arcot, Reagan Moore, Mike Wan, and Wayne Schroeder. "Policy-Based Distributed Data Management Systems." *Journal of Digital Information* 11, no. 1 (2010). <http://journals.tdl.org/jodi/index.php/jodi/article/view/756>

Ramachandran, Rahul, Christopher Lynnes, Kathleen Baynes, Kevin Murphy, Jamie Baker, Jamie Kinney, Ariel Gold, Jed Sundwall, Mark Korver, Allison Lieber, William Vambenepe, Matthew Hancher, Rebecca Moore, Tyler Erickson, Josh Henretig, Brant Zwiefel, Heather Patrick-Ahlstrom, and Matthew J. Smith. "Recommendations to Improve Downloads of Large Earth Observation Data." *Data Science Journal* 17, no. 2. (2018). <http://doi.org/10.5334/dsj-2018-002>

With the volume of Earth observation data expanding rapidly, cloud computing is quickly changing the way these data are processed, analyzed, and visualized. Collocating freely available Earth observation data on a cloud computing infrastructure may create opportunities unforeseen by the original data provider for innovation and value-added data re-use, but existing systems at data centers are not designed for supporting requests for large data transfers. A lack of common methodology necessitates that each data center handle such requests from different cloud vendors differently. Guidelines are needed to support enabling all cloud vendors to utilize a common methodology for bulk-downloading data from data centers, thus preventing the providers from building custom capabilities to meet the needs of individual vendors.

This paper presents recommendations distilled from use cases provided by three cloud vendors (Amazon, Google, and Microsoft) and are based on the vendors' interactions with data systems at different Federal agencies and organizations. These specific recommendations range from obvious steps for improving data usability (such as ensuring the use of standard data formats and commonly supported projections) to non-obvious undertakings important for enabling bulk data downloads at scale. These recommendations can be used to evaluate and improve existing data systems for high-volume data transfers, and their adoption can lead to cloud vendors utilizing a common methodology.

Rambo, Neil. *Research Data Management Roles for Libraries*. New York: Ithaka S+R, 2015. <http://dx.doi.org/10.18665/sr.274643>

Ramírez, Marisa L. "Whose Role Is It Anyway?: A Library Practitioner's Appraisal of the Digital Data Deluge." *Bulletin of the American Society for Information Science & Technology* 37, no. 5 (2011): 21-23. <https://doi.org/10.1002/bult.2011.1720370508>

Rappert, Brian, and Louise Bezuidenhout. "Data Sharing in Low-Resourced Research Environments." *Prometheus* (2017): 1-18. <http://dx.doi.org/10.1080/08109028.2017.1325142>

Rath, Linda L. "Low-Barrier-to-Entry Data Tools: Creating and Sharing Humanities Data." *Library Hi Tech* 34, no. 2 (2016): 268-285. <https://doi.org/10.1108/lht-07-2015-0073>

Ray, Joyce M., ed. *Research Data Management: Practical Strategies for Information Professionals*. West Lafayette, IN: Purdue University Press 2014. <http://www.worldcat.org/oclc/900191339>

Read, Kevin B. "Adapting Data Management Education to Support Clinical Research Projects in an Academic Medical Center." *Journal of the Medical Library Association*. 107, no. 1 (2019): 89-97. <https://dx.doi.org/10.5195%2Fjmla.2019.580>

Read, Kevin B., Jessica Koos, Rebekah S. Miller, Cathryn F. Miller, Gesina A. Phillips, Laurel Scheinfeld, and Alisa Surkis. "A Model for Initiating Research Data Management Services at Academic Libraries." *Journal of the Medical Library Association* 107, no. 3 (2019): 432–441. <https://doi.org/10.5195/jmla.2019.545>

Background: Librarians developed a pilot program to provide training, resources, strategies, and support for medical libraries seeking to establish research data management (RDM) services. Participants were required to complete eight educational modules to provide the necessary background in RDM. Each participating institution was then required to use two of the following three elements: (1) a template and strategies for data interviews, (2) the Teaching Toolkit to teach an introductory RDM class, or (3) strategies for hosting a data class series.

Case Presentation: Six libraries participated in the pilot, with between two and eight librarians participating from each institution. Librarians from each institution completed the online training modules. Each institution conducted between six and fifteen data interviews, which helped build connections with researchers, and taught between one and five introductory RDM classes. All classes received very positive evaluations from attendees. Two libraries conducted a data series, with one bringing in instructors from outside the library.

Conclusion: The pilot program proved successful in helping participating librarians learn about and engage with their research communities, jump-start their teaching of RDM, and develop institutional partnerships around RDM services. The practical, hands-on approach of this pilot proved to be successful in helping libraries with different environments establish RDM services. The success of this pilot provides a proven path forward for libraries that are developing data services at their own institutions.

Read, Kevin B., Alisa Surkis, Catherine Larson, Aileen McCrillis, Alice Graff, Joey Nicholson, and Juanchan Xu. "Starting the Data Conversation: Informing Data Services at an Academic Health Sciences Library." *Journal of the Medical Library Association* 10, no. 3 (2015): 131-135. <https://doi.org/10.3163/1536-5050.103.3.005>

Recker, Astrid, and Stefan Müller. "Preserving the Essence: Identifying the Significant Properties of Social Science Research Data." *New Review of Information Networking* 20, no. 1-2 (2015): 229-235. <http://dx.doi.org/10.1080/13614576.2015.1110404>

Recker, Astrid, Stefan Müller, Jessica Trixa, and Natascha Schumann. "Paving the Way for Data-Centric, Open Science: An Example From the Social Sciences."

Journal of Librarianship and Scholarly Communication 3, no. 2 (2015): eP1227.
<http://doi.org/10.7710/2162-3309.1227>

INTRODUCTION Data has moved into the spotlight as an important scholarly output that should be shared with the scientific community for replication and re-use in new contexts. This has a direct impact on libraries, archives, and other service providers in the data curation and access landscape. **DESCRIPTION OF PROJECT** The GESIS Data Archive for the Social Sciences (DAS) has been curating and disseminating social science research data since 1960. The article presents tools, services, and strategies developed by the DAS to support the research community in adequately responding to the legal, ethical, and practical challenges that the transformation towards data-centric, open science presents. These include GESIS's Secure Data Center, the data publication platform "datorium" and a recent project to create a georeferencing service for survey data. **LESSONS LEARNED** The experiences gained through these activities show that getting involved-now, rather than further down the road-pays off in that it allows service providers to actively shape the ongoing transformation. At the same time, by cooperating with suitable partners, the effort and investment of resources can be kept at a manageable level for individual organizations.

Redkina, N.S. "Current Trends in Research Data Management." *Scientific and Technical Information Processing* 46 (2019): 53–58.
<https://doi.org/10.3103/S0147688219020035>

Reed, Robyn B. "Diving into Data: Planning a Research Data Management Event." *Journal of eScience Librarianship* 4, no. 1 (2015): e1071.
<https://doi.org/10.7191/jeslib.2015.1071>

Reilly, Michele, and Anita R. Dryden. "Building an Online Data Management Plan Tool." *Journal of Librarianship and Scholarly Communication* 1, no. 3 (2013): eP1066. <http://doi.org/10.7710/2162-3309.1066>

Following the 2011 announcement by the National Science Foundation (NSF) that it would begin requiring Data Management Plans with every funding application, the University of Houston Libraries explored ways to support our campus researchers in meeting this requirement. A small team of librarians built an online tool using a Drupal module. The tool includes informational content, an interactive questionnaire, and an extensive FAQ to meet diverse

researcher needs. This easily accessible and locally maintained tool allows us to provide a high level of personalized service to our researchers.

Reimer, Torsten. "Your Name Is Not Good Enough: Introducing the ORCID Researcher Identifier at Imperial College London." *Insights* 28, no. 3 (2016): 76-82. <http://doi.org/10.1629/uksg.268>

The ORCID researcher identifier ensures that research outputs can always reliably be traced back to their authors. ORCID also makes it possible to automate the sharing of research information, thereby increasing data quality, reducing duplication of effort for academics and saving institutions money. In 2014, Imperial College London created ORCID identifiers (iDs) for academic and research staff. This article discusses the implementation project in the context of the role of ORCID in the global scholarly communications system. It shows how ORCID can be used to automate reporting, help with research data publication and support open access (OA).

Renear, Allen H., Carole L. Palmer, and John Unsworth. *Extending Data Curation to the Humanities: Curriculum Development and Recruiting*. Urbana-Champaign: Graduate School of Library and Information Science, University of Illinois at Urbana-Champaign, 2013. <http://hdl.handle.net/2142/42628>

Renwick, Shamin, Marsha Winter, and Michelle Gill. "Managing Research Data at an Academic Library in a Developing Country." *IFLA Journal* 43, no. 1 (2017): 51-64. <https://doi.org/10.1177/0340035216688703>

Reznik-Zellen, Rebecca C., Jessica Adamick, and Stephen McGinty. "Tiers of Research Data Support Services." *Journal of eScience Librarianship* 1, no. 1 (2012): e1002. <http://dx.doi.org/10.7191/jeslib.2012.1002>

Ribeiro, Cristina, Maria Eugénia, and Matos Fernandes. "Data Curation at U. Porto: Identifying Current Practices across Disciplinary Domains." *IASSIST Quarterly* 35, no. 4 (2011): 14-17. <https://doi.org/10.29173/iq893>

Rice, Robin. "Research Data MANTRA: A Labour of Love." *Journal of eScience Librarianship* 3, no. 1 (2014): e1056. <http://dx.doi.org/10.7191/jeslib.2014.1056>

Rice, Robin, Çuna Ekmekcioglu, Jeff Haywood, Sarah Jones, Stuart Lewis, Stuart Macdonald, and Tony Weir. "Implementing the Research Data Management Policy: University of Edinburgh Roadmap." *International Journal of Digital Curation* 8, no. 2 (2013): 194-204. <https://doi.org/10.2218/ijdc.v8i2.283>

This paper discusses work to implement the University of Edinburgh Research Data Management (RDM) policy by developing the services needed to support researchers and fulfil obligations within a changing national and international setting. This is framed by an evolving Research Data Management Roadmap and includes a governance model that ensures cooperation amongst Information Services (IS) managers and oversight by an academic-led steering group. IS has taken requirements from research groups and IT professionals, and at the request of the steering group has conducted pilot work involving volunteer research units within the three colleges to develop functionality and presentation for the key services. The first pilots cover three key services: the data store, a customisation of the Digital Curation Centre's DMPonline tool, and the data repository. The paper will report on the plans, achievements and challenges encountered while we attempt to bring the University of Edinburgh RDM Roadmap to fruition.

Rice, Robin, and Jeff Haywood. "Research Data Management Initiatives at University of Edinburgh." *International Journal of Digital Curation* 6, no. 2 (2011): 232-244. <https://doi.org/10.2218/ijdc.v6i2.199>

Rice, Robin, and John Southall. *The Data Librarian's Handbook*. London: Facet Publishing, 2016. <http://www.worldcat.org/oclc/986613159>

Richards, Julian D. "Digital Preservation and Access." *European Journal of Archaeology* 5, no. 3 (2002): 343-366. <https://doi.org/10.1177/146195702761692347>

———, "Twenty Years Preserving Data: A View from the United Kingdom." *Advances in Archaeological Practice* 5, no. 3 (2017): 227-237. <https://doi.org/10.1017/aap.2017.11>

Rimkus, Kyle, Thomas Padilla, Tracy Popp, and Greer Martin. "Digital Preservation File Format Policies of ARL Member Libraries: An Analysis." *D-Lib Magazine* 20, no. 3/4 (2014). <https://doi.org/10.1045/march2014-rimkus>

Rinehart, Amanda K. "Getting Emotional about Data: The Soft Side of Data Management Services." *College & Research Libraries News* 76, no. 8 (2015): 437-440. <https://doi.org/10.5860/crln.76.8.9364>

Rios, Fernando. "Incorporating Software Curation into Research Data Management Services: Lessons Learned." *International Journal of Digital Curation* 13, no. 1 (2018): 235-247.

Many large research universities provide research data management (RDM) support services for researchers. These may include support for data management planning, best practices (e.g., organization, support, and storage), archiving, sharing, and publication. However, these data-focused services may under-emphasize the importance of the software that is created to analyse said data. This is problematic for several reasons. First, because software is an integral part of research across all disciplines, it undermines the ability of said research to be understood, verified, and reused by others (and perhaps even the researcher themselves). Second, it may result in less visibility and credit for those involved in creating the software. A third reason is related to stewardship: if there is no clear process for how, when, and where the software associated with research can be accessed and who will be responsible for maintaining such access, important details of the research may be lost over time.

This article presents the process by which the RDM services unit of a large research university addressed the lack of emphasis on software and source code in their existing service offerings. The greatest challenges were related to the need to incorporate software into existing data-oriented service workflows while minimizing additional resources required, and the nascent state of software curation and archiving in a data management context. The problem was addressed from four directions: building an understanding of software curation and preservation from various viewpoints (e.g., video games, software engineering), building a conceptual model of software preservation to guide service decisions, implementing software-related services, and documenting and evaluating the work to build expertise and establish a standard service level.

Roche, Dominique G., Robert Lanfear, Sandra A. Binning, Tonya M. Haff, Lisa E. Schwanz, Kristal E. Cain, Hanna Kokko, Michael D. Jennions, and Loeske E. B. Kruuk. "Troubleshooting Public Data Archiving: Suggestions to Increase Participation." *PLOS Biology* 12, no. 1 (2014): e1001779.
<http://dx.doi.org/10.1371/journal.pbio.1001779>

An increasing number of publishers and funding agencies require public data archiving (PDA) in open-access databases. PDA has obvious group benefits for the scientific community, but many researchers are reluctant to share their data publicly because of real or perceived individual costs. Improving participation in PDA will require lowering costs and/or in-creasing benefits for primary data collectors. Small, simple changes can enhance existing

measures to ensure that more scientific data are properly archived and made publicly available: (1) facilitate more flexible embargoes on archived data, (2) encourage communication between data generators and re-users, (3) disclose data re-use ethics, and (4) encourage increased recognition of publicly archived data.

Rousidis, Dimitris, Emmanouel Garoufallou, Panos Balatsoukas, and Miguel-Angel Sicilia. "Metadata for Big Data: A Preliminary Investigation of Metadata Quality Issues in Research Data Repositories." *Information Services & Use* 34, no. 3-4 (2014): 279-286. <https://doi.org/10.3233/isu-140746>

Rowhani-Farid, Anisa, Michelle Allen, and Adrian G. Barnett. "What Incentives Increase Data Sharing in Health and Medical Research? A Systematic Review." *Research Integrity and Peer Review* 2, no. 1 (2017): 4. <https://doi.org/10.1186/s41073-017-0028-9>

Background

The foundation of health and medical research is data. Data sharing facilitates the progress of research and strengthens science. Data sharing in research is widely discussed in the literature; however, there are seemingly no evidence-based incentives that promote data sharing.

Methods

A systematic review (registration: doi.org/10.17605/OSF.IO/6PZ5E) of the health and medical research literature was used to uncover any evidence-based incentives, with pre- and post-empirical data that examined data sharing rates. We were also interested in quantifying and classifying the number of opinion pieces on the importance of incentives, the number observational studies that analysed data sharing rates and practices, and strategies aimed at increasing data sharing rates.

Results

Only one incentive (using open data badges) has been tested in health and medical research that examined data sharing rates. The number of opinion pieces (n=85) out-weighed the number of article-testing strategies (n=76), and the number of observational studies exceeded them both (n=106).

Conclusions

Given that data is the foundation of evidence-based health and medical research, it is paradoxical that there is only one evidence-based incentive to promote data sharing. More well-designed studies are needed in order to increase the currently low rates of data sharing.

Rueda, Laura, Martin Fenner, and Patricia Cruse. "DataCite: Lessons Learned on Persistent Identifiers for Research Data." *International Journal of Digital Curation* 11, no. 2 (2017): 39-47. <https://doi.org/10.2218/ijdc.v11i2.421>

Data are the infrastructure of science and they serve as the groundwork for scientific pursuits. Data publication has emerged as a game-changing breakthrough in scholarly communication. Data form the outputs of research but also are a gateway to new hypotheses, enabling new scientific insights and driving innovation. And yet stakeholders across the scholarly ecosystem, including practitioners, institutions, and funders of scientific research are increasingly concerned about the lack of sharing and reuse of research data. Across disciplines and countries, researchers, funders, and publishers are pushing for a more effective research environment, minimizing the duplication of work and maximizing the interaction between researchers. Availability, discoverability, and reproducibility of research outputs are key factors to support data reuse and make possible this new environment of highly collaborative research.

An interoperable e-infrastructure is imperative in order to develop new platforms and services for to data publication and reuse. DataCite has been working to establish and promote methods to locate, identify and share information about research data. Along with service development, DataCite supports and advocates for the standards behind persistent identifiers (in particular DOIs, Digital Object Identifiers) for data and other research outputs. Persistent identifiers allow different platforms to exchange information consistently and unambiguously and provide a reliable way to track citations and reuse. Because of this, data publication can become a reality from a technical standpoint, but the adoption of data publication and data citation as a practice by researchers is still in its early stages.

Since 2009, DataCite has been developing a series of tools and services to foster the adoption of data publication and citation among the research community. Through the years, DataCite has worked in a close collaboration with interdisciplinary partners on these issues and we have gained insight into

the development of data publication workflows. This paper describes the types of different actions and the lessons learned by DataCite.

Rumsey, Sally, and Neil Jefferies. "Challenges in Building an Institutional Research Data Catalogue." *International Journal of Digital Curation* 8, no. 2 (2013): 205-214. <https://doi.org/10.2218/ijdc.v8i2.284>

The University of Oxford is preparing systems and services to enable members of the university to manage research data produced by its scholars. Much of the work has been carried out under the Jisc-funded Damaro project. This project draws together existing nascent services, adds new systems and services to 'fill the gaps' and provides a wide-ranging infrastructure. Development comprises four parallel strands: endorsement of a university research data management policy; training and guidance in research data management; technical infrastructure; and future sustainability. A key element of the technical infrastructure is DataFinder, a catalogue of Oxford research data outputs. DataFinder's core purposes are to record the existence of Oxford datasets, enable their discovery, and provide details of their location. DataFinder will record metadata about Oxford research data, irrespective of location, discipline or format, and is viewed by the university as a crucial hub for the university's Research Data Management (RDM) infrastructure.

———. "DataFinder: A Research Data Catalogue for Oxford." *Ariadne*, no. 71 (2013). <http://www.ariadne.ac.uk/issue71/rumsey-jefferies>

Sallans, Andrew, and Martin Donnelly. "DMP Online and DMPTool: Different Strategies towards a Shared Goal." *International Journal of Digital Curation* 7, no. 2 (2012): 123-129. <https://doi.org/10.2218/ijdc.v7i2.235>

This paper provides a comparative discussion of the strategies employed in the UK's DMP Online tool and the US's DMPTool, both designed to provide a structured environment for research data management planning (DMP) with explicit links to funder requirements. Following the Sixth International Digital Curation Conference, held in Chicago in December 2010, a number of US institutions partnered with the Digital Curation Centre's DMP Online team to learn from their experiences while developing a US counterpart. DMPTool arrived in beta in August 2011 and released a production version in November 2011. This joint paper will compare and contrast use cases, organizational and national/cultural characteristics that have influenced the

development decisions, outcomes achieved so far, and planned future developments.

