FOSTERING DATA REUSABILITY:

Increasing Impact and Ease in Sharing and Reusing Research Data

Workshop Report and Action Steps

September 3, 2021

IOWA STATE UNIVERSITY

UNIVERSITY of VIRGINIA

NSF

FOSTERING DATA REUSABILITY:

Increasing Impact and Ease in Sharing and Reusing Research Data

Workshop Report and Action Steps

September 3, 2021

WORKSHOP PLANNING TEAM

NSF Co-Investigators

Sarah Nusser, Iowa State University and University of Virginia

Gizem Korkmaz, University of Virginia

Alyssa Mikytuck, University of Virginia and Randolph-Macon College

Facilitator

Joel E. Cutcher-Gershenfeld, Brandeis University and WayMark Analytics

Supporting Staff

Kristen Hanson, Iowa State University

Jeni Maiers, Iowa State University

Roger Suski, Iowa State University

ACKNOWLEDGEMENT

This workshop and material contained in this report was supported in part by the National Science Foundation under NSF EAGER Award 2039677. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

CITATION

Nusser, S. M., Cutcher-Gershenfeld, J. E., Mikytuck, A. M., Korkmaz, G. 2021. *Fostering Data Reusability: Increasing Impact and Ease in Sharing and Reusing Research Data – Workshop Report and Action Steps*. https://doi.org/10.5281/zenodo.5390123

Table of Contents

I. Executive Summary

For centuries, researchers have shared and discussed their findings in articles and convenings, seeking to advance scientific progress and enable scrutiny of their ideas. These practices are rooted in the principle of transparency in research, which promotes rigor and trust in the scientific process, increases the potential for equitable access and collaboration, and accelerates the pace of discovery and society impact. Today, the equivalent practice of open science (or scholarship) recognizes that a more complete suite of research outputs should be made publicly accessible whenever possible. Indeed, our global challenges and the drive to increase equity in access to knowledge urgently depend on the accessibility of data and associated documentation (e.g., provenance, metadata, software code, methodological detail) as the foundational building block to addressing urgent societal problems through the collaborative initiatives of researchers, professional societies, government agencies, private foundations, and international institutions.

This report tackles one of the most significant barriers to progress: the hurdles faced by researchers in producing and reusing publicly accessible research data, both in their research practice and in the surrounding ecosystem shaped by external stakeholders. The central challenge in high quality data sharing is to understand how researchers can increase the downstream value of shared data while reducing burden for both data producers and reusers. To galvanize progress on this problem, we organized a virtual workshop series on *Fostering Data Reusability: Increasing Impact and Ease in Data Sharing and Reuse* in June 2021. The series explored what context data reusers need to evaluate and appropriately reuse the data, identified practices that will improve data reusability and reduce the burden in producing and sharing research data, and used a stakeholder alignment approach to identify actions stakeholders could take to foster progress in reducing burden and increasing impact in data sharing and reuse.

Discussions were held among 35 researchers during two workshop sessions, followed by a third session where researchers were joined by 40 stakeholders, primarily representatives from societies and funding agencies. Collectively these conversations raised up several major themes that require actions by both researchers and stakeholders to accelerate progress in establishing research data sharing and reuse as an expected, well-executed, and rewarded norm across all fields.

The overarching theme of the conversations was the need to ***treat publicly accessible research data as a first-class research product***, whether that pertains to a researcher's practice, the organizational setting supporting the researcher, or the stakeholders supporting the broader research data sharing ecosystem. Recurring themes listed below highlight key issues and actions for researchers and stakeholders to accelerate progress in establishing effective practice and reward systems for data sharing and reuse. This report provides more detailed summaries of the workshop, discussions among participants, and research findings presented to contextualize the discussions.

Actions to improve research practice for data producers and reusers

1. Establish a manageable minimum standard for data sharing (e.g., documentation to include, persistent identifiers, citation information, repository choice, and other elements needed to make data findable, accessible, interoperable, and reusable (FAIR)) to guide data producer planning and practice and to address current shortfalls in materials provided to data reusers.
2. Develop training and toolsets that enable producers to:
   a. efficiently create complete and high-quality contextual documentation that promotes reproducibility and replicability based on the minimum data sharing standard,

   b. share data and software using data-proximate computing approaches via virtual environments that encapsulate the data, documentation and software needed to facilitate low-burden reuse of data, and
   c. evaluate the reusability of a shared data source bundled with its contextual documentation and supporting software packages.
3. Reframe research teams to recognize the essential role of data stewards in data sharing and reuse life cycles by including them in the primary investigative team.

Actions to expand organizational support for researchers and their teams

1. Develop services and templates (e.g., licenses, data use agreements (DUAs)) to ensure appropriate future data uses and data ethics are considered early in the study design and planning phase, particularly for data that need to be restricted (e.g., human subjects, proprietary).
2. Ensure researchers and institutions with fewer resources have access to necessary supporting services and computational infrastructure, e.g., by developing collaborative consortia to share services and infrastructure and reduce the burden of developing and maintaining these resources.
3. Develop organizational recognition and reward systems for researchers and data stewards involved in data sharing and reuse of shared data, including defining the metrics and approaches to document the quality and impact of shared data.
4. Embed data stewardship training earlier in the pipeline, including undergraduate education and STEM teacher professional development.

Actions for stakeholders to accelerate recognition of data as a first-class research product

1. Create and implement a consistent minimum approach to describing and citing data in publications that provides access to data, credit to data producers, and data citations to be tracked, ideally through a collaboration among society journals and other publishing venues.
2. Explore approaches to peer-reviewing a data source to support evaluation of data quality before publishing.
3. Create and promulgate mechanisms that allow future reuse of shared data to be tracked and assessed for quality and impact, with the goals of
   a. ensuring reusers give credit to producers and properly cite shared data used to generate new findings,
   b. providing evidence of accomplishments for promotion and award processes, and
   c. evaluating ROI on funder and publisher data sharing investments.
4. Frame funding opportunities to more completely articulate and support the costs associated with quality data stewardship throughout the data lifecycle, from planning, collection and processing to long-run curation and sharing of a data source.
5. Develop an agile approach to creating flexible community-developed standards for data variables, with an eye towards interoperability and cross-disciplinary collaborations.
6. Track progress in sub-fields and domains with increased sharing of data in order to identify broader impacts of data sharing and reuse.

II. About the Workshop

Motivation

For centuries, researchers have shared and discussed their findings in articles and convenings, seeking to advance scientific progress and enable scrutiny of their ideas. These practices are rooted in the principle of transparency in research, which promotes rigor and trust in the scientific process, increases the potential for equitable access and collaboration, and accelerates the pace of discovery and society impact.

Today, the equivalent practice of open science (or scholarship) recognizes that a more complete suite of research outputs should be made publicly accessible whenever possible. Open science involves a more complete form of transparency by sharing not only publications, but also research data, metadata, software code, methodological detail, and other products that foster collaboration and enable review and reuse of research outputs.

Researchers face many hurdles in adopting data sharing and other open science practices. For example, most researchers do not have a practice for preparing and sharing publicly accessible research data. They lack knowledge on how to make data both useful to others and of a quality that stands up to public scrutiny. In many cases, norms for a discipline or field have not been developed, and in some cases, there is a long-standing tradition of keeping data for one's own use; for data reusers, some fields do not value scholarly work based on others' data. Importantly, the evaluation and promotion process for researchers in academia and other sectors typically does not treat data as a first-class product, creating the absence of a reward for embracing good data sharing practices. While academic institutions and other organizations are establishing necessary infrastructure to support research data sharing, much of the infrastructure is underdeveloped relative to researcher needs. As a result, researchers also face significant time and financial constraints that add to the challenge of sharing data.

To better understand barriers and approaches to increasing the downstream value of shared research data and reducing researcher burden and to identify actions that could accelerate progress in establishing effective data sharing and reuse practices, we organized a National Science Foundation (NSF)-funded workshop series with researchers and ecosystem stakeholders on *Fostering Data Reusability: Increasing Impact and Ease in Data Sharing and Reuse* in June 2021.

Design

The design of the workshop series was informed by findings drawn from on research supported by NSF Award 2039677, *EAGER: Improving the Quality and Reducing the Burden of Producing and Reusing Publicly Accessible Research Data*. The research explores reusability of research data and practices to promote reusability and ease researcher burden, and was conducted by co-investigators Sarah Nusser, Gizem Korkmaz, Alyssa Mikytuck in collaboration with Joel Cutcher-Gershenfeld.

The *Fostering Data Reusability* workshop series was convened in three sessions held on June 14, 15, and 21, 2021. Workshop sessions were designed to address the following workshop goals:

1. Develop an understanding of the value propositions for data producers and reusers.
2. Identify promising research practices and approaches that reduce burden and increase impact in data sharing and reuse.

3. Identify actions researchers and external stakeholders could take to reduce burden and increase impact in data sharing and reuse.

To address these goals, we framed the workshop sessions first on understanding what is needed by another researcher to evaluate a data source's fitness for use in another context and appropriately reuse the data source; this is the "pull" from other researchers. With information on data sharing practices that will improve data reusability, we then focused on practices and infrastructure needed to reduce burden and increase impact for data producers; this is the "push" from researchers who share data.

Workshop participants included approximately 35 researchers from a diverse range of scholarly fields who produce and/or reuse publicly accessible research data, many of whom were nominated by professional societies. In the third session, researchers were joined by approximately 40 data sharing ecosystem stakeholders, primarily representatives from a diverse set of societies and funders. Session agendas and participant lists are available in Appendix A and B, respectively.

During the first two researcher sessions, participants were provided with background and research findings on the data sharing and reuse, followed by breakout groups to develop a draft shared vision of success and identify potential stakeholder value propositions. The remaining time was devoted to breakout conversations exploring challenges and potential solutions for six data sharing and reuse practice topics, including:

1. Re-user process: Discovering, selecting, and accessing shared data source(s)
2. Re-user process: Understanding and working with accessed data
3. Re-user process: Citation, credit, recognition for producer and user
4. Producer process: Planning for data sharing as part of study design
5. Producer process: Approaches to data capture and processing to reduce burden, improve consistency, promote quality
6. Producer process: Approaches to preserving and sharing data to promote impactful reuse and credit

Summaries of these conversations are in section III of this report and detailed notes are in Appendix D and E.

In third session with researchers and stakeholders, a summary of prior researcher discussions was presented, followed by researcher-stakeholder breakout group discussions on ten themes that arose from the first two sessions (descriptions in Appendix C). The topics were:

1. Community-based standards development
2. Community-level culture change
3. Credit and rewards for sharing and reuse
4. Data as a first-class research product
5. Data producer training modules
6. Data reuser training modules
7. Data sharing and reuse research ethics
8. Equity and inclusive data sharing and reuse
9. Journal data sharing policies
10. Planning to share and reuse data

Breakout group discussions were aimed at articulating a vision, forms of stakeholder alignment and misalignment, and potential action steps for advancing progress for the topic. Summaries of these sessions are in section IV of this report and detailed notes are in Appendix F.

III. Overview of Research Findings

To contextualize discussions among researchers and stakeholders, selected findings from research conducted under NSF Award 2039667 were presented. Material was derived from the data reusability literature, interviews conducted with researchers, and a stakeholder survey of scholarly society members.

Data Reusability

The concept of *data reusability*, or the ability of a new user to find, understand, evaluate, and appropriately incorporate the shared data in their analyses, was presented to explain context needed by data reusers and why they need it.

Faniel et al. (2010, 2017, 2019)[1] note that reusers need to execute several tasks in working with publicly shared research data. They need to understand: (1) whether data are relevant to their own study goals; (2) whether the data are credible, including the level of trust they have in the data and the researcher who produced the data, as well as how their research community perceives the data source; and (3) they need to be able to understand the data in detail, including its contents, allowable uses, quality and known limitations, and how the data can be appropriately analyzed or integrated with other data.

To support these tasks, reusers require a variety of contextual information. Reusers look to the reputation of the producer, their institution, and the repository to evaluate factors that affect their trust in the data. They explore other researchers' use of the data to further understand the credibility of the data and how it might be properly analyzed. Documentation on study goals, methodology, provenance, metadata, and other information about the data is essential to developing a deep understanding of a data source. Reusers also need guidance from a data use agreement or license to discern appropriate uses of the data and software code to properly read and analyze the data and understand how the data have been edited and processed.

Researchers who share data are often unaware of the complexity of tasks reusers face and what they need to appropriately reuse shared data. This disconnect needs to be addressed in considering how to increase the potential impact of shared data. While sharing a complete set of documentation with data is consistent with open science and transparent research practices, many producers perceive the level of documentation needed by reusers as added work, which begs for a practical approach to reducing data producer burden in planning and creating contextual documentation.

Researcher Interviews

To gain a deeper understanding of the practices, experiences, and perspectives of researchers who share and/or reuse publicly accessible research data, the NSF investigator team (Nusser, Korkmaz, Mikytuck) conducted virtual semi-structured interviews with 20 researchers during the winter of 2020–2021.

---

[1] Faniel, I.M. and Jacobsen, T.E. (2010). Reusing Scientific Data: how earthquake engineering researchers assess the reusability of colleagues' data. Computer Supported Cooperative Work 19(3):355-375.

Faniel, I.M. and Yakel, E. (2017). Practices do not make perfect: disciplinary data sharing and reuse practices and their implications for repository curation. In Curating Research Data, Volume 1: Practical strategies for your digital repository, p. 103-136.

Faniel, I.M., Frank, R.D. and Yakel, E. (2019), "Context from the data reuser's point of view," Journal of Documentation, Vol. 75 No. 6, pp. 1274-1297. https://doi.org/10.1108/JD-08- 2018-0133

Participating researchers were recruited from institutions involved in the *Accelerating Public Access to Research Data* (APARD) initiative led by the Association of American Universities (AAU) and Association of Public and Land-grant Universities (APLU). The researchers hailed from diverse scholarly fields and varied widely in their career stage, as well as in the size and research intensity of their institutions. Interview transcripts were analyzed using a grounded theory coding process to identify themes arising in researcher responses. Selected themes were shared with workshop participants.

Over half of researchers (11) both shared and used publicly accessible research data, one third only reused data (7), and one-tenth (2) only produced and shared data.

Researchers shared their motivations for sharing and reusing publicly accessible research during the interviews. Motivations for sharing and reusing data mirror key arguments for open science (Table 1), the presence of funder and publisher mandates, a desire to minimize cost and effort in their research practice, and expressed norms about data sharing and reuse in their field.

Table 1. Summary of researcher motivations for sharing and reusing data

Data Producers	Data Reusers
Accelerate discovery	Avoid burden, cost of data collection
Increase return on investment	Expanded inference base by integrating multiple studies
For the public good	Collaboration and network opportunities
Commitment to open science, transparency	Norm in field
Promote rigor and replicability	Support for open science
Required (journal, sponsor)	

Researchers also reported several challenges and concerns in producing and reusing publicly accessible research data (Table 2). Issues face by producers include reputational concerns, and lack of knowledge or resources needed to do proper job of data sharing. For reusers, concerns center on their inability to access, understand, and appropriately use shared data. They experience challenges when producers fail to respond to a data or information request and/or when there is a cost involved in accessing data. Numerous difficulties also arise when integrating multiple data sources that are not harmonized or do not follow standard coding and format schemes. Both producers and reusers experience barriers in the lack of incentives and rewards for data sharing and/or reuse in many fields and inadequate infrastructure access with restricted, complex, or large data sources.

