

# Research Data Sharing and Reuse Bibliography

Charles W. Bailey, Jr.

Houston: Digital Scholarship

2021

The *Research Data Sharing and Reuse Bibliography* includes over 200 selected English-language articles and books that are useful in understanding the sharing and reuse of research data.

For an introduction to data sharing and reuse, see:

Dijkers, Marcel P. "A Beginner's Guide to Data Stewardship and Data Sharing" *Spinal Cord* 57 (2019): 169-182. <https://doi.org/10.1038/s41393-018-0232-6>

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as MP3 files), editorials, e-mail messages, interviews, letters to the editor, presentation slides or transcripts, unpublished e-prints, or weblog postings.

Most sources have been published from January 2009 through December 2021; however, a limited number of earlier key sources are also included. The bibliography has links to included works. Where possible, this bibliography uses Digital Object Identifier System (DOI) URLs. All links are subject to change.

Some publishers may use nontraditional citation elements and patterns, and they may omit standard bibliographic elements and/or substitute new ones.

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 work (see the "Note on the Inclusion of Abstracts" below for more details).

Unless otherwise noted, article abstracts in this bibliography are under a [Creative Commons Attribution 4.0 International License](#). Abstracts are reproduced as written in the source material.

For a in-depth treatment of the curation of digital research data with over 800 references, see:

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

## Dedication

In memory of [Paul Evan Peters](#) (1947-1996), founding Executive Director of the Coalition for Networked Information, whose visionary leadership at the dawn of the Internet era fostered the development of scholarly electronic publishing.



## Bibliography

Abele-Brehm, Andrea E., Mario Gollwitzer, Ulf Steinberg, and Felix D. Schönbrodt. "Attitudes toward Open Science and Public Data Sharing." *Social Psychology* 50, no. 4 (2019): 252-260. <https://doi.org/10.1027/1864-9335/A000384>.

Abella, Alberto, Marta Ortiz-de-Urbina-Criado, and Carmen De-Pablos-Heredero. "The Process of Open Data Publication and Reuse." *Journal of the Association for Information Science and Technology* 70, no. 3 (2019): 296-300. <https://doi.org/https://doi.org/10.1002/asi.24116>

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.

Alma, Bridget. "Perseids: Experimenting with Infrastructure for Creating and Sharing Research Data in the Digital Humanities." *Data Science Journal* 16, no. 19 (2017): p.19. <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.

Alter, George, and Richard Gonzalez. "Responsible Practices for Data Sharing." *American Psychologist* 73, no. 2 (2018): 146-156. <https://doi.org/10.1037/AMP0000258>.

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/>.

Andrea, Sixto Costoya, Robinson Garcia Nicolas, Leeuwen van Thed, and Costas Rodrigo. "Exploring the Relevance of ORCID as a Source of Study of Data Sharing Activities at the Individual-Level: A Methodological Discussion." *Scientometrics* 126, no. 8 (2021): 7149-7165. <https://doi.org/10.1007/s11192-021-04043-5>

Bangani, Siviwe, and Mathew Moyo. "Data Sharing Practices among Researchers at South African Universities." *Data Science Journal*, 18, no. 1 (2019): p.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.

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>

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. <https://doi.org/10.29173/iq119>

Bierer, Barbara E., Mercè Crosas, and Heather H. Pierce. "Data Authorship as an Incentive to Data Sharing." *The New England Journal of Medicine* 376, no. 17 (2017): 1684-1687. <https://doi.org/10.1056/NEJMSB1616595>

Binder, Piotr, and Piotr Filipkowski. "Data Sharing and Archiving Qualitative and QL Data in Poland." *IASSIST Quarterly* 34, no 3-4 (2011): 70. <https://doi.org/10.29173/iq467>

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, Libby, and Arja Kuula-Luomi. "Revisiting Qualitative Data Reuse: A Decade On." *SAGE Open* 7, no. 1 (2017). <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/>.

Bishop, Libby, and Bren Neale. "Sharing Qualitative and Qualitative Longitudinal Data in the UK." *IASSIST Quarterly* 34, no. 3/4 (2011): 23. <https://doi.org/10.29173/iq457>

Bond-Lamberty, B. "Data Sharing and Scientific Impact in Eddy Covariance Research." *Journal of Geophysical Research* 123, no. 4 (2018): 1440-1443. <https://doi.org/10.1002/2018JG004502>

Bonifacio, Flavio. "Differences in Data Sharing Attitudes and Behaviours." *IASSIST Quarterly* 42, no. 3 (2018): 1-40. <https://doi.org/10.29173/iq912>

Borghi, John A., and Ana E. Van Gulick. "Data Management and Sharing: Practices and Perceptions of Psychology Researchers." *PLoS ONE* 16, no. 5 (2021): e0252047. <https://doi.org/10.1371/JOURNAL.PONE.0252047>.

Research data is increasingly viewed as an important scholarly output. While a growing body of studies have investigated researcher practices and perceptions related to data sharing, information about data-related practices throughout the research process (including data collection and analysis) remains largely anecdotal. Building on our previous study of data practices in neuroimaging research, we conducted a survey of data management practices in the field of psychology. Our survey included questions about the type(s) of data collected, the tools used for data analysis, practices related to data organization, maintaining documentation, backup procedures, and long-term archiving of research materials. Our results demonstrate the complexity of managing and sharing data in psychology. Data is collected in multifarious forms from human participants, analyzed using a range of software tools, and archived in formats that may become obsolete. As individuals, our participants demonstrated relatively good data management practices, however they also indicated that there was little standardization within their research group. Participants generally indicated that they were willing to change their current practices in light of new technologies, opportunities, or requirements.

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>

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>

Boté, Juan-José, and Miquel Termens. "Reusing Data Technical and Ethical Challenges." *DESIDOC Journal of Library & Information Technology* 39, no. 6 (2019): 329-337. <https://doi.org/10.14429/DJLIT.39.06.14807>

Boué, Stéphanie, Michael Byrne, A. Wallace Hayes, Julia Hoeng, and Manuel C. Peitsch. "Embracing Transparency through Data Sharing." *International Journal of Toxicology* 37, no. 6 (2018): 466-471. <https://doi.org/10.1177/1091581818803880>

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>

Bull, Susan, Phaik Yeong Cheah, Spencer Denny, Irene Jao, Vicki Marsh, Laura Merson, Neena Shah More, Le Nguyen Thanh Nhan, David Osrin, Decha Tangseefa, Douglas Wassenaar, and Michael Parker. "Best Practices for Ethical Sharing of Individual-Level Health Research Data From Low- and Middle-Income Settings." *Journal of Empirical Research on Human Research Ethics* 10, no. 3 (2015): 302-313. <https://doi.org/10.1177%2F1556264615594606>

Sharing individual-level data from clinical and public health research is increasingly being seen as a core requirement for effective and efficient biomedical research. This article discusses the results of a systematic review and multisite qualitative study of key stakeholders' perspectives on best practices in ethical data sharing in low- and middle-income settings. Our research suggests that for data sharing to be effective and sustainable, multiple social and ethical requirements need to be met. An effective model of data sharing will be one in which considered judgments will need to be made about how best to achieve scientific progress, minimize risks of harm, promote fairness and reciprocity, and build and sustain trust.

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

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 heterogenous environment.

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, 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.

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>

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.

Copeland, Andrea J., Ayoung Yoon, and Sheng Zhang. "Data Reuse Practices and Expectations for Data Resources and Services among Public Library Users." *Public Library Quarterly* 40, no. 4 (2021): 330-345. <https://doi.org/10.1080/01616846.2020.1773749>

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): 636-74. <https://doi.org/10.1353/LIB.2013.0005>

Corti, Louise. "Qualitative Archiving and Data Sharing: Extending the Reach and Impact of Qualitative Data." *IASSIST Quarterly* 29, no. 3 (2006): 8. <https://doi.org/10.29173/iq105>

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>

Couture, Jessica L., Rachael E. Blake, Gavin McDonald, and Colette L. Ward. "A Funder-Imposed Data Publication Requirement Seldom Inspired Data Sharing." *PLOS ONE* 13, no.7 (2018): e0199789. <https://doi.org/10.1371/journal.pone.0199789>

Growth of the open science movement has drawn significant attention to data sharing and availability across the scientific community. In this study, we tested the ability to recover data collected under a particular funder-imposed requirement of public availability. We assessed overall data recovery success, tested whether characteristics of the data or data creator were indicators of recovery success, and identified hurdles to data recovery. Overall the majority of data were not recovered (26% recovery of 315 data projects), a similar result to journal-driven efforts to recover data. Field of research was the most important indicator of recovery success, but neither home agency sector nor age of data were determinants of recovery. While we did not find a relationship between recovery of data and age of data, age did predict whether we could find contact information for the grantee. The main hurdles to data recovery included those associated with communication with the researcher; loss of contact with the data creator accounted for half (50%) of unrecoverable datasets, and unavailability of contact information accounted for 35% of unrecoverable datasets. Overall, our results suggest that funding agencies and journals face similar challenges to enforcement of data requirements. We advocate that funding agencies could improve the availability of the data they fund by dedicating more resources to enforcing compliance with data requirements, providing data-sharing tools and technical support to awardees, and administering stricter consequences for those who ignore data sharing preconditions.

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>

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>

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>

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.

Custers, Bart, Helena Uršič. "Big Data and Data Reuse: A Taxonomy of Data Reuse for Balancing Big Data Benefits and Personal Data Protection." *International Data Privacy Law* 6, no. 1 (2016): 4-15. <https://doi.org/10.1093/idpl/ipv028>

Custers, Bart, Helena U. Vrabec, and Michael Friedewald. "Assessing the Legal and Ethical Impact of Data Reuse." *European Data Protection Law Review* 5, no. 3 (2019): 317-337. <https://doi.org/10.21552/EDPL/2019/3/7>

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): 16-43. <https://doi.org/10.1080/13614576.2014.883936>

Damerow, Joan E., Charuleka Varadharajan, Kristin Boye, Eoin L. Brodie, Madison Burrus, K. Dana Chadwick, Robert Crystal-Ornelas, Hesham Elbashandy, Ricardo J. Eloy Alves, Kim S. Ely, Amy E. Goldman, Ted Haberman, Valerie Hendrix, Zarine Kakalia, Kenneth M. Kemner, Annie B. Kersting, Nancy Merino, Fianna O'Brien, Zach Perzan, Emily Robles, Patrick Sorensen, James C. Stegen, Ramona L. Walls, Pamela Weisenhorn, Mavrik Zavarin, and Deborah Agarwal. "Sample Identifiers and Metadata to Support Data Management and Reuse in Multidisciplinary Ecosystem Sciences." *Data Science Journal* 20, no. 1 (2021): p.11. <https://doi.org/10.5334/DSJ-2021-011>.