Salvatori, Lorenza, Ana Sesartic, Nathalie Lambeng, and Eliane Blumer. "Keep Calm and Fill in Your DMP: Lessons Learnt from a Swiss DMP-Template Initiative." *International Journal of Digital Curation* 13, no. 1 (2018): 215-222. <https://doi.org/10.2218/ijdc.v13i1.617>

Aligning with other funders such as Horizon 2020, the Swiss National Science Foundation (SNSF) requires researchers who apply for project funding to provide a Data Management Plan (DMP) as an integral part of their research proposal. In an attempt to assist and guide researchers filling out this document, and to provide a service as efficient as possible, the libraries of the Ecole Polytechnique Fédérale de Lausanne (EPFL) and ETH Zurich took the lead to elaborate on a DMP template with content suggestions and recommendations. In this practice paper, we will describe the collaborative effort between the two Swiss federal institutes of technology, namely EPFL and ETH Zurich, as well as some partners of the national Data Life Cycle Management (DLCM) project, which resulted in a very helpful document as reported by our researchers.

Sánchez, David, and Viejo Alexandre. "Personalized Privacy in Open Data Sharing Scenarios." *Online Information Review* 41, no. 3 (2017): 298-310. <https://doi.org/10.1108/OIR-01-2016-0011>

Sands, Ashley E., Christine L. Borgman, Sharon Traweek, and Laura A. Wynholds. "We're Working On It: Transferring the Sloan Digital Sky Survey from Laboratory to Library." *International Journal of Digital Curation* 9, no. 2 (2014): 98-110. <https://doi.org/10.2218/ijdc.v9i2.336>

This article reports on the transfer of a massive scientific dataset from a national laboratory to a university library, and from one kind of workforce to another. We use the transfer of the Sloan Digital Sky Survey (SDSS) archive to examine the emergence of a new workforce for scientific research data management. Many individuals with diverse educational backgrounds and domain experience are involved in SDSS data management: domain scientists, computer scientists, software and systems engineers, programmers, and librarians. These types of positions have been described using terms such as research technologist, data scientist, e-science professional, data curator,

and more. The findings reported here are based on semi-structured interviews, ethnographic participant observation, and archival studies from 2011-2013.

The library staff conducting the data storage and archiving of the SDSS archive faced two performance problems. The preservation specialist and the system administrator worked together closely to discover and implement solutions to the slow data transfer and verification processes. The team overcame these slow-downs by problem solving, working in a team, and writing code. The library team lacked the astronomy domain knowledge necessary to meet some of their preservation and curation goals.

The case study reveals the variety of expertise, experience, and individuals essential to the SDSS data management process. A variety of backgrounds and educational histories emerge in the data managers studied. Teamwork is necessary to bring disparate expertise together, especially between those with technical and domain education. The findings have implications for data management education, policy and relevant stakeholders.

This article is part of continuing research on Knowledge Infrastructures.

Sapp Nelson, Megan. "Data Management Outreach to Junior Faculty Members: A Case Study." *Journal of eScience Librarianship* 4, no. 1 (2015): e1076.
<https://doi.org/10.7191/jeslib.2015.1076>

———. "A Pilot Competency Matrix for Data Management Skills: A Step toward the Development of Systematic Data Information Literacy Programs." *Journal of eScience Librarianship* 5, no. 1 (2017): e1096.
<https://doi.org/10.7191/jeslib.2017.1096>

Savage, Caroline J., and Andrew J. Vickers. "Empirical Study of Data Sharing by Authors Publishing in PLoS Journals." *PLoS ONE* 4, no. 9 (2009): e7078.
<http://dx.doi.org/10.1371/journal.pone.0007078>

Background

Many journals now require authors share their data with other investigators, either by depositing the data in a public repository or making it freely available upon request. These policies are explicit, but remain largely untested. We sought to determine how well authors comply with such policies by requesting data from authors who had published in one of two journals with clear data sharing policies.

Methods and Findings

We requested data from ten investigators who had published in either *PLoS Medicine* or *PLoS Clinical Trials*. All responses were carefully documented. In the event that we were refused data, we reminded authors of the journal's data sharing guidelines. If we did not receive a response to our initial request, a second request was made. Following the ten requests for raw data, three investigators did not respond, four authors responded and refused to share their data, two email addresses were no longer valid, and one author requested further details. A reminder of PLoS's explicit requirement that authors share data did not change the reply from the four authors who initially refused. Only one author sent an original data set.

Conclusions

We received only one of ten raw data sets requested. This suggests that journal policies requiring data sharing do not lead to authors making their data sets available to independent investigators.

Savage, James L., and Lauren Cadwallader. "Establishing, Developing, and Sustaining a Community of Data Champions." *Data Science Journal*, 18, no. 1 (2019): 23. <http://doi.org/10.5334/dsj-2019-023>

Supporting good practice in Research Data Management (RDM) is challenging for higher education institutions, in part because of the diversity of research practices and data types across disciplines. While centralised research data support units now exist in many universities, these typically possess neither the discipline-specific expertise nor the resources to offer appropriate targeted training and support within every academic unit. One solution to this problem is to identify suitable individuals with discipline-specific expertise that are already embedded within each unit, and empower these individuals to advocate for good RDM and to deliver support locally. This article focuses on an ongoing example of this approach: the Data Champion Programme at the University of Cambridge, UK. We describe how the Data Champion programme was established; the programme's reach, impact, strengths and weaknesses after two years of operation; and our anticipated challenges and planned strategies for maintaining the programme over the medium- and long-term.

Sayogoa, Djoko Sigit, and Theresa A. Pard. "Exploring the Determinants of Scientific Data Sharing: Understanding the Motivation to Publish Research Data."

Government Information Quarterly 30, no. S1 (2013): S19-S31.
<https://doi.org/10.1016/j.giq.2012.06.011>

Scaramozzino, Jeanine Marie, Marisa L. Ramírez, and Karen J. McGaughey. "A Study of Faculty Data Curation Behaviors and Attitudes at a Teaching-Centered University." *College & Research Libraries* 73, no. 4 (2012): 349-365.
<https://doi.org/10.5860/crl-255>

Schirrwagen, Jochen, Philipp Cimiano, Vidya Ayer, Christian Pietsch, Cord Wiljes, Johanna Vompras, Dirk Pieper, "Expanding the Research Data Management Service Portfolio at Bielefeld University According to the Three-pillar Principle Towards Data FAIRness." *Data Science Journal*, 18, no. 1 (2019): 6. <http://doi.org/10.5334/dsj-2019-006>

Research Data Management at Bielefeld University is considered as a cross-cutting task among central facilities and research groups at the faculties. While initially started as project "Bielefeld Data Informium" lasting over seven years (2010–2015), it is now being expanded by setting up a Competence Center for Research Data. The evolution of the institutional RDM is based on the three-pillar principle: 1. Policies, 2. Technical infrastructure and 3. Support structures. The problem of data quality and the issues with reproducibility of research data is addressed in the project Conquaire. It is creating an infrastructure for the processing and versioning of research data which will finally allow publishing of research data in the institutional repository. Conquaire extends the existing RDM infrastructure in three ways: with a Collaborative Platform, Data Quality Checking, and Reproducible Research.

Schirrwagen, Jochen, Paolo Manghi, Natalia Manola, Lukasz Bolikowski, Najla Rettberg, and Birgit Schmidt. "Data Curation in the OpenAIRE Scholarly Communication Infrastructure." *Information Standards Quarterly* 25, no. 3 (2013): 13-19. <https://www.niso.org/niso-io/2013/09/data-curation-openaire-scholarly-communication-infrastructure>

Schmidt, Birgit, and Jens Dierkes. "New Alliances for Research and Teaching Support: Establishing the Göttingen eResearch Alliance." *Program* 49, no. 4 (2015): 461-474. <http://dx.doi.org/10.1108/PROG-02-2015-0020>

Schmidt, Birgit, Birgit Gemeinholzer, and Andrew Treloar. "Open Data in Global Environmental Research: The Belmont Forum's Open Data Survey." *PLoS ONE* 11, no. 1 (2016): e0146695. <http://dx.doi.org/10.1371/journal.pone.0146695>

This paper presents the findings of the Belmont Forum's survey on Open Data which targeted the global environmental research and data infrastructure community. It highlights users' perceptions of the term "open data", expectations of infrastructure functionalities, and barriers and enablers for the sharing of data. A wide range of good practice examples was pointed out by the respondents which demonstrates a substantial uptake of data sharing through e-infrastructures and a further need for enhancement and consolidation. Among all policy responses, funder policies seem to be the most important motivator. This supports the conclusion that stronger mandates will strengthen the case for data sharing.

Schoöpfel, Joachim, Stéphane Chaudiron, Bernard Jacquemin, H  l  ne Prost, Marta Severo, and Florence Thiault. "Open Access to Research Data in Electronic Theses and Dissertations: An Overview." *Library Hi Tech* 32, no. 4 (2014): 612-627. <https://doi.org/10.1108/lht-06-2014-0058>

Schoöpfel, Joachim, and H  l  ne Prost. "Research Data Management in Social Sciences and Humanities: A Survey at the University of Lille (France)." *LIBREAS*, no. 29 (2016). <http://libreas.eu/ausgabe29/09schoepfel/>

The paper presents results from a campus-wide survey at the University of Lille (France) on research data management in social sciences and humanities. The survey received 270 responses, equivalent to 15% of the whole sample of scientists, scholars, PhD students, administrative and technical staff (research management, technical support services); all disciplines were represented. The responses show a wide variety of practice and usage. The results are discussed regarding job status and disciplines and compared to other surveys. Four groups can be distinguished, i.e. pioneers (20-25%), motivated (25-30%), unaware (30%) and reluctant (5-10%). Finally, the next steps to improve the research data management on the campus are presented.

Schoots, Fieke, Laurents Sesink, Peter Verhaar, and Floor Frederiks. "Implementing a Research Data Policy at Leiden University." *International Journal of Digital Curation* 12, no. 2 (2017): 256-265. <https://doi.org/10.2218/ijdc.v12i2.575>

In this paper, we discuss the various stages of the institution-wide project that lead to the adoption of the data management policy at Leiden University in 2016. We illustrate this process by highlighting how we have involved all

stakeholders. Each organisational unit was represented in the project teams. Results were discussed in a sounding board with both academic and support staff. Senior researchers acted as pioneers and raised awareness and commitment among their peers. By way of example, we present pilot projects from two faculties. We then describe the comprehensive implementation programme that will create facilities and services that must allow implementing the policy as well as monitoring and evaluating it. Finally, we will present lessons learnt and steps ahead. The engagement of all stakeholders, as well as explicit commitment from the Executive Board, has been an important key factor for the success of the project and will continue to be an important condition for the steps ahead.

Schopf, Jennifer M., and Steven Newhouse. "User Priorities for Data: Results from SUPER." *International Journal of Digital Curation* 2, no. 1 (2007): 149-155. <https://doi.org/10.2218/ijdc.v2i1.23>

SUPER, a Study of User Priorities for e-infrastructure for Research, was a six-month effort funded by the UK e-Science Core Programme and JISC. Its aim was to inform investment in order to provide a usable, useful, and accessible e-infrastructure for all researchers and a coherent set of e-infrastructure services that would increase usage by at least a factor of ten by 2010. Through a series of unstructured face-to-face interviews with over 45 participants from 30 different projects, an online survey, together with a day-long workshop at NeSC, we have observed recurring issues relating to the provision of e-infrastructure. In this article we focus on the data-related issues identified during these interactions. We conclude with a prioritised list of future activities for research, development, and adoption in the data space.

Schubert, Carolyn, Yasmeeen Shorish, Paul Frankel, and Kelly Giles. "The Evolution of Research Data: Strategies for Curation and Data Management." *Library Hi Tech News* 30, no. 6 (2013): 1-6. <https://doi.org/10.1108/lhtn-06-2013-0035>

Schumacher, Jaime, and Drew VandeCreek. "Intellectual Capital at Risk: Data Management Practices and Data Loss by Faculty Members at Five American Universities." *International Journal of Digital Curation* 10, no. 2 (2015): 96-109. <https://doi.org/10.2218/ijdc.v10i2.321>

A study of 56 professors at five American universities found that a majority had little understanding of principles, well-known in the field of data curation,

informing the ongoing administration of digital materials and chose to manage and store work-related data by relying on the use of their own storage devices and cloud accounts. It also found that a majority of them had experienced the loss of at least one work-related digital object that they considered to be important in the course of their professional career. Despite such a rate of loss, a majority of respondents expressed at least a moderate level of confidence that they would be able to make use of their digital objects in 25 years. The data suggest that many faculty members are unaware that their data is at risk. They also indicate a strong correlation between faculty members' digital object loss and their data management practices. University professors producing digital objects can help themselves by becoming aware that these materials are subject to loss. They can also benefit from awareness and use of better personal data management practices, as well as participation in university-level programmatic digital curation efforts and the availability of more readily accessible, robust infrastructure for the storage of digital materials.

Schumann, Natascha. "Tried and Trusted: Experiences with Certification Processes at the GESIS Data Archive." *IASSIST Quarterly* 36, no. 3/4 (2012): 24-27. <https://doi.org/10.29173/iq474>.

Schumann, Natascha, and Reiner Mauer. "The GESIS Data Archive for the Social Sciences: A Widely Recognised Data Archive on its Way." *International Journal of Digital Curation* 8, no. 2 (2013): 215-222. <https://doi.org/10.2218/ijdc.v8i2.285>

This paper describes initial experiences in evaluating an established data archive with a long-standing commitment to preservation and dissemination of social science research data against recently formulated standards for trustworthy digital archives. As stakeholders need to be sure that the data they produce, use or fund is treated according to common standards, the GESIS Data Archive decided to start a process of audit and certification within the European Framework of Certification and Audit, starting with the Data Seal of Approval (DSA). This paper gives an overview of workflows within the archive and illustrates some of the steps necessary to obtain the DSA as well as to optimize some of its services. Finally, a short appraisal of the method of the DSA is made.

Schumann, Natascha, and Astrid Recker. "De-mystifying OAIS Compliance: Benefits and Challenges of Mapping the OAIS Reference Model to the GESIS

Data Archive." *IASSIST Quarterly* 36, no. 2 (2012): 6-11.
<https://doi.org/10.29173/iq769>

Schweers, Stefan, Katharina Kinder-Kurlanda, Stefan Müller, and Pascal Siegers. "Conceptualizing a Spatial Data Infrastructure for the Social Sciences: An Example from Germany." *Journal of Map & Geography Libraries* 12, no. 1 (2016): 100-126. <http://dx.doi.org/10.1080/15420353.2015.1100152>

Searle, Samantha, Malcolm Wolski, Natasha Simons, and Joanna Richardson. "Librarians as Partners in Research Data Service Development at Griffith University." *Program* 49, no. 4 (2015): 440-460. <http://dx.doi.org/10.1108/PROG-02-2015-0013>

Semeler, Alexandre Ribas, Adilson Luiz Pinto, and Helen Beatriz Frota Rozados. "Data Science in Data Librarianship: Core Competencies of a Data Librarian." *Journal of Librarianship and Information Science* 51, no. 3 (September 2019): 771–80. <https://doi.org/10.1177/0961000617742465>.

Sesarti, Ana , Andreas Fischlin, and Matthias Töwe. "Towards Narrowing the Curation Gap-Theoretical Considerations and Lessons Learned from Decades of Practice" *ISPRS International Journal of Geo-Information* 5, no. 6 (2016).
<https://doi.org/10.3390/ijgi5060091>

Sesarti, Ana, and Matthias Töwe. "Research Data Services at ETH-Bibliothek." *IFLA Journal* 42, no. 4 (2016): 284–291.
<https://doi.org/10.1177/0340035216674971>

Shadbolt, Anna, Leo Konstantelos, Liz Lyon, and Marieke Guy. "Delivering Innovative RDM Training: The immersiveInformatics Pilot Programme." *International Journal of Digital Curation* 9, no. 1 (2014): 313-323.
<https://doi.org/10.2218/ijdc.v9i1.318>

This paper presents the findings, lessons learned and next steps associated with the implementation of the immersiveInformatics pilot: a distinctive research data management (RDM) training programme designed in collaboration between UKOLN Informatics and the Library at the University of Melbourne, Australia. The pilot aimed to equip a broad range of academic and professional staff roles with RDM skills as a key element of capacity and capability building within a single institution.

Shaffer, Christopher J. "The Role of the Library in the Research Enterprise." *Journal of eScience Librarianship* 2, no. 1 (2013): e1043. <http://dx.doi.org/10.7191/jeslib.2013.1043>

Shankar, Kalpana. "For Want of a Nail: Three Tropes in Data Curation." *Preservation, Digital Technology & Culture* 44, no. 4 (2016): 161-170. <https://doi.org/10.1515/pdte-2015-0019>

Shaon, Arif, Sarah Callaghan, Bryan Lawrence, Brian Matthews, Timothy Osborn, Colin Harpham, and Andrew Woolf. "Opening Up Climate Research: A Linked Data Approach to Publishing Data Provenance." *International Journal of Digital Curation* 7, no. 1 (2012): 163-173. <https://doi.org/10.2218/ijdc.v7i1.223>

Traditionally, the formal scientific output in most fields of natural science has been limited to peer-reviewed academic journal publications, with less attention paid to the chain of intermediate data results and their associated metadata, including provenance. In effect, this has constrained the representation and verification of the data provenance to the confines of the related publications. Detailed knowledge of a dataset's provenance is essential to establish the pedigree of the data for its effective re-use, and to avoid redundant re-enactment of the experiment or computation involved. It is increasingly important for open-access data to determine their authenticity and quality, especially considering the growing volumes of datasets appearing in the public domain. To address these issues, we present an approach that combines the Digital Object Identifier (DOI)—a widely adopted citation technique—with existing, widely adopted climate science data standards to formally publish detailed provenance of a climate research dataset as an associated scientific workflow. This is integrated with linked-data compliant data re-use standards (e.g. OAI-ORE) to enable a seamless link between a publication and the complete trail of lineage of the corresponding dataset, including the dataset itself.

Shaon, Arif, and Andrew Woolf. "Long-Term Preservation for Spatial Data Infrastructures: A Metadata Framework and Geo-portal Implementation." *D-Lib Magazine* 17, no. 9/10 (2012). <https://doi.org/10.1045/september2011-shaon>

Shen, Yi. "Data Sustainability and Reuse Pathways of Natural Resources and Environmental Scientists." *New Review of Academic Librarianship* 24, no. 2 (2018): 136-156. <https://doi.org/10.1080/13614533.2018.1424642>

———. "Research Data Sharing and Reuse Practices of Academic Faculty Researchers: A Study of the Virginia Tech Data Landscape." *International Journal of Digital Curation* 10, no. 2 (2015): 157-175.
<https://doi.org/10.2218/ijdc.v10i2.359>

This paper presents the results of a research data assessment and landscape study in the institutional context of Virginia Tech to determine the data sharing and reuse practices of academic faculty researchers. Through mapping the level of user engagement in "openness of data," "openness of methodologies and workflows," and "reuse of existing data," this study contributes to the current knowledge in data sharing and open access, and supports the strategic development of institutional data stewardship. Asking faculty researchers to self-reflect sharing and reuse from both data producers' and data users' perspectives, the study reveals a significant gap between the rather limited sharing activities and the highly perceived reuse or repurpose values regarding data, indicating that potential values of data for future research are lost right after the original work is done. The localized and sporadic data management and documentation practices of researchers also contribute to the obstacles they themselves often encounter when reusing existing data.