Table 2. Summary of researcher challenges and concerns in sharing and reusing data

Data Producers	Data Reusers
Public scrutiny, exposure (mistakes, criticism)	Unresponsive producers, access costs
Ownership (being scooped, data hoarding)	Insufficient documentation to understand, evaluate, and use data
Effort to prepare data and documentation	Using complex or multiple data sources
Underdeveloped sharing practices, supporting culture	Lack of standardized approach to cite data, credit data producer
Inability to determine downstream impact (reuse, citation, credit)	Field unsupportive of secondary data analyses
Lack of incentives, rewards	Cyberinfrastructure for restricted, large data
Cyberinfrastructure for restricted, large data	

Society Member Stakeholder Survey

Themes from the researcher interviews were used to adapt a stakeholder survey instrument developed by Joel Cutcher-Gershenfeld and WayMark Analytics to elicit information on the perspectives and experiences of researchers engaged in data sharing and reuse activities. The Data Sharing and Reuse Stakeholder Survey was conducted in collaboration with 11 professional societies (see graphic below[2]) who expressed a willingness and ability to survey their members. Four societies provided their members access to the survey via email (AEA, AGE, ASA, ASPB), a newsletter link (APHA, ESA, AAS), or sent a link or email to a subset of their members (IACM, LERA, PON, AoM/CM).

Figure 1. Societies participating in the Data Sharing and Reuse Stakeholder Survey

[2] American Geophysical Union (AGU), American Economic Association (AEA), American Statistical Association (ASA), American Scoeity of Plant Biologists (ASPB), American Public Health Association (APHA), American Astronomical Society (AAS), International Association for Conflict Management (IACM), Labor and Employment Relations Association (LERA), Program on Negotiation at Harvard Law School (PON), Academy of Management/Conflict Management (AoM/CM).

We received 1,596 completed responses during the period late April–late May 2021, which were analyzed by Cutcher-Gershenfeld. Most survey questions involved a statement with a rating of how important it is (0 = not important, 10 = very important) and how easy (0 = very challenging, 10 = very easy) it is to accomplish, or a statement with an agreement scale (0 = strongly disagree, 10 = strongly agree).

The breakout of responses by the participating societies is noted in Figure 1. Of the 1,467 respondents who produced and/or reused data, 62% of researchers both shared and used publicly accessible research data, 22% only reused data, and 16% only produced and shared data. This and many other findings were remarkably similar to the data obtained from researcher interviews.

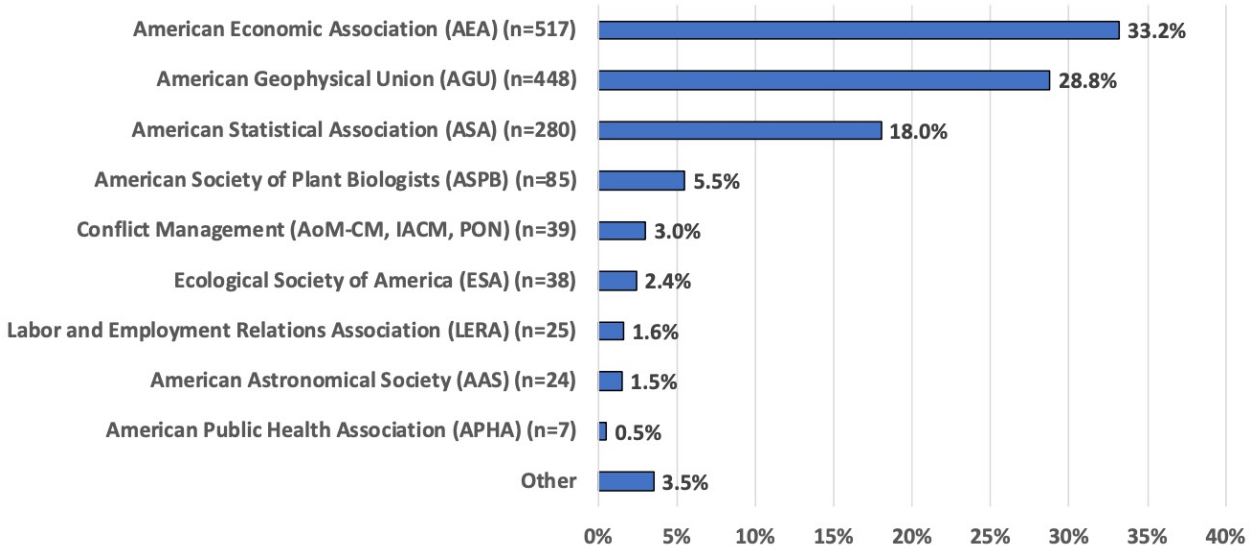


Figure 2. Response levels by participating society (n = number of completed responses, % = percent of completed responses)

To set the stage on producer and reuser perceptions, results were presented on knowledge and tools to reduce burden in sharing data, rewards for sharing data, and support for data reuse; all perceived as areas require significant improvement (Table 2). Society-specific responses to these questions are quite similar, as they were for nearly all questions. The responses in Table 3 point to a major gap in the knowledge and training on the software and tools that can reduce the burden on researchers, a major gap in the rewards and incentive for sharing research data, and deep challenges in building a culture that is supportive of the reuse of research data.

Table 3. Selected society member perceptions on data sharing and reuse

Statements	Result
There is sufficient knowledge and training in my primary field or discipline on software and tools that can reduce burden in producing and documenting research data	23% agree or strongly agree (7-10)
The tenure, promotion, and rewards in my organization recognize and value researchers for sharing research data	10% agree or strongly agree (7-10)
Building a culture in my fields and disciplines that is supportive of the reuse of research data	90% important or very important (7-10) 58% challenging or very challenging (0-3)

A sample of 208 open-ended responses to questions about motivations was coded into categories. The largest motivator by far was the advancement of science (32%), followed by a commitment to the public good (16%). Other motivations included transparency and rigor (12%), efficiency and effectiveness (of data reuse, 11%), availability and quality of data (10%), and requirements (9%).

Stakeholder survey respondents shared the three greatest bottlenecks they experience in their workflow with data (Figure 2). Data producers identified funding to support data sharing work and curating data for use by others as the two greatest barriers, while data reusers cited gaining access to data from researchers or repositories. Other highly ranked barriers included working with restricted data and access to expertise to work with data. All of these barriers were raised multiple times in researcher discussions during the workshop.

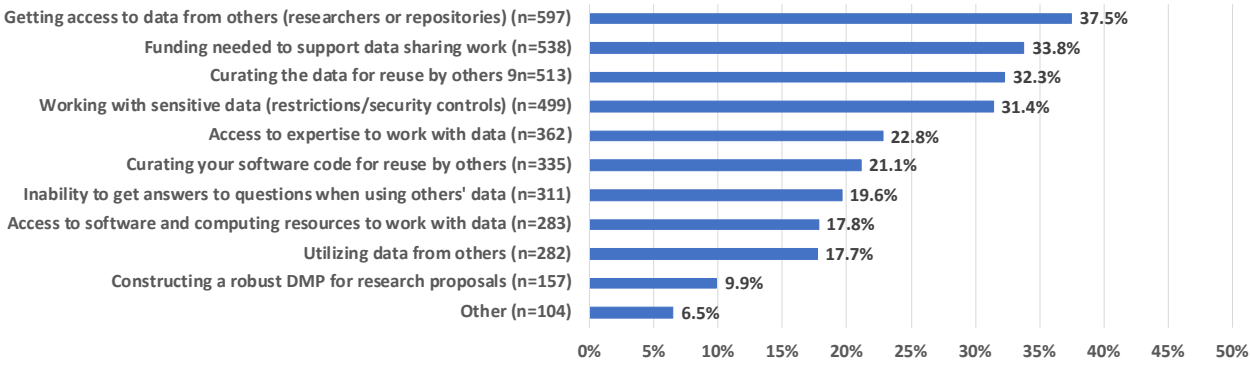


Figure 3. Greatest bottlenecks of sharing/reusing data (n = number of completed responses, % = percent of completed responses)

IV. Shared Vision of Success and Stakeholder Value Propositions

After discussing research findings, researchers brainstormed ideas for a shared vision of success in data sharing and reuse and developed potential stakeholder value propositions for actors in the research data sharing and reuse ecosystem. The outcome of these processes is presented below, with notes provided in Appendix D.

Shared Vision of Success

To guide discussions, researchers developed a draft shared vision based on their ideas of what success would look like for data sharing and reuse. The vision below further refines their draft based on workshop notes.

> *Early in the planning and design of research studies, researchers from all career stages, organizations, fields, and disciplines anticipate the sharing and reuse of research data with a full suite of contextual documentation (e.g., study goals, methods, metadata, software, provenance). During their research process, researchers benefit from practices that reduce burden and promote quality and appropriate use of shared data. Organizational and institutional leaders, publishers, funders, and other stakeholders support the sharing and reuse of research data with enabling infrastructure, aligned rewards and recognition, and documented impacts. As a result of competent and supported data sharing and reuse, scientific discoveries and societal impacts are accelerated.*

Stakeholder Value Propositions

The stakeholder alignment approach used in the workshop involves understanding the perspectives of different stakeholders in the data sharing and reuse ecosystem and identifying areas of alignment and misalignment. To frame the research practice recommendations, participants were introduced to the stakeholder alignment model in Figure 4. This model outlines the connections between value identification for stakeholders, the value proposition for each stakeholder, and delivery of value to stakeholders. In the workshop context, research findings and workshop participants' lived experience all contributed to *value identification*, or the values held in data sharing and reuse. The discussion involved consolidating expressed values by researchers and other stakeholders into *value propositions* for each stakeholder group. Specifying the value propositions is an important step in achieving *value delivery*, or actions that will deliver value to stakeholders in the data sharing and reuse ecosystem.

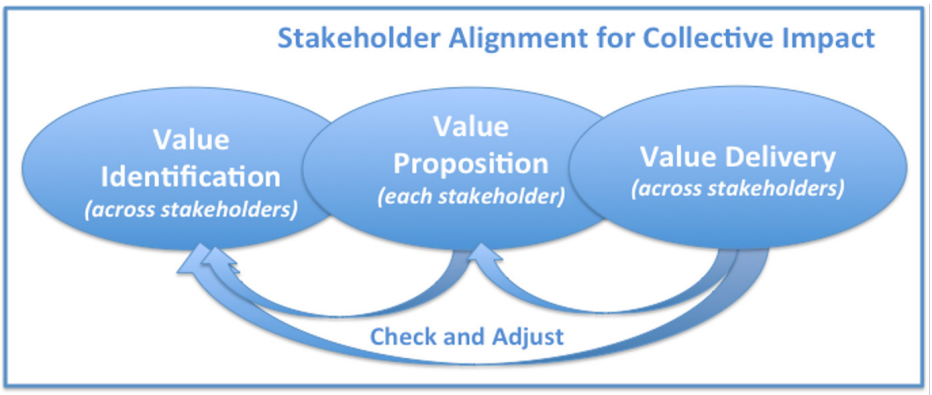


Figure 4. Stakeholder alignment process. (Adapted from: Earll Murman, Tom Allen, Kirkor Bozdogan, Joel Cutcher-Gershenfeld, Hugh McManus, Debbie Nightingale, Eric Rebentisch, Tom Shields, Fred Stahl, Myles Walton, Joyce Warmkessel, Stanley Weiss, & Sheila Widnall. 2002. *Lean Enterprise Value: Insights from MIT's Lean Aerospace Initiative*, New York: Palgrave/Macmillan.)

Data Producers

Data producers will be more motivated to share research data for reuse when they are provided with coherent and consistent approaches and protections, resources, and credit for their efforts.

Data Reusers

Data reusers will be more motivated to reuse research data when they are provided with properly curated data that enables effective and appropriate use of data and assistance when questions arise.

Funders

Funders will be more motivated to support the sharing and reuse of research data when there is evidence that it will increase return on their investment and advance their funding and organizational priorities.

Societies

Societies will be more motivated to support the sharing and reuse of research data when there is evidence that it will advance society priorities of advancing in their field and providing benefits to members.

Publishers

Publishers will be more motivated to support the sharing and reuse of research data when there is evidence that it will advance publisher business models and recruit high-quality scholarship to their publishing venues.

Research Data Services

Research data services in libraries and other organizations will be more motivated to support the sharing and reuse of research data when it presents them with high-impact work and opportunities for professional growth and recognition.

General Public

The public will be more motivated to support the sharing and reuse of research data when they trust the scholarly process and there is an understanding and appreciation of its broader societal benefits.

Note that these value propositions are overlapping, but not identical. Success in the sharing and reuse of research data will involve approaches that deliver on all of these value propositions.

V. Research Data Sharing and Reuse Practice Recommendations

The majority of discussions in workshop sessions 1 and 2 focused on examining six phases in the data sharing and reuse life cycle. For each phase, research findings were presented outlining the challenges and successes experienced by researchers for that phase (listed in Appendix E).

We began with three phases of the data reuse process to better understand the struggles and success factors for reusers, who hold the key to creating downstream impact for research data. We then turned to three phases of the data production and sharing process and how that could be crafted to reduce burden and increase downstream impact. Below is a summary of researcher recommendations from the discussions.

DATA REUSE PRACTICE RECOMMENDATIONS

Topic 1: Discovering, Selecting, and Accessing Shared Data Source(s)

A. **Standards for Sharing Data:** Develop a manageable minimum standard for data sharing (e.g., expected documentation, minimum standard for data dictionaries and ontologies, data variable standards that promote interoperability, desirable repository characteristics).
B. **Incentives to Prepare Data that Fosters Data Discovery:** Provide incentives, in the form of sufficient funding for data preparation and outreach for data producers and others, that concretely provide tools for data discovery and access.
C. **Training and Toolkits to Provide Replicable Frameworks:** Develop training and toolkits that enable data producers to develop documented data sources (e.g., data, study goals, methodology, metadata, software code) with replicable frameworks, particularly if not in an R1 institution.
D. **Awareness and Motivation to Value All Data:** Mechanisms to increase understanding of the value of all research data, including null results.

Topic 2: Understanding and Working with Accessed Data

A. **Community-developed Standards and Formats:** Create mechanisms for developing community-based standards for data and documentation (data, metadata, schema, formats, etc.) to facilitate and reduce burden in data reuse and sharing.
B. **Data-proximate Computing:** Producers should store and/or package data using approaches (e.g., cloud storage, containers) that enable researchers to compute without having to move the data.
C. **Funding for Producing Reusable Data:** Provide adequate funding to support for the time required for producers to make data reusable and interoperable (aligned with funder value propositions and data management plans (DMPs)).
D. **Career Development for Data Workers:** Appreciate and provide professional development to data workers who have the expertise and skills to support quality data sharing and reuse.
E. **Training on Data Usage:** Ensure researchers of all career stages have access to professional development on all aspects of data usage.
F. **Data Sharing Assessment Tools:** Create methods for assessing data to understand what researchers do and don't know about the data or how shareable data might be.

