Journal of eScience Librarianship
putting the pieces together: theory and practice

Full-Length Paper

Bringing All the Stakeholders to the Table: A Collaborative Approach to Data Sharing

Megan N. O'Donnell and Curtis Brundy

Iowa State University, Ames, IA, USA

Abstract

Objective: This paper examines a unique data set disclosure process at a medium sized, land grant, research university and the campus collaboration that led to its creation.

Methods: The authors utilized a single case study methodology, reviewing relevant documents and workflows. As first-hand participants in the collaboration and disclosure process development, their own accounts and experiences also were utilized.

Results: A collaborative approach to enhancing research data sharing is essential, considering the wide array of stakeholders involved across the life cycle of research data. A transparent, inclusive data set disclosure process is a viable route to ensuring research data can be appropriately shared.

Conclusions: Successful sharing of research data impacts a range of university units and individuals. The establishment of productive working relationships and trust between these stakeholders is critical to expanding the sharing of research data and to establishing shared workflows.

Disclosures: The authors report no conflict of interest.

Introduction

There is growing recognition that scientific progress and the effectiveness of the research enterprise are improved and accelerated with increased openness and sharing (National Academies of Sciences, Engineering, and Medicine et al. 2018; International Science Council 2021; The Royal Society 2012). Driven by advances in technology and the expectation that publicly funded research should be freely shared, open science and scholarship have emerged as a growing and important issue for research universities. A recent report from the National Academies of Science, Engineering, and Medicine, *Open Science by Design*, states, "Openness and sharing of information are fundamental to the progress of science and to the effective functioning of the research enterprise" (2018, 17). While the imperative towards increased openness and sharing is being enacted and expressed across the research enterprise world-wide, the recent COVID-19 pandemic has highlighted how much more work needs to be done.

In the face of the public health threat posed by COVID-19, researchers, sponsors, publishers, and the public have all recognized the need to accelerate scientific research and discovery. This consensus has led to a rapid advance in the sharing of COVID-19 research outputs, with practices such as depositing pre-prints and publishing data sets openly becoming the norm (Kupferschmidt 2020). Swift embrace of the practices of open science has helped validate the importance of open scholarship in accelerating knowledge creation.

This article will examine two key aspects of Iowa State's approach to data sharing. First, we will explore Iowa State's Data Sharing Task Force (DSTF), which served as the collaborative center of the university's efforts to identify and implement changes to support and grow the sharing of data generated by the university's research activities. Second, we outline the development and implementation of an innovative data set disclosure process that allows campus data stakeholders to review and approve, or decline, the sharing of data sets submitted to Iowa State's data repository. The collaborative efforts by members of the DSTF and the group's outcomes have proved critical in advancing the awareness and sharing of research data at Iowa State. These efforts will provide helpful examples to libraries considering or actively pursuing similar goals on their own campuses.

Background

In early 2011, the United States' National Science Foundation (NSF) began to require researchers to share research data generated by funded research. This data sharing requirement triggered a cascade of questions and concerns among researchers and universities such as: will data sharing need to be reported?, how will an institution monitor compliance?, who will provide infrastructure?, what do we do with sensitive data?, how long does it need to be kept?, who will pay for it?, etc. Two years later, a directive from the Office of Science and Technology Policy set new requirements for all federal agencies with more than $100 million in research and development expenditures to develop plans to make federally funded

research, data and papers, publicly available (Holdren 2013). This directive heightened the existing concerns of universities regarding data sharing compliance and monitoring as other large federal research funding agencies, such as the U.S. The Department of Agriculture, Department of Energy, and National Institutes of Health would soon have policies similar to the one implemented by NSF in 2011.

The primary stakeholders involved in data sharing on a university campus include researchers, administrators, Information Technology (IT), and library. The concerns of researchers regarding the sharing of research data have been explored and covered in depth (Tenopir et al. 2011; Akers 2013; Kim and Zhang 2015; Digital Science et al. 2019). They can be summarized as concerns of additional burden (especially time and money), lack of reward or incentive, lack of training and support, and concerns of data misuse. Advances addressing these concerns have been made during the past decade but there remains a lot of room for further improvement. In general researcher attitudes have changed "from whether one can and should share data to when, where, and how a researcher can share their data" (Goben and Griffin 2019).