———. "Strategic Planning for a Data-Driven, Shared-Access Research Enterprise: Virginia Tech Research Data Assessment and Landscape Study." *College & Research Libraries* 77, no. 4 (2016): 500-519.
<https://doi.org/10.5860/crl.77.4.500>

Shorish, Yasmeen. "Data Curation Is for Everyone! The Case for Master's and Baccalaureate Institutional Engagement with Data Curation." *Journal of Web Librarianship* 6, no. 4 (2012): 263-273.
<https://doi.org/10.1080/19322909.2012.729394>

Silvello, Gianmaria. "Theory and Practice of Data Citation." *Journal of the Association for Information Science and Technology* 69, no. 1 (2018): 6-20.
<https://doi.org/10.1002/asi.23917>

Simms, Stephanie, and Sarah Jones. "Next-Generation Data Management Plans: Global, Machine-Actionable, FAIR." *International Journal of Digital Curation* 12, no. 1 (2017): 36-45. <https://doi.org/10.2218/ijdc.v12i1.513>

At IDCC 2016 the Digital Curation Centre (DCC) and University of California Curation Center (UC3) at the California Digital Library (CDL)

announced plans to merge our respective data management planning tools, DMPonline and DMPTool, into a single platform. By formalizing our partnership and co-developing a core infrastructure for data management plans (DMPs), we aim to meet the skyrocketing demand for our services in our national, and increasingly international, contexts. The larger goal is to engage with what is now a global DMP agenda and help make DMPs a more useful exercise for all stakeholders in the research enterprise. This year we offer a progress report that encompasses our co-development roadmap and future enhancements focused on implementing use cases for machine-actionable DMPs.

Simms, Stephanie, Sarah Jones, Kevin Ashley, Marta Ribeiro, John Chodacki, Stephen Abrams, and Marisa Strong. "Roadmap: A Research Data Management Advisory Platform." *Research Ideas and Outcomes* 2 (2016): e8649.
<https://doi.org/10.3897/rio.2.e8649>

Simms, Stephanie, Sarah Jones, Tomasz Miksa, Daniel Mietchen, Natasha Simons, Kathryn Unsworth. "A Landscape Survey of #ActiveDMPs." *International Journal of Digital Curation* 13, no. 1 (2018): 204-214.
<https://doi.org/10.2218/ijdc.v13i1.629>

Machine-actionable or 'active' Data Management Plans have gathered a great deal of interest over recent years, with many groups worldwide discussing a vision of how DMPs can enable researchers to manage their data and connect them with service providers for support. Discussions are focused on converting DMPs from a stick to a carrot. Researchers and other stakeholders must come to regard them as a benefit: something useful for doing their research, a manifest of their methods and outputs that can be used for reporting, evaluation and implementation, rather than an annoying administrative burden.

This paper reviews the work underway by different groups to gather user requirements and trial solutions. It notes several international fora where discussions are taking place and lists DMP platforms in active development. We offer a summary of where things are going, who needs to be involved and how we can include them. We conclude with next steps for machine-actionable DMPs that focus on continuing efforts to connect interested parties, share ideas, experiment in multiple directions to test these concepts and turn machine-actionable DMPs into reality.

Simms, Stephanie, Marisa Strong, Sarah Jones, and Marta Ribeiro. "The Future of Data Management Planning: Tools, Policies, and Players." *International Journal of Digital Curation* 11, no. 1 (2016): 208-217. <https://doi.org/10.2218/ijdc.v11i1.413>

DMPonline and the DMPTool are well-established tools for data management planning. As the software of each matures and the user communities grow, we turn our attention to issues of sustainability, culture change, and international collaboration. Here we outline strategies for addressing these issues. We propose to build a new, global framework for data management planning that links plans to researchers, funders, publications, data, and other components of the research lifecycle. By refocusing our efforts from promoting the creation of data management plans (DMPs) to comply with funder requirements to supporting the creation of good DMPs that can be implemented, we seek to further enable the open scholarship revolution, advancing science and society.

Simons, Natasha. "Implementing DOIs for Research Data." *D-Lib Magazine* 18, no. 5/6 (2012). <https://doi.org/10.1045/may2012-simons>

Simons, Natasha, Karen Visser, and Sam Searle. "Growing Institutional Support for Data Citation: Results of a Partnership Between Griffith University and the Australian National Data Service." *D-Lib Magazine* 19, no. 11/12 (2013). <https://doi.org/10.1045/november2013-simons>

Smit, Eefke. "Eloise and Abelard: Why Data and Publications Belong Together." *D-Lib Magazine* 17, no. 1/2 (2011). <https://doi.org/10.1045/january2011-smit>

Smith, Jordan W., William S. Slocumb, Charlynn Smith, and Jason Matney. "A Needs-Assessment Process for Designing Geospatial Data Management Systems within Federal Agencies." *Journal of Map & Geography Libraries* 11, no. 2 (2015): 226-244. <http://dx.doi.org/10.1080/15420353.2015.1048035>

Southall, John, and Catherine Scutt. "Training for Research Data Management at the Bodleian Libraries: National Contexts and Local Implementation for Researchers and Librarians." *New Review of Academic Librarianship* 23, no. 2/3 (2017): 303-322. <http://dx.doi.org/10.1080/13614533.2017.1318766>

Soyka, Heather, Amber Budden, Viv Hutchison, David Bloom, Jonah Duckles, Amy Hodge, Matthew S. Mayernik, Timothée Poisot, Shannon Rauch, Gail Steinhart, Leah Wasser, Amanda L. Whitmire, and Stephanie Wright. "Using Peer Review to Support Development of Community Resources for Research Data

Management." *Journal of eScience Librarianship* 6, no. 2 (2017): e1114.
<https://doi.org/10.7191/jeslib.2017.1114>

Objective: To ensure that resources designed to teach skills and best practices for scientific research data sharing and management are useful, the maintainers of those materials need to evaluate and update them to ensure their accuracy, currency, and quality. This paper advances the use and process of outside peer review for community resources in addressing ongoing accuracy, quality, and currency issues. It further describes the next step of moving the updated materials to an online collaborative community platform for future iterative review in order to build upon mechanisms for open science, ongoing iteration, participation, and transparent community engagement.

Setting: Research data management resources were developed in support of the DataONE (Data Observation Network for Earth) project, which has deployed a sustainable, long-term network to ensure the preservation and access to multi-scale, multi-discipline, and multi-national environmental and biological science data (Michener et al. 2012). Created by members of the Community Engagement and Education (CEE) Working Group in 2011-2012, the freely available Educational Modules included three complementary components (slides, handouts, and exercises) that were designed to be adaptable for use in classrooms as well as for research data management training.

Methods: Because the modules were initially created and launched in 2011-2012, the current members of the (renamed) Community Engagement and Outreach (CEO) Working Group were concerned that the materials could be and / or quickly become outdated and should be reviewed for accuracy, currency, and quality. In November 2015, the Working Group developed an evaluation rubric for use by outside reviewers. Review criteria were developed based on surveys and usage scenarios from previous DataONE projects. Peer reviewers were selected from the DataONE community network for their expertise in the areas covered by one of the 11 educational modules. Reviewers were contacted in March 2016, and were asked to volunteer to complete their evaluations online within one month of the request, by using a customized Google form.

Results: For the 11 modules, 22 completed reviews were received by April 2016 from outside experts. Comments on all three components of each

module (slides, handouts, and exercises) were compiled and evaluated by the postdoctoral fellow attached to the CEO Working Group. These reviews contributed to the full evaluation and revision by members of the Working Group of all educational modules in September 2016. This review process, as well as the potential lack of funding for ongoing maintenance by Working Group members or paid staff, provoked the group to transform the modules to a more stable, non-proprietary format, and move them to an online open repository hosting platform, GitHub. These decisions were made to foster sustainability, community engagement, version control, and transparency.

Conclusion: Outside peer review of the modules by experts in the field was beneficial for highlighting areas of weakness or overlap in the education modules. The modules were initially created in 2011-2012 by an earlier iteration of the Working Group, and updates were needed due to the constant evolving practices in the field. Because the review process was lengthy (approximately one year) comparative to the rate of innovations in data management practices, the Working Group discussed other options that would allow community members to make updates available more quickly. The intent of migrating the modules to an online collaborative platform (GitHub) is to allow for iterative updates and ongoing outside review, and to provide further transparency about accuracy, currency, and quality in the spirit of open science and collaboration. Documentation about this project may be useful for others trying to develop and maintain educational resources for engagement and outreach, particularly in communities and spaces where information changes quickly, and open platforms are already in common use.

This work is licensed under a Creative Commons 1.0 Universal Public Domain Dedication, <https://creativecommons.org/publicdomain/zero/1.0/>.

Stamatoplos, Anthony, Tina Neville, and Deborah Henry. "Analyzing the Data Management Environment in a Master's-level Institution." *The Journal of Academic Librarianship* 42, no. 2 (2016): 109-190.
<http://dx.doi.org/10.1016/j.acalib.2015.11.004>

Starr, Joan, Eleni Castro, Mercè Crosas, Michel Dumontier, Robert R. Downs, Ruth Duerr, Laurel L. Haak, Melissa Haendel, Ivan Herman, Simon Hodson, Joe Hourclé, John Ernest Kratz, Jennifer Lin, Lars Holm Nielsen, Amy Nurnberger, Stefan Proel, Andreas Rauber, Simone Sacchi, Arthur Smith, Mike Taylor, and Tim Clark. "Achieving Human and Machine Accessibility of Cited Data in

Scholarly Publications." *PeerJ Computer Science* 1 (2015): e1.
<http://dx.doi.org/10.7717/peerj-cs.1>

Reproducibility and reusability of research results is an important concern in scientific communication and science policy. A foundational element of reproducibility and reusability is the open and persistently available presentation of research data. However, many common approaches for primary data publication in use today do not achieve sufficient long-term robustness, openness, accessibility or uniformity. Nor do they permit comprehensive exploitation by modern Web technologies. This has led to several authoritative studies recommending uniform direct citation of data archived in persistent repositories. Data are to be considered as first-class scholarly objects, and treated similarly in many ways to cited and archived scientific and scholarly literature. Here we briefly review the most current and widely agreed set of principle-based recommendations for scholarly data citation, the Joint Declaration of Data Citation Principles (JDDCP). We then present a framework for operationalizing the JDDCP; and a set of initial recommendations on identifier schemes, identifier resolution behavior, required metadata elements, and best practices for realizing programmatic machine actionability of cited data. The main target audience for the common implementation guidelines in this article consists of publishers, scholarly organizations, and persistent data repositories, including technical staff members in these organizations. But ordinary researchers can also benefit from these recommendations. The guidance provided here is intended to help achieve widespread, uniform human and machine accessibility of deposited data, in support of significantly improved verification, validation, reproducibility and re-use of scholarly/scientific data.

This work is licensed under a Creative Commons 1.0 Universal Public Domain Dedication, <https://creativecommons.org/publicdomain/zero/1.0/>.

Starr, Joan, and Angela Gastl. "isCitedBy: A Metadata Scheme for DataCite." *D-Lib Magazine* 17, no. 1/2 (2011).
<http://www.dlib.org/dlib/january11/starr/01starr.html>

Starr, Joan, Perry Willett, Lis Federer, Horning Claudia, and Mary Linn Bergstrom. "A Collaborative Framework for Data Management Services: The Experience of the University of California." *Journal of eScience Librarianship* 1, no. 2 (2012): e1014. <http://dx.doi.org/10.7191/jeslib.2012.1014>

Štebe, Janez. (2020). "Examining Barriers for Establishing a National Data Service." *IASSIST Quarterly* 43, no. 4 (2019): 1–14. <https://doi.org/10.29173/iq960>

Steeleworthy, Michael. "Research Data Management and the Canadian Academic Library: An Organizational Consideration of Data Management and Data Stewardship." *Partnership: the Canadian Journal of Library and Information Practice and Research* 9, no. 1 (2014).
<https://doi.org/10.21083/partnership.v9i1.2990>

Steinhart, Gail. "DataStaR: A Data Sharing and Publication Infrastructure to Support Research." *Agricultural Information Worldwide: An International Journal for the Information Specialists in Agriculture, Natural Resources, and the Environment* 4, no. 1 (2011).
<http://journals.sfu.ca/iaald/index.php/aginfo/article/view/199>

———. "Libraries as Distributors of Geospatial Data: Data Management Policies as Tools for Managing Partnerships." *Library Trends* 55, no. 2 (2006): 264-284.
<https://doi.org/10.1353/lib.2006.0063>

Steinhart, Gail, Eric Chen, Florio Arguillas, Dianne Dietrich, and Stefan Kramer. "Prepared to Plan? A Snapshot of Researcher Readiness to Address Data Management Planning Requirements." *Journal of eScience Librarianship* 1, no. 2 (2012): e1008. <http://dx.doi.org/10.7191/jeslib.2012.1008>

Steinhart, Gail, Dianne Dietrich, and Ann Green. "Establishing Trust in a Chain of Preservation: The TRAC Checklist Applied to a Data Staging Repository (DataStaR)." *D-Lib Magazine* 16, no. 9/10 (2009).
<https://doi.org/10.1045/september2009-steinhart>

Stigler, Stephen M. "Data Have a Limited Shelf Life." *Harvard Data Science Review* 1, no. 2 (2019). <https://doi.org/10.1162/99608f92.f9a1e510>

Data, unlike some wines, do not improve with age. The contrary view, that data are immortal, a view that may underlie the often-observed tendency to recycle old examples in texts and presentations, is illustrated with three classical examples and rebutted by further examination. Some general lessons for data science are noted, as well as some history of statistical worries about the effect of data selection on induction and related themes in recent histories of science.

Stockhause, Martina, and Michael Lautenschlager. "CMIP6 Data Citation of Evolving Data." *Data Science Journal* 16, no. 30 (2017).
<http://doi.org/10.5334/dsj-2017-030>

Data citations have become widely accepted. Technical infrastructures as well as principles and recommendations for data citation are in place but best practices or guidelines for their implementation are not yet available. On the other hand, the scientific climate community requests early citations on evolving data for credit, e.g. for CMIP6 (Coupled Model Intercomparison Project Phase 6). The data citation concept for CMIP6 is presented. The main challenges lie in limited resources, a strict project timeline and the dependency on changes of the data dissemination infrastructure ESGF (Earth System Grid Federation) to meet the data citation requirements. Therefore a pragmatic, flexible and extendible approach for the CMIP6 data citation service was developed, consisting of a citation for the full evolving data superset and a data cart approach for citing the concrete used data subset. This two citation approach can be implemented according to the RDA recommendations for evolving data. Because of resource constraints and missing project policies, the implementation of the second part of the citation concept is postponed to CMIP7.

Strasser, Carly. *Research Data Management: A Primer Publication of the National Information Standards Organization*. Baltimore, MD: NISO, 2015.
<https://www.niso.org/publications/primer-research-data-management>

Strasser, Carly, Stephen Abrams, and Patricia Cruse. "DMPTool 2: Expanding Functionality for Better Data Management Planning." *International Journal of Digital Curation* 9, no. 1 (2014): 324-330. <https://doi.org/10.2218/ijdc.v9i1.319>

Scholarly researchers today are increasingly required to engage in a range of data management planning activities to comply with institutional policies, or as a precondition for publication or grant funding. The latter is especially true in the U.S. in light of the recent White House Office of Science and Technology Policy (OSTP) mandate aimed at maximizing the availability of all outputs—data as well as the publications that summarize them—resulting from federally-funded research projects.

To aid researchers in creating effective data management plans (DMPs), a group of organizations—California Digital Library, DataONE, Digital Curation Centre, Smithsonian Institution, University of Illinois Urbana-

Champaign, and University of Virginia Library—collaborated on the development of the DMPTool, an online application that helps researchers create data management plans. The DMPTool provides detailed guidance, links to general and institutional resources, and walks a researcher through the process of generating a comprehensive plan tailored to specific DMP requirements. The uptake of the DMPTool has been positive: to date, it has been used by over 6,000 researchers from 800 institutions, making use of more than 20 requirements templates customized for funding bodies.

With support from the Alfred P. Sloan Foundation, project partners are now engaged in enhancing the features of the DMPTool. The second version of the tool has enhanced functionality for plan creators and institutional administrators, as well as a redesigned user interface and an open RESTful application programming interface (API).

New administrative functions provide the means for institutions to better support local research activities. New capabilities include support for plan co-ownership; workflow provisions for internal plan review; simplified maintenance and addition of DMP requirements templates; extensive capabilities for the customization of guidance and resources by local institutional administrators; options for plan visibility; and UI refinements based on user feedback and focus group testing. The technical work undertaken for the DMPTool Version 2 has been accompanied by a new governance structure and the growth of a community of engaged stakeholders who will form the basis for a sustainable path forward for the DMPTool as it continues to play an important role in research data management activities.

Strasser, Carly, and Stephanie E. Hampton. "The Fractured Lab Notebook: Undergraduates and Ecological Data Management Training in the United States." *Ecosphere* 3, no. 12 (2012): art116. <http://dx.doi.org/10.1890/es12-00139.1>

Data management is a timely and increasingly important topic for ecologists. Recent funder mandates requiring data management plans, combined with the data deluge that faces scientists, make education about data management critical for any future ecologist. In this study, we surveyed instructors of general ecology courses at 48 major institutions in the United States. We chose instructors at institutions that are likely to train future ecologists, and therefore, are most likely to influence the trajectory of data management education in this field. The survey queried instructors about institution and course characteristics, the extent to which data-related topics are included in

their courses, the barriers to their teaching these topics, and their own personal beliefs and values associated with data management and stewardship. We found that, in general, data management topics are not being covered in undergraduate ecology courses for a wide range of reasons. Most often, instructors cited a lack of time and a lack of resources as barriers to teaching data management. Although data are used for instruction at some point in the majority of the courses surveyed, good data management practices and a thorough understanding of the importance of data stewardship are not being taught. We offer potential explanations for this and suggestions for improvement.

This work is licensed under a Creative Commons Attribution 3.0 Unported License, <https://creativecommons.org/licenses/by/3.0/>.

Strasser, Carly, John Kunze, Stephen Abrams, and Patricia Cruse. "DataUp: A Tool to Help Researchers Describe and Share Tabular Data." *F1000Research* 3, no. 6 (2014). <http://dx.doi.org/10.12688/f1000research.3-6.v2>

Scientific datasets have immeasurable value, but they lose their value over time without proper documentation, long-term storage, and easy discovery and access. Across disciplines as diverse as astronomy, demography, archeology, and ecology, large numbers of small heterogeneous datasets (i.e., the long tail of data) are especially at risk unless they are properly documented, saved, and shared. One unifying factor for many of these at-risk datasets is that they reside in spreadsheets. In response to this need, the California Digital Library (CDL) partnered with Microsoft Research Connections and the Gordon and Betty Moore Foundation to create the DataUp data management tool for Microsoft Excel. Many researchers creating these small, heterogeneous datasets use Excel at some point in their data collection and analysis workflow, so we were interested in developing a data management tool that fits easily into those work flows and minimizes the learning curve for researchers. The DataUp project began in August 2011. We first formally assessed the needs of researchers by conducting surveys and interviews of our target research groups: earth, environmental, and ecological scientists. We found that, on average, researchers had very poor data management practices, were not aware of data centers or metadata standards, and did not understand the benefits of data management or sharing. Based on our survey results, we composed a list of desirable components and requirements and solicited feedback from the community to prioritize potential features of the DataUp tool. These requirements were then relayed

to the software developers, and DataUp was successfully launched in October 2012.