Topic 3: Citation, Credit, and Recognition for Producer and User

A. **Funding for Data Curation:** Provide adequate funding for quality data curation, which is as important as the data creation process.
B. **Data Citation Expectations:** Data providers should specify appropriate citation format and other aspects of appropriate use (e.g., DUA, license).
C. **Consistent Identifiers:** Encourage and expect a consistent set of identifiers and their related metadata (e.g., DOIs and PIDs for data, researcher, institution, funder).
D. **Journal Requirements:** Work with journals to develop consistent approaches and require authors to give cite data and credit data producers.
E. **Recognition:** Institutions should develop approaches to recognizing good data sharing and reuse practices in promotion, tenure, and career paths.

DATA SHARING PRACTICE RECOMMENDATIONS

Topic 4: Planning for data sharing as part of study design

A. **Equity in Resources for Data Sharing:** Increased equity among large and small institutions by considering consortia for developing and supporting infrastructure and resources needed for quality data sharing.
B. **Advance Repository Consultations:** Consult with institutional or domain repository before launching into a proposal and DMP to better anticipate what will be needed for curating and depositing data that will maximize reuse by others.
C. **Data Governance:** Include the Institutional Review Board (IRB) and other offices in advance planning and education on the types of data that will be generated and used to ensure appropriate use and confidentiality protections (e.g., other forms of restricted data such as proprietary data, issues of data governance, DUAs, licensing).
D. **IRB Standards:** facilitate greater standardization of IRB practices and expectations across institutions and other organizations.
E. **Understand Potential Data Reusers:** Talk to potential data reusers in advance of beginning the project to anticipate what they will need.
F. **Community-based Practice and Data Standards:** Utilize community standards for practices and data to extent possible, anticipate need for data integration.
G. **Funders and DMPs:** Add more "teeth" to the DMP evaluation in the review process, including choice of repositories and documentation plans that facilitate downstream reuse, and require updates on DMP as part of the reporting process.

Topic 5: Approaches to data capture and processing to reduce burden, improve consistency, promote quality

A. **Organizational Leadership:** Publishers, societies, and funding organizations should support data publication with guidance on domain standards and via incentives (e.g., teeth in DMPs, consideration in tenure and promotion).

B. **Flexible Community Standards:** Establish standards developed through communities (domain specific) for data to be recoverable and machine readable, with a focus on "good enough" or minimum requirements.
C. **Training and Resources:** Provide institutional support for training in data sharing and reuse practice (e.g., leading practices and tools, study planning for data sharing, ethics training) for researchers of all career stages and with an understanding that a mix of formats will be needed (e.g., components of courses, regular "booster shots," modular lessons, asynchronous, just-in-time modules).
D. **Modernize IRB Approaches:** Research offices should increase awareness of and work with IRBs to appreciate the need for compliance with data sharing mandates and balanced share requirements against risk (e.g., evaluate templates in use by IRBs, such as provisions to destroy data after five years).

Topic 6: Approaches to preserving and sharing data to promote impactful reuse and credit

A. **Templates to Simplify Sharing:** Provide guidelines, templates, and checklists to help data producers evaluate compliance with requirements (e.g., DMP, confidentiality) and best practices in sharing data (e.g., open-source options, selecting repository).
B. **Lowering Barriers:** Develop templates to assist data producers (in addition to what comes from IRBs).
C. **Promotion and Tenure:** Organizations (all sectors) should develop policies and culture that meaningfully recognizes data at all stages of a researcher's career.
D. **Review DMPs:** Review DMPs to ensure compliance with commitments stated in DMPs.
E. **Curriculum for Data Ethics and Handling:** Teach researchers (faculty, staff, students, scientists) about ethical and other considerations around data and open-source sharing, which promotes increased data literacy in broader society.
F. **Data Management Careers:** Actively develop and reward a pipeline of professionals in research data handling and management in support of research.
G. **Support FAIR Data and Reproducibility/Replicable Practices:** Establish a closer dialog between funders, publishers, and repositories to ensure publication of data and materials for reuse, reproducibility, and replicability are required and support producers/authors in meeting these requirements (e.g., with code checking, removing limits on methods section).

The full listing of recommendations reveals the scale and scope of change needed to accelerate discovery in fields and disciplines through the sharing and reuse of research data. This listing indicates that there isn't a simple one-time policy change or single intervention that will accomplish the task. Rather, it will take a sustained change management process, with many stakeholders contributing in various ways. It was in that spirit that the third workshop focused on the stakeholder alignment needed for action and results.

VI. Stakeholder Alignment and Action Plans

The third session brought an opportunity for researchers and stakeholders to consider action plans for going forward. The shared vision, value propositions, and researcher recommendations for data sharing and reuse practices developed during the first two workshop sessions were used to identify 10 issues for discussion in this third session:

1. Community-based standards development
2. Community-level culture change
3. Credit and rewards for sharing and reuse
4. Data as a first-class research product
5. Data producer training modules
6. Data reuser training modules
7. Data sharing and reuse research ethics
8. Equity and inclusive data sharing and reuse
9. Journal data sharing policies
10. Planning to share and reuse data

Relevant research findings were also identified for each topic to provide context and set the stage for discussions among researchers and stakeholders.

With this information, each researcher-stakeholder breakout group used a standard template (Appendix F) to engage in a process that refined the issue and defined a vision, and then identified alignments and misalignments among stakeholders and what areas would lend themselves to near-term actions that would promote progress in increasing impact and ease in sharing and reusing research data. Detailed breakout group notes and related research findings for each topic are outlined in Appendix F. Below is an edited version of what was generated during the workshop, with key points compressed and some integration of points across topics.

The original template format is based on change management tools when launching a change process, with the aim of summarizing the initiative on a single page. This can then be shared with key stakeholders to literally get them "on the same page" with respect to strengths on which to build and challenges to be addressed. This summary format can then be used in outreach and education when introducing the change initiative. It is hoped that this will be helpful on follow-on initiatives that build on the workshop sessions.

Topic 1. Community-based standards development

<table>
<tr><td>Issue</td><td colspan="2">For many fields, the lack of appropriate standards (e.g., data, formats, metadata, expected content, vocabularies) or awareness of those standards hampers effective sharing and reuse, particularly when working across disciplines.</td></tr>
<tr><td>Vision</td><td colspan="2">There is an agile way to create and update standards that enable harmonization and interoperability among data sources. As a result, core uniform standards and specific subject-based standards are established and valued by researchers, who are in turn incentivized to use them. The community focuses on adoption as a higher priority than optimization of standards.</td></tr>
<tr><td rowspan="2">Stakeholders</td><td colspan="2">Starting: researchers, funders, community builders, publishers, service providers, public
Added: service providers that facilitate standards development and sharing</td></tr>
<tr><td>Alignment strengths:
• The importance of standards to facilitate understanding and use of data at later time</td><td>Alignment challenges:
• Priority given to standards that enable FAIR data, reduce costs, and increase efficiency in data reuse</td></tr>
<tr><td>Prposed Action Steps</td><td colspan="2">1. Identify exemplar standards and convene discussions of how and why certain standards become widely used.
2. Identify specific problems/tasks where standards are lacking and which communities want to develop these standards.
3. Identify an agile approach to creating and updating standards to foster greater creation, updating, and use by researchers and others.</td></tr>
<tr><td>Potential Partners</td><td colspan="2">NISO, DataCite, CHORUS, AGU (and other societies), NASEM/BRDI, Funders</td></tr>
</table>

Topic 2. Community-level culture change

<table>
<tr><td>Issue</td><td colspan="2">Many researchers do not support data sharing and reuse and/or do not have supportive communities of practice or research cultures that recognize the importance and potential impact of data sharing and reuse.</td></tr>
<tr><td>Vision</td><td colspan="2">Data sharing and reuse is valued, expected, practiced, supported, and rewarded widely across disciplines, domains, and stakeholders. Broad support for data sharing and reuse is expressed through adequate resources to engage in quality data sharing and reuse (e.g., time, money, personnel, training, sustained and shared infrastructure, professional recognition of open science). In addition, all stakeholders care about and are willing to pay for future use of data.</td></tr>
<tr><td rowspan="2">Stakeholders</td><td colspan="2">Starting: researchers, funders, community builders, publishers, service providers, public
Added: business (benefit from shared data), Congress (influence funding agency budgets)</td></tr>
<tr><td>Alignment strengths:
Accelerating advancement of scientific knowledge
Importance of shared data that can have a high impact on societal needs</td><td>Alignment challenges:
Data producers not recognizing the value or investing the effort in data sharing relative to what data users need to reuse the data
Costs associated with long-term sustainable infrastructure versus funder priorities for new research, appropriate business models to support data sharing</td></tr>
<tr><td>Proposed Action Steps</td><td colspan="2">1. Evaluation: develop recommended wording or approach for documenting contribution, quality, impact of shared data (societies), promote to academic administration.
2. Infrastructure: develop sustainable models for trusted repositories (public/private partnerships).
3. Knowledge: develop best practice documents and workshops on how to create supportable data management approaches.
4. Pipeline: create data handling and stewardship training for undergraduates and STEM teachers.</td></tr>
<tr><td>Potential Partners</td><td colspan="2">Societies, ORFG (nonprofits supportive of open science)</td></tr>
</table>

Topic 3. Credit and rewards for sharing and reuse

<table>
<tr><td>Issue</td><td colspan="2">Many researchers do not receive credit or rewards for sharing data. Researchers struggle to document the impact of shared data because they cannot easily identify when and how their data is used by others. In some fields, reuse of data has low value relative to generating original data.</td></tr>
<tr><td>Vision</td><td colspan="2">Producing and sharing data is viewed as comparable to publishing a research paper. This is manifested through the capacity for shared data to stand on its own without a publication; via published papers that promote data sources that are valued for immediate, additional and ongoing uses; and because data is consistently recognized as a contribution that receives credit comparable to publications. Because data are valued, DOIs are assigned to all data outputs to improve discovery and facilitate credit and citation when data is reused. Funders support this vision by requiring and promoting open science and data sharing practices and rewarding data reuse.</td></tr>
<tr><td rowspan="2">Stakeholders</td><td colspan="2">Starting: researchers, funders, community builders, publishers, service providers, public
Added: disciplinary research data service and repositories, open-source repositories and software/system providers (e.g., Docker, R), facilitators (e.g., CHORUS, DataCite)</td></tr>
<tr><td>Alignment strengths:
• The importance of advancing science</td><td>Alignment challenges:
• Prioritization of funding to support for the effort required to do a quality job of preparing and sharing data
• Disagreement on the need to shift incentives from allowing data hoarding towards rewarding data sharing</td></tr>
<tr><td>Proposed Action Steps</td><td colspan="2">1. Explore potential funding for post-grant data curation and early career researcher training.
2. Identify leaders in the field who can actively promote data sharing.
3. Develop standards for data generation, sharing and reuse via cross-pollination across disciplinary societies.</td></tr>
<tr><td>Potential Partners</td><td colspan="2">FASEB (convenings), AGU (automated citation credit), National Academies (research leaders), ORFG (incentives, practices), STM, Crossref, DataCite, CHORUS</td></tr>
</table>

Topic 4. Data as a first-class product

<table>
<tr><td>Issue</td><td colspan="2">Data are increasingly a primary output of research process. In some cases, key data sets are surrounded by communities of practice and associated articles, but these are exceptions. More often, data is just viewed as an input into research products (articles, presentations, etc.).</td></tr>
<tr><td>Vision</td><td colspan="2">Data are treated as having a life story, emanating and continuing to evolve from a flow-based research and reuse process. Sharing and reuse is championed by individuals, and researchers prioritize quality data sharing in response to requirements and being rewarded with credit and recognition. Data are so valuable that talks, posters, and articles begin with links to data and metadata that can be immediately accessed. Researchers are supported with role-based training and sufficient resources and services to make data sharing a seamless part of the research process. This results in shared data that is machine actionable and interoperable with other data sources.</td></tr>
<tr><td rowspan="2">Stakeholders</td><td colspan="2">Starting: researchers, funders, community builders, publishers, service providers, public
Added: disciplinary data service providers/repositories; open source repositories, software, system providers; other services (CHORUS, DataCite, et al.); institutions (departments, research development office, IP office)</td></tr>
<tr><td>Alignment strengths:
The value provided in selected fields when data sharing is expected/valued and credit is received
The need for increasing institutional support for research data services (e.g., evolution of libraries)</td><td>Alignment challenges:
Researchers required to share data, but are not rewarded for sharing data
Current funding models versus resources needed to support necessary components that promote sharing over the full data life cycle
Conflicting messages within academic institutions and the lack of equity across institutions
Societies supporting open science versus their non-open access journals</td></tr>
<tr><td>Proposed Action Steps</td><td colspan="2">1. Clarify credit researchers want for data sharing, meaningful metrics for quality and impact
2. Identify and share current examples and practices where data sharing value/credit align
3. Develop peer-review of data, including stand-alone data sources
4. Funders: create interactive tools for robust data management planning, include evidence of well-shared data in prior research expectations and in ranking proposals, give new awardees a data sharing how-to guide</td></tr>
<tr><td>Potential Partners</td><td colspan="2">Societies, funders, AAU/APLU</td></tr>
</table>

Topic 5. Data producer training modules

<table>
<tr><td>Issue</td><td colspan="2">There is a need for training modules to help data producers at all career stages build skills that reduce burden in their research process and increase the impact of their shared data. Training needs include accessing and working with data, preparing metadata and documentation, understanding the ethics and data protection, and giving credit.</td></tr>
<tr><td>Vision</td><td colspan="2">Best practices for producing and sharing data is so well understood and internalized by researchers that ad hoc specialized training is not necessary. As a result, researchers are more innovative in their discoveries because they can readily access a complete array of past and current research and they have saved time previously spent in data discovery and interoperability. In addition, researchers are so motivated to be open and share their research outputs that requirements are no longer needed.</td></tr>
<tr><td rowspan="2">Stakeholders</td><td colspan="2">Starting: researchers, funders, community builders, publishers, service providers, public
Added: research data service providers (e.g., libraries)</td></tr>
<tr><td>Alignment strengths:
Support for scientific progress, transparency, and rigor</td><td>Alignment challenges:
Availability and/or discoverability of relevant training material versus actual need for training
The desire for data to be viewed as a first-class product and the reuse of shared data versus lack reward/impact on career for shared data</td></tr>
<tr><td>Proposed Action Steps</td><td colspan="2">1. Identify experts on training development.
2. Develop a coalition to work on training.
3. Design and develop modules prioritizing researcher needs .
4. Convene to develop instructor guides and train trainers.
5. Build into graduate education in research methods.</td></tr>
<tr><td>Potential Partners</td><td colspan="2">Societies (AGU, FABBS, AStatA), institutions (libraries and other research data services), funders, related efforts (Data Management Training Clearing House, FSCI, RDA, carpentries, openscapes)</td></tr>
</table>