Campus IT, tasked with providing technology solutions and security to their campus, is poorly equipped to address data sharing and preservation needs. As Salo observes, digital preservation, a requirement of most data sharing policies, is not part of a Campus IT department's mission as "digital preservation goes far beyond mere provisioning of digital *storage*" (2020, 220). However, the security concerns of IT are directly correlated to data sharing as they are responsible for making sure that storage systems, local and remote, are secure and appropriate for the data housed there. In contrast the campus library is heavily involved in the dissemination of research results and has "extensive experience with selection, metadata, collections, institutional repositories, preservation, curation and access" (Erway 2013, 10). For this reason, campus libraries are often key members and leaders at their campuses for organizing and facilitating data sharing and preservation efforts. In this space Libraries and IT are partners as the library can design and manage technology systems but relies on the expertise of IT to comply with data security and storage policies and standards as well as infrastructure maintenance.[1]

Research data concerns of university administrators, such as those who work in offices that oversee and manage sponsored programs, intellectual property, research integrity and ethics, institutional review boards, and legal counsel (Erway 2013) are not well documented. Among the concerns institutional administrators face, the most important may be the liabilities sharing research data may expose for the university. This is not an easy question to answer given the complex nature of modern research regulation and compliance. Still other concerns could include: How will this affect the competitiveness of our researchers? What other policies or regulations govern the data? How does intellectual property law intersect? Is it legal or ethical to share data? How will agencies enforce or monitor compliance?

---

1   Some universities in the U.S. have organized the campus library under campus IT, which may reflect some of the interdependencies of these units.

How do we educate our researchers and students to responsibly share data? Do we invest in infrastructure or rely on others? These concerns must be addressed before university administrators can encourage, guide, or require their researchers to systematically share data.

Now, more than a decade after NSF's data sharing policy was first implemented, 18 U.S. federal agencies have data sharing policies in place (SPARC n.d.), multiple private research funders have implemented the same, and a growing number of professional societies and journals require that data supporting articles be available at the same time as the article.

Institutional profile

At Iowa State University, efforts are underway to advance open access and the sharing of research data. This work is in alignment with the university's mission statement, which states Iowa State should share the knowledge it creates to make Iowa and the world a better place (Iowa State University Office of the President n.d.). Iowa State is a medium-sized, land-grant, research university with over 35,000 enrolled students (U.S. Department of Education. Institute of Education Sciences, National Center for Education Statistics. n.d.). While the university does not have a medical school, it does host a veterinary medicine program and multiple federal research centers including Ames Laboratory, a U.S. Department of Energy national laboratory. In fiscal year 2021 the university received $154.8 million in federal research funding with the largest amounts coming from the Department of Agriculture, National Science Foundation, and Department of Energy (Office of the Vice President for Research July, 10 2021).

The Iowa State University Library actively advances open scholarship and the university's mission statement through its work developing and piloting new open access business models with scholarly publishers; leading campus and statewide initiatives to increase the creation and use of open educational resources; and expanding the support for sharing meaningful research data. The library's efforts to advance the sharing of research data are done in collaboration with key campus stakeholders. This collaborative approach has proven essential to the university's progress to date.

Data sharing at Iowa State

As a public land-grant university, Iowa State University operates under an expectation that the knowledge emerging from its research enterprise will be shared and put to practical use. However, it was not until open access began to gain momentum that the expectation of sharing scholarly research outputs beyond the academy was more fully put into practice.

The open access movement for publications began to take hold at Iowa State in 2012 when the library, in partnership with the Office of the Provost, launched the Digital Repository, Iowa State's first institutional repository. The Digital Repository

continues to have a high rate of participation from faculty and has helped raise open access awareness on campus. In 2017, the Iowa State Faculty Senate passed an open access resolution that recognized the many benefits of openly sharing research publications ("Resolution Adopting the Principles of Open Access of Research by the Iowa State University Faculty Senate" 2017). And more recently the University Library has taken a leadership role nationally in converting traditional, paywalled subscription agreements to open access agreements that allow Iowa State authors to retain copyright and make their articles free to read.[2] The sharing of research data, however, has moved more slowly and required a collaborative, cross-campus approach to achieve lasting progress.