This work is licensed under a Creative Commons Attribution 3.0 Unported License, <https://creativecommons.org/licenses/by/3.0/>.

Sturges, Paul, Marianne Bamkin, Jane H.S. Anders, Bill Hubbard, Azhar Hussain, and Melanie Heeley. "Research Data Sharing: Developing a Stakeholder-Driven Model for Journal Policies." *Journal of the Association for Information Science and Technology* 66, no. 12 (2015): 2445-2455 <https://doi.org/10.1002/asi.23336>

Sun, Guangyuan, Christopher Soo, and Guan Khoo. "Social Science Research Data Curation: Issues of Reuse." *Libellarium: Journal for the Research of Writing, Books, and Cultural Heritage Institutions* 9, no. 2 (2016). <http://dx.doi.org/10.15291/libellarium.v9i2.291>

Data curation is attracting a growing interest in the library and information science community. The main purpose of data curation is to support data reuse. This paper discusses the issues of reusing quantitative social science data from three perspectives of searching and browsing for datasets, evaluating the reusability of datasets (including evaluating topical relevance, utility and data quality), and integrating datasets, by comparing dataset searching with online database searching. The paper also discusses using knowledge representation techniques of metadata and ontology, and a graphical visualization interface to support users in browsing, assessing and integrating datasets.

Sunikka, Anne. "Organising RDM and Open Science Services: Case Finland and Aalto University." *International Journal Of Digital Curation* 14 no. 1 (2019): 180-193. <https://doi.org/10.2218/ijdc.v14i1.641>

This paper describes how the Finnish Ministry of Education and Culture launched an initiative on research data management and open data, open access publishing, and open and collaborative ways of working in 2014. Most of the universities and research institutions took part in the collaborative initiative building new tools and training material for the Finnish research needs. Measures taken by one university, Aalto University, are described in detail and analysed, and compared with the activities taking place in other universities.

The focus of this paper is in the changing roles of experts at Aalto University, and organisational transformation that offers possibilities to serve academic personnel better. Various ways of building collaboration and arranging services are described, and their benefits and drawbacks are discussed.

Surkis, Alisa, Aileen McCrillis, Richard McGowan, Jeffrey Williams, Brian L. Schmidt, Markus Hardt, and Neil Rambo. "Informationist Support for a Study of the Role of Proteases and Peptides in Cancer Pain." *Journal of eScience Librarianship* 2, no. 1 (2013): e1029. <http://dx.doi.org/10.7191/jeslib.2013.1029>

Swanson, Juleah, and Amanda K. Rinehart. "Data in Context: Using Case Studies to Generate A Common Understanding of Data in Academic Libraries." *The Journal of Academic Librarianship* 42, no. 1 (2016): 97-101. <http://dx.doi.org/10.1016/j.acalib.2015.11.005>

Sweeney, Latanya, Mercè Crosas, and Michael Bar-Sinai. "Sharing Sensitive Data with Confidence: The Datatags System." *Technology Science*. (16 October 2015). <http://techscience.org/a/2015101601>

Sweetkind, Julie, Mary Lynette Larsgaard, and Tracey Erwin. "Digital Preservation of Geospatial Data." *Library Trends* 55, no. 2 (2006): 304-314. <https://doi.org/10.1353/lib.2006.0065>

Tamaro, Anna Maria, Krystyna K. Matusiak, Frank Andreas Sposito, Vittore Casarosa. "Data Curator's Roles and Responsibilities: An International Perspective." *Libri* 69, no. 2 (2019): 89-104. <https://doi.org/10.1515/libri-2018-0090>

Tananbaum, Greg. *Implementing an Open Data Policy: A Primer for Research Funders*. Washington, DC: Scholarly Publishing and Academic Resources Coalition, 2013. <http://sparcopen.org/our-work/implementing-an-open-data-policy/>

Tang, Rong, and Zhan Hu. "Providing Research Data Management (RDM) Services in Libraries: Preparedness, Roles, Challenges, and Training for RDM Practice." *Data and Information Management* 3, no. 2, (2019): 84-101. <https://doi.org/10.2478/dim-2019-0009>

Tarver, Hannah, and Mark Phillips. "Integrating Image-based Research Datasets into an Existing Digital Repository Infrastructure." *Cataloging & Classification Quarterly* 51, no. 1-3 (2013): 238-250. <https://doi.org/10.1080/01639374.2012.732203>

Teal, Tracy K., Karen A. Cranston, Hilmar Lapp, Ethan White, Greg Wilson, Karthik Ram, and Aleksandra Pawlik. "Data Carpentry: Workshops to Increase Data Literacy for Researchers." *International Journal of Digital Curation* 10, no. 1 (2015): 135-143. <https://doi.org/10.2218/ijdc.v10i1.351>

In many domains the rapid generation of large amounts of data is fundamentally changing how research is done. The deluge of data presents great opportunities, but also many challenges in managing, analyzing and sharing data. However, good training resources for researchers looking to develop skills that will enable them to be more effective and productive researchers are scarce and there is little space in the existing curriculum for courses or additional lectures. To address this need we have developed an introductory two-day intensive workshop, Data Carpentry, designed to teach basic concepts, skills, and tools for working more effectively and reproducibly with data.

These workshops are based on Software Carpentry: two-day, hands-on, bootcamp style workshops teaching best practices in software development, that have demonstrated the success of short workshops to teach foundational research skills. Data Carpentry focuses on data literacy in particular, with the objective of teaching skills to researchers to enable them to retrieve, view, manipulate, analyze and store their and other's data in an open and reproducible way in order to extract knowledge from data.

Tenopir, Carol, Suzie Allard, Kimberly Douglass, Arsev Umur Aydinoglu, Lei Wu, Eleanor Read, Maribeth Manoff, and Mike Frame. "Data Sharing by Scientists: Practices and Perceptions." *PLoS ONE* 6, no. 6 (2011): e21101. <https://doi.org/10.1371/journal.pone.0021101>

Background

Scientific research in the 21st century is more data intensive and collaborative than in the past. It is important to study the data practices of researchers— data accessibility, discovery, re-use, preservation and, particularly, data sharing. Data sharing is a valuable part of the scientific method allowing for verification of results and extending research from prior results.

Methodology/Principal Findings

A total of 1329 scientists participated in this survey exploring current data sharing practices and perceptions of the barriers and enablers of data sharing.

Scientists do not make their data electronically available to others for various reasons, including insufficient time and lack of funding. Most respondents are satisfied with their current processes for the initial and short-term parts of the data or research lifecycle (collecting their research data; searching for, describing or cataloging, analyzing, and short-term storage of their data) but are not satisfied with long-term data preservation. Many organizations do not provide support to their researchers for data management both in the short- and long-term. If certain conditions are met (such as formal citation and sharing reprints) respondents agree they are willing to share their data. There are also significant differences and approaches in data management practices based on primary funding agency, subject discipline, age, work focus, and world region.

Conclusions/Significance

Barriers to effective data sharing and preservation are deeply rooted in the practices and culture of the research process as well as the researchers themselves. New mandates for data management plans from NSF and other federal agencies and world-wide attention to the need to share and preserve data could lead to changes. Large scale programs, such as the NSF-sponsored DataNET (including projects like DataONE) will both bring attention and resources to the issue and make it easier for scientists to apply sound data management principles.

Tenopir, Carol, Suzie Allard, Priyanki Sinha, Danielle Pollock, Jess Newman, Elizabeth Dalton, Mike Frame, Lynn Baird. "Data Management Education from the Perspective of Science Educators." *International Journal of Digital Curation* 11, no. 1 (2016): 232-251. <https://doi.org/10.2218/ijdc.v11i1.389>

In order to better understand the current state of data management education in multiple fields of science, this study surveyed scientists, including information scientists, about their data management education practices, including at what levels they are teaching data management, which topics they covering, and what barriers they experience in teaching these topics. We found that a handful of scientists are teaching data management in undergraduate, graduate, and other types of courses, as well as outside of classroom settings. Commonly taught data management topics included quality control, protecting data, and management planning. However, few instructors felt they were covering data management topics thoroughly, and respondents cited barriers such as lack of time, lack of necessary expertise,

and lack of information for teaching data management. We offer some potential explanations for the existing state of data management education and suggest areas for further research.

Tenopir, Carol, Ben Birch, and Suzie Allard. *Academic Libraries and Research Data Services: Current Practices and Plans for the Future*. Chicago: Association of College and Research Libraries, 2012. <http://www.worldcat.org/oclc/821933602>

Tenopir, Carol, Dane Hughes, Suzie Allard, Mike Frame, Ben Birch, Lynn Baird, Robert Sandusky, Madison Langseth, and Andrew Lundeen. "Research Data Services in Academic Libraries: Data Intensive Roles for the Future?" *Journal of eScience Librarianship* 4, no. 2 (2015): e1085. <https://doi.org/10.7191/jeslib.2015.1085>

Tenopir, Carol, Robert J. Sandusky, Suzie Allard, and Ben Birch. "Academic Librarians and Research Data Services: Preparation and Attitudes." *IFLA Journal* 39, no. 1 (2013): 70-78. <https://doi.org/10.1177/0340035212473089>

———. "Research Data Management Services in Academic Research Libraries and Perceptions of Librarians." *Library & Information Science Research* 36, no. 2 (2014): 84-90. <https://doi.org/10.1016/j.lisr.2013.11.003>

Tenopir, Carol, Sanna Talja, Wolfram Horstmann, Elina Late, Dane Hughes, Danielle Pollock, Birgit Schmidt, Lynn Baird, Robert Sandusky, and Suzie Allard. "Research Data Services in European Academic Research Libraries." *LIBER Quarterly* 27, no. 1 (2017): 23–44. <http://doi.org/10.18352/lq.10180>

Research data is an essential part of the scholarly record, and management of research data is increasingly seen as an important role for academic libraries. This article presents the results of a survey of directors of the Association of European Research Libraries (LIBER) academic member libraries to discover what types of research data services (RDS) are being offered by European academic research libraries and what services are planned for the future. Overall, the survey found that library directors strongly agree on the importance of RDS. As was found in earlier studies of academic libraries in North America, more European libraries are currently offering or are planning to offer consultative or reference RDS than technical or hands-on RDS. The majority of libraries provide support for training in skills related to RDS for their staff members. Almost all libraries collaborate with other organizations inside their institutions or with outside institutions in order to offer or develop policy related to RDS. We discuss the implications of the current state of

RDS in European academic research libraries, and offer directions for future research.

Teperek, Marta , Maria J. Cruz, Ellen Verbakel, Jasmin Böhmer, and Alastair Dunning. "Data Stewardship Addressing Disciplinary Data Management Needs." *International Journal of Digital Curation* 13, no. 1 (2018): 141-149. <http://www.ijdc.net/article/view/604>

One of the biggest challenges for multidisciplinary research institutions which provide data management support to researchers is addressing disciplinary differences (Akers and Doty, 2013). Centralised services need to be general enough to cater for all the different flavours of research conducted in an institution. At the same time, focusing on the common denominator means that subject-specific differences and needs may not be effectively addressed. In 2017, Delft University of Technology (TU Delft) embarked on an ambitious Data Stewardship project, aiming to comprehensively address data management needs across a multi-disciplinary campus. In this article we describe the principles behind the Data Stewardship project at TU Delft, the progress so far, identify the key challenges and explain our plans for the future.

Teplitzky, Samantha. "Open Data, [Open] Access: Linking Data Sharing and Article Sharing in the Earth Sciences." *Journal of Librarianship and Scholarly Communication* 5, no. 1 (2017): eP2150. <http://doi.org/10.7710/2162-3309.2150>

INTRODUCTION The norms of a research community influence practice, and norms of openness and sharing can be shaped to encourage researchers who share in one aspect of their research cycle to share in another. Different sets of mandates have evolved to require that research data be made public, but not necessarily articles resulting from that collected data. In this paper, I ask to what extent publications in the Earth Sciences are more likely to be open access (in all of its definitions) when researchers open their data through the Pangaea repository. **METHODS** Citations from Pangaea data sets were studied to determine the level of open access for each article. **RESULTS** This study finds that the proportion of gold open access articles linked to the repository increased 25% from 2010 to 2015 and 75% of articles were available from multiple open sources. **DISCUSSION** The context for increased preference for gold open access is considered and future work linking researchers' decisions to open their work to the adoption of open access mandates is proposed.

Terry, Robert F., Katherine Littler, and Piero L. Olliaro. "Sharing Health Research Data—the Role of Funders in Improving the Impact." *F1000Research* 7 (2018): 1641 <https://doi.org/10.12688/f1000research.16523.2>

Recent public health emergencies with outbreaks of influenza, Ebola and Zika revealed that the mechanisms for sharing research data are neither being used, or adequate for the purpose, particularly where data needs to be shared rapidly.

A review of research papers, including completed clinical trials related to priority pathogens, found only 31% (98 out of 319 published papers, excluding case studies) provided access to all the data underlying the paper—65% of these papers give no information on how to find or access the data. Only two clinical trials out of 58 on interventions for WHO priority pathogens provided any link in their registry entry to the background data.

Interviews with researchers revealed a reluctance to share data included a lack of confidence in the utility of the data; an absence of academic-incentives for rapid dissemination that prevents subsequent publication and a disconnect between those who are collecting the data and those who wish to use it quickly. The role of the funders of research needs to change to address this. Funders need to engage early with the researchers and related stakeholders to understand their concerns and work harder to define the more explicitly the benefits to all stakeholders. Secondly, there needs to be a direct benefit to sharing data that is directly relevant to those people that collect and curate the data. Thirdly more work needs to be done to realise the intent of making data sharing resources more equitable, ethical and efficient. Finally, a checklist of the issues that need to be addressed when designing new or revising existing data sharing resources should be created. This checklist would highlight the technical, cultural and ethical issues that need to be considered and point to examples of emerging good practice that can be used to address them.

Thanos, Costantino, "Research Data Reusability: Conceptual Foundations, Barriers and Enabling Technologies." *Publications* 5, no. 2 (2017): 2. <https://doi.org/10.3390/publications5010002>

High-throughput scientific instruments are generating massive amounts of data. Today, one of the main challenges faced by researchers is to make the best use of the world's growing wealth of data. Data (re)usability is becoming a distinct characteristic of modern scientific practice. By data (re)usability, we

mean the ease of using data for legitimate scientific research by one or more communities of research (consumer communities) that is produced by other communities of research (producer communities). Data (re)usability allows the reanalysis of evidence, reproduction and verification of results, minimizing duplication of effort, and building on the work of others. It has four main dimensions: policy, legal, economic and technological. The paper addresses the technological dimension of data reusability. The conceptual foundations of data reuse as well as the barriers that hamper data reuse are presented and discussed. The data publication process is proposed as a bridge between the data author and user and the relevant technologies enabling this process are presented.

Thanos, Costantino, Friederike Klan, Kyriakos Kritikos, and Leonardo Candela. "White Paper on Research Data Service Discoverability." *Publications* 5, no. 1 (2017): 1. <https://doi.org/10.3390/publications5010001>

This White Paper reports the outcome of a Workshop on "Research Data Service Discoverability" held in the island of Santorini (GR) on 21-22 April 2016 and organized in the context of the EU funded Project "RDA-E3." The Workshop addressed the main technical problems that hamper an efficient and effective discovery of Research Data Services (RDSs) based on appropriate semantic descriptions of their functional and non-functional aspects. In the context of this White Paper, by RDSs are meant those data services that manipulate/transform research datasets for the purpose of gaining insight into complicated issues. In this White Paper, the main concepts involved in the discovery process of RDSs are defined; the RDS discovery process is illustrated; the main technologies that enable the discovery of RDSs are described; and a number of recommendations are formulated for indicating future research directions and making an automatic RDS discovery feasible.

Thelwall, Mike, and Kousha Kayvan. "Do Journal Data Sharing Mandates Work? Life Sciences Evidence from Dryad." *Aslib Journal of Information Management* 69, no. 1 (2017): 36-45. <https://doi.org/10.1108/AJIM-09-2016-0159>

Thessen, Anne E., and David J. Patterson. "Data Issues in the Life Sciences." *ZooKeys* 150 (2011): 15-51. <http://dx.doi.org/10.3897/zookeys.150.1766>

Thielen, Joanna, and Amanda Nichols Hess. "Advancing Research Data Management in the Social Sciences: Implementing Instruction for Education

Graduate Students into a Doctoral Curriculum." *Behavioral & Social Sciences Librarian* 36, no. 1 (2017): 16-30. <https://doi.org/10.1080/01639269.2017.1387739>

Thoegersen, Jennifer L. "Examination of Federal Data Management Plan Guidelines." *Journal of eScience Librarianship* 4, no. 1 (2015): e1072. <https://doi.org/10.7191/jeslib.2015.1072>

Thompson, Kristi, and Guoying Liu. "Lives in Data: Prominent Data Librarians, Archivists and Educators Share Their Thoughts." *International Journal of Librarianship* 2, no. 1 (2017). <https://doi.org/10.23974/ijol.2017.vol2.1.35>

We asked several data librarians, archivists and educators who have had prominent and interesting careers if they would be willing to let us profile them and share some of their thoughts on the field. Six graciously agreed to be interviewed via email. Many of our respondents played key roles in developing data services and infrastructure in their respective countries, while others are involved in building the future of the field through education, advancing standards, and advocacy.

Our virtual panel includes Tuomas J. Alaterä, Finland; Ann Green and Jian Qin, United States; Guangjing Li, China; Wendy Watkins, Canada; and Lynn Woolfrey, South Africa.

Thompson, Kristi, and Shenqin Yin. "The Development of Academic Data Services in Canada and China: Profiles of Data Services at Fudan University and the University of Windsor." *International Journal of Librarianship* 2, no. 1 (2017): 73-78. <https://doi.org/10.23974/ijol.2017.vol2.1.32>

Thomson, Sara Day. *Preserving Transactional Data*. Glasgow: Digital Preservation Coalition, 2016. <http://dx.doi.org/10.7207/twr16-02>

Tomaszewski, Robert. "Citations to Chemical Databases in Scholarly Articles: To Cite or Not to Cite?" *Journal of Documentation* 75 no. 6, (2019): 1317-1332. <https://doi.org/10.1108/JD-12-2018-0214>

Toups, Megan, and Michael Hughes. "When Data Curation Isn't: A Redefinition for Liberal Arts Universities." *Journal of Library Administration* 53, no. 4 (2013): 223-233. <https://doi.org/10.1080/01930826.2013.865386>

Treloar, Andrew. "Design and Implementation of the Australian National Data Service." *International Journal of Digital Curation* 4, no. 1 (2009): 125-137. <https://doi.org/10.2218/ijdc.v4i1.83>

This paper will describe the genesis and realisation of the Australian National Data Service (ANDS). It will commence by outlining the context within which ANDS was conceived, both in the international research and Australian research support domains. It will then describe the process that brought about the ANDS vision and the principles that informed the realisation of that vision. The paper will then outline each of the four ANDS programs (Developing Frameworks, Providing Utilities, Seeding the Commons, and Building Capabilities) while also discussing particular items of note about the approach ANDS is taking. The paper concludes by briefly examining related work in the UK and US.