Topic 6. Data reuser training modules

<table>
<tr><td>Issue</td><td colspan="2">Data reusers at all career stages need a way to build skills to ensure appropriate and effective reuse of research data. Training needed includes finding/accessing data, importing data, manipulating and integrating data, statistical analysis and visualization, and ethical responsibilities, including understanding terms of reuse and working with IRB/human subjects data.</td></tr>
<tr><td>Vision</td><td colspan="2">Modules, books, and other training material and workshops at different levels (beginner to more advanced) exist, are vetted and are easily available for different fields and learning styles, and researchers have supported time to learn data access and analysis tools such as R and Python (e.g., summer salary support). Communities of practice exist to support one another with shared tools and data analysis resources.</td></tr>
<tr><td rowspan="2">Stakeholders</td><td colspan="2">Starting: researchers, funders, community builders, publishers, service providers, public</td></tr>
<tr><td>Alignment strengths:
Importance of open access to suite of quality small-bite training modules
Need for consortium/grants to develop and assess training materials</td><td>Alignment challenges:
Availability and/or discoverability of relevant training material versus actual need for training</td></tr>
<tr><td>Proposed Action Steps</td><td colspan="2">1. Inventory and assess existing training modules (see list in Appendix E, Topic 6 notes)
2. Prioritize gaps, plan initial modules
3. Identify content creators/curators and instructors to develop/deliver modules
4. Identify vehicles for dissemination and promotion of modules
5. Assess shared training modules
6. Build into graduate education in research methods</td></tr>
<tr><td>Potential partners</td><td colspan="2">Societies, Funders (federal, nonprofit)</td></tr>
</table>

Topic 7. Data sharing and reuse ethics

<table>
<tr><td>Issue</td><td colspan="2">Data sharing involves ethical considerations around data sharing and reuse, including protecting confidentiality, appropriately reusing of data both statistically and in compliance with data use agreements or licenses, ensuring producers and data contributors are credited, and related equity issues (e.g., in access, resources, sensitivities). Many researchers lack awareness and understanding of data ethics and associated issues, as well as appropriate practices to support ethical data sharing and reuse.</td></tr>
<tr><td>Vision</td><td colspan="2">Researchers enthusiastically share and reuse data using leading practices that are well-established and support ethical considerations. Data reusers are knowledgeable about how to appropriately use shared data, and producers are respected and credited by reusers. The rights of study participants and other entities that contribute data are also respected and contributors support sharing of their data.</td></tr>
<tr><td rowspan="2">Stakeholders</td><td colspan="2">Starting: researchers, funders, community builders, publishers, service providers, public
Added: study participants and other data contributors</td></tr>
<tr><td>Alignment strengths:
The need for ensuring data security and privacy when sharing and reusing data, and providing credit to producers and contributors</td><td>Alignment challenges:
The tension between security/privacy and requirements for open access
The importance of ethics versus the availability of guidance, services and infrastructure to address ethical considerations</td></tr>
<tr><td>Proposed Action Steps</td><td colspan="2">1. Develop material that outlines what needs to be considered in the research planning process to embed ethics and good data sharing/reuse practices.
2. Develop guidelines that facilitate appropriate sharing and reuse of sensitive data.
3. Clarify the connection between openness/access and protection/security and identify mechanisms for sharing/access that respect both.</td></tr>
<tr><td>Potential Partners</td><td colspan="2">Societies, funders</td></tr>
</table>

Topic 8: Equity and inclusive data sharing and reuse

<table>
<tr><td>Issue</td><td colspan="2">Issues of equity and inclusion are woven throughout data sharing and reuse. For example, some research institutions lack resources to provide full support for data sharing, some repositories require extra resources to access, some populations are studied without access to their own data, and some research study approaches contain bias.</td></tr>
<tr><td>Vision</td><td colspan="2">Researchers everywhere are able to find and access data and documentation (even if restricted) from studies that have been designed to avoid bias in data collection and analysis. Researchers have the knowledge and skills to appropriately use the data, even if they are restricted. Resources needed to share and access/reuse data are shared across institutions to facilitate equity across institutions and fields.</td></tr>
<tr><td rowspan="2">Stakeholders</td><td colspan="2">Starting: researchers, funders, community builders, publishers, service providers, public
Added: study participants and their populations, under-resourced and underrepresented groups</td></tr>
<tr><td>Alignment strengths:
The need for inclusion and consideration of underrepresented voices in decision making about data sharing and reuse
The need for equity in study design and analysis approaches</td><td>Alignment challenges:
Underpopulated pathways for underrepresented groups to pursue research data and computing careers versus need for more representation in these fields
Current practice versus data contributor’s autonomy over their data, the right to privacy, and the right to be forgotten
Community control over its own data</td></tr>
<tr><td>Proposed Action Steps</td><td colspan="2">1. Audit current data sharing processes and their impacts on equity.
2. Identify underrepresented voices and invite them to participate in developing an action plan to address issues from audit.
3. Construct and ask stakeholders to review an action plan to address equity issues.
4. Draft policy based on refined action plan that can be shared across scientific disciplines.</td></tr>
<tr><td>Potential Partners</td><td colspan="2">Federal funding agencies (NSF, NIH), nonprofit funders (Gordon and Betty Moore Foundation)</td></tr>
</table>

Topic 9: Journal data sharing policies

<table>
<tr><td>Issue</td><td colspan="2">Journals vary in how they handle data associated with articles. For example, differences arise in whether and how data are shared with the article (e.g., paywall repositories, allowing authors to make data available upon request, no policy), whether and how data are formally described and cited (e.g., whether links are available to immediately access the data), and whether the journal enforces data access when data sharing is expressed as “data available on request”. As a result, data reusers experience difficulties in finding, accessing, and using shared data, and data producers fail to receive the credit and recognition they deserve.</td></tr>
<tr><td>Vision</td><td colspan="2">Data shared through journal articles are readily accessible and well-documented, allowing researchers to replicate and build on published research. In addition, publication data are documented with a common set of persistent identifiers and citation formats that enable producers to receive credit and provide users with immediate access to the data or detailed information about data if restricted. Journal policies are consistent with and promote open science practices and FAIR data.</td></tr>
<tr><td rowspan="2">Stakeholders</td><td colspan="2">Starting: researchers, funders, community builders, publishers, service providers, public</td></tr>
<tr><td>Alignment strengths:
High-quality scholarship published in journals that arises from rigorously collected and prepared research data</td><td>Alignment challenges:
Business model for publishers versus need to make data and associated article publicly accessible
Differences in perceived roles for enforcing ethical and appropriate access to research data supporting publications</td></tr>
<tr><td>Proposed Action Steps</td><td colspan="2">1. Develop COPDESS (Coalition for Publishing Data in the Earth and Space Sciences)-like process for other disciplines, which supports FAIR data principles.
2. Develop community-based discipline-specific practices in data sharing, building toward data standards across journals in a discipline.
3. Develop tools and resources for researchers (e.g., template for documentation to accompany data archive for publication) to clarify expectations for researchers in sharing data associated with an article.
4. Identify characteristics for a data editor role.</td></tr>
<tr><td>Potential Partners</td><td colspan="2">Federal agency funding to support disciplinary convenings, NASEM Roundtable on Incentives for Open Science, STM and publisher groups</td></tr>
</table>

Topic 10: Planning to share and reuse data

<table>
<tr><td>Issue</td><td colspan="2">Producers experience many challenges in preparing data, with proper documentation, when they have not planned ahead. Too often, they do not know what materials need to be shared with data to support its effective and appropriate reuse or the most effective way to reduce burden and promote quality and consistency in data acquisition and processing.</td></tr>
<tr><td>Vision</td><td colspan="2">Researchers value data sharing and are motivated to start early in planning their data products and sharing strategy. In addition, data capture that facilitates sharing is easy and readily embedded in the research process has become the default practice. As a result, data are collected and documented with a view towards future use and potential users can easily discover, access, evaluate, and appropriately use shared data. Costs are lower for both producers and reusers and easily supported by research funding and institutions.</td></tr>
<tr><td rowspan="2">Stakeholders</td><td colspan="2">Starting: researchers, funders, community builders, publishers, service providers, public</td></tr>
<tr><td>Alignment strengths:
Shared interest in rigor, replicability, reproducibility, and transparency, as well as public trust in research findings</td><td>Alignment challenges:
Extra work required in planning and executing work required to produce and share quality data and documentation vs. available funding, infrastructure and knowledge to embed best practices in study planning and data handling
Requirement to develop DMPs versus actual detail required to manage data, the DMP's minimal role in the downstream research process, and lack of approach to updating plans</td></tr>
<tr><td>Proposed Action Steps</td><td colspan="2">1. Inventory knowledge and gaps in best practices for data sharing that can form the basis for a planning approach for the research process that results in making data and documentation FAIR and publicly accessible.
2. Adapt findings from proposed action step 1 to specific situations arising with disciplines, restricted data, and other settings that require special procedures.</td></tr>
<tr><td>Potential Partners</td><td colspan="2">Societies, funders (nonprofit, federal), NIST Research Data Framework (RDaF), data sharing advocates (e.g., RDA, CODATA, GO-FAIR), data services and repositories (see notes), academic advocates (AAU, APLU, ARL)</td></tr>
</table>

Appendix A: Workshop Series Agenda

FOSTERING DATA REUSABILITY:
Increasing Impact and Ease in Sharing and Reusing Research Data

RESEARCHER WORKSHOP SESSIONS 1 AND 2:
JUNE 14 AND 15, 2021

RESEARCHER–STAKEHOLDER WORKSHOP SESSION 3:
JUNE 22, 2021

ALL DAYS: 12:00-3:30 ET / 11:00-2:30 CT / 10:00-1:30 MT / 9:00-12:30 PT

WORKSHOP GOALS

- Develop an understanding of the value propositions for data producers and reusers.
- Identify promising research practices and approaches that reduce burden and increase impact in data sharing and reuse.
- Identify actions researchers and external stakeholders could take to reduce burden and increase impact in data sharing and reuse.

IOWA STATE UNIVERSITY

UNIVERSITY of VIRGINIA

SPONSOR

This workshop is funded in part by NSF award 2039677.

FACILITATOR

Joel Cutcher-Gershenfeld – Brandeis University

ORGANIZERS

Sarah Nusser – Iowa State University
Gizem Korkmaz – University of Virginia
Alyssa Mikytuck – University of Virginia

SUPPORT

Kristen Hanson – Iowa State University

For more information: **https://bit.ly/2SWrrjy**

AGENDA

Session 1 – Researcher Discussions: June 14

All Times in Eastern Daylights Savings Time

12:00 p.m. **Welcome, Goals and Overview**

Welcome – Martin Halbert, National Science Foundation
Workshop Goals – Sarah Nusser, PI
Workshop Structure and Process – Joel Cutcher-Gershenfeld, Facilitator

12:15 p.m. **Overview of Research Findings – Qualitative and Quantitative**

Big Picture from ISU-UVA Project – Sarah Nusser
Big Picture from Society Surveys – Joel Cutcher-Gershenfeld

12:35 p.m. **Discussion**

12:45 p.m. **Break**

12:55 p.m. **Elements of a Value Proposition for Data Sharing and Data Reuse**

Brainstorm Potential Elements of a Data Reuse Value Proposition

1:35 p.m. **Break**

1:45 p.m. **Data Relevant for Breakout Groups 1-3**

Relevant Interview and Survey Findings for Each Breakout Group – Gizem Korkmaz, Joel Cutcher-Gershenfeld

2:05 p.m. **Data Reuse: Parallel Breakout Groups**

Topic 1: Re-user Process: Discovering, Selecting and Accessing Shared Data Source(s)
Topic 2: Re-user Process: Understanding and Working with Accessed Data
Topic 3: Re-user Process: Citation, Credit, Recognition for Producer and User

Group Process:

- Introductions
- Individual Stories of Success and Frustration
- 3-5 Leading Policy and Practice Recommendations
- Implementation Considerations (Stakeholders, Resources, etc.)

3:05 p.m. **Group Reports**

3:25 p.m. **Concluding Comments**

AGENDA

Session 2 – Researcher Discussions: June 15

All Times in Eastern Daylights Savings Time

12:00 p.m. **Welcome, Overview and Check-in**

12:30 p.m. **Data Relevant for Breakout Groups 4-6**

Relevant Interview and Survey Findings for Each Breakout Group – Alyssa Mikytuck, Joel Cutcher-Gershenfeld

12:50 p.m. **Break**

1:00 p.m. **Data Sharing: Parallel Breakout Groups – Leading Practices and Barriers**

Topic 4: Producer Process: Planning for Data Sharing as Part of Study Design
Topic 5: Producer Process: Approaches to Data Capture and Processing to Reduce Burden, Improve Consistency, Promote Quality
Topic 6: Producer Process: Approaches to Preserving and Sharing Data to Promote Impactful Reuse and Credit

Group Process:

- Introductions
- Individual Stories of Success and Frustration
- 3-5 leading Policy and Practice Recommendations
- Implementation Considerations (Stakeholders, Resources, etc.)

2:00 p.m. **Break**

2:10 p.m. **Group Reports**

2:30 p.m. **Return to Edited Versions of the Value Propositions for Data Sharing and Data Reuse**

2:45 p.m. **Overall Success Vision**

Brainstorming Potential Elements of a Data Sharing and Reusability Success Vision

3:05 p.m. **Prepare for Session 3**

3:25 p.m. **Concluding Comments**

AGENDA

Session 3 – Researcher-Stakeholder Discussions: June 22

All Times in Eastern Daylights Savings Time

12:00 p.m. **Welcome, Goals, and Overview**

12:05 p.m. **Summary of Researcher Sessions**

Research Highlights – Sarah Nusser, PI and Joel Cutcher-Gershenfeld, Facilitator
Stakeholder Propositions – Joel Cutcher-Gershenfeld
Vision for Success – Joel Cutcher-Gershenfeld
Researcher Recommendations – Gizem Korkmaz and Alyssa Mikytuck, co-PIs

12:35 p.m. **Discussion of Researcher Sessions**

Breakout Discussions
Plenary Discussion

1:10 p.m. **Break**

1:20 p.m. **Breakouts: Implementation Principles, Overview of Topics, and Group Process**

Community-based standards development
Community-level culture change
Credit and rewards for sharing and reuse
Data as a first-class research product
Data producer training modules
Data reuser training modules
Data sharing and reuse ethics
Equity and participation in data sharing and reuse
IRB alignment on sharing and reuse
Journal data sharing policies
Planning to share and reuse data

2:30 p.m. **Break**

2:40 p.m. **Group Reports**

3:15 p.m. **Next Steps**

3:25 p.m. **Concluding Comments**

Appendix B: Workshop Participants

First Name	Last Name	Institution/Organization
RESEARCHERS		
Karen	Adolph	New York University
Nan	Bernstein Ratner	University of Maryland
Nazli	Bhatia	University of Pennsylvania
Claire	Bowen	Urban Institute
Sylvie	Brouder	Purdue University/ASA
Judy	Brusslan	California State University, Long Beach
Atul	Butte	UCSF
Mine	Cetinkaya-Rundel	Duke University
Susan	Chen	Virginia Tech
Renato	Corbetta	University of Alabama at Birmingham
Theresa	Crimmins	USA National Phenology Network, University of Arizona
Tom	DeSutter	North Dakota State University
Stephen	Diggs	UCSD
Shawn	Dorius	Iowa State University
Stephen	Eubank	University of Virginia
Rick	Gilmore	Penn State University
Hannah	Gunderman	Carnegie Mellon University
Martin	Halbert	National Science Foundation
Bryan	Heidorn	University of Arizona
Adina	Howe	Iowa State University
Christine	Kirkpatrick	San Diego Supercomputer Center, UC San Diego