This work accelerated in the fall of 2017, when a campus-wide group to advance the sharing of research data on campus was created. Leadership from Iowa State's then Vice President for Research, Dr. Sarah Nusser, was essential in the creation and success of the group. At the time, Dr. Nusser was serving in leadership roles in the Association of American Universities and Association of Public and Land-grant Universities initiative on Accelerating Public Access to Research Data.[3] She brought significant expertise, interest, and urgency to the local efforts at Iowa State.

Data sharing task force

The Data Sharing Task Force (DSTF) was jointly sponsored by the University Library, Office of the Vice President for Research, and the Office of the Chief Information Officer. The DSTF was charged with "collectively considering the set of actions and guidance needed to support researchers and the institution in providing public access to research data (Office of the Vice President for Research 2017). To achieve this goal, the make-up of the task force was very important. Membership included a targeted group of faculty, staff, and administrative stakeholders.[4] Key areas of representation included the Office of Intellectual Property and Technology Transfer (since renamed to "Office of Innovation Commercialization"), University Counsel, University Library, Office of the CEO, and the Office of the Vice President for Research. Broad and inclusive representation was essential for tackling the sort of cross-discipline, cross-department issues the DSTF needed to address.

The DSTF worked across four sometimes overlapping areas: policy, systems and services, research practice, and compliance. The Policy group focused on developing research data guidelines for campus. The guidelines were moved forward and adopted as official university policy (Iowa State University 2021) in early 2021. The new policy provides clarity around issues such as research data ownership, retention, transfers, and responsibilities.

---

2   A current list of active agreements is available at:
    https://open.lib.iastate.edu/open-access/agreements.

3   See https://www.aau.edu/accelerating-public-access-research-data for an overview of the initiative and related publications.

4   The task force's charge (Office of the Vice President for Research 2017) contains more information about task force members and deliverables.

The main accomplishment of the systems and services group was supporting the launch of an institutional data repository. Before the task force convened, the University Library had already reached the conclusion that the Digital Repository's current platform, bepress, was not a good solution for managing and sharing research data. The library's review of other software as a service (SAAS) platforms was incorporated into the DSTF under the systems and services group. This group guided initial policies, processes, and scope of the repository, which led to the soft launch of DataShare, an institutional instance of Figshare, in 2018. Since exiting its beta test period in 2019, DataShare has experienced steady growth. For example, 2020 had a 55% increase in the number of data sharing requests ($n=45$) compared to 2019 ($n=29$) though there was some decrease in 2021 ($n=35$).[5] The launch of DataShare also provided needed infrastructure to support the data outputs of citizen science projects such as the Lakeside Lab Dark Data project,[6] which seeks to make historic Iowa species records and specimen images publicly available.

The purpose of the research practice group was to benchmark and enhance data sharing practices at Iowa State. A 2018 National Science Foundation grant to a DSTF faculty member provided funding to investigate the practices and attitudes of Iowa State faculty around the sharing of research data. This work is ongoing and will provide important insight, highlighting areas for outreach and improvement.

The last focus area for the DSTF was compliance with federal, state, and local policies and regulations. This group's membership consisted of staff from the Office of Research Ethics, Legal Counsel, Office of Intellectual Property and Technology Transfer, and the Office of the Vice President of Research. Work by this sub-group led to the creation of the data set disclosure process, which will be more fully explored below.

The DSTF's work was largely successful when the group was sunsetted in 2020. Its final report included several recommendations that will further enhance data sharing at Iowa State. Among the recommendations was the creation of a campus-wide data portal to provide a one-stop shop for university data resources and services. Another recommendation was to create a new campus level initiative focused on accelerating open science and scholarship. As with open access, the culture change needed to advance data sharing should take place in the broader context of open science and scholarship practices.

Data set disclosure process

One of the major outputs of the DSTF was the creation of the data set disclosure process. This process was developed to address the data sharing concerns of university administrative offices that manage risk, liability, research integrity,

---

5   The authors suspect there may be a correlation to the continuing impact of the COVID-19 pandemic on non-medical research.

6   The data outputs of this project will begin to be available on DataShare in 2022. For an overview of the project see: https://www.zooniverse.org/projects/lbiederman/lakeside-dark-data.

research ethics, and intellectual property. The disclosure process acts as a filter: its purpose is not to restrict data sharing but to screen for potential conflicts and problems before data is shared. At this time, the disclosure process only applies to DataShare and is a separate process from DataShare's data curation process, which is managed by the library. Very few data set disclosures have been declined (*n*=2) and a similarly low number of data sets have needed alteration before being shared. To date, 121 disclosures have been reviewed and are associated with over 200 data sets[7] on DataShare.