Physical samples are foundational entities for research across biological, Earth, and environmental sciences. Data generated from sample-based analyses are not only the basis of individual studies, but can also be integrated with other data to answer new and broader-scale questions. Ecosystem studies increasingly rely on multidisciplinary team-science to study climate and environmental changes. While there are widely adopted conventions within certain domains to describe sample data, these have gaps when applied in a multidisciplinary context. In this study, we reviewed existing practices for identifying, characterizing, and linking related environmental samples. We then tested practicalities of assigning persistent identifiers to samples, with standardized metadata, in a pilot field test involving eight United States Department of Energy projects. Participants collected a variety of sample types, with analyses



conducted across multiple facilities. We address terminology gaps for multidisciplinary research and make recommendations for assigning identifiers and metadata that supports sample tracking, integration, and reuse. Our goal is to provide a practical approach to sample management, geared towards ecosystem scientists who contribute and reuse sample data.

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>

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. <https://datascience.codata.org/articles/abstract/44/>

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/>.

Devriendt, Thijs, Mahsa Shabani, and Pascal Borry. "Data Sharing in Biomedical Sciences: A Systematic Review of Incentives." *Biopreservation and Biobanking* 19, no. 3 (2021): 219-227. <https://doi.org/10.1089/bio.2020.0037>.

Dijkers, Marcel P. "A Beginner's Guide to Data Stewardship and Data Sharing" *Spinal Cord* 57 (2019): 169-182. <https://doi.org/10.1038/s41393-018-0232-6>

Donaldson, Devan Ray, Shawn Martin, and Thomas Proffen. "Understanding Perspectives on Sharing Neutron Data at Oak Ridge National Laboratory." *Data Science Journal* 16 (2017): p.35. <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.

Doorn, Peter, Ingrid Dillo, and René van Horik. "Lies, Damned Lies and Research Data: Can Data Sharing Prevent Data Fraud?" *International Journal of Digital Curation* 8, no. 1 (2013): 229-243. <https://doi.org/10.2218/IJDC.V8I1.256>.

After a spectacular case of data fraud in the field of social psychology surfaced in The Netherlands in September 2011, the Dutch research community was confronted with a number of questions. Is this an isolated case or is scientific fraud with data more common? Is the scientific method robust enough to uncover the results of misconduct and to withstand the breach of trust that fraud causes? How responsible and reliable are researchers when they collect, process, analyse and report on data? How can we prevent data fraud? Do we need to adapt the codes of conduct for researchers or do we need stricter rules for data management and data sharing?

This paper discusses the conclusions and recommendations of two reports that were published recently in consequence of this data fraud. The reports are relevant for scientific integrity and trustworthy treatment of research data. Next, this paper reports on the outcomes of enquiries in data cultures in a number of scientific disciplines. The concluding section of this paper contains a number of examples that show that the approach towards data sharing is improving gradually. The data fraud case can be regarded as a wake-up call.

Dorta-González, Pablo, Sara M. González-Betancor, and María Isabel Dorta-González. "To What Extent Is Researchers' Data-Sharing Motivated by Formal Mechanisms of Recognition and Credit?" *Scientometrics* 126, no. 3 (2021): 2209-2225. <https://doi.org/10.1007/s11192-021-03869-3>

Dosch, Brianne, and Tyler Martindale. "Reading the Fine Print: A Review and Analysis of Business Journals' Data Sharing Policies." *Journal of Business & Finance Librarianship* 25, no. 3-4 (2020): 261-280. <https://doi.org/10.1080/08963568.2020.1847549>

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>

Downey, Moira, Sophia Lafferty-Hess, Patrick Charbonneau, and Angela Zoss. "Engaging Researchers in Data Dialogues: Designing Collaborative Programming to Promote Research Data Sharing." *Journal of eScience Librarianship* 10, no. 2 (2021): e1193. <https://doi.org/10.7191/jeslib.2021.1193>

A range of regulatory pressures emanating from funding agencies and scholarly journals increasingly encourage researchers to engage in formal data sharing practices. As academic libraries continue to refine their role in supporting researchers in this data sharing space, one particular challenge has been finding new ways to meaningfully engage with campus researchers. Libraries help shape norms and encourage data sharing through education and training, and there has been significant growth in the services these institutions are able to provide and the ways in which

library staff are able to collaborate and communicate with researchers. Evidence also suggests that within disciplines, normative pressures and expectations around professional conduct have a significant impact on data sharing behaviors (Kim and Adler 2015; Sigit Sayogo and Pardo 2013; Zenk-Moltgen et al. 2018). Duke University Libraries' Research Data Management program has recently centered part of its outreach strategy on leveraging peer networks and social modeling to encourage and normalize robust data sharing practices among campus researchers. The program has hosted two panel discussions on issues related to data management—specifically, data sharing and research reproducibility. This paper reflects on some lessons learned from these outreach efforts and outlines next steps.

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

Duke, Clifford S., and John H. Porter. "The Ethics of Data Sharing and Reuse in Biology." *Bioscience* 63, no. 6 (2013): 483-489. <https://doi.org/10.1525/BIO.2013.63.6.10>

Elsayed, Amany M., and Emad I. Saleh. "Research Data Management and Sharing among Researchers in Arab Universities: An Exploratory Study." *IFLA Journal* 44, no. 4 (2018): 281-299. <https://doi.org/10.1177/0340035218785196>

Emam, Khaled El, Sam Rodgers, and Bradley Malin. "Anonymising and Sharing Individual Patient Data." *BMJ* 350, no. 6 (2015): 337-343. <https://doi.org/10.1136/BMJ.H1139>

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>

Faniel, Ixchel M., Adam Kriesberg, and Elizabeth Yakel. "Social Scientists' Satisfaction with Data Reuse." *Journal of the Association for Information Science and Technology* 67 (2016): 1404-1416. <https://doi.org/10.1002/ASI.23480>.

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>

There is almost universal agreement that scientific data should be shared for use beyond the purposes for which they were initially collected. Access to data enables system-level science, expands the instruments and products of research to new communities, and advances solutions to complex human problems. While demands for data are not new, the vision of open access to data is increasingly ambitious. The aim is to make data accessible and usable to anyone, anytime, anywhere, and for any purpose. Until recently, scholarly investigations related to data sharing and reuse were sparse. They have become more common as technology and instrumentation have advanced, policies that mandate sharing have been implemented, and research has become more interdisciplinary. Each of these factors has contributed to what is commonly referred to as the "data deluge". Most discussions about increases in the scale of sharing and reuse have focused on growing amounts of data. There are other issues related to open access to data that also concern scale which have not been as widely discussed: broader participation in data sharing and reuse, increases in the number and types of intermediaries, and more digital data products. The purpose of this paper is to develop a research agenda for scientific data sharing and reuse that considers these three areas.

Farrell, Shannon L., Lois G. Hendrickson, Kristen L. Mastel, and Julia A. Kelly. "Historical Scientific Analog Data: Life Sciences Faculty's Perspectives on Management, Reuse and Preservation." *Data Science Journal* 19, no. 1 (2020): p.51. <https://doi.org/10.5334/dsj-2020-051>

Older data in paper or analog format (e.g., field/lab notebooks, photos, maps) held in labs, offices, and archives across research institutions are an often overlooked resource for potential reuse in new scientific studies. However, there are few mechanisms to help researchers find existing analog data in order to reuse it. Yet, in the literature, reuse of historical data is particularly important in studies of biodiversity and climate change.

We surveyed life science researchers at the University of Minnesota to understand and explore current and potential future use of historical data, attitudes around sharing and reusing data, and preservation of the data. Large amounts of historical data existed on our campus. Most researchers had reused or shared it, and many continued to add to their data sets. Some data had been scanned, over half of researchers have re-keyed some of their data into machine-readable format, and nearly all that were converted to a digital format were stored on unstable platforms and legacy formats. Researchers also expressed concerns about long-term preservation plans, or who to contact for assistance in planning for the future of the data, since much of these data are at risk for loss. Currently produced digital data sets are subject to guidelines and requirements developed at a national level. Solutions for historical analog data could benefit from a similar high-level treatment, and it will take experts from various fields to lead this effort. Given libraries' expertise in data management and preservation, librarians are in a position to collaborate on devising cross-disciplinary solutions.

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>

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.

Federer, Lisa M., Christopher W. Belter, Douglas J. Joubert, Alicia Livinski, Ya-Ling Lu, Lissa N. Snyders, and Holly Thompson. "Data Sharing in PLoS ONE: An Analysis of Data Availability Statements." *PLoS ONE* 13, no. 5 (2018): e0194768. <https://doi.org/10.1371/JOURNAL.PONE.0194768>

A number of publishers and funders, including PLOS, have recently adopted policies requiring researchers to share the data underlying their results and publications. Such policies help increase the reproducibility of the published literature, as well as make a larger body of data available for reuse and re-analysis. In this study, we evaluate the extent to which authors have complied with this policy by analyzing Data Availability Statements from 47,593 papers published in PLoS ONE between March 2014 (when the policy went into effect) and May 2016. Our analysis shows that compliance with the policy has increased, with a significant decline over time in papers that did not include a Data Availability Statement. However, only about 20% of statements indicate that data are deposited in a repository, which the PLOS policy states is the preferred method. More commonly, authors state that their data are in the paper itself or in the supplemental information, though it is unclear whether these data meet the level of sharing required in the PLOS policy. These findings suggest that additional review of Data Availability Statements or more stringent policies may be needed to increase data sharing.

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

Federer, Lisa M., Ya-Ling Lu, Douglas J. Joubert, Judith Welsh, and Barbara Brandys. "Biomedical Data Sharing and Reuse: Attitudes and Practices of Clinical and Scientific

## Background

Significant efforts are underway within the biomedical research community to encourage sharing and reuse of research data in order to enhance research reproducibility and enable scientific discovery. While some technological challenges do exist, many of the barriers to sharing and reuse are social in nature, arising from researchers' concerns about and attitudes toward sharing their data. In addition, clinical and basic science researchers face their own unique sets of challenges to sharing data within their communities. This study investigates these differences in experiences with and perceptions about sharing data, as well as barriers to sharing among clinical and basic science researchers.

## Methods

Clinical and basic science researchers in the Intramural Research Program at the National Institutes of Health were surveyed about their attitudes toward and experiences with sharing and reusing research data. Of 190 respondents to the survey, the 135 respondents who identified themselves as clinical or basic science researchers were included in this analysis. Odds ratio and Fisher's exact tests were the primary methods to examine potential relationships between variables. Worst-case scenario sensitivity tests were conducted when necessary.

## Results and Discussion

While most respondents considered data sharing and reuse important to their work, they generally rated their expertise as low. Sharing data directly with other researchers was common, but most respondents did not have experience with uploading data to a repository. A number of significant differences exist between the attitudes and practices of clinical and basic science researchers, including their motivations for sharing, their reasons for not sharing, and the amount of work required to prepare their data.