First Name	Last Name	Institution/Organization
Carolyn	Lawrence-Dill	Iowa State University
David	LeBauer	University of Arizona
Kerstin	Lehnert	Columbia University
Chris	Lenhardt	University of North Carolina at Chapel Hill
Serghei	Mangul	University of Southern California
Matthew	Mayernik	National Center for Atmospheric Research (NCAR)
Rachel	McCrary	National Center for Atmospheric Research (NCAR)
Seth	McGinnis	National Center for Atmospheric Research (NCAR)
Joshua	Rosenbloom	Iowa State University
Sandra	Tang	University of Michigan
Sherry	Towers	Towers Consulting, LLC
Kevin	Tyle	University at Albany State University of New York
Kathy	Yeater	ASA-CSSA-SSSA
STAKEHOLDERS		
Tom	Arrison	National Academy of Sciences
Juliane	Baron	FABBS
Lauren	Cadwallader	PLOS
Deborah	Cai	Temple University
Marie	Connolly	University of Quebec in Montreal
Regina	Davis Moss	American Public Health Association
Erik	Gjesfjeld	John Templeton Foundation
Silvia	Glick	Harvard Law School
Daniel	Goroff	Alfred P. Sloan Foundation

First Name	Last Name	Institution/Organization
Susan	Gregurick	National Institutes of Health
Corinna	Gries	Environmental Data Initiative
Robert	Hanisch	NIST
Brooks	Hanson	American Geophysical Union
Angie	Hunter	American Chemical Society
Scott	Kahn	Helmsley Charitable Trust (Consultant)
Sallie	Keller	University of Virginia
Varsha	Khodiyar	Springer Nature
Frank	Krause	FASEB
Donna	LaLonde	American Statistical Association
Leah	McEwen	Cornell University
Emily	Miller	Association of American Universities
August	Muench	American Astronomical Society
Belinda	Orland	American Heart Association
Deepa	Prasad	Alex's Lemonade Stand Foundation
Howard	Ratner	CHORUS
Kacy	Redd	APLU
Carly	Robinson	Department of Energy
Doug	Schuster	National Center for Atmospheric Research (NCAR)
Jerry	Sheehan	Office of Science and Technology Policy
Emily	Smith	University of Illinois at Urbana-Champaign
Tobin	Smith	Association of American Universities
Shelley	Stall	American Geophysical Union

First Name	Last Name	Institution/Organization
Ann	Stapleton	USDA
George	Strawn	National Academy of Sciences
Greg	Tananbaum	Open Research Funders Group (ORFG)
Crispin	Taylor	American Society of Plant Biologists
Carolyn	Vasques Scalera	American Sociological Association
Valda	Vinson	Science
Ron	Wasserstein	American Statistical Association
WORKSHOP TEAM		
Joel	Cutcher-Gershenfeld	Brandeis University and WayMark
Gizem	Korkmaz	University of Virginia
Alyssa	Mikytuck	University of Virginia
Sarah	Nusser	Iowa State University
REUSABILITY PROJECT/DATA SCIENCE FOR THE PUBLIC GOOD TEAM MEMBERS		
Emily	Kurtz	University of Virginia
Aditi	Mahabal	University of Virginia
Akilesh	Ramakrishna	University of Virginia
Jack	Studier	Iowa State University
Adisak	Sukul	Iowa State University
Sonyta	Ung	Iowa State University
Saul	Varshavsky	Iowa State University
David	Walker	Iowa State University
Tiancheng	Zhou	Iowa State University

Appendix C: Researcher-Stakeholder Acton Plan Topics

1. Community-based standards development

The lack of appropriate standards (e.g., formats, expected content, vocabularies) for many fields hampers effective sharing and reuse, particularly when working across disciplines. Interview and survey results say that this is important but hard do to. What initiatives will advance community-based actions to create or update standards that enable work within and across fields?

2. Community-level culture change

Many researchers do not have supportive communities of practice/research cultures that recognize the importance and potential of data sharing and reuse. Interview and survey results say that this is important but hard do to. What can we do to build community among practitioners in ways that are supportive of data sharing and reuse practices? This includes challenging embedded assumptions that undercut sharing and reuse, such as proprietary views of data.

3. Credit and rewards for sharing and reuse

Many researchers do not receive credit or rewards for sharing data. Researchers struggle to document the impact of shared data because they cannot easily identify when and how their data is used by others. Reuse of data is not always valued the same as generating original data. Interview and survey results say that this is important but hard do to. What actions are needed to provide evidence of data reuse and impact? What can stakeholders do to encourage and reward the sharing and reuse data?

4. Data as a first-class research product

Data are increasingly a primary output of research process. In some cases, key data sets are surrounded by communities of practice and associated articles, but these are exceptions. More often, data is just viewed as an input into research products (articles, presentations, etc.). What actions are needed to ensure that data is valued a first-class product of research projects?

5. Data producer training modules

Data producers need a way to build skills that reduce burden in their research process and increase the impact of their shared data. This is true for researchers at all stages of their careers. Interview and survey results say that this is important but hard do to. Single-point modules (30–45 minutes) can be particularly impactful. What are examples of modules that should be developed for data producers?

6. Data reuser training modules

Data reusers need a way to build skills to ensure appropriate and effective reuse of research data. This is true for researchers at all stages of their careers. Interview and survey results say that this is important but hard do to. Single-point modules (30-45 min.) can be particularly impactful. What are examples of modules that should be developed for data reusers?

7. Data sharing and reuse ethics

Data sharing involves ethical considerations around data sharing and reuse. For example, this includes protecting confidentiality, ensuring appropriate reuse of data, and enabling early career researchers to make appropriate use of data before sharing. What actions are needed to codify ethical guidelines, ensure awareness, and track ethical compliance in data sharing and reuse?

8. Equity and participation in data sharing and reuse

Issues of equity and inclusion are woven throughout data sharing and reuse. For example, some research institutions lack resources to provide full support for data sharing, some repositories require extra resources to access (e.g., proprietary repositories, secure data centers such as the Federal Statistical Data Research Centers), and some populations are studied without access to their own data. What initiatives will advance equitable access to data, support for inclusive data sharing and reuse, and propel broader impacts in society through data sharing and reuse?

9. Journal data sharing policies

Journals vary in how they handle data associated with articles (e.g., paywall repositories, allowing authors to make data available upon request, no policy, etc.). Journals may or may not support researchers in depositing research data in disciplinary or other repositories. What models will support public access to data associated with journal publications while minimizing data sharing costs for journals?

10. Planning to share and reuse data

Producers experience challenges in preparing data, with proper documentation, when they have not planned ahead. Too often, they do not know what materials need to be shared with data to support its effective and appropriate reuse. Interview and survey results say that this is important but hard do to. What actions can be taken to promote advance planning for the sharing and reuse of data?

Appendix D: Success Vision and Value Proposition Brainstorm Notes

BRAINSTORMED ELEMENTS FOR A SHARED VISION OF SUCCESS

Researchers brainstormed elements for a shared vision of success at the beginning of the workshop session 1, with an additional discussion at the end of session 2. These include:

- Data sharing is a rigorously executed and expected part of the research process across fields and disciplines.
  - An integral part of the undergraduate experience; early use in graduate careers; a requirement for graduate students to do replication studies with data.
- Increased infrastructure for automation of meta-data generation, built into the work of instrument developers (lab notebooks, etc.).
- IRBs presume that data will be shared, unless specified otherwise.
  - A menu of data sharing options that researchers can choose from.
- Read a paper that uses a data set with a DOI that is a live link and within 5 minutes be able to load into "R" and use it (reusable in real time)—the scientific paper of the future.
- Creation and curation of data and sharing is valued as much as the clever analysis of the data.
- Attention to the way this work is valued and the gendered nature of the work (in comparison to data analytics, for example).
- Data scientists are equal partners at the table with domain scientists. It is the science of data, not just data work.
- Instead of cutting grants by "x" percent, augment grants for data.
- More common use of previously published data in statement of hypotheses and methods.
- Expanding knowledge through interoperable data.
- Note that not all data can be shared and save, requires standards for sharing and retention.
- No more need for workshops like this based on progress achieved.

VALUE PROPOSITION SUMMARIES AND ASSOCIATED BRAINSTORMING

Data Producers

Data producers will be more motivated to share research data for reuse when they are provided with coherent and consistent approaches and protections, resources, and credit for their efforts. These include:

- clear standards for data and standardization of sharing practice.
- education and training (at all career stages) to
  - increase understanding of underlying issues associated with data sharing (culture) and
  - build practices for meeting the standardized approaches (practice).
- support systems and tools for sharing provided by universities (or other employer) and service providers such as funding agencies, professional societies, publishers, repositories, libraries, and other key stakeholders.

- protections against improper use of the data.
- requirements for data sharing by funders and publishers.
- resources commensurate with the time and cost involved with producing data.
- mechanisms for users to provide and producers to receive credit for reuse.
- mechanisms for coordination with IRBs in order to support future data sharing and reuse.
- recognition for sharing and reuse reflected in promotion, tenure, and career advancement for faculty, staff, and students (as appropriate) and for professionals in other sectors.
- feedback on frequency of use and on downstream impacts of data reuse.
- increased use of shared data in data-informed policy decisions.
- demonstrated progress against key societal challenges.

Data Reusers

Data reusers will be more motivated to reuse research data when they are provided with properly curated data that enables effective and appropriate use of data and assistance when questions arise. These include:

- tools and methods to find the right data, including training (wikis).
- supporting information (metadata, codebooks, methods, provenance, intended use, software, etc.) in order to make proper use of the data.
- information and evidence on the quality and reliability of the data.
- opportunities for the data to be interoperable with other data sets (e.g., through use of standards/ standardized approaches).
- formats for the data that do not depend on proprietary software, systems, or other additional costs.
- recognition for reuse reflected in promotion, tenure, and career advancement for faculty, staff, and students (as appropriate) and for professionals in other sectors.
- increased reuse of data in data-informed policy decisions.
- demonstrated progress against key societal challenges.

Funders

Funders will be more motivated to support the sharing and reuse of research data when there is evidence that it will increase return on their investment and advance their funding and organizational priorities. This includes:

- compliance with federal mandates and community standards.
- fostering cultures of sharing and reuse among relevant communities of practice.
- indications that sharing and reuse will result in cost savings or economies of scale.
- indications that sharing and reuse expands interdisciplinary approaches and collaborations, fosters innovation and accelerates scholarly discoveries.
- indications that sharing and reuse broaden participation and involvement with science and technology.
- increased use of shared data in data-informed policy decisions.
- demonstrated progress against key societal challenges.

Societies

Societies will be more motivated to support the sharing and reuse of research data when there is evidence that it will advance society priorities of advancing in their field and providing benefits to members. This includes:

- accelerating progress and coordination across relevant fields and disciplines.
- increasing coordination and integration among communities of practice (e.g., to develop standards and approaches for data sharing, foster resources such as community repositories or integrated data resources).
- creating an inclusive culture that values data sharing and reuse at all stages of careers
- reducing the burden on members through clear standards, education, and training for all career levels.
- compliance with federal mandates and community standards.
- shifting the culture toward increased collaboration.
- increased use of shared data in data-informed policy decisions.
- demonstrated progress against key societal challenges.

Publishers

Publishers will be more motivated to support the sharing and reuse of research data when there is evidence that it will advance publisher business models and recruit high-quality scholarship to their publishing venues. This includes:

- increased citations leading to increased impact factors.
- increased individual and library subscriptions.
- reduced cost and complexity through partnerships with repositories and other resources.
- increased appeal to authors for submissions (and reduced risk of authors turning away).
- compliance with federal mandates and community standards.
- emergence of new business models associated centered on data as a first-class research product (with articles attached)
- demonstrated progress against key societal challenges.

Research Data Services

Research data services in libraries and other organizations will be more motivated to support the sharing and reuse of research data when it presents them with high-impact work and opportunities for professional growth and recognition. This includes:

- research data services as a necessary and supported part of the infrastructure for science
- more reliable and longer-term funding to ensure data services are sustainable.
- opportunities to collaborate with leading experts on some of the most challenging software, middleware, computation, storage, security, and other problems facing the profession.
- collaboration in standards development—community of practice among service providers—and buy in on the use of standards
- supporting policies by publishers and funders

- community buy-in on services resulting in impacts on the sciences
- evidence of network effects through interoperable and extensible software and services ("n" services generating "n-squared" capabilities).
- clear standards and guidance on commercialization and open source considerations.
- rewarding career paths within universities, government facilities, and industry.
- indications that the sharing and reuse will result in cost savings or economies of scale.
- indications that the sharing and reuse will foster innovation and accelerate scientific discoveries.
- demonstrated progress against key societal challenges.

General Public

The public will be more motivated to support the sharing and reuse of research data when they trust the scholarly process and there is an understanding and appreciation of its broader societal benefits. This includes:

- demonstrated progress against key societal challenges.
- indications that the sharing and reuse will result in return on investment via cost savings or economies of scale.
- indications that the sharing and reuse will foster innovation and accelerate scientific discoveries.
- indications that the sharing and reuse will broaden involvement with science and technology.
- increased trust, appreciation, and critical thinking about science and scientists.
- Increase in data-informed policy decisions
- increased access to data by lay users/citizen scientists.

Appendix E: Findings Presented for Research Practice Topics

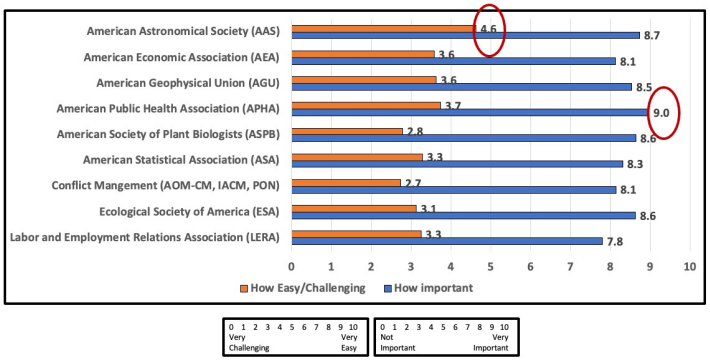
TOPIC 1: DISCOVERING, SELECTING, ACCESSING DATA

Interviews

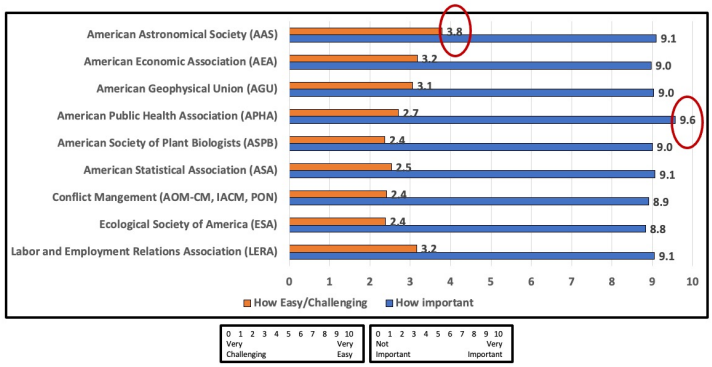
Challenges	Approaches
1. Discovering: not challenging other than assessing for trustworthy, reliable sources 2. Selecting: insufficient info on proprietary data (cost, documentation gaps), understanding variables and populations included in data 3. Accessing/housing data: unresponsive producers, unenforced journal policies, restricted data access facilities (FSRDCs), infrastructure needs for restricted data, and computational challenges due to data format and size	1. Discovering: learn about data availability in the field (publications, consortiums, conferences, other researchers); data deemed credible/trustworthy based on journal, researcher, institution's reputation 2. Selecting: use data exploration tools provided by producers to assess if data is relevant; review summary data 3. Accessing/housing data: contact producer or repository; access to computational resources; institutional legal & IRB support to work with restricted data

Survey

Data Reuse: Finding and getting access to data that has been made available for reuse by others. (cont.)