Data set disclosures are submitted through an online form.[8] The form asks for information about the research such as author names, data subject matter and contents, funding sources, assurances and details regarding human and animal subjects, biohazards, select agents, export control, and Controlled Unclassified Information. Once submitted, the information is stored in an online spreadsheet in the cloud application Smartsheet (smartsheet.com). Different fields of the spreadsheet have been programmed to automatically email a review request to staff when the contents of a disclosure meet a set of defined criteria. For example, the Office of Research Ethic's associate director is alerted to a new disclosure if any of the questions pertaining to confidential or proprietary information, Controlled Unclassified Information, or export control are answered "yes" (Figure 1) while the Office of Intellectual Property and Technology Transfer is alerted as soon as a new disclosure is received.
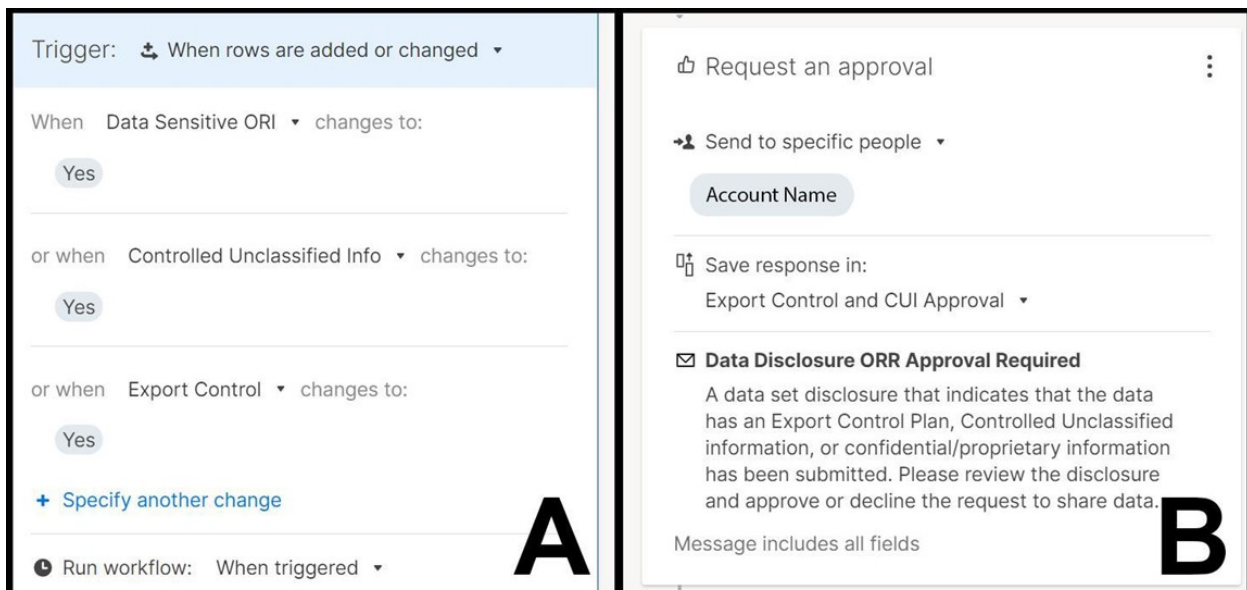

Figure 1: A screenshot of Smartsheet showing A) the conditions needed to trigger the approve/decline email message and B) the text of the email message and how the response is stored in the spreadsheet.

---

7   Disclosures may cover multiple "data sets" as they are presented on DataShare as the data files are organized dependent upon multiple factors including researcher preference. Thirty-six is currently the largest number of "data sets" represented by one disclosure.