Ferguson, Adam R., Jessica L. Nielson, Melissa H. Cragin, Anita E. Bandrowski, and Maryann E. Martone. "Big Data from Small Data: Data-Sharing in the 'Long Tail' of Neuroscience." *Nature Neuroscience* 17, no. 11 (2014): 1442-1447.  
<https://doi.org/10.1038/NN.3838>

Figueiredo, Ana Sofia. "Data Sharing: Convert Challenges into Opportunities." *Frontiers in Public Health* 5, no. 327 (2017): 327. <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.

Frank, Rebecca D., Kara Suzuka, Eric Johnson, and Elizabeth Yakel. "Tool Selection among Qualitative Data Reusers." *International Journal of Digital Curation* 15, no. 1 (2020).  
<https://doi.org/10.2218/ijdc.v15i1.710>

This paper explores the tension between the tools that data reusers in the field of education prefer to use when working with qualitative video data and the tools that repositories make available to data reusers. Findings from this mixed-methods study show that data reusers utilizing qualitative video data did not use repository-based tools. Rather, they valued common, widely available tools that were collaborative and easy to use.

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://datascience.codata.org/articles/abstract/10.5334/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.

Garrison, Nanibaa' A., Nila A. Sathe, Armand H. Matheny Antommara, Ingrid A. Holm, Saskia C. Sanderson, Maureen E. Smith, Melissa L. McPheeters, and Ellen W. Clayton. "A Systematic Literature Review of Individuals' Perspectives on Broad Consent and Data Sharing in the United States." *Genetics in Medicine* 18, no. 7 (2016): 663-671. <https://doi.org/10.1038/GIM.2015.138>

Gil, Yolanda, Cédric H. David, Ibrahim Demir, Bakinam T. Essawy, Robinson W. Fulweiler, Jonathan L. Goodall, Leif Karlstrom, Huikyo Lee, Heath J. Mills, Ji-Hyun Oh, Suzanne A. Pierce, Allen Pope, Mimi W. Tzeng, Sandra R. Villamizar, and Xuan Yu. "Toward the Geoscience Paper of the Future: Best Practices for Documenting and Sharing Research from Data to Software to Provenance." *Earth and Space Science* 3, no. 10 (2016): 388-415. <https://doi.org/10.1002/2015EA000136>

Gorman, Dennis M. "Availability of Research Data in High-Impact Addiction Journals with Data Sharing Policies." *Science and Engineering Ethics* 26, no. 3 (2020): 1625-1632. <https://doi.org/10.1007/s11948-020-00203-7>

Grabus, Sam, and Jane Greenberg. "The Landscape of Rights and Licensing Initiatives for Data Sharing." *Data Science Journal*, 18, no. 1 (2019): p.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 undertaken, and comments on the potential for using license-specific metadata supplements and developing data-centric rights and licensing ontologies.

Groth, Paul, Helena Cousijn, Tim Clark, and Carole A. Goble. "FAIR Data Reuse—The Path through Data Citation." *Data Intelligence* 2, no. 1-2 (2020): 78-86. [https://doi.org/10.1162/dint\\_a\\_00030](https://doi.org/10.1162/dint_a_00030).

One of the key goals of the FAIR guiding principles is defined by its final principle—to optimize data sets for reuse by both humans and machines. To do so, data providers need to implement and support consistent machine readable metadata to describe their data sets. This can seem like a daunting task for data providers, whether it is determining what level of detail should be provided in the provenance metadata or figuring out what common shared vocabularies should be used. Additionally, for existing data sets it is often unclear what steps should be taken to enable maximal, appropriate reuse. Data citation already plays an important role in making data findable and accessible, providing persistent and unique identifiers plus metadata on over 16 million data sets. In this paper, we discuss how data citation and its underlying infrastructures, in particular associated metadata, provide an important pathway for enabling FAIR data reuse.

We explored the characteristics and motivations of people who, having obtained their genetic or genomic data from Direct-To-Consumer genetic testing (DTC-GT) companies, voluntarily decide to share them on the publicly accessible web platform openSNP. The study is the first attempt to describe open data sharing activities undertaken by individuals without institutional oversight. In the paper we provide a detailed overview of the distribution of the demographic characteristics and motivations of people engaged in genetic or genomic open data sharing. The geographical distribution of the respondents showed the USA as dominant. There was no significant gender divide, the age distribution was broad, educational background varied and respondents with and without children were equally represented. Health, even though prominent, was not the respondents' primary or only motivation to be tested. As to their motivations to openly share their data, 86.05% indicated wanting to learn about themselves as relevant, followed by contributing to the advancement of medical research (80.30%), improving the predictability of genetic testing (76.02%) and considering it fun to explore genotype and phenotype data (75.51%). Whereas most respondents were well aware of the privacy risks of their involvement in open genetic data sharing and considered the possibility of direct, personal repercussions troubling, they estimated the risk of this happening to be negligible. Our findings highlight the diversity of DTC-GT consumers who decide to openly share their data. Instead of focusing exclusively on health-related aspects of genetic testing and data sharing, our study emphasizes the importance of taking into account benefits and risks that stretch beyond the health spectrum. Our results thus lend further support to the call for a broader and multi-faceted conceptualization of genomic utility.

Hamish A. Campbell, Mariana A. Micheli-Campbell, and Vinay Udyawer. "Early Career Researchers Embrace Data Sharing." *Trends in Ecology & Evolution* 34, no. 2 (2019): 95-98. <https://doi.org/10.1016/j.tree.2018.11.010>

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>

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>

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.

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>

Hood, Amelia S. C., and William J. Sutherland. "The Data-Index: An Author-Level Metric That Values Impactful Data and Incentivizes Data Sharing." *Ecology and Evolution* (2021): 14344-14350. <https://doi.org/10.1002/ECE3.8126>

Author-level metrics are a widely used measure of scientific success. The h-index and its variants measure publication output (number of publications) and research impact (number of citations). They are often used to influence decisions, such as allocating funding or jobs. Here, we argue that the emphasis on publication output and impact hinders scientific progress in the fields of ecology and evolution because it disincentivizes two fundamental practices: generating impactful (and therefore often long-term) datasets and sharing data. We describe a new author-level metric, the data-index, which values both dataset output (number of datasets) and impact (number of data-index citations), so promotes generating and sharing data as a result. We discuss how it could be implemented and provide user guidelines. The data-index is designed to complement other metrics of scientific success, as scientific contributions are diverse and our value system should reflect that both for the benefit of scientific progress and to create a value system that is more equitable, diverse, and inclusive. Future work should focus on promoting other scientific contributions, such as communicating science, informing policy, mentoring other scientists, and providing open-access code and tools.

Houtkoop, Bobby Lee, Chris Chambers, Malcolm Macleod, Dorothy V. M. Bishop, Thomas E. Nichols, and Eric-Jan Wagenmakers. "Data Sharing in Psychology: A Survey on Barriers and Preconditions." *Advances in Methods and Practices in Psychological Science* 1, no. 1 (2018): 70-85. <https://doi.org/10.1177%2F2515245917751886>

Despite its potential to accelerate academic progress in psychological science, public data sharing remains relatively uncommon. In order to discover the perceived barriers to public data sharing and possible means for lowering them, we conducted a survey, which elicited responses from 600 authors of articles in psychology. The results confirmed that data are shared only infrequently. Perceived barriers included respondents' belief that sharing is not a common practice in their fields, their preference to share data only upon request, their perception that sharing requires extra work, and their lack of training in sharing data. Our survey suggests that strong encouragement from institutions, journals, and funders will be particularly effective in overcoming these barriers, in combination with educational materials that demonstrate where and how data can be shared effectively.

Hrynaszkiewicz, Iain, James Harney, and Lauren Cadwallader. "A Survey of Researchers' Needs and Priorities for Data Sharing." *Data Science Journal* 20, no. 1 (2021): p.31. <https://doi.org/10.5334/dsj-2021-031>

One of the ways in which the publisher PLOS supports open science is via a stringent data availability policy established in 2014. Despite this policy, and more data sharing policies being introduced by other organizations, best practices for data sharing are adopted by a minority of researchers in their publications. Problems with effective research data sharing persist and these problems have been quantified by previous research as a lack of time, resources, incentives, and/or skills to share data.

In this study we built on this research by investigating the importance of tasks associated with data sharing, and researchers' satisfaction with their ability to complete these tasks. By investigating these factors we aimed to better understand opportunities for new or improved solutions for sharing data.

In May-June 2020 we surveyed researchers from Europe and North America to rate tasks associated with data sharing on (i) their importance and (ii) their satisfaction with their ability to complete them. We received 617 completed responses. We calculated mean importance and satisfaction scores to highlight potential opportunities for new solutions to and compare different cohorts.



Tasks relating to research impact, funder compliance, and credit had the highest importance scores. 52% of respondents reuse research data but the average satisfaction score for obtaining data for reuse was relatively low. Tasks associated with sharing data were rated somewhat important and respondents were reasonably well satisfied in their ability to accomplish them. Notably, this included tasks associated with best data sharing practice, such as use of data repositories. However, the most common method for sharing data was in fact via supplemental files with articles, which is not considered to be best practice.

We presume that researchers are unlikely to seek new solutions to a problem or task that they are satisfied in their ability to accomplish, even if many do not attempt this task. This implies there are few opportunities for new solutions or tools to meet these researcher needs. Publishers can likely meet these needs for data sharing by working to seamlessly integrate existing solutions that reduce the effort or behaviour change involved in some tasks, and focusing on advocacy and education around the benefits of sharing data.

There may however be opportunities—unmet researcher needs—in relation to better supporting data reuse, which could be met in part by strengthening data sharing policies of journals and publishers, and improving the discoverability of data associated with published articles.

Huang, Xiaolei, Bradford A. Hawkins, Fumin Lei, Gary L. Miller, Colin Favret, Ruiling Zhang, and Gexia Qiao. "Willing or Unwilling to Share Primary Biodiversity Data: Results and Implications of an International Survey." *Conservation Letters* 5, no. 5 (2012): 399-406. <https://doi.org/10.1111/J.1755-263X.2012.00259.X>

Hübner, Andreas. "Earth Science and Biodiversity Journals Can Improve Support for Data Sharing." *Data Science Journal* 19, no. 1 (2020): p.37. <https://doi.org/10.5334/dsj-2020-037>

This study reviews research data policies and author instructions of 31 journals from the Earth sciences and from Biodiversity that are published by German learned societies or research institutions. 12 journals don't address data sharing at all. The statements on data sharing of the journal's data policies/author guidelines were matched to 14 defined features of journal research data policies. A brief discussion on quality of data policies is presented to raise awareness of German learned societies/research institutions and to guide them towards improved data policies of their journals.