Data Reuse: Having confidence in the quality of data made available for reuse by others. (cont.)



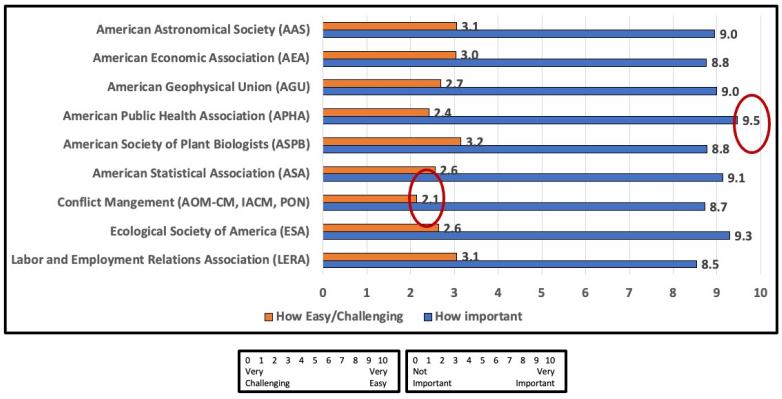
TOPIC 2: UNDERSTANDING AND WORKING WITH DATA

Interviews

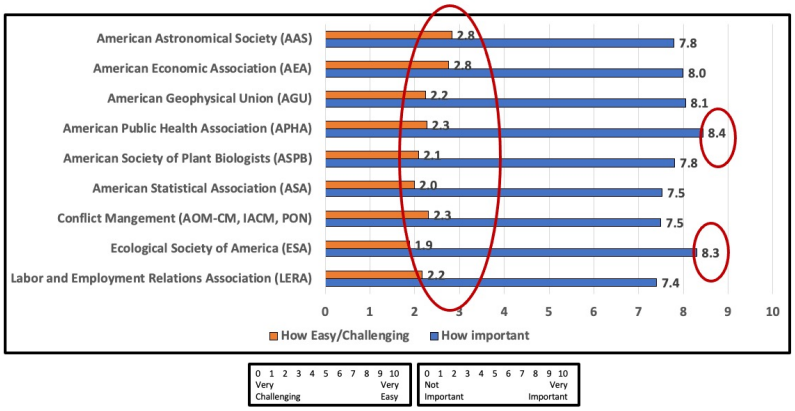
Challenges	Approaches
1. Start-up cost of working with new data (time and effort), which increases with inadequate metadata standards and/or limited documentation 2. Accessing/understanding documentation of methods, data source, including limitations and impact on analyses/conclusions 3. Multiple data sources: data integration and harmonizing inconsistent variables & granularity 4. Technical knowledge needed to work with specialized data and documentation formats	1. Contact with producer or responsive repository staff as needed; use of supplemental documents; review publications that use the data source 2. Prioritize sources with clear documentation 3. Coordination among data producers; standards and controlled vocabulary in the field 4. Use data in common/non-proprietary format; lean on others (including graduate students) for training/ knowledge

Survey

Data Reuse: Ensuring that there is clear, accurate and complete metadata and documentation associated with shared data. (cont.)



Data Reuse: The interoperability of data made available for reuse by others (so multiple data sets can be combined). (cont.)



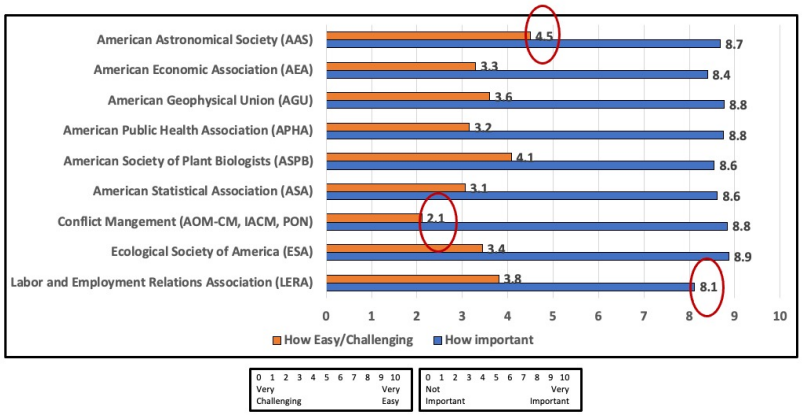
TOPIC 3: CITATION, CREDIT, RECOGNITION FOR PRODUCERS AND USERS

Interviews

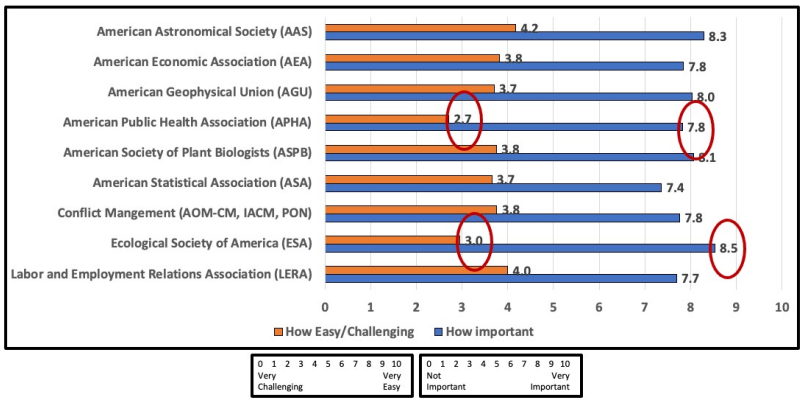
Challenges	Approaches
1. Users: Lack of awareness of the need to cite, as well as a lack of knowledge, standards or guidance on how to cite data source 2. Producers: Unable to track reuse of data, repositories don't track reuse 3. Reuse culture: Field doesn't value secondary data analysis (lack of funding, hard to publish secondary data analyses in top-tier journals)	1. Citation through references, annotations in figures, or in-text mentions 2. Registration required to download (user info collected), google for citations, informal mentions of data use (Twitter, conferences, etc.) 3. Culture shift: model after fields that accept and routinely use secondary data, senior researchers establish community repository to legitimize secondary analysis

Survey

Data Reuse: Building a culture in my fields and disciplines that is supportive of the reuse of research data. (cont.)



Data Sharing: Getting credit or recognition when your data is used by others in their research. (cont.)



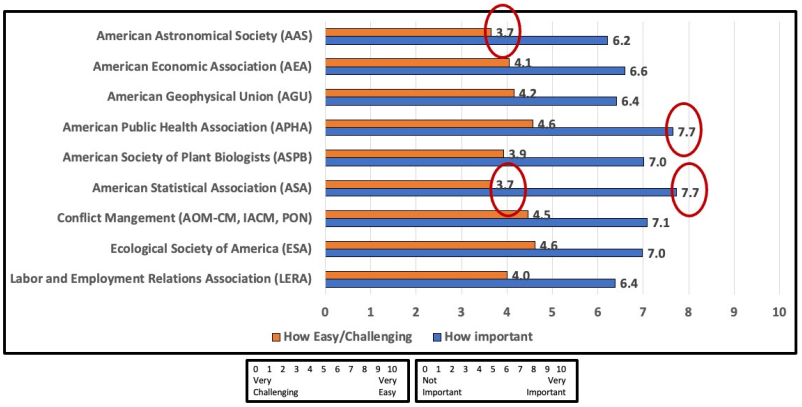
TOPIC 4: PROSPECTIVE PLANNING FOR DATA SHARING

Interviews

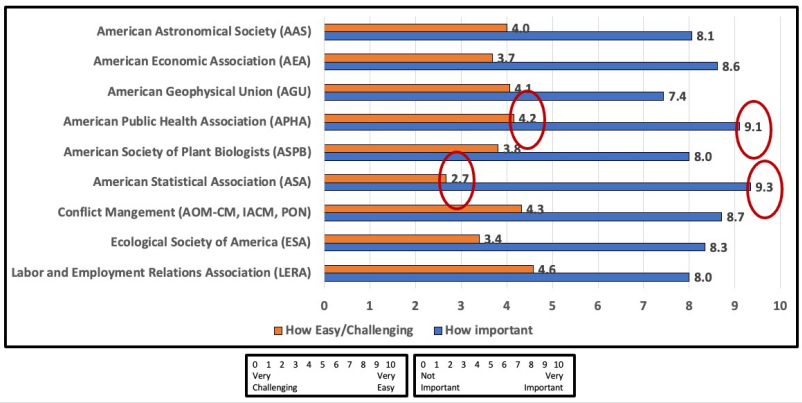
Challenges	Approaches
1. When planning occurs: time sink when it is not done early enough, inability to share human subjects research data if not planned for in IRB/consent process 2. Insufficient consideration of how documentation will be captured 3. Lack of knowledge or awareness of need to plan for restricted data (human subjects/IRB, proprietary data, use agreements) 4. Understanding or having skills to handle technical requirements for the project (data agreements, software/tools, skills building)	1. Plan before data collection, expect to update plans. Established detailed workflows, standards, pipelines for lab. 2. Borrowing models from lab fields with documentation sharing experience (Jupyter notebooks). 3. Negotiate data sharing conditions early (address IRB questions, what's sharable from proprietary sources). 4. Leverage institutional support and resources as well as others' knowledge/training (e.g., grad students)

Survey

Data Sharing: Planning for how the data will be documented and shared before you start your research study. (cont.)



Data Sharing: Knowing how to protect sensitive data before sharing. (cont.)



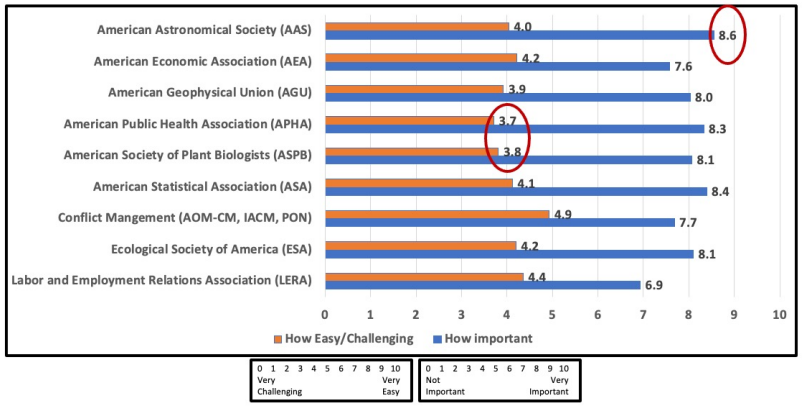
TOPIC 5: INCREASING EASE, REPLICABILITY, DATA QUALITY IN COLLECTING AND PROCESSING DATA

Interviews

Challenges	Approaches
1. Evaluating and addressing data quality issues before sharing (volume, effort) 2. Documenting data and processing for others to understand (time and effort required, knowledge of approach and tools that would help) 3. Data standards issues: lacking, outdated, multiple options, insufficient for cross-disciplinary projects	1. Prioritize resolving errors that will impact analyses 2. Within-lab sharing processes facilitates beyond lab sharing process (document for own future reuse first). Using workflows, pipelines, automation, borrowing/updating algorithms 3. Field and/or repository-established standards, culture of standards creation

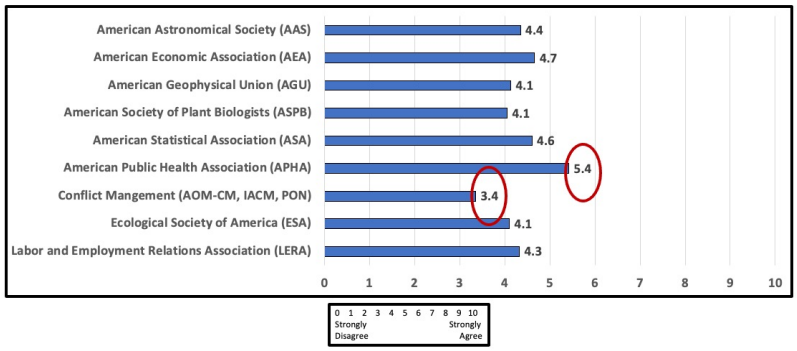
Survey

Data Sharing: Learning the skills and tools needed to prepare and document data for reuse. (cont.)



Views on fields, disciplines, and organizational support

There is currently sufficient knowledge and training in my primary field or discipline on software and tools that can reduce burden in producing and documenting research data.



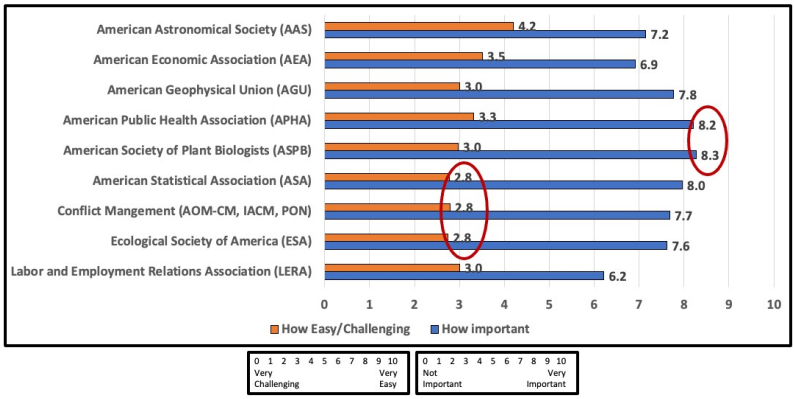
TOPIC 6: PROMOTING REUSE AND IMPACT IN DATA SHARING

Interviews

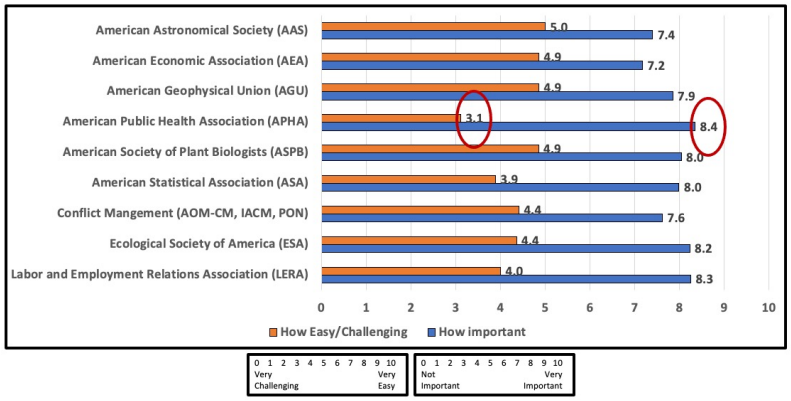
Challenges	Approaches
1. Not knowing what to share: many items created during a research process (e.g., data, code, protocols, codebook, metadata, etc.), lack of guidance on what to share 2. How to document methods, provenance, etc.; for large and complex sources, what documentation is needed for novice users 3. How to support data access (restricted data), choices of user burden vs. tracking data downloads 4. Inability to measure reuse and impact of shared data	1. Following repository, funder or journal requirements 2. Docker and other technologies for easing reuser burden in accessing and using data 3. Institutional support in review and sharing process, handling restricted data and licenses 4. Provide citation mechanism/formats for users, persistent/unique identifiers, ad-hoc tracking via google scholar, track downloads, create data paper to fit inside a journal reference list

Survey

Data Sharing: The availability of standards and controlled vocabularies to facilitate effective sharing and reuse of your data. (cont.)