8   The current form is available at:
    https://app.smartsheet.com/b/form/de43938cc6f34868930c619d568e2dca.

Up to three different university units may be involved in a data set disclosure review (Figure 2). Every disclosure is reviewed by the Office of Intellectual Property and Technology Transfer to make sure that the shared research does not have commercial potential for software or database licensing, patents, and other forms of intellectual property. Staff at the Office of Research Ethics may also review a disclosure if the researcher indicated that the research involved human or animal subjects, biohazards, Controlled Unclassified Information, and more on the form. Lastly, if the research is associated with Ames Laboratory then laboratory staff are also asked to review the disclosure. Once all applicable offices have
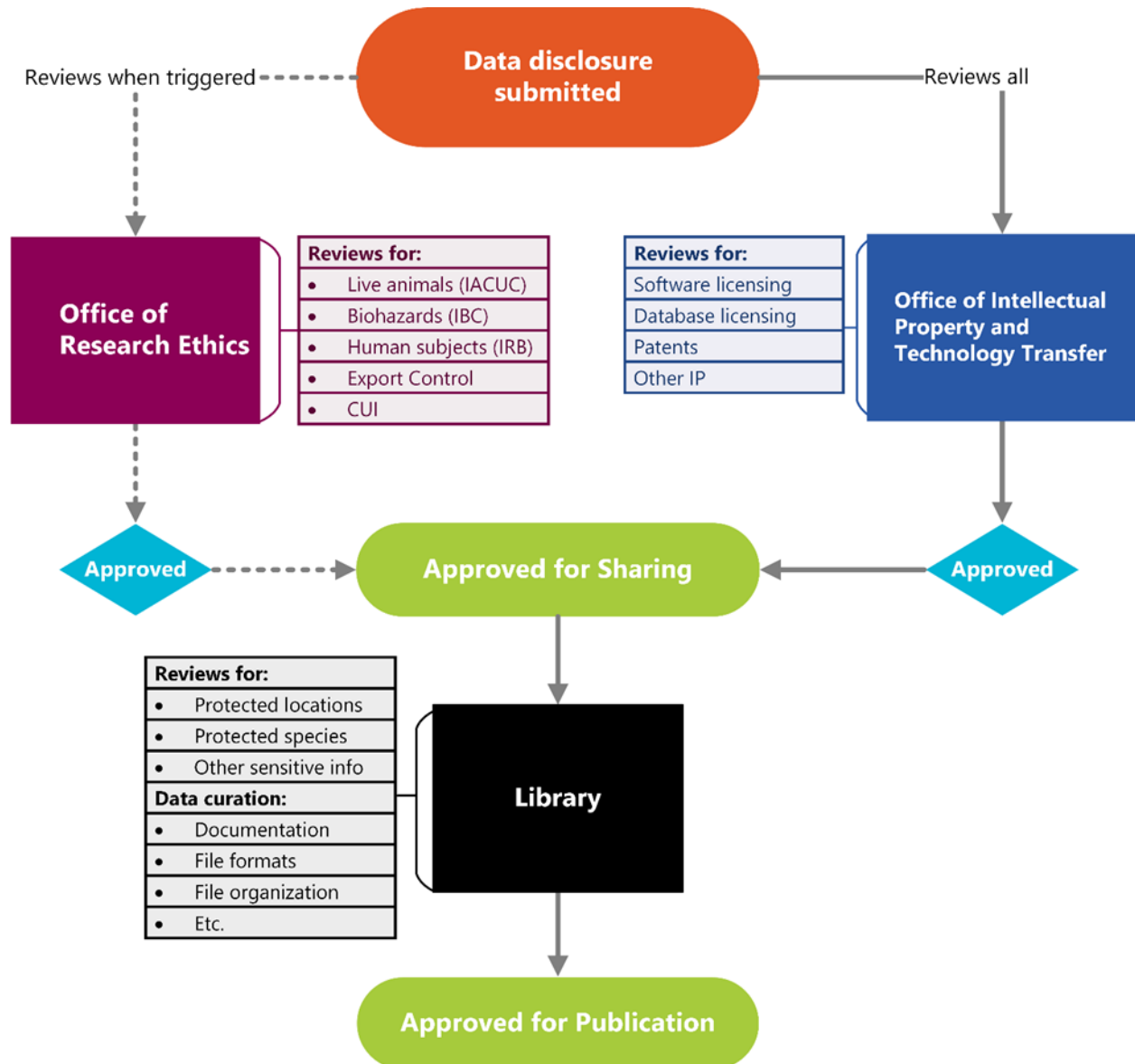
Figure 2: A diagram overview of the data set disclosure process showing how the Office of Research Ethics and Office of Intellectual Property and Technology Transfer, and the University Library work together to review data set disclosures and data sets.

reviewed and approved a disclosure the data is considered "approved for sharing". If any stakeholder involved in the review process "declines" a sharing request, then the data is considered "unapproved for sharing" and cannot be published on DataShare.