Hulsen, Tim. "Sharing Is Caring—Data Sharing Initiatives in Healthcare." *International Journal of Environmental Research and Public Health* 17, no. 9 (2020): 3046. <https://doi.org/10.3390/ijerph17093046>

Hutchings, Elizabeth, Max Loomes, Phyllis Butow, and Frances M. Boyle. "A Systematic Literature Review of Attitudes towards Secondary Use and Sharing of Health Administrative and Clinical Trial Data: A Focus on Consent." *Systematic Reviews* 10, no. 1 (2021): 1-44. <https://doi.org/10.1186/s13643-021-01663-z>

## Background

We aimed to synthesise data on issues related to stakeholder perceptions of consent for the use of secondary data. To better understand the current literature available, we conducted a systematic literature review of healthcare consumer attitudes towards the secondary use and sharing of health administrative and clinical trial data.

## Methods

EMBASE/MEDLINE, Cochrane Library, PubMed, CINAHL, Informit Health Collection, PROSPERO Database of Systematic Reviews, PsycINFO and ProQuest databases were searched. Eligible articles included those reporting qualitative or quantitative original research and published in English. No restrictions were placed on publication dates, study design or disease setting. One author screened articles for eligibility and two authors were involved in the full-text review process. Conflicts were resolved by consensus. Quality and bias were assessed using the QualSyst criteria for qualitative studies.

## Results

This paper focuses on a subset of 47 articles identified from the wider search and focuses on the issue of consent. Issues related to privacy, trust and transparency, and attitudes of healthcare professionals and researchers to secondary use and sharing of data have been dealt with in previous publications. Studies included a total of 216,149 respondents. Results indicate that respondents are generally supportive of using health data for research, particularly if the data is de-identified or anonymised. The requirement by participants to obtain consent prior to the use of health data for research was not universal, nor is the requirement for this always supported by legislation. Many respondents believed that either no consent or being informed of the research, but not providing additional consent, were sufficient.

## Conclusions

These results indicate that individuals should be provided with information and choice about how their health data is used and, where feasible, a mechanism to opt-out should be provided. To increase the acceptability of using health data for research, health organisations and data custodians must provide individuals with concise information about data protection mechanisms and under what circumstances their data may be used and by whom.

Imker, Heidi J., Hoa Luong, William H. Mischo, Mary C. Schlembach, and Chris Wiley. "An Examination of Data Reuse Practices within Highly Cited Articles of Faculty at a Research University." *The Journal of Academic Librarianship* 47, no. 4 (2021): 102369.  
<https://doi.org/10.1016/J.ACALIB.2021.102369>

Data sharing and reuse are regarded as important components of the research workflow and key elements in open science. While reuse is well-documented in some circumstances, the utility of data sharing for all domains is less clear, and limited evidence of wide-spread demand can make it challenging to justify effort and funds required to format, document, share, and preserve data. This paper describes a project that: (1) surveyed authors of highly cited papers published in 2015 at the University of Illinois at Urbana-Champaign in nine STEM disciplines to determine if data were generated for their article and their knowledge of reuse by other researchers, and (2) surveyed authors who cited these 2015 articles to ascertain whether they reused data from the original article and how that data was obtained. The project goal was to better understand data reuse in practice and to explore if research data from an initial publication was reused in subsequent publications. While the results revealed reuse in many situations (and deemed important in these cases), the survey results and researcher supplied comments also indicated that data does not play the same role in all studies or even in studies that build on previous ones.

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>

Climate and Weather are of increasing interest to the scientific community and the general public. Data curation and stewardship are essential building blocks in the science community's quest to better understand how natural climate and weather systems behave and how activities of human civilization are altering the natural system. Rudimentary observations of the atmosphere and ocean have been collected for over one hundred years and proxy measurements of the climate can trace our planet's climatic history for millions of years. These observations coupled with the rapid advances in technology, such as powerful computers, rapid access to massive amounts of data, and satellite observations, have allowed innovative techniques to be used to understand and predict the planet's climate and weather.

Jeong, Geum Hee. "Status of the Data Sharing Policies of Scholarly Journals Published in Brazil, France, and Korea and Listed in Both the 2018 Scimago Journal and Country Ranking and the Web of Science." *Science Editing* 7, no. 2 (2020): 136-141.  
<https://doi.org/10.6087/kcse.208>

## Purpose

The present study analyzed the current status of the data sharing policies of journals published in Brazil, France, and Korea that were listed in the 2018 Scimago Journal and Country Ranking and Web of Science Core Collection.

## Methods

Web of Science journals were selected from the 2018 Scimago Journal and Country Ranking. The homepages of all target journals were searched for the presence of statements on data sharing policies, including clinical trial data sharing policies, the level of the policies, and actual statements of data availability in articles.

## Results

Out of 565 journals from these three countries, 118 (20.9%) had an optional data sharing policy, and one had a mandatory data sharing policy. Harvard Dataverse was the repository of one journal. The number of journals that had adopted a data sharing policy was 11 (6.7%) for Brazil, 64 (27.6%) for France, and 44 (25.9%) for Korea. One journal from Brazil and 20 journals from Korea had adopted clinical trial data sharing policies in accordance with the International Committee of Medical Journal Editors. Statements of data sharing were found in articles from two journals.

## Conclusion

Journals from France and Korea adopted data sharing policies more actively than those from Brazil. However, the actual implementation of these policies through descriptions of data availability in articles remains rare. In many journals that appear to have data sharing policies, those policies may just reflect a standard description by the publisher, especially in France. Actual data sharing was not found to be frequent.

Johnson, Jeremiah N., Keith A. Hanson, Caleb A. Jones, Ramesh Grandhi, Jaime Guerrero, and Jesse S. Rodriguez. "Data Sharing in Neurosurgery and Neurology Journals." *Cureus* 10, no. 5 (2018): e2680. <https://dx.doi.org/10.7759/cureus.2680>

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>

Kaye, Jane, Sharon F. Terry, Eric Juengst, Sarah Coy, Jennifer R. Harris, Don Chalmers, Edward S. Dove, Isabelle Budin-Ljøsne, Clement Adebamowo, Emilomo Ogbe, Louise Bezuidenhout, Michael Morrison, Joel T. Minion, Madeleine J. Murtagh, Jusaku Minari, Harriet Teare, Rosario Isasi, Kazuto Kato, Emmanuelle Rial-Sebbag, Patricia Marshall, Barbara Koenig, and Anne Cambon-Thomsen. "Including All Voices in International Data-Sharing Governance." *Human Genomics* 12, no. 13 (2018). <https://doi.org/10.1186/s40246-018-0143-9>

## Background

Governments, funding bodies, institutions, and publishers have developed a number of strategies to encourage researchers to facilitate access to datasets. The rationale behind this approach is that this will bring a number of benefits and enable advances in healthcare and medicine by allowing the maximum returns from the investment in research, as well as reducing waste and promoting transparency. As this approach gains momentum, these data-sharing practices have implications for many kinds of research as they become standard practice across the world.

## Main text

The governance frameworks that have been developed to support biomedical research are not well equipped to deal with the complexities of international data sharing. This system is nationally based and is dependent upon expert committees for

oversight and compliance, which has often led to piece-meal decision-making. This system tends to perpetuate inequalities by obscuring the contributions and the important role of different data providers along the data stream, whether they be low- or middle-income country researchers, patients, research participants, groups, or communities. As research and data-sharing activities are largely publicly funded, there is a strong moral argument for including the people who provide the data in decision-making and to develop governance systems for their continued participation.

## Conclusions

We recommend that governance of science becomes more transparent, representative, and responsive to the voices of many constituencies by conducting public consultations about data-sharing addressing issues of access and use; including all data providers in decision-making about the use and sharing of data along the whole of the data stream; and using digital technologies to encourage accessibility, transparency, and accountability. We anticipate that this approach could enhance the legitimacy of the research process, generate insights that may otherwise be overlooked or ignored, and help to bring valuable perspectives into the decision-making around international data sharing.

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>

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, Nushrat, Mike Thelwall, and Kayvan Kousha. "Measuring the Impact of Biodiversity Datasets: Data Reuse, Citations and Altmetrics." *Scientometrics* 126, no. 4 (2021): 3621-3639. <https://doi.org/10.1007/s11192-021-03890-6>

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, Jihyun, Soon Kim, Hye-Min Cho, Jae Hwa Chang, and Soo Young Kim. "Data Sharing Policies of Journals in Life, Health, and Physical Sciences Indexed in Journal Citation Reports." *PeerJ* 8 (2020): e9924 <https://doi.org/10.7717/peerj.9924>

## Background

Many scholarly journals have established their own data-related policies, which specify their enforcement of data sharing, the types of data to be submitted, and their procedures for making data available. However, except for the journal impact factor and the subject area, the factors associated with the overall strength of the data sharing policies of scholarly journals remain unknown. This study examines how factors, including impact factor, subject area, type of journal publisher, and geographical location of the publisher are related to the strength of the data sharing policy.

## Methods

From each of the 178 categories of the Web of Science's 2017 edition of Journal Citation Reports, the top journals in each quartile (Q1, Q2, Q3, and Q4) were selected in December 2018. Of the resulting 709 journals (5%), 700 in the fields of life, health, and physical sciences were selected for analysis. Four of the authors independently reviewed the results of the journal website searches, categorized the journals' data sharing policies, and extracted the characteristics of individual journals. Univariable multinomial logistic regression analyses were initially conducted to determine whether there was a relationship between each factor and the strength of the data sharing policy. Based on the univariable analyses, a multivariable model was performed to further investigate the factors related to the presence and/or strength of the policy.

## Results

Of the 700 journals, 308 (44.0%) had no data sharing policy, 125 (17.9%) had a weak policy, and 267 (38.1%) had a strong policy (expecting or mandating data sharing). The impact factor quartile was positively associated with the strength of the data sharing policies. Physical science journals were less likely to have a strong policy relative to a weak policy than Life science journals (relative risk ratio [RRR], 0.36; 95% CI [0.17-0.78]). Life science journals had a greater probability of having a weak policy relative to no policy than health science journals (RRR, 2.73; 95% CI [1.05-7.14]). Commercial publishers were more likely to have a weak policy relative to no policy than non-commercial publishers (RRR, 7.87; 95% CI, [3.98-15.57]). Journals by publishers in Europe, including the majority of those located in the United Kingdom and the Netherlands, were more likely to have a strong data sharing policy than a weak policy (RRR, 2.99; 95% CI [1.85-4.81]).

## Conclusions

These findings may account for the increase in commercial publishers' engagement in data sharing and indicate that European national initiatives that encourage and mandate data sharing may influence the presence of a strong policy in the associated journals. Future research needs to explore the factors associated with varied degrees in the strength of a data sharing policy as well as more diverse characteristics of journals related to the policy strength.