———. "The Research Data Alliance: Globally Co-ordinated Action against Barriers to Data Publishing and Sharing." *Learned Publishing* 27, no. 5 (2014): 9-13. <https://doi.org/10.1087/20140503>

Treloar, Andrew, David Groenewegen, and Cathrine Harboe-Ree. "The Data Curation Continuum: Managing Data Objects in Institutional Repositories." *D-Lib Magazine* 13, no. 9/10 (2007). <https://doi.org/10.1045/september2007-treloar>

Treloar, Andrew, and Jens Klump. "Updating the Data Curation Continuum." *International Journal Of Digital Curation* 14 no. 1 (2019): 87-101. <https://doi.org/10.2218/ijdc.v14i1.643>

The Data Curation Continuum was developed as a way of thinking about data repository infrastructure. Since its original development over a decade ago, a number of things have changed in the data infrastructure domain. This paper revisits the thinking behind the original data curation continuum and updates it to respond to changes in research objects, storage models, and the repository landscape in general.

Treloar, Andrew, and Ross Wilkinson. "Access to Data for eResearch: Designing the Australian National Data Service Discovery Services." *International Journal of Digital Curation* 3, no. 2 (2008): 151-158. <https://doi.org/10.2218/ijdc.v3i2.66>

Much work on data repositories has derived from effort on document repositories. It is our contention that people do not access research data for the same reasons that they access research publications. We argue that it is

valuable to understand information needs, both immediate and contextual, in establishing both what information should be collected, what metadata are captured, and what discovery services should be established. We report on the information needs that we have collected in our efforts in establishing the Australian National Data Service. These needs cover much more than data – there are needs for information about the data, their creators, a need for overviews, and further requirements to do with proof, collaboration, and innovation. We provide an analysis of those needs, and a set of conclusions that has led to some implementation decisions for ANDS.

Treloar, Andrew, and Mingfang Wu. "Provenance in Support of ANDS' Four Transformations." *International Journal of Digital Curation* 11, no. 1 (2016): 183-194. <https://doi.org/10.2218/ijdc.v11i1.416>

This article introduces the provenance activities that are being carried out at the Australia National Data Services (ANDS). Since its beginning, ANDS has been promoting four data transformations so that Australia's research data become more valuable and reusable by researchers. Among many other activities that enable the four transformations, ANDS has been encouraging ANDS partners to capture and describe rich context at the time when a data collection is created. In 2015, ANDS funded a number of external projects that had provenance components. In addition, ANDS is working on the interoperability between the schema that is used by the ANDS research data registration and discovery service—Research Data Australia (RDA)—and the W3C recommended provenance standard, Provenance Ontology (PROV-O), and investigating how to enrich the schema to access provenance information. The article concludes by discussing the lessons we learnt and our future planned activity.

Trimble, Leanne, Cheryl Woods, Francine Berish, Daniel Jakubek, and Sarah Simpkin. "Collaborative Approaches to the Management of Geospatial Data Collections in Canadian Academic Libraries: A Historical Case Study." *Journal of Map & Geography Libraries* 11, no. 3 (2015): 330-358. <https://doi.org/10.1080/15420353.2015.1043067>

Tsoi, Ah Chung, Jeff McDonell, Andrew Treloar, and Ian Atkinson. "Dataset Acquisition, Accessibility, Annotation, E-Research Technologies (DART) Project." *International Journal on Digital Libraries* 7, no. 1/2 (2007): 53-55. <https://doi.org/10.1007/s00799-007-0019-4>

Tuyl, Steve Van, and Gabrielle Michalek. "Assessing Research Data Management Practices of Faculty at Carnegie Mellon University." *Journal of Librarianship and Scholarly Communication* 3, no. 3 (2015): eP1258. <http://doi.org/10.7710/2162-3309.1258>

INTRODUCTION Recent changes to requirements for research data management by federal granting agencies and by other funding institutions have resulted in the emergence of institutional support for these requirements. At CMU, we sought to formalize assessment of research data management practices of researchers at the institution by launching a faculty survey and conducting a number of interviews with researchers. **METHODS** We submitted a survey on research data management practices to a sample of faculty including questions about data production, documentation, management, and sharing practices. The survey was coupled with in-depth interviews with a subset of faculty. We also make estimates of the amount of research data produced by faculty. **RESULTS** Survey and interview results suggest moderate level of awareness of the regulatory environment around research data management. Results also present a clear picture of the types and quantities of data being produced at CMU and how these differ among research domains. Researchers identified a number of services that they would find valuable including assistance with data management planning and backup/storage services. We attempt to estimate the amount of data produced and shared by researchers at CMU. **DISCUSSION** Results suggest that researchers may need and are amenable to assistance with research data management. Our estimates of the amount of data produced and shared have implications for decisions about data storage and preservation. **CONCLUSION** Our survey and interview results have offered significant guidance for building a suite of services for our institution.

Ulbricht, Damian, Kirsten Elger, Roland Bertelmann, and Jens Klump. "panMetaDocs, eSciDoc, and DOIDB—An Infrastructure for the Curation and Publication of File-Based Datasets for GFZ Data Services." *ISPRS International Journal of Geo-Information* 5, no. 3 (2016): 25. <http://dx.doi.org/10.3390/ijgi5030025>

Ure, Jenny, Tasneem Irshad, Janet Hanley, Angus Whyte, Claudia Pagliari, Hilary Pinnock, and Brian McKinstry. "Curating Complex, Dynamic and Distributed Data: Telehealth as a Laboratory for Strategy." *International Journal of Digital Curation* 6, no. 2 (2011): 128-145. <https://doi.org/10.2218/ijdc.v6i2.207>

Telehealth monitoring data is now being collected across large populations of patients with chronic diseases such as stroke, hypertension, COPD and dementia. These large, complex and heterogeneous datasets, including distributed sensor and mobile datasets, present real opportunities for knowledge discovery and re-use, however they also generate new challenges for curation. This paper uses qualitative research with stakeholders in two nationally-funded telehealth projects to outline the perceptions, practices and preferences of different stakeholders with regard to data curation. Telehealth provides a living laboratory for the very different challenges implicit in designing and managing data infrastructure for embedded and ubiquitous computing. Here, technical and human agents are distributed, and interaction and state change is a central component of design, rather than an inconvenient challenge to it. The authors argue that there are lessons to be learned from other domains where data infrastructure has been radically rethought to address these challenges.

Valentino, Maura, and Michael Boock. "Data Management Services in Academic Libraries: A Case Study at Oregon State University." *Practical Academic Librarianship: The International Journal of the SLA Academic Division* 5, no. 2 (2015): 77-91. <https://journals.tdl.org/pal/index.php/pal/article/view/7001>

Libraries have been asked to provide many new services over the past several decades. This paper aims to show how data management services were incorporated into the services that Oregon State University provides to faculty and graduate students. The lessons learned are general and applicable to any research institute that needs to manage data or help others with managing data.

This work is licensed under a Creative Commons Attribution 2.5 License, <https://creativecommons.org/licenses/by/2.5/legalcode>.

Van de Sandt, Stephanie, Sünje Dallmeier-Tiessen, Artemis Lavasa, Vivien Petras. "The Definition of Reuse." *Data Science Journal* 18, no. 1 (2019): 22. <http://doi.org/10.5334/dsj-2019-022>

The ability to reuse research data is now considered a key benefit for the wider research community. Researchers of all disciplines are confronted with the pressure to share their research data so that it can be reused. The demand for data use and reuse has implications on how we document, publish and share research in the first place, and, perhaps most importantly, it affects how

we measure the impact of research, which is commonly a measurement of its use and reuse. It is surprising that research communities, policy makers, etc. have not clearly defined what use and reuse is yet.

We postulate that a clear definition of use and reuse is needed to establish better metrics for a comprehensive scholarly record of individuals, institutions, organizations, etc. Hence, this article presents a first definition of reuse of research data. Characteristics of reuse are identified by examining the etymology of the term and the analysis of the current discourse, leading to a range of reuse scenarios that show the complexity of today's research landscape, which has been moving towards a data-driven approach. The analysis underlines that there is no reason to distinguish use and reuse. We discuss what that means for possible new metrics that attempt to cover Open Science practices more comprehensively. We hope that the resulting definition will enable a better and more refined strategy for Open Science.

Van den Eynden, Veerle, and Louise Corti. "Advancing Research Data Publishing Practices for the Social Sciences: From Archive Activity to Empowering Researchers." *International Journal on Digital Libraries* 18, no. 2 (2017): 113-121. <https://doi.org/10.1007/s00799-016-0177-3>

Sharing and publishing social science research data have a long history in the UK, through long-standing agreements with government agencies for sharing survey data and the data policy, infrastructure, and data services supported by the Economic and Social Research Council. The UK Data Service and its predecessors developed data management, documentation, and publishing procedures and protocols that stand today as robust templates for data publishing. As the ESRC research data policy requires grant holders to submit their research data to the UK Data Service after a grant ends, setting standards and promoting them has been essential in raising the quality of the resulting research data being published. In the past, received data were all processed, documented, and published for reuse in-house. Recent investments have focused on guiding and training researchers in good data management practices and skills for creating shareable data, as well as a self-publishing repository system, ReShare. ReShare also receives data sets described in published data papers and achieves scientific quality assurance through peer review of submitted data sets before publication. Social science data are reused for research, to inform policy, in teaching and for methods learning. Over a 10 years period, responsive developments in system workflows, access control options, persistent identifiers, templates, and checks, together with

targeted guidance for researchers, have helped raise the standard of self-publishing social science data. Lessons learned and developments in shifting publishing social science data from an archivist responsibility to a researcher process are showcased, as inspiration for institutions setting up a data repository.

Van Deventer, Martie, and Heila Pienaar. "Research Data Management in a Developing Country: A Personal Journey." *International Journal of Digital Curation* 10, no. 2 (2015): 33-47. <https://doi.org/10.2218/ijdc.v10i2.380>

This paper explores our own journey to get to grips with research data management (RDM). It also mentions the overlap between our own 'journeys' and that of the country. We share the lessons that we learnt along the way—the most important lesson being that you can learn many wonderful and valuable RDM lessons from the international trend setters, but in the end you need to get your hands dirty and get the work done yourself. You must, within the set parameters, implement the RDM practice that is both appropriate and acceptable for and to your own set of researchers—who may be conducting research in a context that may be very dissimilar to that of international peers.

Van Horik, René, and Dirk Roorda. "Migration to Intermediate XML for Electronic Data (MIXED): Repository of Durable File Format Conversions." *International Journal of Digital Curation* 6, no. 2 (2011): 245-252. <https://doi.org/10.2218/ijdc.v6i2.200>

Van Loon, James E., Katherine G. Akers, Cole Hudson, and Alexandra Sarkozy. "Quality Evaluation of Data Management Plans at a Research University." *IFLA Journal* 43, no. 1 (2017): 98-104. <https://doi.org/10.1177/0340035216682041>

Van Tuyl, Steven, and Amanda L. Whitmire. "Water, Water, Everywhere: Defining and Assessing Data Sharing in Academia." *PLOS ONE* 11, no. 2 (2016): e0147942. <https://doi.org/10.1371/journal.pone.0147942>

Sharing of research data has begun to gain traction in many areas of the sciences in the past few years because of changing expectations from the scientific community, funding agencies, and academic journals. National Science Foundation (NSF) requirements for a data management plan (DMP) went into effect in 2011, with the intent of facilitating the dissemination and sharing of research results. Many projects that were funded during 2011 and 2012 should now have implemented the elements of the data management plans required for their grant proposals. In this paper we define 'data sharing'

and present a protocol for assessing whether data have been shared and how effective the sharing was. We then evaluate the data sharing practices of researchers funded by the NSF at Oregon State University in two ways: by attempting to discover project-level research data using the associated DMP as a starting point, and by examining data sharing associated with journal articles that acknowledge NSF support. Sharing at both the project level and the journal article level was not carried out in the majority of cases, and when sharing was accomplished, the shared data were often of questionable usability due to access, documentation, and formatting issues. We close the article by offering recommendations for how data producers, journal publishers, data repositories, and funding agencies can facilitate the process of sharing data in a meaningful way.

Van Zeeland, Hilde, and Jacquelijnn Ringersma. "The Development of a Research Data Policy at Wageningen University & Research: Best Practices as a Framework." *LIBER Quarterly* 27, no. 1 (2017): 153-170.
<https://doi.org/10.18352/lq.10215>

The current case study describes the development of a Research Data Management policy at Wageningen University & Research, the Netherlands. To develop this policy, an analysis was carried out of existing frameworks and principles on data management (such as the FAIR principles), as well as of the data management practices in the organisation. These practices were defined through interviews with research groups. Using criteria drawn from the existing frameworks and principles, certain research groups were identified as 'best-practices': cases where data management was meeting the most important data management criteria. These best-practices were then used to inform the RDM policy. This approach shows how engagement with researchers can not only provide insight into their data management practices and needs, but directly inform new policy guidelines.

Vanden-Hehir, Sally , Helena Cousijn, and Hesham Attalla. "Research Data Management Practices: Synergies and Discords between Researchers and Institutions." *International Journal of Digital Curation* 13, no. 1 (2018): 73-90.
<http://www.ijdc.net/article/view/499>

The aim of this study was to explore the synergies and discords in attitudes towards research data management (RDM) drivers and barriers for both researchers and institutions. Previous work has studied RDM from a single perspective, but not compared researchers' and institutions' perspectives. We

carried out qualitative interviews with researchers as well as institutional representatives to identify drivers and barriers, and to explore synergies and discords of both towards RDM. We mapped these to a data lifecycle model and found that the contradictions occur at early stages in the lifecycle of data and the synergies occur at the later stages. This means that for future successful RDM, the points of discord at the start of the data lifecycle must be overcome. Finally, we conclude by proposing key recommendations that could help institutions when addressing both researcher and institutional RDM needs.

Vandepitte, Leen, Samuel Bosch, Lennert Tyberghein, Filip Waumans, Bart Vanhoorne, Francisco Hernandez, Olivier De Clerck, and Jan Mees. "Fishing for Data and Sorting the Catch: Assessing the Data Quality, Completeness and Fitness for Use of Data in Marine Biogeographic Databases." *Database: The Journal of Biological Databases and Curation* (2015): bau125.
<https://doi.org/10.1093/database/bau125>

Vardakosta, Ifigenia, and Sarantos Kapidakis. "Geospatial Data Collection Policies, Technology and Open Source in Websites of Academic Libraries Worldwide." *The Journal of Academic Librarianship* 42, no. 4 (2016): 319-328.
<http://dx.doi.org/10.1016/j.acalib.2016.05.014>

Vardigan, Mary, Darrell Donakowski, Pascal Heus, Sanda Ionescu, and Julia Rotondo. "Creating Rich, Structured Metadata: Lessons Learned in the Metadata Portal Project." *IASSIST Quarterly* 38, no. 3 (2014): 15-20.
<https://doi.org/10.29173/iq123>

Vardigan, Mary, Pascal Heus, and Wendy Thomas. "Data Documentation Initiative: Toward a Standard for the Social Sciences." *International Journal of Digital Curation* 3, no. 1 (2008): 107-113.
<http://www.ijdc.net/index.php/ijdc/article/view/66/45>

Vardigan, Mary, and Cole Whiteman. "ICPSR Meets OAIS: Applying the OAIS Reference Model to the Social Science Archive Context." *Archival Science* 7, no. 1 (2007): 73-87. <https://doi.org/10.1007/s10502-006-9037-z>

Varvel, Virgil E., Jr., and Yi Shen. "Data Management Consulting at The Johns Hopkins University." *New Review of Academic Librarianship* 19, no. 3 (2013): 224-245. <https://doi.org/10.1080/13614533.2013.768277>

Vasilevsky, Nicole A., Jessica Minnier, Melissa A. Haendel, and Robin E. Champieux. "Reproducible and Reusable Research: Are Journal Data Sharing Policies Meeting the Mark?" *PeerJ* 5 (April 25, 2017): e3208.
<https://doi.org/10.7717/peerj.3208>

Background

There is wide agreement in the biomedical research community that research data sharing is a primary ingredient for ensuring that science is more transparent and reproducible. Publishers could play an important role in facilitating and enforcing data sharing; however, many journals have not yet implemented data sharing policies and the requirements vary widely across journals. This study set out to analyze the pervasiveness and quality of data sharing policies in the biomedical literature.

Methods

The online author's instructions and editorial policies for 318 biomedical journals were manually reviewed to analyze the journal's data sharing requirements and characteristics. The data sharing policies were ranked using a rubric to determine if data sharing was required, recommended, required only for omics data, or not addressed at all. The data sharing method and licensing recommendations were examined, as well any mention of reproducibility or similar concepts. The data was analyzed for patterns relating to publishing volume, Journal Impact Factor, and the publishing model (open access or subscription) of each journal.

Results

A total of 11.9% of journals analyzed explicitly stated that data sharing was required as a condition of publication. A total of 9.1% of journals required data sharing, but did not state that it would affect publication decisions. 23.3% of journals had a statement encouraging authors to share their data but did not require it. A total of 9.1% of journals mentioned data sharing indirectly, and only 14.8% addressed protein, proteomic, and/or genomic data sharing. There was no mention of data sharing in 31.8% of journals. Impact factors were significantly higher for journals with the strongest data sharing policies compared to all other data sharing criteria. Open access journals were not more likely to require data sharing than subscription journals.

Discussion

Our study confirmed earlier investigations which observed that only a minority of biomedical journals require data sharing, and a significant association between higher Impact Factors and journals with a data sharing requirement. Moreover, while 65.7% of the journals in our study that required data sharing addressed the concept of reproducibility, as with earlier investigations, we found that most data sharing policies did not provide specific guidance on the practices that ensure data is maximally available and reusable.

Verbaan, E., and A.M. Cox. "Occupational Sub-Cultures, Jurisdictional Struggle and Third Space: Theorising Professional Service Responses to Research Data Management." *The Journal of Academic Librarianship* 40, no. 3-4 (2014): 211-219. <http://dx.doi.org/10.1016/j.acalib.2014.02.008>

Verhaar, Peter, Fieke Schoots, Laurents Sesink, and Floor Frederiks. "Fostering Effective Data Management Practices at Leiden University." *LIBER Quarterly* 27, no. 1 (2017): 1–22. <http://doi.org/10.18352/lq.10185>

At Leiden University, it is increasingly recognised that effective data management forms an integral component of responsible research. To actively promote the stewardship of all the research data that are produced at Leiden University, a comprehensive, institution-wide programme was launched in 2015, which centrally aims to encourage its researchers to carefully plan the temporal storage, long-term preservation and potential reuse of their data. This programme, which is managed centrally by the Department of Academic Affairs, and which receives important contributions from academic staff, from Leiden University Libraries, and from the University's central ICT organisation, basically consists of three parts. Firstly, a basic central policy has been formulated, containing clear guidelines for activities before, during and after research projects. The central aim of this institutional policy is to ensure that all Leiden-based research projects can effectively comply with the most common requirements stipulated by funding agencies, academic publishers, the Dutch standard evaluation protocol and the European data protection directive. As a second part of the data management programme, faculties have organised workshops and meetings, concentrating on the rationale and on the technical and organisational practicalities of effective data management in order to bring about a discipline-specific protocol. Data librarians employed by Leiden University Libraries have developed educational materials and provide training for PhDs in the principles and benefits of good data management. Thirdly, to ensure that

scholars can genuinely make a reasoned selection among the many tools that are currently available, a central catalogue was developed which lists and characterises the most relevant data management services. The catalogue currently provides information about, amongst many other aspects, the organisations behind these services, the main academic disciplines which are targeted and the accepted file formats and metadata formats. The various aspects of these facilities have been classified using terminology provided by conceptual models developed by the UKDA, ANDS and the DCC. Using Leiden University's policy guidelines as criteria, the overall suitability of each service has also been evaluated. Leiden University's data management programme has a total duration of three years, and its basic objective is to offer a comprehensive form of support, in which the data management policy which is propagated centrally is complemented by various forms of assistance which ought to make it easier for scholars to adhere to this policy. The catalogue of data management services also aims to bolster the implementation of an adequate technical infrastructure, as the qualitative evaluations of the services enable policy-makers and developers to quickly establish gaps or other shortcomings within existing facilities.