Data Sharing: Knowing which repository or repositories are trusted and appropriate for your data. (cont.)



Appendix F: Stakeholder Alignment and Action Plan Breakout Notes

The following 10 action plans use a "Stakeholder Alignment and Action" format, which includes the elements in the template below. The form is designed to help achieve sufficient alignment during breakout group discussions for actions to be developed. It builds situational awareness around a success vision, a landscape of stakeholders and interests, an identification of points of alignment and misalignment, and an action plan moving forward.

Stakeholder Alignment and Action: *A Living Document v2.0 (July 2021)*

Issue	[A statement of the issues or set of issues that are the focus – essentially a well posed problem statement]
Vision	[Brainstorming elements of a vision of success on this issue or set of issues, begun with the phrase: ***Imagine a world where…***]
Stakeholders	[A starter list of all relevant types of stakeholders, to be augmented during discussions of relevant stakeholders] • Researchers – academic (faculty and staff, graduate and undergraduate students) • Researchers – government or nonprofit • Researchers – private industry • Researchers – citizen scientists • Funders – government funding agencies • Funders – nonprofit/NGO funders • Funders – private entities • Community/culture builders – societies and consortia • Community/culture builders – academic departments, academic leadership • Publishers – society journals • Publishers – other nonprofit journals • Publishers – for-profit journals • Propellers – disciplinary research data service providers and repositories • Propellers – commercial research data service providers (AWS, and others) • Propellers – open-source repositories (e.g., GitHub, Kaggle, OSF, and others) • Propellers – open-source software/system providers (e.g., Docker, R, and others) • Propellers – other service providers (CHORUS, DataCite, and others) • Public interest – policy makers • Public interest – general public

<table>
<tr><td>Interests</td><td colspan="3">[A listing of what is “at stake” for the stakeholders – not specific positions, but underlying hopes, fears, concerns, and other matters associated with their value proposition with respect to the issue or issues]</td></tr>
<tr><td rowspan="2">Alignment (Focusing on researchers and funders)</td><td>Misaligned Interests</td><td>Moderate Alignment</td><td>Well-aligned Interests</td></tr>
<tr><td>• [A list of interests for which there is considerable misalignment – in Red]</td><td>• [A list of interests for which there is moderate alignment – in Black]</td><td>• [A list of interests for which there is considerable alignment – in Green]</td></tr>
<tr><td>Actions</td><td colspan="3">[Milestones for implementation on a quarterly or annual basis]
• 3Q 2021
o
• 4Q 2021
o
• 2022
o</td></tr>
</table>

For each topic below, the **Stakeholder Action Plan Notes** for each topic are presented, followed by **Related Findings** that contextualized the action plan discussion. Related findings were drawn from Research Practice Recommendations from Section V (p. 16–18) and Data Sharing and Reuse Stakeholder Survey findings.

1. COMMUNITY-BASED STANDARDS DEVELOPMENT

Stakeholder Action Plan Notes

Stakeholder Alignment and Action: *A Living Document v2.0 (July 2021)*

Issue	The lack of appropriate standards (e.g., formats, expected content, vocabularies), or awareness of those standards, for many fields hampers effective sharing and reuse, particularly when working across disciplines. Interview and survey results say that this is important but hard to do. What initiatives will advance community-based actions to create or update standards that enable work within and across fields?
Vision	***Imagine a world where…*** It is possible to harmonize standards. Standards are interoperable. Standards are adopted, people see the value and incentives support standards use. Core standards are adopted and used, and specific subject-based standards are also used. Standards that are adopted are proven to be useful for the purpose for which they were designed. Adoption is the priority over optimization.
Stakeholders	• Researchers – academic (faculty and staff, graduate and undergraduate students) • Researchers – government or nonprofit • Researchers – private industry • Researchers – citizen scientists • Funders – government funding agencies • Funders – nonprofit/NGO funders • Funders – private entities • Community/culture builders – professional societies and consortia • Community/culture builders – academic departments, academic leadership • Publishers – society journals • Publishers – other nonprofit journals • Publishers – for-profit journals • Propellers – disciplinary research data service providers and repositories • Propellers – commercial research data service providers (AWS, and others) • Propellers – open-source repositories (GitHub, Kaggle, OSF, and others) • Propellers – open-source software/system providers (Docker, R, and others) • Propellers – other service providers, including standards bodies (ISO, NISO, W3C, NIST, CHORUS, DataCite, and others) • Public interest – policy makers • Public interest – general public

<table>
<tr><td>Interests</td><td colspan="3">• Data reuse is enabled by standards.
• Reduce economic costs and increase efficiency in data use/reuse.
• Identifying positive and null findings, by sharing and publishing data.
• Prevent technology gap (understanding data later in time).
• Enabling meta-analysis.
• Enable all aspects of FAIR.
• The need to return data back to research participants/communities in a standardized way.</td></tr>
<tr><td rowspan="2">Alignment</td><td>Misaligned Interests</td><td>Moderate Alignment</td><td>Well-aligned Interests</td></tr>
<tr><td>• Reducing economic costs and increasing efficiency in data use/reuse</td><td>• Data reuse is enabled by standards
• Enable all aspects of FAIR
• Identifying positive and null findings, by sharing and publishing data
• Enabling meta-analysis
• The need to return data back to research participants/communities in a standardized way</td><td>• Prevent technology gap (understanding data later in time)</td></tr>
<tr><td>Actions</td><td colspan="3">• 3Q 2021
◦ Identify exemplars, convene discussions of how/why certain standards became widely used.
• 4Q 2021
◦ Identify specific problems/tasks where standards are lacking and which communities want to accomplish these tasks.
◦ Study which data sharing/reuse related standards are being used within specific research communities.
• 1Q 2022
◦ Workshops/conferences showcasing the exemplars (society conferences, etc.).
◦ Outreach to scale up the number of creators and institutions recorded in dataset/software metadata.
◦ Do gap analyses on dataset/software metadata.</td></tr>
</table>

Potential partners/sponsors for implementation on this topic:

- NISO
- DataCite
- CHORUS
- AGU
- AAU/APLU
- NASEM/BRDI
- Funders

Related Findings

Research Practice Recommendations

1.A. **Standards for Sharing Data:** Develop a manageable minimum standard for data sharing (e.g., expected documentation, minimum standard for data dictionaries and ontologies, data variable standards that promote interoperability, desirable repository characteristics).

2.A **Community-developed Standards and Formats:** Create mechanisms for developing community-based standards for data and documentation (data, metadata, schema, formats, etc.) to facilitate and reduce burden in data reuse and sharing.

4.F. **Community-based Practice and Data Standards:** Utilize community standards for practices and data to extent possible, anticipate need for data integration.

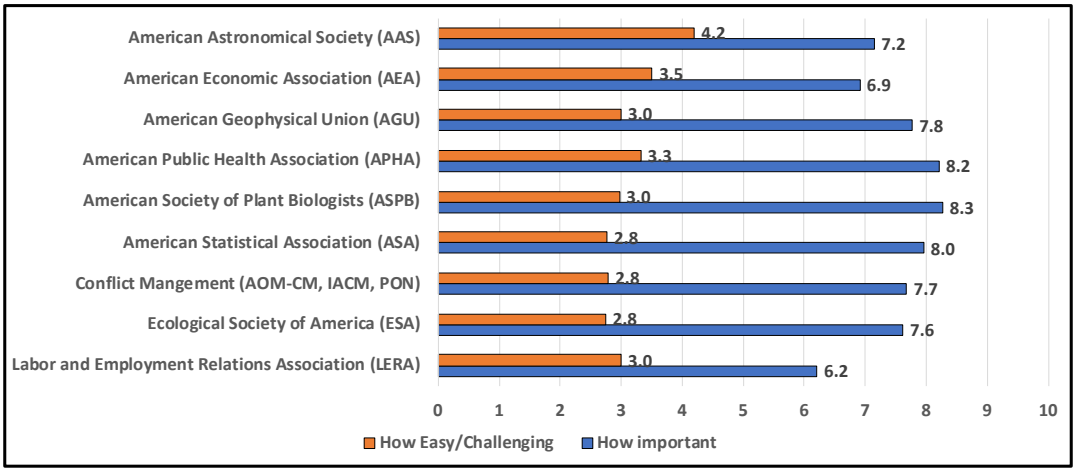
5.B. **Flexible Community Standards:** Establish standards developed through communities (domain specific) for data to be recoverable and machine readable, with a focus on "good enough" or minimum requirements.

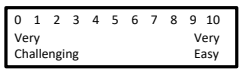
Survey Results

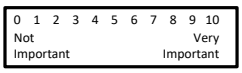
Indicator Issue	Important/ Very Important	Challenging/ Very Challenging
"The availability of standards and controlled vocabularies to facilitate effective sharing and reuse of your data"	74%	64%

Note: *Important/Very Important responses are 7-10 on a scale from 0-10 with 0=Not Important and 10=Very Important; Challenging/Very Challenging are 0-3 on a scale form 0-10 with 0=Very Challenging and 10=Very Easy.*

"The availability of standards and controlled vocabularies to facilitate effective sharing and reuse of your data"



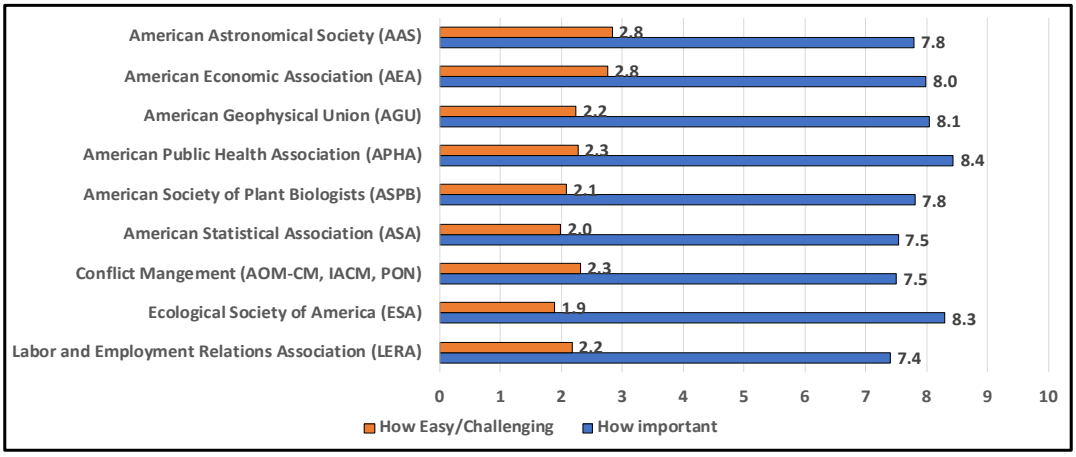


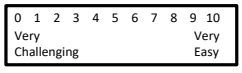


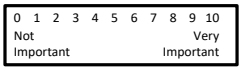
Indicator Issue	Important/ Very Important	Challenging/ Very Challenging
"The interoperability of data made available for reuse by others (so multiple data sets can be combined)."	80%	76%

Note: *Important/Very Important responses are 7-10 on a scale from 0-10 with 0=Not Important and 10=Very Important; Challenging/Very Challenging are 0-3 on a scale form 0-10 with 0=Very Challenging and 10=Very Easy.*

"The interoperability of data made available for reuse by others (so multiple data sets can be combined)."







2. COMMUNITY-LEVEL CULTURE CHANGE

Stakeholder Action Plan Notes

Stakeholder Alignment and Action: *A Living Document v2.0 (July 2021)*

Issue	Many researchers do not support and/or do not have supportive communities of practice/ research cultures that recognize the importance and potential of data sharing and reuse. Interview and survey results say that this is important but hard to do. What can we do to build community among practitioners in ways that are supportive of data sharing and reuse practices? This includes challenging embedded assumptions that undercut sharing and reuse, such as proprietary views of data?
Vision	***Imagine a world where…*** • Data sharing and reusing is valued, expected, practiced, and supported widely across STEM disciplines, other domains, and research stakeholders (listed below). • Supported/valued through time, money, personnel, training, sustained/shared infrastructure, professional recognition of open science. • All stakeholders care about and pay for future use of data.
Stakeholders	• Researchers – academic (faculty and staff, graduate and undergraduate students) • Researchers – government. • Researchers – private industry • Researchers – citizen scientists • Funders – government funding agencies • Funders – nonprofit/NGO funders • Funders – private entities • Funders – universities • Community/culture builders – professional societies and consortia • Community/culture builders – academic departments, academic leadership • Community/culture builders – university administrators/leaders • Publishers – society journals • Publishers – other nonprofit journals • Publishers – for-profit journals • Propellers – disciplinary research data service providers and repositories • Propellers – commercial research data service providers (AWS, and others) • Propellers – open-source repositories (GitHub, Kaggle, OSF, and others) • Propellers – open-source software/system providers (Docker, R, and others) • Propellers – other service providers (DataCite, and others) • Public interest – policy makers • Public interest – general public • Business – use their tools and they use products of research including data • Congress – funding agencies

Interests	• Researchers: value of creating of shareable data—recognition of work, accelerate scientific advancement, increased effort and time for publishing data • Funders: highest impact, public funding is required to publish data (with exceptions) • Publishers: increase readership, attract high quality publications, increase citations (documentary evidence for this last point would be helpful) • Propellers: monetary gain • Public: transparency increases trust in science, decision making is based in data • Accuracy, completeness, interpretability of data • Generational change in perspective		
Alignment	**Misaligned Interests**	**Moderate Alignment**	**Well-aligned Interests**
	• Data providers not aligned with data users in recognizing the value and effort • Cost of infrastructure and publication efforts versus open access • Expectation of data publishing but no willingness to fund long-term maintenance • Tension between scientific openness and data security	• Transparency desired by most stakeholders versus not in private sector • Misalignment of some funding agency's goal of new research and long-term infrastructure funding	• Accelerated advancement of scientific knowledge • If data have high impact all stakeholders win
Actions	• 4Q 2021 ◦ Working group to develop standards for data trust (for public/private partnerships) ◦ Getting data sharing on the agenda of university governing bodies, e.g., workshops/ meeting plenary sessions for presidents, provosts, VPs in higher education associations • 1Q 2022 ◦ Societies develop recommended wording for promotion requirements ◦ Societies develop best practices documents/workshops for how to create supportable data management plans • 2Q 2022 ◦ Add data handling training to basic STEM education at undergraduate level (where?) STEM Teacher Organization		

Potential partners/sponsors for implementation on this topic:

- Foundations such as Moore (science, environment) and Gates (health and agriculture) with interests in solving particular societal problems
- Open Research Funders Group
- Professional societies, ASPB, AGU, ESA, etc.
- Establishment of Data trusts – connect public and private enterprise
- Standards for developing data trusts

Individual commitments to support implementation on this topic:

- American Society of Plant Biologists

Related Findings

Research Practice Recommendations

1.B. **Incentives to Prepare Data that Fosters Data Discovery:** Provide incentives, in the form of sufficient funding for data preparation and outreach for data producers and others, that concretely provide tools for data discovery and access.