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>

———. "A Study of the Determinants of Psychologists' Data Sharing and Open Data Badge Adoption." *Learned Publishing* 34, no. 4 (2021): 499-509. <https://doi.org/10.1002/LEAP.1388>

———. "A Study of the Roles of Metadata Standard and Data Repository in Science, Technology, Engineering and Mathematics Researchers' Data Reuse." *Online Information Review* 45, no. 7 (2021): 1306-1321. <https://doi.org/10.1108/OIR-09-2020-0431>

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 (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 (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://utpjournals.press/doi/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>

King, Gary. "An Introduction to the Dataverse Network as an Infrastructure for Data Sharing." *Sociological Methods & Research* 36, no. 2 (2007): 173-199. <https://doi.org/10.1177/0049124107306660>

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): p.43. <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>

Koesten, Laura, Pavlos Vougiouklis, Elena Simperl, and Paul Groth. "Dataset Reuse: Toward Translating Principles to Practice." *Patterns* 1, no.8 (2020): 100136. <https://doi.org/10.1016/j.patter.2020.100136>

The web provides access to millions of datasets that can have additional impact when used beyond their original context. We have little empirical insight into what makes a dataset more reusable than others and which of the existing guidelines and frameworks, if any, make a difference. In this paper, we explore potential reuse features through a literature review and present a case study on datasets on GitHub, a popular open platform for sharing code and data. We describe a corpus of more than 1.4 million data files, from over 65,000 repositories. Using GitHub's engagement metrics as proxies for dataset reuse, we relate them to reuse features from the literature and devise an initial model, using deep neural networks, to predict a dataset's reusability. This demonstrates the practical gap between principles and actionable insights that allow data publishers and tools designers to implement functionalities that provably facilitate reuse.

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 Literacy for Researchers and Data Librarians." *Journal of Librarianship and Information Science* 49, no. 1 (2017): 3-14.  
<https://doi.org/10.1177%2F0961000615616450>

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>

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.

Kurata, Keiko, Mamiko Matsubayashi, and Shinji Mine. "Identifying the Complex Position of Research Data and Data Sharing Among Researchers in Natural Science." *SAGE Open* (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.

Kush, R. D., D. Warzel, M. A. Kush, A. Sherman, E. A. Navarro, R. Fitzmartin, F. Pétavy, J. Galvez, L. B. Becnel, F. L. Zhou, N. Harmon, B. Jauregui, T. Jackson, and L. Hudson. "FAIR Data Sharing: The Roles of Common Data Elements and Harmonization." *Journal of Biomedical Informatics* 107 (2020): 103421. <https://doi.org/10.1016/J.JBI.2020.103421>.

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/>.

Law, Margaret. "Reduce, Reuse, Recycle: Issues in the Secondary Use of Research Data." *IASSIST Quarterly* 29, no. 1 (2006): 5. <https://doi.org/10.29173/IQ599>

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: The Journal of the Association of European Research Libraries* 27, no. 1 (2017): 58-72. <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.

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>

Mannheimer, Sara. "Data Curation Implications of Qualitative Data Reuse and Big Social Research." *Journal of eScience Librarianship* 10, no. 4 (2021): e1218. <https://doi.org/10.7191/jeslib.2021.1218>

Objective:

Big social data (such as social media and blogs) and archived qualitative data (such as interview transcripts, field notebooks, and diaries) are similar, but their respective communities of practice are under-connected. This paper explores shared challenges



in qualitative data reuse and big social research and identifies implications for data curation.

#### Methods:

This paper uses a broad literature search and inductive coding of 300 articles relating to qualitative data reuse and big social research. The literature review produces six key challenges relating to data use and reuse that are present in both qualitative data reuse and big social research—context, data quality, data comparability, informed consent, privacy & confidentiality, and intellectual property & data ownership.

#### Results:

This paper explores six key challenges related to data use and reuse for qualitative data and big social research and discusses their implications for data curation practices.

#### Conclusions:

Data curators can benefit from understanding these six key challenges and examining data curation implications. Data curation implications from these challenges include strategies for: providing clear documentation; linking and combining datasets; supporting trustworthy repositories; using and advocating for metadata standards; discussing alternative consent strategies with researchers and IRBs; understanding and supporting deidentification challenges; supporting restricted access for data; creating data use agreements; supporting rights management and data licensing; developing and supporting alternative archiving strategies. Considering these data curation implications will help data curators support sounder practices for both qualitative data reuse and big social research.

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>.

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>

Mbuagbaw, Lawrence, Gary Foster, Ji Cheng, and Lehana Thabane. "Challenges to Complete and Useful Data Sharing." *Trials* 18, no. 71 (2017). <https://doi.org/10.1186/s13063-017-1816-8>

Data sharing from clinical trials is one way of promoting fair and transparent conduct of clinical trials. It would maximise the use of data and permit the exploration of additional hypotheses. On the other hand, the quality of secondary analyses cannot always be ascertained, and it may be unfair to investigators who have expended resources to collect data to bear the additional burden of sharing. As the discussion on the best modalities of sharing data evolves, some of the practical issues that may arise need to be addressed. In this paper, we discuss issues which impede the use of data even when sharing should be possible: (1) multicentre studies requiring consent from all the investigators in each centre; (2) remote access platforms with software limitations and Internet requirements; (3) on-site data analysis when data cannot be moved; (4) governing bodies for data generated in one jurisdiction and analysed in another; (5) using programmatic data collected as part of routine care; (6) data collected in multiple languages; (7) poor data quality. We believe these issues apply to all primary data and cause undue difficulties in conducting analysis even when there is some willingness to share. They can be avoided by anticipating the possibility of sharing any clinical data and pre-emptively removing or addressing restrictions that limit complete sharing. These issues should be part of the data sharing discussion.

Melero, Remedios, and Carolina Navarro-Molina. "Researchers' Attitudes and Perceptions towards Data Sharing and Data Reuse in the Field of Food Science and Technology." *Learned Publishing* 33, no. 2 (2020): 163-179. <https://doi.org/10.1002/LEAP.1287>

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.

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.

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>

Moody, Bryony, Tom Dye, Keith May, Holly Wright, and Caitlin Buck. "Digital Chronological Data Reuse in Archaeology: Three Case Studies with Varying Purposes and Perspectives." *Journal of Archaeological Science Reports, Part A* 40 (2021): 103188. <https://doi.org/10.1016/J.JASREP.2021.103188>

Mozersky, 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 (2020): 1-23. <https://doi.org/10.29173/IQ952>

Naudet, Florian, Charlotte Sakarovitch, Perrine Janiaud, Ioana Cristea, Daniele Fanelli, David Moher, and John P. A. Ioannidis. "Data Sharing and Reanalysis of Randomized Controlled Trials in Leading Biomedical Journals with a Full Data Sharing Policy: Survey of Studies Published in *The BMJ* and *PLoS Medicine*." *BMJ* 360, no. 8141 (2018): k400. <https://doi.org/10.1136/BMJ.K400>

Naudet, Florian, Maximilian Siebert, Claude Pellen, Jeanne Gaba, Cathrine Axfors, Ioana Cristea, Valentin Danchev, Ulrich Mansmann, Christian Ohmann, Joshua D. Wallach, David Moher, and John P. A. Ioannidis. "Medical Journal Requirements for Clinical Trial Data Sharing: Ripe for Improvement." *PLoS Medicine* 18, no. 10 (2021): e1003844. <https://doi.org/10.1371/JOURNAL.PMED.1003844>

Efficient sharing and reuse of data from clinical trials are critical in advancing medical knowledge and developing improved treatments.,

We believe that the International Committee of Medical Journal Editors (ICMJE) clinical trial data sharing policy is currently inadequate.,

Although data sharing plans help increase transparency, they do not ensure that data are shared, and they are often inadequately implemented.,

We believe that the ICMJE should adapt a stronger policy on data sharing that is enforced rigorously in all ICMJE members and affiliated journals.,

The policy should include a strong evaluation component to ensure that all clinical trial data are shared, their value maximized, and data producers incentivized.,

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>

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>

Panhuis, Willem G van, Proma Paul, Claudia Emerson, John Grefenstette, Richard Wilder, Abraham J Herbst, David Heymann, and Donald S Burke. "A Systematic Review of Barriers to Data Sharing in Public Health." *BMC Public Health* 14, no. 1144 (2014). <https://doi.org/10.1186/1471-2458-14-1144>

## Background

In the current information age, the use of data has become essential for decision making in public health at the local, national, and global level. Despite a global commitment to the use and sharing of public health data, this can be challenging in reality. No systematic framework or global operational guidelines have been created for data sharing in public health. Barriers at different levels have limited data sharing but have only been anecdotally discussed or in the context of specific case studies. Incomplete systematic evidence on the scope and variety of these barriers has limited opportunities to maximize the value and use of public health data for science and policy.

## Methods

We conducted a systematic literature review of potential barriers to public health data sharing. Documents that described barriers to sharing of routinely collected public health data were eligible for inclusion and reviewed independently by a team of experts. We grouped identified barriers in a taxonomy for a focused international dialogue on solutions.

## Results

Twenty potential barriers were identified and classified in six categories: technical, motivational, economic, political, legal and ethical. The first three categories are deeply rooted in well-known challenges of health information systems for which structural solutions have yet to be found; the last three have solutions that lie in an international dialogue aimed at generating consensus on policies and instruments for data sharing.

## Conclusions

The simultaneous effect of multiple interacting barriers ranging from technical to intangible issues has greatly complicated advances in public health data sharing. A systematic framework of barriers to data sharing in public health will be essential to accelerate the use of valuable information for the global good.

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>

Park, Hyoungjoo, Sukjin You, and Dietmar Wolfram. "Informal Data Citation for Data Sharing and Reuse Is More Common than Formal Data Citation in Biomedical Fields." *Journal of the Association for Information Science and Technology* 69, no. 11 (2018): 1346-1354. <https://doi.org/10.1002/ASI.24049>.

Parsons, Rebecca, and Summers, Scott. "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>

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): p.8. <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.

Pavlenko, Elena, Daniel Streh, and Holger Langhof. "Implementation of Data Access and Use Procedures in Clinical Data Warehouses. A Systematic Review of Literature and Publicly Available Policies." *BMC Medical Informatics and Decision Making* 20, no. 1 (2020): 1-13. <https://doi.org/10.1186/s12911-020-01177-z>

## Background

The promises of improved health care and health research through data-intensive applications rely on a growing amount of health data. At the core of large-scale data integration efforts, clinical data warehouses (CDW) are also responsible for data governance, managing data access and (re)use. As the complexity of the data flow increases, greater transparency and standardization of criteria and procedures are required in order to maintain objective oversight and control. Therefore, the development of practice oriented and evidence-based policies is crucial. This study assessed the spectrum of data access and use criteria and procedures in clinical data warehouses governance internationally.

## Methods

We performed a systematic review of (a) the published scientific literature on CDW and (b) publicly available information on CDW data access, e.g., data access policies. A qualitative thematic analysis was applied to all included literature and policies.

## Results

Twenty-three scientific publications and one policy document were included in the final analysis. The qualitative analysis led to a final set of three main thematic categories: (1) requirements, including recipient requirements, reuse requirements, and formal requirements; (2) structures and processes, including review bodies and review values; and (3) access, including access limitations.

## Conclusions

The description of data access and use governance in the scientific literature is characterized by a high level of heterogeneity and ambiguity. In practice, this might limit the effective data sharing needed to fulfil the high expectations of data-intensive approaches in medical research and health care. The lack of publicly available information on access policies conflicts with ethical requirements linked to principles of transparency and accountability.