Vidal-Infer, Antonio, Beatriz Tarazona, Adolfo Alonso-Arroyo, and Rafael Aleixandre-Benavent. "Public Availability of Research Data in Dentistry Journals Indexed in Journal Citation Reports." *Clinical Oral Investigations* (2017). <https://doi.org/10.1007/s00784-017-2108-0>

Vitale, Cynthia R. H., Brianna Marshall, and Amy Nurnberger. "You're in Good Company: Unifying Campus Research Data Services." *Bulletin of the Association for Information Science and Technology* 41, no. 6 (2015): 26-28. <https://doi.org/10.1002/bult.2015.1720410611>

Vlaeminck, Sven. "Data Management in Scholarly Journals and Possible Roles for Libraries—Some Insights from EDaWaX." *LIBER Quarterly* 23, no. 1 (2013): 48-79. <http://doi.org/10.18352/lq.8082>

In this paper we summarize the findings of an empirical study conducted by the EDaWaX Project. 141 economics journals were examined regarding the quality and extent of data availability policies that should support replications of published empirical results in economics. This paper suggests criteria for such policies that aim to facilitate replications. These criteria were also used for analysing the data availability policies we found in our sample and to identify best practices for data policies of scholarly journals in economics. In

addition, we also evaluated the journals' data archives and checked the percentage of articles associated with research data. To conclude, an appraisal as to how scientific libraries might support the linkage of publications to underlying research data in cooperation with researchers, editors, publishers and data centres is presented.

Vlaeminck, Sven, and Gert G. Wagner. "On the Role of Research Data Centres in the Management of Publication-Related Research Data " *LIBER Quarterly* 23, no. 4 (2014): 336-357. <http://doi.org/10.18352/lq.9356>

This paper summarizes the findings of an analysis of scientific infrastructure service providers (mainly from Germany but also from other European countries). These service providers are evaluated with regard to their potential services for the management of publication-related research data in the field of social sciences, especially economics. For this purpose we conducted both desk research and an online survey of 46 research data centres (RDCs), library networks and public archives; almost 48% responded to our survey. We find that almost three-quarters of all respondents generally store externally generated research data—which also applies to publication-related data. Almost 75% of all respondents also store and host the code of computation or the syntax of statistical analyses. If self-compiled software components are used to generate research outputs, only 40% of all respondents accept these software components for storing and hosting. Eight out of ten institutions also take specific action to ensure long-term data preservation. With regard to the documentation of stored and hosted research data, almost 70% of respondents claim to use the metadata schema of the Data Documentation Initiative (DDI); Dublin Core is used by 30 percent (multiple answers were permitted). Almost two-thirds also use persistent identifiers to facilitate citation of these datasets. Three in four also support researchers in creating metadata for their data. Application programming interfaces (APIs) for uploading or searching datasets currently are not yet implemented by any of the respondents. Least common is the use of semantic technologies like RDF.

Concluding, the paper discusses the outcome of our survey in relation to Research Data Centres (RDCs) and the roles and responsibilities of publication-related data archives for journals in the fields of social sciences.

Voß, Viola, and Hamrin, Göran. "Quadcopters or Linguistic Corpora: Establishing RDM Services for Small-Scale Data Producers at Big Universities." *LIBER Quarterly* 28, no. 1 (2018): 1-58. <http://doi.org/10.18352/lq.10255>

During an international library conference in 2017 the authors had many productive exchanges about similarities and differences in Swedish and German higher-education libraries. Since research data management (RDM) is an emerging topic on both sides of the Baltic Sea, we find it valuable to compare strategies, services, and workflows to learn from each other's practices.

Aim: In this paper, we aim to compare the practices and needs of small-scale data producers in engineering and the humanities. In particular, we try to answer the following research questions:

What kind of data do the small-scale data producers produce?

What do these producers need in terms of RDM support?

What then can we librarians help them with?

Hypothesis: Our research hypothesis is that small-scale data producers have similar needs in engineering and the humanities. This hypothesis is based on the similarities in demands from funding agencies on (open) research data and on the assumption that research in different subjects often creates results which are different in content but similar in structure.

Method: We study the current strategies, practices, and services of our respective universities (KTH Royal Institute of Technology Stockholm and Westfälische Wilhelms-Universität Münster). We also study the work and initiatives done on a more advanced level by universities, libraries, and other organisations in Sweden and Germany.

Results: The paper will give an overview of how we did the groundwork for the initial services provided by our libraries. We focus on what we are doing and why we are doing it. We find that we are following in the leading footsteps of other university libraries. The experiences shared by colleagues help us to adapt their best practices to our local demands, making them better practices for KTH and WWU researchers.

Limitation: We restrict ourselves to studying only researchers who create data on a small scale, since the large-scale data producers handle the RDM on their own.

Waddington, Simon, Jun Zhang, Gareth Knight, Mark Hedges, Jens Jensen, and Roger Downing. "Kindura: Repository Services for Researchers Based on Hybrid Clouds." *Journal of Digital Information* 13, no. 1 (2012).
<http://journals.tdl.org/jodi/index.php/jodi/article/view/5877>

Waide, Robert B., James W. Brunt, and Mark S. Servilla. "Demystifying the Landscape of Ecological Data Repositories in the United States." *BioScience* 67 no. 12 (2017): 1044-1051. <https://doi.org/10.1093/biosci/bix117>

Walling, David, and Maria Esteva. "Automating the Extraction of Metadata from Archaeological Data Using iRods Rules." *International Journal of Digital Curation* 6, no. 2 (2011): 253-264. <https://doi.org/10.2218/ijdc.v6i2.201>

The Texas Advanced Computing Center and the Institute for Classical Archaeology at the University of Texas at Austin developed a method that uses iRods rules and a Jython script to automate the extraction of metadata from digital archaeological data. The first step was to create a record-keeping system to classify the data. The record-keeping system employs file and directory hierarchy naming conventions designed specifically to maintain the relationship between the data objects and map the archaeological documentation process. The metadata implicit in the record-keeping system is automatically extracted upon ingest, combined with additional sources of metadata, and stored alongside the data in the iRods preservation environment. This method enables a more organized workflow for the researchers, helps them archive their data close to the moment of data creation, and avoids error prone manual metadata input. We describe the types of metadata extracted and provide technical details of the extraction process and storage of the data and metadata.

Wallis, Jillian. "Data Producers Courting Data Reusers: Two Cases from Modeling Communities." *International Journal of Digital Curation* 9, no. 1 (2014): 98-109. <https://doi.org/10.2218/ijdc.v9i1.304>

Data sharing is a difficult process for both the data producer and the data reuser. Both parties are faced with more disincentives than incentives. Data producers need to sink time and resources into adding metadata for data to be findable and usable, and there is no promise of receiving credit for this effort.

Making data available also leaves data producers vulnerable to being scooped or data misuse. Data reusers also need to sink time and resources into evaluating data and trying to understand them, making collecting their own data a more attractive option. In spite of these difficulties, some data producers are looking for new ways to make data sharing and reuse a more viable option. This paper presents two cases from the surface and climate modeling communities, where researchers who produce data are reaching out to other researchers who would be interested in reusing the data. These cases are evaluated as a strategy to identify ways to overcome the challenges typically experienced by both data producers and data reusers. By working together with reusers, data producers are able to mitigate the disincentives and create incentives for sharing data. By working with data producers, data reusers are able to circumvent the hurdles that make data reuse so challenging.

Wallis, Jillian C., Christine L. Borgman, Matthew S. Mayernik, and Alberto Pepe. "Moving Archival Practices Upstream: An Exploration of the Life Cycle of Ecological Sensing Data in Collaborative Field Research." *International Journal of Digital Curation* 3, no. 1 (2008): 114-126. <https://doi.org/10.2218/ijdc.v3i1.46>

The success of eScience research depends not only upon effective collaboration between scientists and technologists but also upon the active involvement of data archivists. Archivists rarely receive scientific data until findings are published, by which time important information about their origins, context, and provenance may be lost. Research reported here addresses the life cycle of data from collaborative ecological research with embedded networked sensing technologies. A better understanding of these processes will enable archivists to participate in earlier stages of the life cycle and to improve curation of these types of scientific data. Evidence from our interview study and field research yields a nine-stage life cycle. Among the findings are the cumulative effect of decisions made at each stage of the life cycle; the balance of decision-making between scientific and technology research partners; and the loss of certain types of data that may be essential to later interpretation.

Wallis, Jillian C., Elizabeth Rolando, and Christine L. Borgman. "If We Share Data, Will Anyone Use Them? Data Sharing and Reuse in the Long Tail of Science and Technology." *PLoS ONE* 8, no. 7 (2013): e67332. <https://doi.org/10.1371/journal.pone.0067332>

Research on practices to share and reuse data will inform the design of infrastructure to support data collection, management, and discovery in the long tail of science and technology. These are research domains in which data tend to be local in character, minimally structured, and minimally documented. We report on a ten-year study of the Center for Embedded Network Sensing (CENS), a National Science Foundation Science and Technology Center. We found that CENS researchers are willing to share their data, but few are asked to do so, and in only a few domain areas do their funders or journals require them to deposit data. Few repositories exist to accept data in CENS research areas. Data sharing tends to occur only through interpersonal exchanges. CENS researchers obtain data from repositories, and occasionally from registries and individuals, to provide context, calibration, or other forms of background for their studies. Neither CENS researchers nor those who request access to CENS data appear to use external data for primary research questions or for replication of studies. CENS researchers are willing to share data if they receive credit and retain first rights to publish their results. Practices of releasing, sharing, and reusing of data in CENS reaffirm the gift culture of scholarship, in which goods are bartered between trusted colleagues rather than treated as commodities.

Walton, David, Roy Lowry, and Sarah Callaghan. "Data Citation and Publication by NERC's Environmental Data Centres." *Ariadne*, no. 68 (2012). <http://www.ariadne.ac.uk/issue68/callaghan-et-al>

Wanchoo, Lalit, Nathan James, and Hampapuram K. Ramapriyan. "NASA EOSDIS Data Identifiers: Approach and System." *Data Science Journal* 16, no. 15 (2017). <http://doi.org/10.5334/dsj-2017-015>

NASA's Earth Science Data and Information System (ESDIS) Project began investigating the use of Digital Object Identifiers (DOIs) in 2010 with the goal of assigning DOIs to various data products. These Earth science research data products produced using Earth observations and models are archived and distributed by twelve Distributed Active Archive Centers (DAACs) located across the United States. Each data center serves a different Earth science discipline user community and, accordingly, has a unique approach and process for generating and archiving a variety of data products. These varied approaches present a challenge for developing a DOI solution. To address this challenge, the ESDIS Project has developed processes, guidelines, and several models for creating and assigning DOIs. Initially the DOI assignment and registration process was started as a prototype but now it is fully operational.

In February 2012, the ESDIS Project started using the California Digital Library (CDL) EZID for registering DOIs. The DOI assignments were initially labor-intensive. The system is now automated, and the assignments are progressing rapidly. As of February 28, 2017, over 50% of the data products at the DAACs had been assigned DOIs. Citations using the DOIs increased from about 100 to over 370 between 2015 and 2016.

Wang, Wei Min, Tobias Göpfert, and Rainer Stark. "Data Management in Collaborative Interdisciplinary Research Projects—Conclusions from the Digitalization of Research in Sustainable Manufacturing." *ISPRS International Journal of Geo-Information* 5, no. 4 (2016): 41.
<http://dx.doi.org/10.3390/ijgi5040041>

Wanga, Minglu, and Bonnie L. Fonga. "Embedded Data Librarianship: A Case Study of Providing Data Management Support for a Science Department." *Science & Technology Libraries* 34, no. 3 (2015): 228-240.
<https://doi.org/10.1080/0194262x.2015.1085348>

Ward, Catharine, Lesley Freiman, Sarah Jones, Laura Molloy, and Kellie Snow. "Making Sense: Talking Data Management with Researchers." *International Journal of Digital Curation* 6, no. 2 (2011): 265-273.
<https://doi.org/10.2218/ijdc.v6i2.202>

Wasner, Catharina, Ingo Barkow, Fabian Odoni. Enhancing the Research Data Management of Computer-Based Educational Assessments in Switzerland. *Data Science Journal*, 17 (2018): 18. <http://doi.org/10.5334/dsj-2018-018>

Since 2006 the education authorities in Switzerland have been obliged by the Constitution to harmonize important benchmarks in the educational system throughout Switzerland. With the development of national educational objectives in four disciplines an important basis for the implementation of this constitutional mandate was created. In 2013 the Swiss National Core Skills Assessment Program. . . was initiated to investigate the skills of students, starting with three of four domains: mathematics, language of teaching and first foreign language in grades 2, 6 and 9. ÜGK uses a computer-based test and a sample size of 25.000 students per year.

A huge challenge for computer-based educational assessment is the research data management process. Data from several different systems and tools existing in different formats has to be merged to obtain data products researchers can utilize. The long term preservation has to be adapted as well.

In this paper, we describe our current processes and data sources as well as our ideas for enhancing the data management.

Weber, Andreas, and Claudia Piesche. "Requirements on Long-Term Accessibility and Preservation of Research Results with Particular Regard to Their Provenance." *ISPRS International Journal of Geo-Information* 5, no. 4 (2016): 49. <http://dx.doi.org/10.3390/ijgi5040049>

Weber, Nicholas M., Carole L. Palmer, and Tiffany C. Chao. "Current Trends and Future Directions in Data Curation Research and Education." *Journal of Web Librarianship* 6, no. 4 (2012): 305-320. <https://doi.org/10.1080/19322909.2012.730358>

Weber, Nicholas M., Andrea K. Thomer, Matthew S. Mayernik, Bob Dattore, Zaihua Ji, and Steve Worley. "The Product and System Specificities of Measuring Curation Impact." *International Journal of Digital Curation* 8, no. 2 (2013): 223-234. <https://doi.org/10.2218/ijdc.v8i2.286>

Using three datasets archived at the National Center for Atmospheric Research (NCAR), we describe the creation of a 'data usage index' for curation-specific impact assessments. Our work is focused on quantitatively evaluating climate and weather data used in earth and space science research, but we also discuss the application of this approach to other research data contexts. We conclude with some proposed future directions for metric-based work in data curation.

Weller, Travis, and Amalia Monroe-Gulick. "Differences in the Data Practices, Challenges, and Future Needs of Graduate Students and Faculty Members." *Journal of eScience Librarianship* 4, no. 1 (2015): e1070. <https://doi.org/10.7191/jeslib.2015.1070>

———. "Understanding Methodological and Disciplinary Differences in the Data Practices of Academic Researchers." *Library Hi Tech* 32., no. 3 (2014): 467-482. <https://doi.org/10.1108/lht-02-2014-0021>

Wessels, Bridgette, Rachel L. Finn, Peter Linde, Paolo Mazzetti, Stefano Nativi, Susan Riley, Rod Smallwood, Mark J. Taylor, Victoria Tsoukala, Kush Wadhwa, and Sally Wyatt. "Issues in the Development of Open Access to Research Data." *Prometheus: Critical Studies in Innovation* 32, no. 1 (2014): 49-66. <https://doi.org/10.1080/08109028.2014.956505>

Westra, Brian. "Data Services for the Sciences: A Needs Assessment." *Ariadne*, no. 64 (2010). <http://www.ariadne.ac.uk/issue64/westra/>

Westra, Brian, Marisa Ramirez, Susan Wells Parham, and Jeanine Marie Scaramozzino. "Science and Technology Resources on the Internet: Selected Internet Resources on Digital Research Data Curation." *Issues in Science and Technology Librarianship*, no. 63 (2010). <http://www.istl.org/10-fall/internet2.html>

Wheeler, Jonathan, and Karl Benedict. "Functional Requirements Specification for Archival Asset Management: Identification and Integration of Essential Properties of Services-Oriented Architecture Products." *Journal of Map & Geography Libraries* 11, no. 2 (2015): 155-179.
<http://dx.doi.org/10.1080/15420353.2015.1035474>

White, Hollie C. "Considering Personal Organization: Metadata Practices of Scientists." *Journal of Library Metadata* 10, no. 2/3 (2010): 156-172.
<https://doi.org/10.1080/19386389.2010.506396>

———. "Descriptive Metadata for Scientific Data Repositories: A Comparison of Information Scientist and Scientist Organizing Behaviors." *Journal of Library Metadata* 14, no. 1 (2014): 24-51. <https://doi.org/10.1080/19386389.2014.891896>

Whitlock, Michael C. "Data Archiving in Ecology and Evolution: Best Practices." *Trends in Ecology & Evolution* 26, no. 2 (2011): 61-65.
<http://dx.doi.org/10.1016/j.tree.2010.11.006>

Whitmire, Amanda L. "Implementing a Graduate-Level Research Data Management Course: Approach, Outcomes, and Lessons Learned." *Journal of Librarianship and Scholarly Communication* 3, no. 2 (2015): eP1246.
<http://doi.org/10.7710/2162-3309.1246>

INTRODUCTION As data-driven research becomes the norm, practical knowledge in data stewardship is critical for researchers. Despite its growing importance, formal education in research data management (RDM) is rare at the university level. Academic librarians are now playing a leadership role in developing and providing RDM training and support to faculty and graduate students. This case study describes the development and implementation of a new, credit-bearing course in RDM for graduate students from all disciplines. **DESCRIPTION OF PROGRAM** The purpose of the course was to enable students to acquire foundational knowledge and skills in RDM that would support long-term habits in the planning, management, preservation, and

sharing of research data. The pedagogical approach for the course combined outcomes centered course design with active learning techniques. Periodic course assessment was performed through anonymous student surveys, with the objective of gauging course efficacy and quality, and to obtain suggested modifications or improvements. These assessment results indicated that the course content and scope were appropriate and that the active learning approach was effective. Assessments of student learning demonstrated that all major learning objectives were achieved. NEXT STEPS Information derived from the student surveys was used to determine how the course could be modified to improve student experience and the overall quality of the course and the instruction.