The library has three roles in the data set disclosure process. First, the data services librarian is responsible for managing the disclosure review workflow. Second, the library, as the manager of DataShare, has taken point on communicating the importance and purpose of the disclosure process to the rest of campus. Third, the curation of the data files by the library before publication acts as a final screening for potentially sensitive information before its publication (Figure 2).

As usage of DataShare grew, and the variety of data being submitted increased, four additional screening questions were added to the form to help identify data with sensitive subjects such as human-subject data and data about protected species and private or protected spaces. In conjunction with other information entered through the form the new questions help identify when the research may be subject to regulation or additional oversight but the data being shared isn't, and vice versa. Before this section was added "false alerts" were occasionally received that the data had sensitive information. For example, a research project that interviewed farmers and collected data about farms will have an IRB protocol associated with the research but if the data set being shared contains only information about farms and has no human-subject information then the data set does need to adhere to the research's associated IRB protocol.

When a disclosure indicates that the data includes sensitive subjects, or if the library finds potentially sensitive information in the data, the data services librarian and Office of Research Ethics director may review the data and either approve the data as-is or work with the researchers to further obscure or remove the sensitive information. This workflow is not considered a permanent or ideal solution as all stakeholders recognize that it would be preferable to have an expert trained in statistical analysis and advanced data obfuscation techniques on staff. However, DataShare was never meant to host all the university's publicly shared research data and, as an open access repository, data that need more screening or restricted access are referred to off-campus services that specialize in these functions. It is also worth mentioning that the most common potentially sensitive information found in reviewed data sets is information about locations, which may not be governed by research contracts or agreements. For example, it's not uncommon to find the name of a farm or the coordinates of a field site in a readme file or spreadsheet. Double-checking with the authors is usually enough to have any inappropriate information removed or obscured in these cases.

The data set disclosure process has given the campus a better understanding of the type, scale, and variety of research data being shared by its researchers. Of the 121 disclosures processed only one data set had commercialization and licensing potential and nearly all of the data sets flagged with having identifying or

sensitive subjects were able to be shared on DataShare after discussions with the authors and changes to the data. This shows that, to date, the majority of shared research data can be considered to be of "low" or "very low" risk as it was approved for publication on DataShare, an open access repository. The library's data curation process has also benefited from the disclosure process as it collects information such as funding numbers and author emails, which lets curators improve the quality of metadata records.

The data set disclosure process is unique to Iowa State University but is relevant to research institutions facing the same problems and concerns regarding the sharing of research data. It was developed by consensus and collaboration between units with very different goals and cultures united by a common goal: share research data as openly as possible, as responsibly as possible, and with minimum harm. The disclosure process was not a run-away success when it was launched. Researchers found the form confusing, laborious, and tedious. The workflow would bottleneck if researchers didn't answer emails and it was hard to tell when a review was complete. The library—overseeing the workflow but not understanding all of the stakeholder's perspectives and concerns—did not know how to effectively communicate its importance and value but was tasked with explaining and justifying it. It took approximately 18-months, and continuous communication and cooperation among the stakeholders to establish a more user-friendly form, a smart and effective workflow, and to learn to communicate the value of the disclosure process both as a mechanism for risk management and as a tool for improving data quality.

Conclusions

The success of the data set disclosure process, DataShare, and the DSTF moved the conversation at Iowa State from "is this a good idea?" to "how do we do this?". There are still many questions left to resolve, such as how DataShare and the disclosure process will scale up (i.e. staff, and infrastructure costs) and how it will be funded. Fortunately, the track record of collaboration and accomplishments established by the DSTF has provided a strong foundation of trust between the different research data stakeholders on campus. The establishment of productive working relationships and trust is perhaps the greatest accomplishment of the efforts at Iowa State. Efforts to expand the sharing of research data impact a range of university units and individuals and bringing these key stakeholders together to share their concerns and priorities is an essential step to advancing research data sharing.

Supplemental Content

Appendix
An online supplement to this article can be found at https://doi.org/10.7191/jeslib.2022.1224 under "Additional Files".