CDW should publicly disclose by whom and under which conditions data can be accessed, and provide designated governance structures and policies to increase transparency on data access. The results of this review may contribute to the development of practice-oriented minimal standards for the governance of data access, which could also result in a stronger harmonization, efficiency, and effectiveness of CDW.

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](http://theastrodata.org), and we analyze the uptake of that system to-date.

Perrier, Laure, Erik Blondal, and Heather MacDonald. "The Views, Perspectives, and Experiences of Academic Researchers with Data Sharing and Reuse: A Meta-Synthesis." *PLoS ONE* 15, no. 2 (2020): e0229182. <https://doi.org/10.1371/JOURNAL.PONE.0229182>

Funding agencies and research journals are increasingly demanding that researchers share their data in public repositories. Despite these requirements, researchers still withhold data, refuse to share, and deposit data that lacks annotation. We conducted a meta-synthesis to examine the views, perspectives, and experiences of academic researchers on data sharing and reuse of research data.

## Methods

We searched the published and unpublished literature for studies on data sharing by researchers in academic institutions. Two independent reviewers screened citations and abstracts, then full-text articles. Data abstraction was performed independently by two investigators. The abstracted data was read and reread in order to generate codes. Key concepts were identified and thematic analysis was used for data synthesis.

## Results

We reviewed 2005 records and included 45 studies along with 3 companion reports. The studies were published between 2003 and 2018 and most were conducted in North America (60%) or Europe (17%). The four major themes that emerged were data integrity, responsible conduct of research, feasibility of sharing data, and value of sharing data. Researchers lack time, resources, and skills to effectively share their data in public repositories. Data quality is affected by this, along with subjective decisions around what is considered to be worth sharing. Deficits in infrastructure also impede the availability of research data. Incentives for sharing data are lacking.

## Conclusion

Researchers lack skills to share data in a manner that is efficient and effective. Improved infrastructure support would allow them to make data available quickly and seamlessly. The lack of incentives for sharing research data with regards to academic appointment, promotion, recognition, and rewards need to be addressed.

Phillips, Mark. "International Data-Sharing Norms: From the OECD to the General Data Protection Regulation (GDPR)." *Human Genetics* 137, no.8 (2018): 575-582. <https://doi.org/10.1007/s00439-018>

The evolution of genomic research and its integration into clinical practice, as they become international—even global—endeavors, has brought us to a place where scientists and clinicians may now only ignore the rules governing international data sharing at their own peril. Open data policies, on the one hand, increasingly require custodians of others' genomic data to make it as widely available as feasible, including to researchers in other countries. Data protection law, on the other, has become a significant hurdle to the sharing of personal data across jurisdictional borders. The space between these two competing duties is narrowing. In contrast with the other texts in this volume, which explore the present and future of data sharing and data protection, this article's focus is on the past. It centres on the historical development of the data protection rules regarding the international transfer of personal data up to the present. The article's aim is to bring into focus the underlying objectives that have influenced and that will continue to influence the way that data protection rules are applied to the fields of genomics and health, as well as future developments in data protection generally. The first part of this article describes the development of international data-sharing data protection rules since 1970. The second considers difficulties in applying general data protection rules to the specific context of genomics and health. The third and final part compares the options available to comply with the international transfer restrictions set out in the standard-setting EU General Data Protection Regulation from a genomics perspective.

Pisani, Elizabeth, and Carla AbouZahr. "Sharing Health Data: Good Intentions Are Not Enough." *Bulletin of The World Health Organization* 88, no. 6 (2010): 462-466. <https://doi.org/10.2471/BLT.09.074393>

Piwowar, Heather A. "Data Reuse and the Open Data Citation Advantage." *PeerJ* 1, no. 1 (2013): e175. <https://doi.org/10.7717/PEERJ.175>.

**Background.** Attribution to the original contributor upon reuse of published data is important both as a reward for data creators and to document the provenance of research findings. Previous studies have found that papers with publicly available datasets receive a higher number of citations than similar studies without available data. However, few previous analyses have had the statistical power to control for the many variables known to predict citation rate, which has led to uncertain estimates of the "citation benefit". Furthermore, little is known about patterns in data reuse over time and across datasets.

**Method and Results.** Here, we look at citation rates while controlling for many known citation predictors and investigate the variability of data reuse. In a multivariate regression on 10,555 studies that created gene expression microarray data, we found that studies that made data available in a public repository received 9% (95% confidence interval: 5% to 13%) more citations than similar studies for which the data was not made available. Date of publication, journal impact factor, open access status, number of authors, first and last author publication history, corresponding author country, institution citation history, and study topic were included as covariates. The citation benefit varied with date of dataset deposition: a citation benefit was most clear for papers published in 2004 and 2005, at about 30%. Authors published most papers using their own datasets within two years of their first publication on the dataset, whereas data reuse papers published by third-party investigators continued to accumulate for at least six years. To study patterns of data reuse directly, we compiled 9,724 instances of third party data reuse via mention of GEO or ArrayExpress accession numbers in the full text of papers. The level of third-party data use was high: for 100 datasets deposited in year 0, we estimated that 40 papers in PubMed reused a dataset by year 2, 100 by year 4, and more than 150 data reuse papers had been published by year 5. Data reuse was distributed across a broad base of datasets: a very conservative estimate found that 20% of the datasets deposited between 2003 and 2007 had been reused at least once by third parties.

**Conclusion.** After accounting for other factors affecting citation rate, we find a robust citation benefit from open data, although a smaller one than previously reported. We conclude there is a direct effect of third-party data reuse that persists for years beyond the time when researchers have published most of the papers reusing their own data.

Other factors that may also contribute to the citation benefit are considered. We further conclude that, at least for gene expression microarray data, a substantial fraction of archived datasets are reused, and that the intensity of dataset reuse has been steadily increasing since 2003.

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.

This correlation between publicly available data and increased literature impact may further motivate investigators to share their detailed research data.

Pm, Naushad Ali, and Sidra Saeed. "Research Data Management and Data Sharing among Research Scholars of Life Sciences and Social Sciences." *DESIDOC Journal of Library & Information Technology* 39, no. 6 (2019): 290-299. <https://doi.org/10.14429/djlit.39.06.14997>

## 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.

Pronk, Tessa E. "The Time Efficiency Gain in Sharing and Reuse of Research Data." *Data Science Journal*, 18, no. 1 (2019): p.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>

Pryor, Graham. "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.

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>

Read, Kevin, Jessica Athens, Ian Lamb, Joey Nicholson, Sushan Chin, Junchuan Xu, Neil Rambo, and Alisa Surkis. "Promoting Data Reuse and Collaboration at an Academic Medical Center." *International Journal of Digital Curation* 10, no. 1 (2015): 260-267. <https://doi.org/10.2218/IJDC.V10I1.366>.

A need was identified by the Department of Population Health (DPH) for an academic medical center to facilitate research using large, externally funded datasets. Barriers identified included difficulty in accessing and working with the datasets, and a lack of knowledge about institutional licenses. A need to facilitate sharing and reuse of datasets generated by researchers at the institution (internal datasets) was also recognized. The library partnered with a researcher in the DPH to create a catalog of external datasets, which provided detailed metadata and access instructions. The catalog listed researchers at the medical center and the main campus with expertise in using these external datasets in order to facilitate research and cross-campus collaboration. Data description standards were reviewed to create a set of metadata to facilitate access to both externally generated datasets, as well as the internally generated datasets that would constitute the next phase of development of the catalog. Interviews with a range of investigators at the institution identified DPH researchers as most interested in data sharing, therefore targeted outreach to this group was undertaken. Initial outreach resulted in additional external datasets being described, new local experts volunteering, proposals for additional functionality, and interest from researchers in inclusion of their internal datasets in the catalog. Despite limited outreach, the catalog has had ~250 unique page views in the three months since it went live. The establishment of the catalog also led to partnerships with the medical center's data management core and the main university library. The Data Catalog in its present state serves a direct user need from the Department of Population Health to describe large, externally funded datasets. The library will use this initial strong community of users to expand the catalog and include internally

generated research datasets. Future expansion plans will include working with DataCore and the main university library.

Riedel, Nico, Miriam Kip, and Evgeny Bobrov. "Oddpub —A Text-Mining Algorithm to Detect Data Sharing in Biomedical Publications." *Data Science Journal* 19, no. 1 (2020): p.42. <https://doi.org/10.5334/dsj-2020-042>

Open research data are increasingly recognized as a quality indicator and an important resource to increase transparency, robustness and collaboration in science. However, no standardized way of reporting Open Data in publications exists, making it difficult to find shared datasets and assess the prevalence of Open Data in an automated fashion.

We developed ODDPub (Open Data Detection in Publications), a text-mining algorithm that screens biomedical publications and detects cases of Open Data. Using English-language original research publications from a single biomedical research institution (n = 8689) and randomly selected from PubMed (n = 1500) we iteratively developed a set of derived keyword categories. ODDPub can detect data sharing through field-specific repositories, general-purpose repositories or the supplement. Additionally, it can detect shared analysis code (Open Code).

To validate ODDPub, we manually screened 792 publications randomly selected from PubMed. On this validation dataset, our algorithm detected Open Data publications with a sensitivity of 0.73 and specificity of 0.97. Open Data was detected for 11.5% (n = 91) of publications. Open Code was detected for 1.4% (n = 11) of publications with a sensitivity of 0.73 and specificity of 1.00. We compared our results to the linked datasets found in the databases PubMed and Web of Science.

Our algorithm can automatically screen large numbers of publications for Open Data. It can thus be used to assess Open Data sharing rates on the level of subject areas, journals, or institutions. It can also identify individual Open Data publications in a larger publication corpus. ODDPub is published as an R package on GitHub.

Renaut, Sébastien, Amber E. Budden, Dominique Gravel, Timothée Poisot, and Pedro Peres-Neto. "Management, Archiving, and Sharing for Biologists and the Role of Research Institutions in the Technology-Oriented Age." *BioScience* 68, no. 6 (2018): 400-411. <https://doi.org/10.1093/BIOSCI/BIY038>

Riedel, Nico, Miriam Kip, and Evgeny Bobrov. "Oddpub —A Text-Mining Algorithm to Detect Data Sharing in Biomedical Publications." *Data Science Journal* 19, no. 1 (2020): p.42. <https://doi.org/10.5334/dsj-2020-042>

Open research data are increasingly recognized as a quality indicator and an important resource to increase transparency, robustness and collaboration in science. However, no standardized way of reporting Open Data in publications exists, making it difficult to find shared datasets and assess the prevalence of Open Data in an automated fashion.

We developed ODDPub (Open Data Detection in Publications), a text-mining algorithm that screens biomedical publications and detects cases of Open Data. Using English-language original research publications from a single biomedical research institution (n = 8689) and randomly selected from PubMed (n = 1500) we iteratively developed a set of derived keyword categories. ODDPub can detect data sharing through field-specific repositories, general-purpose repositories or the supplement. Additionally, it can detect shared analysis code (Open Code).