Whitmire, Amanda L., Michael Boock, and Shan C. Sutton. "Variability in Academic Research Data Management Practices: Implications for Data Services Development from a Faculty Survey." *Program* 49, no. 4 (2015): 382-407. <https://doi.org/10.1108/prog-02-2015-0017>

Whyte, Angus, Dominic Job, Stephen Giles, and Stephen Lawrie. "Meeting Curation Challenges in a Neuroimaging Group." *International Journal of Digital Curation* 3, no. 1 (2008): 171-181. <https://doi.org/10.2218/ijdc.v3i1.53>

Whyte, Angus, and Graham Pryor. "Open Science in Practice: Researcher Perspectives and Participation." *International Journal of Digital Curation* 6, no. 1 (2011): 199-213. <https://doi.org/10.2218/ijdc.v6i1.182>

We report on an exploratory study consisting of brief case studies in selected disciplines, examining what motivates researchers to work (or want to work) in an open manner with regard to their data, results and protocols, and whether advantages are delivered by working in this way. We review the policy background to open science, and literature on the benefits attributed to open data, considering how these relate to curation and to questions of who participates in science. The case studies investigate the perceived benefits to researchers, research institutions and funding bodies of utilizing open scientific methods, the disincentives and barriers, and the degree to which there is evidence to support these perceptions. Six case study groups were selected in astronomy, bioinformatics, chemistry, epidemiology, language technology and neuroimaging. The studies identify relevant examples and issues through qualitative analysis of interview transcripts. We provide a typology of degrees of open working across the research lifecycle, and conclude that better support for open working, through guidelines to assist

research groups in identifying the value and costs of working more openly, and further research to assess the risks, incentives and shifts in responsibility entailed by opening up the research process are needed.

Wicherts, Jelte M., Marjan Bakker, and Dylan Molenaar. "Willingness to Share Research Data Is Related to the Strength of the Evidence and the Quality of Reporting of Statistical Results." *PLoS ONE* 6, no. 11 (2011): e26828. <https://doi.org/10.1371/journal.pone.0026828>

Background

The widespread reluctance to share published research data is often hypothesized to be due to the authors' fear that reanalysis may expose errors in their work or may produce conclusions that contradict their own. However, these hypotheses have not previously been studied systematically.

Methods and Findings

We related the reluctance to share research data for reanalysis to 1148 statistically significant results reported in 49 papers published in two major psychology journals. We found the reluctance to share data to be associated with weaker evidence (against the null hypothesis of no effect) and a higher prevalence of apparent errors in the reporting of statistical results. The unwillingness to share data was particularly clear when reporting errors had a bearing on statistical significance.

Conclusions

Our findings on the basis of psychological papers suggest that statistical results are particularly hard to verify when reanalysis is more likely to lead to contrasting conclusions. This highlights the importance of establishing mandatory data archiving policies.

Wilcox, David, "Supporting FAIR Data Principles with Fedora." *LIBER Quarterly*, 28, no. 1 (2018): 1-8. <http://doi.org/10.18352/lq.10247>

Making data findable, accessible, interoperable, and re-usable is an important but challenging goal. From an infrastructure perspective, repository technologies play a key role in supporting FAIR data principles. Fedora is a flexible, extensible, open source repository platform for managing, preserving, and providing access to digital content. Fedora is used in a wide

variety of institutions including libraries, museums, archives, and government organizations. Fedora provides native linked data capabilities and a modular architecture based on well-documented APIs and ease of integration with existing applications. As both a project and a community, Fedora has been increasingly focused on research data management, making it well-suited to supporting FAIR data principles as a repository platform. Fedora provides strong support for persistent identifiers, both by minting HTTP URIs for each resource and by allowing any number of additional identifiers to be associated with resources as RDF properties. Fedora also supports rich metadata in any schema that can be indexed and disseminated using a variety of protocols and services. As a linked data server, Fedora allows resources to be semantically linked both within the repository and on the broader web. Along with these and other features supporting research data management, the Fedora community has been actively participating in related initiatives, most notably the Research Data Alliance. Fedora representatives participate in a number of interest and working groups focused on requirements and interoperability for research data repository platforms. This participation allows the Fedora project to both influence and be influenced by an international group of Research Data Alliance stakeholders. This paper will describe how Fedora supports FAIR data principles, both in terms of relevant features and community participation in related initiatives.

Wiley, Christie A. "An Analysis of Datasets within Illinois Digital Environment for Access to Learning and Scholarship (IDEALS), the University of Illinois Urbana-Champaign Repository." *Journal of eScience Librarianship* 4, no. 2 (2015): e1081. <https://doi.org/10.7191/jeslib.2015.1081>

———. "Assessing Research Data Deposits and Usage Statistics within IDEALS." *Journal of eScience Librarianship* 6, no. 2 (2017): e1112. <https://doi.org/10.7191/jeslib.2017.1112>

Objectives: This study follows up on previous work that began examining data deposited in an institutional repository. The work here extends the earlier study by answering the following lines of research questions: (1) What is the file composition of datasets ingested into the University of Illinois at Urbana-Champaign (UIUC) campus repository? Are datasets more likely to be single-file or multiple-file items? (2) What is the usage data associated with these datasets? Which items are most popular?

Methods: The dataset records collected in this study were identified by filtering item types categorized as "data" or "dataset" using the advanced search function in IDEALS. Returned search results were collected in an Excel spreadsheet to include data such as the Handle identifier, date ingested, file formats, composition code, and the download count from the item's statistics report. The Handle identifier represents the dataset record's persistent identifier. Composition represents codes that categorize items as single or multiple file deposits. Date available represents the date the dataset record was published in the campus repository. Download statistics were collected via a website link for each dataset record and indicates the number of times the dataset record has been downloaded. Once the data was collected, it was used to evaluate datasets deposited into IDEALS.

Results: A total of 522 datasets were identified for analysis covering the period between January 2007 and August 2016. This study revealed two influxes occurring during the period of 2008-2009 and in 2014. During the first timeframe a large number of PDFs were deposited by the Illinois Department of Agriculture. Whereas, Microsoft Excel files were deposited in 2014 by the Rare Books and Manuscript Library. Single-file datasets clearly dominate the deposits in the campus repository. The total download count for all datasets was 139,663 and the average downloads per month per file across all datasets averaged 3.2.

Conclusion: Academic librarians, repository managers, and research data services staff can use the results presented here to anticipate the nature of research data that may be deposited within institutional repositories. With increased awareness, content recruitment, and improvements, IRs can provide a viable cyberinfrastructure for researchers to deposit data, but much can be learned from the data already deposited. Awareness of trends can help librarians facilitate discussions with researchers about research data deposits as well as better tailor their services to address short-term and long-term research needs.

———. "Data Sharing and Engineering Faculty: An Analysis of Selected Publications." *Science & Technology Libraries*, 37, no. 4 (2018): 409-419. <https://doi.org/10.1080/0194262X.2018.1516596>

———. "Metadata Use in Research Data Management." *Bulletin of the American Society for Information Science and Technology* 40, no. 6 (2014): 38-40. <https://doi.org/10.1002/bult.2014.1720400612>

Wilkinson, Mark D., Michel Dumontier, Susanna-Assunta Sansone, Luiz Olavo Bonino da Silva Santos, Mario Prieto, Dominique Batista, Peter McQuilton, Tobias Kuhn, Philippe Rocca-Serra, Mercè Crosas, and Erik Schultes. "Evaluating FAIR Maturity through a Scalable, Automated, Community-Governed Framework." *Scientific Data* 6, no. 174 (2019). <https://doi.org/10.1038/s41597-019-0184-5>

Willaert, Tom, Jacob Cottyn, Ulrike Kenens, Thomas Vandendriessche, Demmy Verbeke, and Roxanne Wyns. "Research Data Management and the Evolutions of Scholarship: Policy, Infrastructure and Data Literacy at KU Leuven." *LIBER Quarterly*. 29, no. 1 (2019): 1–19. <http://doi.org/10.18352/lq.10272>

Williams, Mary, Jacqueline Bagwell, and Meredith Nahm Zozus. "Data Management Plans: The Missing Perspective." *Journal of Biomedical Informatics* 71, no. Supplement C (2017): 130-142. <https://doi.org/10.1016/j.jbi.2017.05.004>

Williams, Sarah C. "Data Practices in the Crop Sciences: A Review of Selected Faculty Publications." *Journal of Agricultural & Food Information* 13, no. 4 (2012): 308-325. <https://doi.org/10.1080/10496505.2012.717846>

———. "Data Sharing Interviews with Crop Sciences Faculty: Why They Share Data and How the Library Can Help." *Issues in Science and Technology Librarianship*, no. 72 (2013). <http://www.istl.org/13-spring/refereed2.html>

———. "Gathering Feedback from Early-Career Faculty: Speaking with and Surveying Agricultural Faculty Members about Research Data." *Journal of eScience Librarianship* 2, no. 2 (2013): e1048. <http://dx.doi.org/10.7191/jeslib.2013.1048>

———. "Using a Bibliographic Study to Identify Faculty Candidates for Data Services." *Science & Technology Libraries* 32, no. 2 (2013): 202-209. <https://doi.org/10.1080/0194262x.2013.774622>

Willis, Craig, Jane Greenberg, and Hollie White. "Analysis and Synthesis of Metadata Goals for Scientific Data." *Journal of the American Society for Information Science and Technology* 63, no. 8 (2012): 1505-152. <https://doi.org/10.1002/asi.22683>

Willmes, Christian, Daniel Kürner, and Georg Bareth. "Building Research Data Management Infrastructure using Open Source Software." *Transactions in GIS*, 18 (2014): 496-509. <http://dx.doi.org/10.1111/tgis.12060>

Willoughby, Cerys, and Frey Jeremy. "Encouraging and Facilitating Laboratory Scientists to Curate at Source." *International Journal of Digital Curation* 12, no. 2 (2017): 1-25. <https://doi.org/10.2218/ijdc.v12i2.514>

Computers and computation have become essential to scientific activity and significant amounts of data are now captured digitally or even "born digital". Consequently, there is more and more incentive to capture the full experiment records using digital tools, such as Electronic Laboratory Notebooks (ELNs), to enable the effective linking and publication of experiment design and methods with the digital data that is generated as a result. Inclusion of metadata for experiment records helps with providing access, effective curation, improving search, and providing context, and further enables effective sharing, collaboration, and reuse.

Regrettably, just providing researchers with the facility to add metadata to their experiment records does not mean that they will make use of it, or if they do, that the metadata they add will be relevant and useful. Our research has clearly indicated that researchers need support and tools to encourage them to create effective metadata. Tools, such as ELNs, provide an opportunity to encourage researchers to curate their records during their creation, but can also add extra value, by making use of the metadata that is generated to provide capabilities for research management and Open Science that extend far beyond what is possible with paper notebooks.

The Southampton Chemical Information group, has, for over fifteen years, investigated the use of the Web and other tools for the collection, curation, dissemination, reuse, and exploitation of scientific data and information. As part of this activity we have developed a number of ELNs, but a primary concern has been how best to ensure that the future development of such tools is both usable and useful to researchers and their communities, with a focus on curation at source. In this paper, we describe a number of user research and user studies to help answer questions about how our community makes use of tools and how we can better facilitate the capture and curation of experiment records and the related resources.

Wilson, Andrew. "How Much Is Enough: Metadata for Preserving Digital Data." *Journal of Library Metadata* 10, no. 2/3 (2010): 205-217. <https://doi.org/10.1080/19386389.2010.506395>

Wilson, James A. J., Michael A. Fraser, Luis Martinez-Urbe, Paul Jeffreys, Meriel Patrick, Asif Akram, and Tahir Mansoori. "Developing Infrastructure for Research Data Management at the University of Oxford." *Ariadne*, no. 65 (2010).

<http://www.ariadne.ac.uk/issue65/wilson-et-al>

Wilson, James A. J., and Paul Jeffreys. "Towards a Unified University Infrastructure: The Data Management Roll-Out at the University of Oxford." *International Journal of Digital Curation* 8, no. 2 (2013): 235-246.

<https://doi.org/10.2218/ijdc.v8i2.287>

Since presenting a paper at the International Digital Curation Conference 2010 conference entitled 'An Institutional Approach to Developing Research Data Management Infrastructure', the University of Oxford has come a long way in developing research data management (RDM) policy, tools and training to address the various phases of the research data lifecycle. Work has now begun on integrating these various elements into a unified infrastructure for the whole university, under the aegis of the Data Management Roll-out at Oxford (Damaro) Project.

This paper will explain the process and motivation behind the project, and describes our vision for the future. It will also introduce the new tools and processes created by the university to tie the individual RDM components together. Chief among these is the 'DataFinder'—a hierarchically-structured metadata cataloguing system which will enable researchers to search for and locate research datasets hosted in a variety of different datastores from institutional repositories, through Web 2 services, to filing cabinets standing in department offices. DataFinder will be able to pull and associate research metadata from research information databases and data management plans, and is intended to be CERIF compatible. DataFinder is being designed so that it can be deployed at different levels within different contexts, with higher-level instances harvesting information from lower-level instances enabling, for example, an academic department to deploy one instance of DataFinder, which can then be harvested by another at an institutional level, which can then in turn be harvested by another at a national level.

The paper will also consider the requirements of embedding tools and training within an institution and address the difficulties of ensuring the sustainability of an RDM infrastructure at a time when funding for such endeavours is limited. Our research shows that researchers (and indeed departments) are at present not exposed to the true costs of their (often suboptimal) data

management solutions, whereas when data management services are centrally provided the full costs are visible and off-putting. There is, therefore, the need to sell the benefits of centrally-provided infrastructure to researchers. Furthermore, there is a distinction between training and services that can be most effectively provided at the institutional level, and those which need to be provided at the divisional or departmental level in order to be relevant and applicable to researchers. This is being addressed in principle by Oxford's research data management policy, and in practice by the planning and piloting aspects of the Damaro Project.

Wilson, James A. J., Luis Martinez-Uribe, Michael A. Fraser, and Paul Jeffreys. "An Institutional Approach to Developing Research Data Management Infrastructure." *International Journal of Digital Curation* 6, no. 2 (2011): 274-287. <https://doi.org/10.2218/ijdc.v6i2.203>

Witt, Michael. "Co-designing, Co-developing, and Co-implementing an Institutional Data Repository Service." *Journal of Library Administration* 52, no. 2 (2012): 172-188. <https://doi.org/10.1080/01930826.2012.655607>

———. "Institutional Repositories and Research Data Curation in a Distributed Environment." *Library Trends* 57, no. 2 (2008): 191-201. <https://doi.org/10.1353/lib.0.0029>

Wittenburg, Peter. "From Persistent Identifiers to Digital Objects to Make Data Science More Efficient." *Data Intelligence* 1, no. 1 (2019): 6–21. https://doi.org/10.1162/dint_a_00004

Wolski, Malcolm, Louise Howard, and Joanna Richardson. "A Trust Framework for Online Research Data Services." *Publications* 5, no. 2 (2017): 14. <https://doi.org/10.3390/publications5020014>

There is worldwide interest in the potential of open science to increase the quality, impact, and benefits of science and research. More recently, attention has been focused on aspects such as transparency, quality, and provenance, particularly in regard to data. For industry, citizens, and other researchers to participate in the open science agenda, further work needs to be undertaken to establish trust in research environments. Based on a critical review of the literature, this paper examines the issue of trust in an open science environment, using virtual laboratories as the focus for discussion. A trust framework, which has been developed from an end-user perspective, is

proposed as a model for addressing relevant issues within online research data services and tools.

Woolfrey, H. "Innovations for the Curation and Sharing of African Social Survey Data." *Data Science Journal* 12 (2013): WDS185-WDS188.
<https://doi.org/10.2481/dsj.wds-031>

A substantial amount of data is collected through surveys conducted in Africa by national statistics offices, international donor organisations, research institutions, and the private sector. Data management at African national statistics offices is hampered by limited resources. An option for data curation in African countries is the establishment of dedicated institutions for data preservation and dissemination, such as survey data archives, and research data centres. DataFirst, at the University of Cape Town, has established an African data service and is helping to improve African data curation practices through providing data, promoting free curation tools, and undertaking data management training in African countries.

Woollard, Matthew. "Embracing the 'Data Revolution': Opportunities and Challenges for Research." *IASSIST Quarterly* 40(, no. 2 (2017): 28.
<https://doi.org/10.29173/iq784>

Wright, Andrea. "Electronic Resources for Developing Data Management Skills and Data Management Plans." *Journal of Electronic Resources in Medical Libraries* 13, no. 1 (2016): 43-48. <https://doi.org/10.1080/15424065.2016.1146640>

Wright, Sarah J., Wendy A. Kozlowski, Dianne Dietrich, Huda J. Khan, and Gail S. Steinhart. "Using Data Curation Profiles to Design the Datastar Dataset Registry." *D-Lib Magazine* 19, no. 7/8 (2013). <https://doi.org/10.1045/july2013-wright>

Wright, Stephanie, Amanda Whitmire, Lisa Zilinski, and David Minor. "Collaboration and Tension between Institutions and Units Providing Data Management Support." *Bulletin of the American Society for Information Science and Technology* 40, no. 6 (2014): 18-21.
<https://doi.org/10.1002/bult.2014.1720400607>

Wu, Mingfang, Fotis Psomopoulos, Siri Jodha Khalsa, and Anita de Waard. "Data Discovery Paradigms: User Requirements and Recommendations for Data Repositories." *Data Science Journal*, 18, no. 1 (2019): 3.
<http://doi.org/10.5334/dsj-2019-003>

As data repositories make more data openly available it becomes challenging for researchers to find what they need either from a repository or through web search engines. This study attempts to investigate data users' requirements and the role that data repositories can play in supporting data discoverability by meeting those requirements. We collected 79 data discovery use cases (or data search scenarios), from which we derived nine functional requirements for data repositories through qualitative analysis. We then applied usability heuristic evaluation and expert review methods to identify best practices that data repositories can implement to meet each functional requirement. We propose the following ten recommendations for data repository operators to consider for improving data discoverability and user's data search experience:

1. Provide a range of query interfaces to accommodate various data search behaviours.
2. Provide multiple access points to find data.
3. Make it easier for researchers to judge relevance, accessibility and reusability of a data collection from a search summary.
4. Make individual metadata records readable and analysable.
5. Enable sharing and downloading of bibliographic references.
6. Expose data usage statistics.
7. Strive for consistency with other repositories.
8. Identify and aggregate metadata records that describe the same data object.
9. Make metadata records easily indexed and searchable by major web search engines.
10. Follow API search standards and community adopted vocabularies for interoperability.

Wynholds, Laura. "Linking to Scientific Data: Identity Problems of Unruly and Poorly Bounded Digital Objects." *International Journal of Digital Curation* 6, no. 1 (2011): 214-225. <https://doi.org/10.2218/ijdc.v6i1.183>

Within information systems, a significant aspect of search and retrieval across information objects, such as datasets, journal articles, or images, relies on the identity construction of the objects. This paper uses identity to refer to the qualities or characteristics of an information object that make it definable and recognizable, and can be used to distinguish it from other objects. Identity, in this context, can be seen as the foundation from which citations, metadata and identifiers are constructed.

In recent years the idea of including datasets within the scientific record has been gaining significant momentum, with publishers, granting agencies and libraries engaging with the challenge. However, the task has been fraught with questions of best practice for establishing this infrastructure, especially in regards to how citations, metadata and identifiers should be constructed. These questions suggests a problem with how dataset identities are formed, such that an engagement with the definition of datasets as conceptual objects is warranted.