References

Akers, Katherine G. 2013. "Disciplinary Differences in Faculty Research Data Management Practices and Perspectives." *International Journal of Digital Curation* 8(2). https://doi.org/10.2218/ijdc.v8i2.263

Digital Science, Briony Fane, Paul Ayris, Mark Hahnel, Iain Hrynaszkiewicz, Grace Baynes, and Emily Farrell. 2019. "The State of Open Data Report 2019." *Digital Science*. https://doi.org/10.6084/m9.figshare.9980783.v1

Erway, Ricky. 2013. "Starting the Conversation: University-Wide Research Data Management Policy." *OCLC Research*. https://www.ncbi.nlm.nih.gov/pubmed/26298936

Goben, Abigail, and Tina Griffin. 2019. "In Aggregate: Trends, Needs, and Opportunities from Research Data Management Surveys." *College and Research Libraries* 80(7): 903. https://crl.acrl.org/index.php/crl/article/view/23610/30923

Holdren, John P. 2013. "Expanding Public Access to the Results of Federally Funded Research." Office of Science and Technology Policy. https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/ostp_public_access_memo_2013.pdf

International Science Council. 2021. "Opening the Record of Science: Making Scholarly Publishing Work for Science in the Digital Era." *International Science Council*. https://doi.org/10.24948/2021.01

Iowa State University. 2021. "Research Data." *Policy Library*, February 1, 2021. https://www.policy.iastate.edu/policy/researchdata

Iowa State University Office of the President. n.d. "Mission and Vision." Iowa State University Office of the President. Accessed August 13, 2021. https://www.president.iastate.edu/projects/mission

Kim, Youngseek, and Ping Zhang. 2015. "Understanding Data Sharing Behaviors of STEM Researchers: The Roles of Attitudes, Norms, and Data Repositories." *Library & Information Science Research* 37(3): 189–200. https://doi.org/10.1016/j.lisr.2015.04.006

Kupferschmidt, Kai. 2020. "'A Completely New Culture of Doing Research.' Coronavirus Outbreak Changes How Scientists Communicate." *Science*, February 26, 2020. https://doi.org/10.1126/science.abb4761

National Academies of Sciences, Engineering, and Medicine, Policy and Global Affairs, Board on Research Data and Information, and Committee on Toward an Open Science Enterprise. 2018. *Open Science by Design: Realizing a Vision for 21st Century Research*. Washington (DC): National Academies Press (US). https://doi.org/10.17226/25116

Office of the Vice President for Research. 2021. "FY21 Federal Research Funding Overview." https://www.research.iastate.edu/wp-content/uploads/2021/07/FY21_federal-funding-overview_FINAL.pdf

Office of the Vice President for Research. 2017. "Supporting Public Access to Research Data at Iowa State University." Iowa State University. https://www.research.iastate.edu/wp-content/uploads/2021/02/Iowa-State-University-Data-Sharing-Task-Force.pdf

"Resolution Adopting the Principles of Open Access of Research by the Iowa State University Faculty Senate." 2017. https://www-facsen.sws.iastate.edu/sites/default/files/uploads/16-17%20Docket%20Calendar/S16-24%20-%20Open%20Access%20at%20ISU%20-%20Senate%20resolution.pdf

Salo, Dorothea. 2020. "Is There a Text in These Data? The Digital Humanities and Preserving the Evidence." In *Reassembling Scholarly Communications: Histories, Infrastructures, and Global Politics of Open Access*, edited by Martin Paul Eve and Jonathan Gray, 215–28. MIT Press. https://doi.org/10.7551/mitpress/11885.001.0001

SPARC. n.d. "Browse Data Sharing Requirements by Federal Agency." *SPARC*. Accessed August 11, 2021. http://datasharing.sparcopen.org/data

Tenopir, Carol, Suzie Allard, Kimberly Douglass, Arsev Umur Aydinoglu, Lei Wu, Eleanor Read, Maribeth Manoff, and Mike Frame. 2011. "Data Sharing by Scientists: Practices and Perceptions." *PloS One* 6(6): e21101. https://doi.org/10.1371/journal.pone.0021101

The Royal Society. 2012. "Science as an Open Enterprise." DES24782. The Royal Society Science Policy Centre. https://royalsociety.org/-/media/policy/projects/sape/2012-06-20-saoe.pdf

U.S. Department of Education. Institute of Education Sciences, National Center for Education Statistics. n.d. "Iowa State University: Institution Profile." IPEDS Data Center. Accessed August 13, 2021. https://nces.ed.gov/ipeds/datacenter/institutionprofile.aspx?unitId=153603