To validate ODDPub, we manually screened 792 publications randomly selected from PubMed. On this validation dataset, our algorithm detected Open Data publications with a sensitivity of 0.73 and specificity of 0.97. Open Data was detected for 11.5% (n = 91) of publications. Open Code was detected for 1.4% (n = 11) of publications with a sensitivity of 0.73 and specificity of 1.00. We compared our results to the linked datasets found in the databases PubMed and Web of Science.

Our algorithm can automatically screen large numbers of publications for Open Data. It can thus be used to assess Open Data sharing rates on the level of subject areas,

journals, or institutions. It can also identify individual Open Data publications in a larger publication corpus. ODDPub is published as an R package on GitHub.

Ross, Joseph S. "Clinical Research Data Sharing: What an Open Science World Means for Researchers Involved in Evidence Synthesis." *Systematic Reviews* 5, no. 159 (2016). <https://doi.org/10.1186/S13643-016-0334-1>

The International Committee of Medical Journal Editors (ICMJE) recently announced a bold step forward to require data generated by interventional clinical trials that are published in its member journals to be responsibly shared with external investigators. The movement toward a clinical research culture that supports data sharing has important implications for the design, conduct, and reporting of systematic reviews and meta-analyses. While data sharing is likely to enhance the science of evidence synthesis, facilitating the identification and inclusion of all relevant research, it will also pose key challenges, such as requiring broader search strategies and more thorough scrutiny of identified research. Furthermore, the adoption of data sharing initiatives by the clinical research community should challenge the community of researchers involved in evidence synthesis to follow suit, including the widespread adoption of systematic review registration, results reporting, and data sharing, to promote transparency and enhance the integrity of the research process.

Ross, Joseph S., Joanne Waldstreicher, Stephen Bamford, Jesse A. Berlin, Karla Childers, Nihar R. Desai, Ginger Gamble, Cary P. Gross, Richard Kuntz, Richard Lehman, Peter Lins, Sandra A. Morris, Jessica D. Ritchie, and Harlan M. Krumholz. "Overview and Experience of the YODA Project with Clinical Trial Data Sharing after 5 Years." *Scientific Data* 5, no. 180268 (2018). <https://doi.org/10.1038/SDATA.2018.268>

The Yale University Open Data Access (YODA) Project has facilitated access to clinical trial data since 2013. The purpose of this article is to provide an overview of the Project, describe key decisions that were made when establishing data sharing policies, and suggest how our experience and the experiences of our first two data generator partners, Medtronic, Inc. and Johnson & Johnson, can be used to enhance other ongoing or future initiatives.

Rousi, Antti Mikael, and Mikael Laakso. "Journal Research Data Sharing Policies: A Study of Highly-Cited Journals in Neuroscience, Physics, and Operations Research." *Scientometrics* 124, no. 1 (2020): 131-152. <https://doi.org/10.1007/s11192-020-03467-9>

The practices for if and how scholarly journals instruct research data for published research to be shared is an area where a lot of changes have been happening as science policy moves towards facilitating open science, and subject-specific repositories and practices are established. This study provides an analysis of the research data sharing policies of highly-cited journals in the fields of neuroscience, physics, and operations research as of May 2019. For these 120 journals, 40 journals per subject category, a unified policy coding framework was developed to capture the most central elements of each policy, i.e. what, when, and where research data is instructed to be shared. The results affirm that considerable differences between research fields remain when it comes to policy existence, strength, and specificity. The findings revealed that one of the most important factors influencing the dimensions of what, where and when of research data policies was whether the journal's scope included specific data types related to life sciences which have established methods of sharing through community-endorsed public repositories. The findings surface the future research potential of approaching policy analysis on the publisher-level as well as on the journal-level. The collected data and coding framework is provided as open data to facilitate future research and journal policy monitoring.

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. 4 (2017). <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](https://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) outweighed 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.

Safran, C. "Update on Data Reuse in Health Care." *Yearbook of Medical Informatics* 26, no. 1 (2017): 24-27. <https://doi.org/10.15265/IY-2017-013>.

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>

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.

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>

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. <https://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.

Scoulas, Jung Mi, Sandra L. De Groote, and Paula R. Dempsey. "Learning from Data Reuse: Successful and Failed Experiences in a Large Public Research University Library." *IASSIST Quarterly* 44, no. 1-2 (2020): 1-15. <https://doi.org/10.29173/iq966>

Shahin, Mohamed H., Sanchita Bhattacharya, Diego Silva, Sarah Kim, Jackson Burton, Jagdeep Podichetty, Klaus Romero, and Daniela J. Conrado. "Open Data Revolution in Clinical Research: Opportunities and Challenges." *Clinical and Translational Science* 13, no. 4 (2020): 665-674. <https://doi.org/10.1111/cts.12756>

Efforts for sharing individual clinical data are gaining momentum due to a heightened recognition that integrated data sets can catalyze biomedical discoveries and drug development. Among the benefits are the fact that data sharing can help generate and investigate new research hypothesis beyond those explored in the original study. Despite several accomplishments establishing public systems and guidance for data sharing in clinical trials, this practice is not the norm. Among the reasons are ethical challenges, such as privacy of individuals, data ownership, and control. This paper creates awareness of the potential benefits and challenges of sharing individual clinical data, how to overcome these challenges, and how as a clinical pharmacology community we can shape future directions in this field.

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.

Siebert, Maximilian, Jeanne Fabiola Gaba, Laura Caquelin, Henri Gouraud, Alain Dupuy, David Moher, and Florian Naudet. "Data-Sharing Recommendations in Biomedical Journals and Randomised Controlled Trials: An Audit of Journals Following the Icmje Recommendations." *BMJ Open* 10, no. 5 (2020): e038887. <https://doi.org/10.1136/bmjopen-2020-038887>

Sielemann, Katharina, Alenka Hafner, and Boas Pucker. "The Reuse of Public Datasets in the Life Sciences: Potential Risks and Rewards." *PeerJ* 8 (2020): e9954. <https://doi.org/10.7717/PEERJ.9954>

The 'big data' revolution has enabled novel types of analyses in the life sciences, facilitated by public sharing and reuse of datasets. Here, we review the prodigious potential of reusing publicly available datasets and the associated challenges, limitations and risks. Possible solutions to issues and research integrity considerations are also discussed. Due to the prominence, abundance and wide distribution of sequencing data, we focus on the reuse of publicly available sequence datasets. We define 'successful reuse' as the use of previously published data to enable novel scientific findings. By using selected examples of successful reuse from different disciplines, we illustrate the enormous potential of the practice, while acknowledging the respective limitations and risks. A checklist to determine the reuse value and potential of a particular dataset is also provided. The open discussion of data reuse and the establishment of this practice as a norm has the potential to benefit all stakeholders in the life sciences.

Sobotkova, Adela. "Sociotechnical Obstacles to Archaeological Data Reuse." *Advances in Archaeological Practice* 6, no. 2 (2018): 117-124. <https://doi.org/10.1017/AAP.2017.37>

Stein, Ayla, and Elise Dunham. "Meaningful Data Sharing: Developing the Illinois Data Bank Metadata Framework." *Journal of Library Metadata* 18, no. 2 (2018): 59-83. <https://doi.org/10.1080/19386389.2018.1488561>.

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>

Suhr, Bettina, Johanna Dungal, and Alexander Stocker. "Search, Reuse and Sharing of Research Data in Materials Science and Engineering—A Qualitative Interview Study." *PLoS ONE* 15, no. 9 (2020): e0239216. <https://doi.org/10.1371/JOURNAL.PONE.0239216>.

Open research data practices are a relatively new, thus still evolving part of scientific work, and their usage varies strongly within different scientific domains. In the literature, the investigation of open research data practices covers the whole range of big empirical studies covering multiple scientific domains to smaller, in depth studies analysing a single field of research. Despite the richness of literature on this topic, there is still a lack of knowledge on the (open) research data awareness and practices in materials science and engineering. While most current studies focus only on some aspects of open research data practices, we aim for a comprehensive understanding of all practices with respect to the considered scientific domain. Hence this study aims at 1) drawing the whole picture of search, reuse and sharing of research data 2) while focusing on materials science and engineering. The chosen approach allows to explore the connections between different aspects of open research data practices, e.g. between data sharing and data search. In depth interviews with 13 researchers in this field were conducted, transcribed verbatim, coded and analysed using content analysis. The main findings characterised research data in materials science and engineering as extremely diverse, often generated for a very specific research focus and needing a precise description of the data and the complete generation process for possible reuse. Results on research data search and reuse showed that the interviewees intended to reuse data but were mostly unfamiliar with (yet interested in) modern methods as dataset search engines, data journals or searching public repositories. Current research data sharing is not open, but bilaterally and usually encouraged by supervisors or employers. Project funding does affect data sharing in two ways: some researchers argue to share their data openly due to their funding agency's policy, while others face legal restrictions for sharing as their projects are partly funded by industry. The time needed for a precise description of the data and their generation process is named as biggest obstacle for data sharing. From these findings, a precise set of actions is derived suitable to support Open Data, involving training for researchers and introducing rewards for data sharing on the level of universities and funding bodies.

Sweeney, Latanya, Mercè Crosas, and Michael Bar-Sinai. "Sharing Sensitive Data with Confidence: The Datatags System." *Technology Science*, no. 2015101601 (October, 15

2015). <http://techscience.org/a/2015101601>

Tedersoo, Leho, Rainer Küngas, Ester Oras, Kajar Köster, Helen Eenmaa, Äli Leijen, Margus Pedaste, Marju Raju, Anastasiya Astapova, and Heli Lukner. 2021. "Data Sharing Practices and Data Availability upon Request Differ across Scientific Disciplines." *Scientific Data* 8, no. 192 (2021). <https://doi.org/10.1038/s41597-021-00981-0>

Data sharing is one of the cornerstones of modern science that enables large-scale analyses and reproducibility. We evaluated data availability in research articles across nine disciplines in *Nature* and *Science* magazines and recorded corresponding authors' concerns, requests and reasons for declining data sharing. Although data sharing has improved in the last decade and particularly in recent years, data availability and willingness to share data still differ greatly among disciplines. We observed that statements of data availability upon (reasonable) request are inefficient and should not be allowed by journals. To improve data sharing at the time of manuscript acceptance, researchers should be better motivated to release their data with real benefits such as recognition, or bonus points in grant and job applications. We recommend that data management costs should be covered by funding agencies; publicly available research data ought to be included in the evaluation of applications; and surveillance of data sharing should be enforced by both academic publishers and funders. These cross-discipline survey data are available from the plutoF repository.