This paper explores some of the ways in which scientific data is an unruly and poorly bounded object, and goes on to propose that in order for datasets to fulfill the roles expected for them, the following identity functions are essential for scholarly publications: (i) the dataset is constructed as a semantically and logically concrete object, (ii) the identity of the dataset is embedded, inherent and/or inseparable, (iii) the identity embodies a framework of authorship, rights and limitations, and (iv) the identity translates into an actionable mechanism for retrieval or reference.

Xia, Jingfeng. "Mandates and the Contributions of Open Genomic Data." *Publications* 1, no. 3 (2013): 99-112.
<http://dx.doi.org/10.3390/publications1030099>

Yang, Seungwon, Boryung Ju, and Haeyong Chung. "Identifying Topical Coverages of Curricula using Topic Modeling and Visualization Techniques: A Case of Digital and Data Curation." *International Journal of Digital Curation* 14 no. 1 (2019): 62-87. <https://doi.org/10.2218/ijdc.v14i1.586>

Digital/data curation curricula have been around for a couple of decades. Currently, several ALA-accredited LIS programs offer digital/data curation courses and certificate programs to address the high demand for professionals with the knowledge and skills to handle digital content and research data in an ever-changing information environment. In this study, we aimed to examine

the topical scopes of digital/data curation curricula in the context of the LIS field. We collected 16 syllabi from the digital/data curation courses, as well as textual descriptions of the 11 programs and their core courses offered in the U.S., Canada, and the U.K. The collected data were analyzed using a probabilistic topic modeling technique, Latent Dirichlet Allocation, to identify both common and unique topics. The results are the identification of 20 topics both at the program- and course-levels. Comparison between the program- and course-level topics uncovered a set of unique topics, and a number of common topics. Furthermore, we provide interactive visualizations for digital/data curation programs and courses for further analysis of topical distributions. We believe that our combined approach of a topic modeling and visualizations may provide insight for identifying emerging trends and co-occurrences of topics among digital/data curation curricula in the LIS field.

Yang, Yanyan, Omer F. Rana, David W. Walker, Roy Williams, Christos Georgousopoulos, Massimo Caffaro, and Giovanni Aloisio. "An Agent Infrastructure for On-Demand Processing of Remote-Sensing Archives." *International Journal on Digital Libraries* 5, no. 2 (2005): 120-132.
<https://doi.org/10.1007/s00799-003-0054-8>

Yoon, Ayoung. "Data Reusers' Trust Development." *Journal of the Association for Information Science and Technology* 68, no. 4 (2017): 946-956.
<http://dx.doi.org/10.1002/asi.23730>

———. "End Users' Trust in Data Repositories: Definition and Influences on Trust Development." *Archival Science* 14, no. 1 (2014): 17-34.
<https://doi.org/10.1007/s10502-013-9207-8>

Yoon, Ayoung, and Youngseek Kim. "Social Scientists' Data Reuse Behaviors: Exploring the Roles of Attitudinal Beliefs, Attitudes, Norms, and Data Repositories." *Library & Information Science Research* 39, no. 3 (2017): 224-233.
<https://doi.org/10.1016/j.lisr.2017.07.008>

Yoon, Ayoung, and Teresa Schultz. "Research Data Management Services in Academic Libraries in the US: A Content Analysis of Libraries' Websites." *College & Research Libraries* 78, no. 7 (2017): 920-933.
<https://doi.org/10.5860/crl.78.7.920>

Yoon, Ayoung, and Helen Tibbo. "Examination of Data Deposit Practices in Repositories with the OAIS Model." *IASSIST Quarterly* 35, no. 4 (2011): 6-13.
<https://doi.org/10.29173/iq892>

Yoon, JungWon, EunKyung Chung, Jae Yun Lee, and Jihyun Kim. "How Research Data Is Cited in Scholarly Literature: A Case Study of HINTS." *Learned Publishing* 32 (2019): 199-206. <https://doi.org/10.1002/leap.1213>

York, Jeremy, Myron Gutmann, and Francine Berman. "What Do We Know about the Stewardship Gap." *Data Science Journal* 17, no. 19 (2018). <http://doi.org/10.5334/dsj-2018-019>

In the 21st century, digital data drive innovation and decision-making in nearly every field. However, little is known about the total size, characteristics, and sustainability of these data. In the scholarly sphere, it is widely suspected that there is a gap between the amount of valuable digital data that is produced and the amount that is effectively stewarded and made accessible. The Stewardship Gap Project (<http://bit.ly/stewardshipgap>) investigates characteristics of, and measures, the stewardship gap for sponsored scholarly activity in the United States. This paper presents a preliminary definition of the stewardship gap based on a review of relevant literature and investigates areas of the stewardship gap for which metrics have been developed and measurements made, and where work to measure the stewardship gap is yet to be done. The main findings presented are 1) there is not one stewardship gap but rather multiple "gaps" that contribute to whether data is responsibly stewarded; 2) there are relationships between the gaps that can be used to guide strategies for addressing the various stewardship gaps; and 3) there are imbalances in the types and depths of studies that have been conducted to measure the stewardship gap.

Yu, Fei, Rebecca Deuble, and Helen Morgan. "Designing Research Data Management Services Based on the Research Lifecycle—A Consultative Leadership Approach." *Journal of the Australian Library and Information Association* 66, no. 3 (2017): 287-298. <http://www.tandfonline.com/doi/abs/10.1080/24750158.2017.1364835>

Yu, Holly, H. "The Role of Academic Libraries in Research Data Service (RDS) Provision: Opportunities and Challenges." *The Electronic Library* 35, no. 4 (2017): 783-797. <https://doi.org/10.1108/EL-10-2016-0233>

Yu, Siu Hong. "Research Data Management: A Library Practitioner's Perspective." *Public Services Quarterly* 13, no. 1 (2017): 48-54. <https://doi.org/10.1080/15228959.2016.1223475>

Yu, Janice Chen Kung, and Sandy Campbell. "What Not to Keep: Not All Data Have Future Research Value." *Journal of the Canadian Health Libraries Association* 37, no. 2 (2016): 53-57. <https://doi.org/10.5596/c16-013>

Zborowski, Mary. "Data Management Activities of Canada's National Science Library—2010 Update and Prospective." *Data Science Journal* 9 (2011): 100-106. <https://doi.org/10.2481/dsj.009-026>

NRC-CISTI serves Canada as its National Science Library (as mandated by Canada's Parliament in 1924) and also provides direct support to researchers of the National Research Council of Canada (NRC). By reason of its mandate, vision, and strategic positioning, NRC-CISTI has been rapidly and effectively mobilizing Canadian stakeholders and resources to become a lead player on both the Canadian national and international scenes in matters relating to the organization and management of scientific research data. In a previous communication (CODATA International Conference, 2008), the orientation of NRC-CISTI towards this objective and its short- and medium-term plans and strategies were presented. Since then, significant milestones have been achieved. This paper presents NRC-CISTI's most recent activities in these areas, which are progressing well alongside a strategic organizational redesign process that is realigning NRC-CISTI's structure, mission, and mandate to better serve its clients. Throughout this transformational phase, activities relating to data management remain vibrant.

Zenk-Möltgen, Wolfgang, and Greta Lepthien. "Data Sharing in Sociology Journals." *Online Information Review* 38, no. 6 (2014): 709-722. <https://doi.org/10.1108/oir-05-2014-0119>

Zhang, Yun, Weina Hua, and Shunbo Yuan. "Mapping the Scientific Research on Open Data: A Bibliometric Review." *Learned Publishing* 31, no. 2 (2017): 95-106. <https://doi.org/10.1002/leap.1110>

Zielinski, Tomasz, Johnny Hay, and Andrew J Millar. "The Grant Is Dead, Long Live the Data—Migration as a Pragmatic Exit Strategy for Research Data Preservation." *Wellcome Open Research* 4 (2019): 104. <https://doi.org/10.12688/wellcomeopenres.15341.2>

Open research, data sharing and data re-use have become a priority for publicly- and charity-funded research. Efficient data management naturally requires computational resources that assist in data description, preservation and discovery. While it is possible to fund development of data management

systems, currently it is more difficult to sustain data resources beyond the original grants. That puts the safety of the data at risk and undermines the very purpose of data gathering.

PlaSMo stands for 'Plant Systems-biology Modelling' and the PlaSMo model repository was envisioned by the plant systems biology community in 2005 with the initial funding lasting until 2010. We addressed the sustainability of the PlaSMo repository and assured preservation of these data by implementing an exit strategy. For our exit strategy we migrated data to an alternative, public repository with secured funding. We describe details of our decision process and aspects of the implementation. Our experience may serve as an example for other projects in a similar situation.

We share our reflections on the sustainability of biological data management and the future outcomes of its funding. We expect it to be a useful input for funding bodies.

Zilinski, Lisa D., Amy Barton, Tao Zhan, Line Pouchard, and Pete Pascuzzi. "RDAP Review: Research Data Integration in the Purdue Libraries." *Bulletin of the Association for Information Science and Technology* 42, no. 2 (2016): 33-37. <https://doi.org/10.1002/bul2.2016.1720420212>

Zilinski, Lisa D., Abigail Gobens, and Kristin Briney. "University Data Policies and Library Data Services: Who Owns Your Data?" *Bulletin of the Association for Information Science and Technology* 41, no. 6 (2015): 32-34. <https://doi.org/10.1002/bult.2015.1720410613>

Zilinski, Lisa D., David Scherer, Darcy Bullock, Deborah Horton, Courtney Matthews. "Evolution of Data Creation, Management, Publication, and Curation in the Research Process." *Transportation Research Record: Journal of the Transportation Research Board* 2414 (2014): 9-19. <https://doi.org/10.3141/2414-02>

Zimmerman, Ann. "Not by Metadata Alone: The Use of Diverse Forms of Knowledge to Locate Data for Reuse." *International Journal on Digital Libraries* 7, no. 1/2 (2007): 5-16. <https://doi.org/10.1007/s00799-007-0015-8>

Zozus, Meredith Nahm. *The Data Book: Collection and Management of Research Data*. Boca Raton: CRC Press, 2017. <http://www.worldcat.org/oclc/959965201>

Related Works

Bailey, Charles W., Jr. *Digital Curation and Preservation Bibliography 2010*. Houston: Digital Scholarship, 2011. <http://digital-scholarship.org/dcpb/dcpb2010.htm>

———. *Digital Curation Bibliography: Preservation and Stewardship of Scholarly Works*. Houston: Digital Scholarship, 2012. <http://digital-scholarship.org/dcbw/dcb.htm>

———. *Digital Curation Bibliography: Preservation and Stewardship of Scholarly Works, 2012 Supplement* Houston: Digital Scholarship, 2013. <http://digital-scholarship.org/dcbw/s1/dcbw-s1.htm>

About the Author

Charles W. Bailey, Jr. is the publisher of [Digital Scholarship](#) and a noncommercial [digital artist](#).

Bailey has over 44 years of information technology, digital publishing, and instructional technology experience, including 24 years of managerial experience in academic libraries. From 2004 to 2007, he was the Assistant Dean for Digital Library Planning and Development at the University of Houston Libraries. From 1987 to 2003, he served as Assistant Dean/Director for Systems at the University of Houston Libraries.

Previously, he served as Head, Systems and Research Services at the Health Sciences Library, The University of North Carolina at Chapel Hill; Systems Librarian at the Milton S. Eisenhower Library, The Johns Hopkins University; User Documentation Specialist at the OCLC Online Computer Library Center; and Media Library Manager at the Learning Resources Center, SUNY College at Oswego.

Bailey has discussed his career in an [interview](#) in *Preservation, Digital Technology & Culture*. See [Bailey's vita](#) for more details.

Bailey has been an [open access publisher](#) for over 31 years. In 1989, Bailey established [PACS-L](#), a discussion list about public-access computers in libraries, and [The Public-Access Computer Systems Review](#), the first open access journal in

the field of library and information science. He served as PACS-L Moderator until November 1991 and as Editor-in-Chief of *The Public-Access Computer Systems Review* until the end of 1996.

In 1990, Bailey and Dana Rooks established *Public-Access Computer Systems News*, an electronic newsletter, and Bailey co-edited this publication until 1992.

In 1992, he founded the *PACS-P* mailing list for announcing the publication of selected e-serials, and he moderated this list until 2007.

In 1996, he established the *Scholarly Electronic Publishing Bibliography (SEPB)*, an open access book that was updated 80 times.

In 2001, he added the *Scholarly Electronic Publishing Weblog*, which announced relevant new publications, to *SEPB*.

In 2001, he was selected as a team member of *Current Cites*, and he has subsequently been a frequent contributor of reviews to this monthly e-serial.

In 2005, he published the *Open Access Bibliography: Liberating Scholarly Literature with E-prints and Open Access Journals* with the Association of Research Libraries (also a [website](#)).

In 2005, Bailey established Digital Scholarship (<http://digital-scholarship.org/>), which provides information and commentary about digital copyright, digital curation, digital repository, open access, research data management, scholarly communication, and other digital information issues. Digital Scholarship's digital publications are open access. Its publications are under Creative Commons licenses.

At that time, he also established *DigitalKoans*, a weblog that covers the same topics as Digital Scholarship.

From April 2005 through December 2019, Digital Scholarship had **over 20 million visitors** from 242 Internet country domains, over 100.7 million file requests, and over 76 million page views. Excluding **spiders**, there were **over 12.2 million visitors** from 242 Internet country domains, over 59 million file requests, and over 35.9 million page views.

From April 2005 through May 2021, Bailey published the following books and book supplements: the *Scholarly Electronic Publishing Bibliography: 2008 Annual*

Edition (2009), *Digital Scholarship 2009* (2010), *Transforming Scholarly Publishing through Open Access: A Bibliography* (2010), the *Scholarly Electronic Publishing Bibliography 2010* (2011), the *Digital Curation and Preservation Bibliography 2010* (2011), the *Institutional Repository and ETD Bibliography 2011* (2011), the *Digital Curation Bibliography: Preservation and Stewardship of Scholarly Works* (2012), the *Digital Curation Bibliography: Preservation and Stewardship of Scholarly Works, 2012 Supplement* (2013), and the *Research Data Curation and Management Bibliography* (2021).

He also published and updated the following bibliographies and weblibliographies as websites with links to freely available works: the *Scholarly Electronic Publishing Bibliography* (1996-2011), the *Electronic Theses and Dissertations Bibliography* (2005-2012), the *Google Books Bibliography* (2005-2011), the *Institutional Repository Bibliography* (2009-2011), the *Open Access Journals Bibliography* (2010), the *Digital Curation and Preservation Bibliography* (2010-2011), the *E-science and Academic Libraries Bibliography* (2011), the *Digital Curation Resource Guide* (2012), the *Research Data Curation Bibliography* (2012-2021), the *Altmetrics Bibliography* (2013), the *Transforming Peer Review Bibliography* (2014), and the *Academic Library as Scholarly Publisher Bibliography* (2018).

In 2011, he established the LinkedIn Digital Curation Group.

For more details, see the "[Digital Scholarship Publications Overview](#)" and "[A Look Back at 31 Years as an Open Access Publisher.](#)"

In 2010, Bailey was given a **Best Content by an Individual Award** by *The Charleston Advisor*. In 2003, he was named as one of *Library Journal's* "Movers & Shakers." In 1993, he was awarded the first LITA/Library Hi Tech Award For Outstanding Communication for Continuing Education in Library and Information Science. In 1992, Bailey received a Network Citizen Award from the Apple Library.

In 1973, Bailey won a **Wallace Stevens Poetry Award**. He is the author of *The Cave of Hypnos: Early Poems*, which includes several poems that won that award.

Bailey has written over 30 papers about digital copyright, expert systems, institutional repositories, open access, scholarly communication, and other topics.

He has served on the editorial boards of *Information Technology and Libraries*, *Library Software Review*, and *Reference Services Review*. He was the founding Vice-Chairperson of the [LITA Imagineering Interest Group](#).

Bailey is a [digital artist](#), and he has made [over 600 digital artworks](#) freely available on social media sites, such as [Flickr](#), under Creative Commons Attribution-NonCommercial licenses. A [list of his artworks that includes links to high resolution JPEG images](#) on Flickr is available.

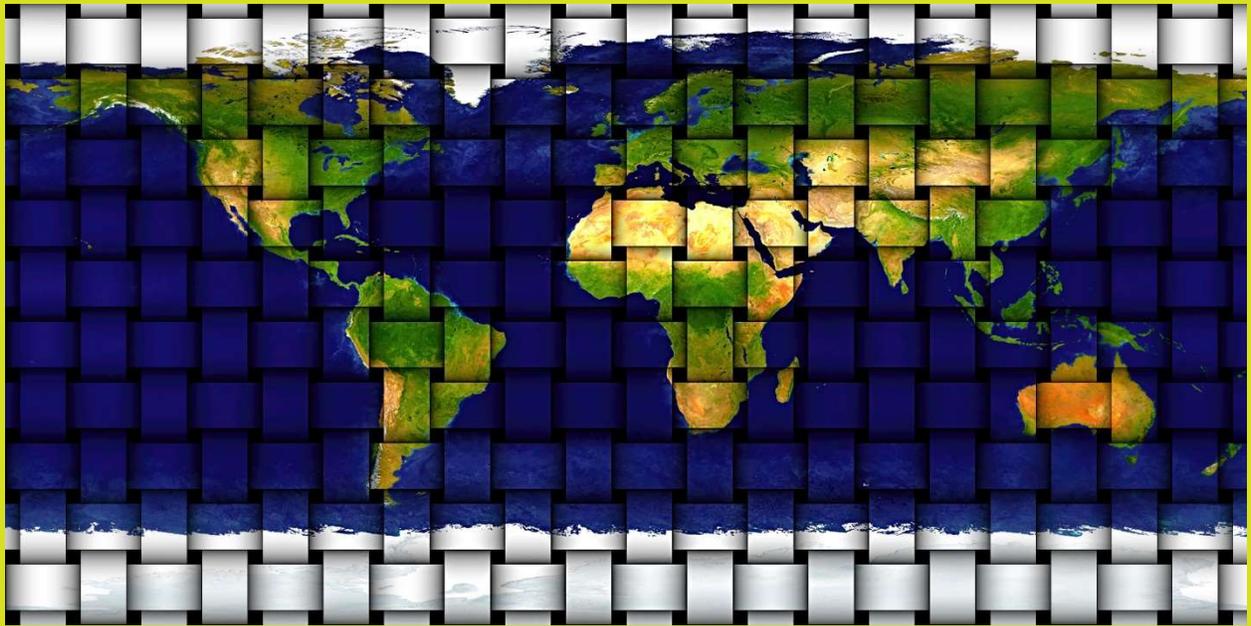
He holds master's degrees in information and library science and instructional media and technology.

You can contact him at: [publisher at digital-scholarship.org](mailto:publisher@digital-scholarship.org).

You can follow Bailey at these URLs:

- Digital Artist weblog: <http://www.charleswbaileyjr.name/> and RSS feed (<http://www.charleswbaileyjr.name/feed/>)
- *DigitalKoans* weblog: <http://digital-scholarship.org/digitalkoans/>
- Flickr: <https://www.flickr.com/photos/charleswbaileyjr/>
- Twitter (*DigitalKoans*): <https://twitter.com/DigitalKoans>
- Twitter mobile devices (*DigitalKoans*): <https://mobile.twitter.com/DigitalKoans>

Digital Scholarship



<http://www.digital-scholarship.org/>