Tegbaru, Dawit, Lisa Braverman, Anthony L. Zietman, Sue S. Yom, W. Robert Lee, Robert C. Miller, Isabel L. Jackson, Todd McNutt, and Andre Dekker. "ASTRO Journals' Data Sharing Policy and Recommended Best Practices." *Advances in Radiation Oncology* 4, no. 4 (2019): 551-558. <https://doi.org/https://doi.org/10.1016/j.adro.2019.08.002>

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, Lisa Christian, Suzie Allard, and Josh Borycz. "Research Data Sharing: Practices and Attitudes of Geophysicists." *Earth and Space Science* 5, no. 12 (2018): 891-902. <https://doi.org/10.1029/2018EA000461>

Tenopir, Carol, Elizabeth D. Dalton, Suzie Allard, Mike Frame, Ivanka Pjesivac, Ben Birch, Danielle Pollock, and Kristina Dorsett. "Changes in Data Sharing and Data Reuse Practices and Perceptions among Scientists Worldwide." *PLoS ONE* 10, no. 8 (2015): e0134826. <https://doi.org/10.1371/JOURNAL.PONE.0134826>

The incorporation of data sharing into the research lifecycle is an important part of modern scholarly debate. In this study, the DataONE Usability and Assessment working group addresses two primary goals: To examine the current state of data sharing and reuse perceptions and practices among research scientists as they compare to the 2009/2010 baseline study, and to examine differences in practices and perceptions across age groups, geographic regions, and subject disciplines. We distributed surveys to a multinational sample of scientific researchers at two different time periods (October 2009 to July 2010 and October 2013 to March 2014) to observe current states of data sharing and to see what, if any, changes have occurred in the past 3-4 years. We also looked at differences across age, geographic, and discipline-based groups as they currently exist in the 2013/2014 survey. Results point to increased acceptance of and willingness to engage in data sharing, as well as an increase in actual data sharing behaviors. However, there is also increased perceived risk associated with data sharing, and specific barriers to data sharing persist. There are also differences across age groups, with younger respondents feeling more favorably toward data sharing and reuse, yet making less of their data available than older respondents. Geographic differences exist as well, which can in part be understood in terms of collectivist and individualist cultural differences. An examination of subject disciplines shows that the constraints and enablers of data sharing and reuse manifest differently across disciplines. Implications of these findings include the continued need to build infrastructure that promotes data sharing while recognizing the needs of different research communities. Moving into the future, organizations such as DataONE will continue to assess, monitor, educate, and provide the infrastructure necessary to support such complex grand science challenges.

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

Tenopir, Carol, Natalie M. Rice, Suzie Allard, Lynn Baird, Josh Borycz, Lisa Christian, Bruce Grant, Robert Olendorf, and Robert J. Sandusky. "Data Sharing, Management, Use, and Reuse: Practices and Perceptions of Scientists Worldwide." *PLoS ONE* 15, no. 3 (2020): e0229003. <https://doi.org/10.1371/JOURNAL.PONE.0229003>

## Background

With data becoming a centerpiece of modern scientific discovery, data sharing by scientists is now a crucial element of scientific progress. This article aims to provide an in-depth examination of the practices and perceptions of data management, including data storage, data sharing, and data use and reuse by scientists around the world.

## Methods

The Usability and Assessment Working Group of DataONE, an NSF-funded environmental cyberinfrastructure project, distributed a survey to a multinational and multidisciplinary sample of scientific researchers in a two-waves approach in 2017-2018. We focused our analysis on examining the differences across age groups, sub-disciplines of science, and sectors of employment.

## Findings

Most respondents displayed what we describe as high and mediocre risk data practices by storing their data on their personal computer, departmental servers or USB drives. Respondents appeared to be satisfied with short-term storage solutions; however, only half of them are satisfied with available mechanisms for storing data beyond the life of the process. Data sharing and data reuse were viewed positively:



over 85% of respondents admitted they would be willing to share their data with others and said they would use data collected by others if it could be easily accessed. A vast majority of respondents felt that the lack of access to data generated by other researchers or institutions was a major impediment to progress in science at large, yet only about a half thought that it restricted their own ability to answer scientific questions. Although attitudes towards data sharing and data use and reuse are mostly positive, practice does not always support data storage, sharing, and future reuse. Assistance through data managers or data librarians, readily available data repositories for both long-term and short-term storage, and educational programs for both awareness and to help engender good data practices are clearly needed.

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

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.

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>

Treloar, Andrew. "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>

Van de Sandt, Stephanie, Sünje Dallmeier-Tiessen, Artemis Lavasa, Vivien Petras. "The Definition of Reuse." *Data Science Journal* 18, no. 1 (2019): p.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 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.

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 (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.

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): 275-280. <https://doi.org/10.1007/s00784-017-2108-0>

Vines, Timothy H., Arianne Y. K. Albert, Rose L. Andrew, Florence Débarre, Dan G. Bock, Michelle T. Franklin, Kimberly J. Gilbert, Jean-Sébastien Moore, Sébastien Renaut, and Diana J. Rennison. "The Availability of Research Data Declines Rapidly with Article Age." *Current Biology* 24 (2014): 94-97. <https://doi.org/10.1016/j.cub.2013.11.014>

Volk, Carol J., Yasmin Lucero, and Katie Barnas. "Why Is Data Sharing in Collaborative Natural Resource Efforts so Hard and What Can We Do to Improve It." *Environmental Management* 53 (2014): 883-893. <https://doi.org/10.1007/S00267-014-0258-2>

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., 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.

Wang, Xiaoguang, Qingyu Duan, and Mengli Liang. "Understanding the Process of Data Reuse: An Extensive Review." *Journal of the Association for Information Science and Technology* 72, no. 9 (2021): 1161-1182. <https://doi.org/10.1002/asi.24483>

Wiley, Chris. "Data Sharing: An Analysis of Medical Faculty Journals and Articles." *Science & Technology Libraries* 40, no. 1 (2021): 104-115. <https://doi.org/10.1080/0194262X.2020.1781740>.

Wiley, Christie A. "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>

Williams, Sarah C. "Data Sharing Interviews with Crop Sciences Faculty: Why They Share Data and How the Library Can Help." *Issues in Science and Technology Librarianship*, no.

Winkler, Christa E., and Rebecca Fay Berenbon. "Validation of a Survey for Measuring Scientists' Attitudes toward Data Reuse." *Journal of the Association for Information Science and Technology* 72, no. 4 (2021): 449-453. <https://doi.org/10.1002/asi.24412>

Woolfrey, H. "Innovations for the Curation and Sharing of African Social Survey Data." *Data Science Journal* 12 (2013): pp.WDS185-WDS188. <https://datascience.codata.org/articles/abstract/137/>

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.

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>

Yoon, Ayoung, and Yoo Young Lee. "Factors of Trust in Data Reuse." *Online Information Review* 43, no. 7 (2019): 1245-1262. <https://doi.org/10.1108/OIR-01-2019-0014>.,

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>

Zhu, Claire S., Paul F. Pinsky, James E. Moler, Andrew Kukwa, Jerome Mabie, Joshua M. Rathmell, Tom Riley, Philip C. Prorok, and Christine D. Berg. "Data Sharing in Clinical Trials: An Experience with Two Large Cancer Screening Trials." *PLoS Medicine* 14, no. 5 (2017): e1002304. <https://doi.org/10.1371/JOURNAL.PMED.1002304>.,

Broad sharing of clinical trial data is important for ensuring reproducibility, transparency, and maximal use of the data by the research community. However, in practice, such data sharing typically requires planning, effort, and resources.,

Here, we describe a web-based data sharing system, the Cancer Data Access System (CDAS), developed for two large cancer screening trials: the Prostate, Lung, Colorectal and Ovarian (PLCO) cancer screening trial and the National Lung Screening Trial (NLST).,,

Deidentified individual participant data were organized into standard datasets readily downloadable from CDAS via a simple web-based application process that involves minimal scientific review. CDAS provides a "one-stop shop" for access requests, review, and data downloads.,

Since the launch of CDAS in November 2012 and through October 2016, 215 requests were received for PLCO data, of which 199 (93%) were approved, and 240 requests were received for NLST, of which 216 (90%) were approved.,

The estimated cost of CDAS was around US\$300,000 for the initial development, plus additional maintenance and user-support costs of about US\$26,000 per month. Because of its modular nature, additional studies can be added to CDAS with relatively little additional cost.,

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

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>

Zinner, Darren E., Genevieve Pham-Kanter, and Eric G. Campbell. "The Changing Nature of Scientific Sharing and Withholding in Academic Life Sciences Research: Trends From National Surveys in 2000 and 2013." *Academic Medicine* 91. no. 3 (2016): 433-440. <https://doi.org/10.1097/ACM.0000000000001028>

Zuiderwijk, Anneke, and Helen Spiers. "Sharing and Re-Using Open Data: A Case Study of Motivations in Astrophysics." *International Journal of Information Management* 49 (2019): 228-241. <https://doi.org/10.1016/J.IJINFOMGT.2019.05.024>

## Note on the Inclusion of Abstracts

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.

Abstracts for works under the following types of Creative Commons Licenses (and their national/international variations) are not included:

Attribution-NoDerivs

Attribution-NonCommercial

Attribution-NonCommercial-NoDerivs

Attribution-NonCommercial-ShareAlike

Attribution-ShareAlike

See the Creative Commons' [Frequently Asked Questions](#) for a discussion of how documents under different Creative Commons licenses can be combined.

## About the Author

Charles W. Bailey, Jr. is the publisher of [Digital Scholarship](#) and a noncommercial digital artist (ORCID ID: <https://orcid.org/0000-0001-8453-4402>).

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 32 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 been a frequent contributor of reviews to this monthly e-serial until 2020.

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 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 webliographies as websites with links to freely available works: the [Scholarly Electronic Publishing Bibliography](#) (1996-2011), the [Electronic Theses and Dissertations Bibliography](#) (2005-2021), 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](#) (2012-2019), the [Altmetrics Bibliography](#) (2013), the [Transforming Peer Review Bibliography](#) (2014), the [Academic Library as Scholarly Publisher Bibliography](#) (2018-2021), and the [Research Data Sharing and Reuse Bibliography](#) (2021).

In 2011, he established the LinkedIn Digital Curation Group.

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 artificial intelligence, digital copyright, 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.

You can follow Bailey at these URLs:

Digital Artist weblog: <http://charleswbaileyjr.name/> and RSS feed (<http://charleswbaileyjr.name/feed/>)

*DigitalKoans* weblog: <http://digital-scholarship.org/digitalkoans/>

Flickr: <https://www.flickr.com/photos/charleswbaileyjr/>

Twitter (*DigitalKoans*): <https://twitter.com/DigitalKoans>

## Citation

Charles W. Bailey, Jr., *Research Data Sharing and Reuse Bibliography* (Houston: Digital Scholarship, 2021), <http://digital-scholarship.org/rdsr/sharing.htm>.

Bailey, Charles W., Jr. *Research Data Sharing and Reuse Bibliography*. Houston: Digital Scholarship, 2021. <http://digital-scholarship.org/rdsr/sharing.htm>.

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/).