2.C. **Funding for Producing Reusable Data:** Provide adequate funding to support for the time required for producers to make data reusable and interoperable (aligned with funder value propositions and data management plans (DMPs)).

3.A. **Funding for Data Curation:** Provide adequate funding for quality data curation, which is as important as the data creation process.

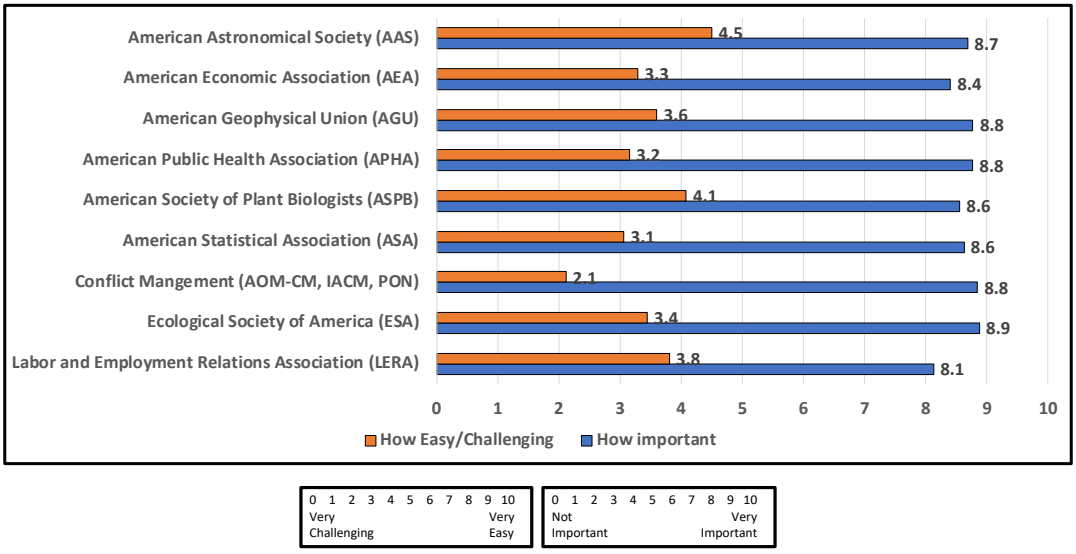
4.G. **Funders and DMPs:** Add more "teeth" to the DMP evaluation in the review process, including choice of repositories and documentation plans that facilitate downstream reuse, and require updates on DMP as part of the reporting process.

Survey Results

Indicator Issue	Important/ Very Important	Challenging/ Very Challenging
"Building a culture in my fields and disciplines that is supportive of the reuse of research data."	90%	58%

Note: *Important/Very Important responses are 7-10 on a scale from 0-10 with 0=Not Important and 10=Very Important; Challenging/Very Challenging are 0-3 on a scale form 0-10 with 0=Very Challenging and 10=Very Easy.*

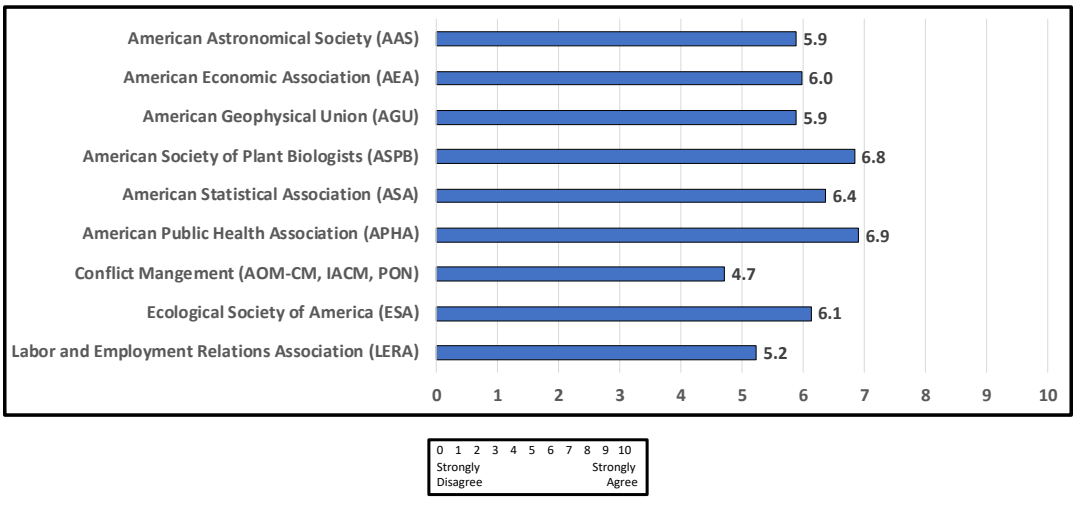
"Building a culture in my fields and disciplines that is supportive of the reuse of research data."



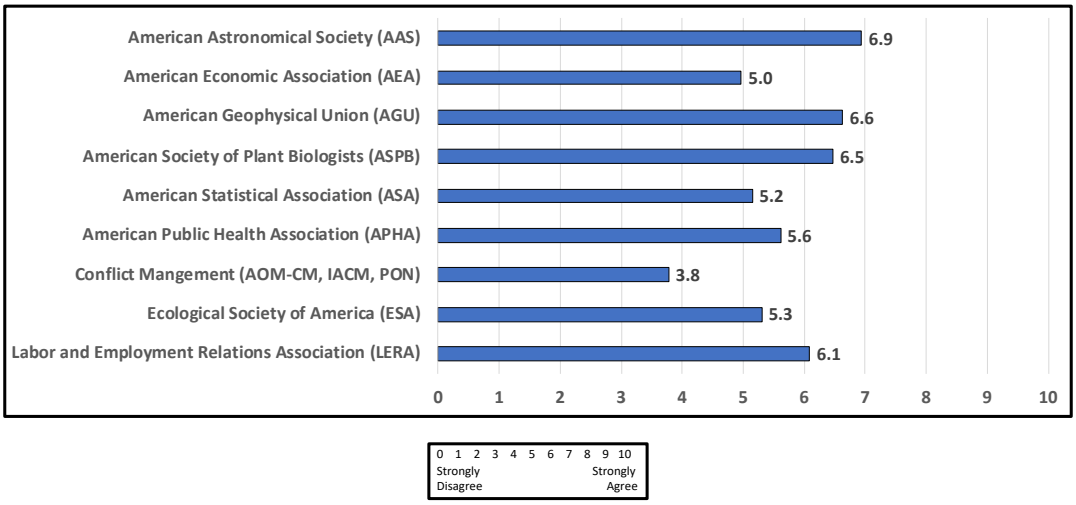
Statements	Strongly Agree
"There is currently a high degree of cooperation and sharing of software, tools, and methods that facilitate understanding and using shared data in my primary field or discipline."	46%
"There is currently a high degree of cooperation and sharing of research data in my primary field or discipline."	49%
"There is currently sufficient data storage, software and computational infrastructure in my organization to facilitate data sharing and reuse."	47%

Note: *Agree/Strongly Agree are 7-10 on a scale from 0-10 with 0=Strongly Disagree and 10=Strongly Agree.*

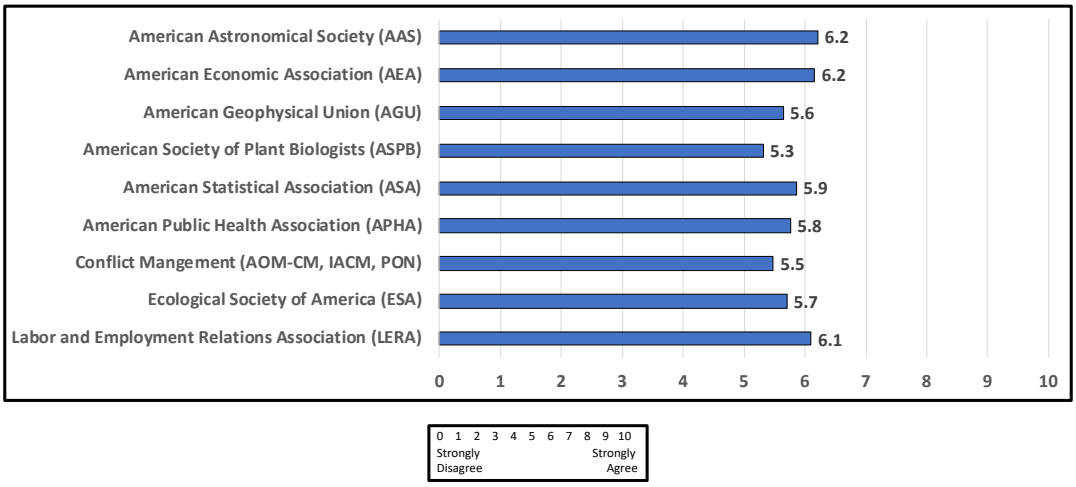
"There is currently a high degree of cooperation and sharing of software, tools, and methods that facilitate understanding and using shared data in my primary field or discipline."



"There is currently a high degree of cooperation and sharing of research data in my primary field or discipline."



"There is currently sufficient data storage, software and computational infrastructure in my organization to facilitate data sharing and reuse."



3. CREDIT AND REWARDS FOR SHARING AND REUSE

Stakeholder Action Plan Notes

Stakeholder Alignment and Action: *A Living Document v2.0 (July 2021)*

Issue	Many researchers do not receive credit or rewards for sharing data. Researchers struggle to document the impact of shared data because they cannot easily identify when and how their data is used by others. Reuse of data is not always valued the same as generating original data. Interview and survey results say that this is important but hard to do. What actions are needed to provide evidence of data reuse and impact? What can stakeholders do to encourage and reward the sharing and reuse data? What are the mechanisms to make this possible? How can we ensure that they actually deliver results with credit for sharing and reuse? What is the scope of this issue—home institutions, funders, research community, etc.?

Vision	***Imagine a world where…***
	• DOIs are assigned to all data outputs—improves discovery, easier to credit (citation) when reused.
	• Production of a dataset is treated as approximately equivalent to the production of a research paper.
	  ◦ Data stands on its own without needing an associated publication.
	  ◦ Published papers are promoting datasets which are valued for its immediate, additional, and ongoing uses.
	  ◦ During tenure and promotion review, data is recognized as a contribution to advance the scientific field/given comparable credit as publications.
	• Funders having open science policies that require/promote open data.
	• Funders provide reward mechanisms for data reuse.

Stakeholders	• Researchers – academic graduate and undergraduate students • Researchers – academic staff • Researchers – academic faculty • Researchers – government or nonprofit • Researchers – private industry • Researchers – citizen scientists • Funders – government funding agencies • Funders – nonprofit/NGO funders • Funders – private entities • Community/culture builders – professional societies and consortia • Community/culture builders – academic departments, academic leadership

<table>
<tr><td></td><td colspan="3">
• Publishers – society journals
• Publishers – other nonprofit journals
• Publishers – for-profit journals
• Propellers – disciplinary research data service providers and repositories
• Propellers – commercial research data service providers (AWS, and others)
• Propellers – open-source repositories (GitHub, Kaggle, OSF, and others)
• Propellers – open-source software/system providers (Docker, R, and others)
• Propellers – other service providers (CHORUS, DataCite, and others)
• Public interest – policy makers
• Public interest – general public
</td></tr>
<tr><td>Interests</td><td colspan="3">
• Researchers
  ◦ Reward mechanisms for data reuse
  ◦ Persistent identifiers for credit
  ◦ Tracking reuse (also important to funders)
• Advancing/accelerating science
• Broader impacts – training, outreach, benefit to society
  ◦ More inclusive and equitable science
  ◦ Trust in science – reproducibility, transparency
  ◦ Social impacts (and ability to track, quantify)
• FAIR principles in the use and reuse of data
</td></tr>
<tr><td rowspan="2">Alignment (Focusing on researchers and funders)</td><td>Misaligned Interests</td><td>Moderate Alignment</td><td>Well-aligned Interests</td></tr>
<tr><td>
• Incentives to hoard data – self-interest to extract all analyses from dataset before sharing
• Timing requirements for making data and publication open
</td><td>
• Senior researchers (and others) don't have time to keep up with current best practices, standards, etc.
• Prioritizing data curation versus publications
• Group level: data sharing is a good thing, individual level it takes work
• Match between funder RFP scope/intent and underlying culture of reviewer (need PMs to keep this aligned)
• Aligning publisher and funder requirements for data publication
</td><td>
• Advancing science
• Understanding reuse and impact
• Additional funding for data curation (potentially after the original grant closes)
</td></tr>
</table>

Actions	

	• 3Q 2021
	  ◦ Explore potential of funding from post-grant data curation; combined with early career researcher training.
	  ◦ Signaling support from leaders in the field.
	  ◦ Associations establish standards for data generation and reuse, including data sharing. Lots of this can be moved forward by cross pollination between associations.
	• 4Q 2021
	  ◦ Funders can enhance carrots and sticks for the sharing of data in their ecosystems

Potential partners/sponsors for implementation on this topic:

- FASEB will invest resources to move these forward
  - — Series of sessions like these, with more focus on specific issues—standards, best practices, toward culture change
  - — 5–10-year horizon
  - — Earth / Space science 5-10 years ahead
- National Academies
- Identify key leaders, get them to clearly state their support/vision for this
- ORFG to coordinate funder incentives/best practices
- STM as a partner to help the journals support these objective (stm-assoc.org)
- AGU is working on the challenges around data citation (and software) specific to enabling more automated credit. You can add us as supportive of this objective and happy to share the work we are doing. Also, Crossref, DataCite, and CHORUS

Related Findings

Research Practice Recommendations

3.E. **Recognition:** Institutions should develop approaches to recognizing good data sharing and reuse practices in promotion, tenure, and career paths.

6.C. **Promotion and Tenure:** Organizations (all sectors) should develop policies and culture that meaningfully recognizes data at all stages of a researcher's career.

2.D. **Career Development for Data Workers:** Appreciate and provide professional development to data workers who have the expertise and skills to support quality data sharing and reuse.

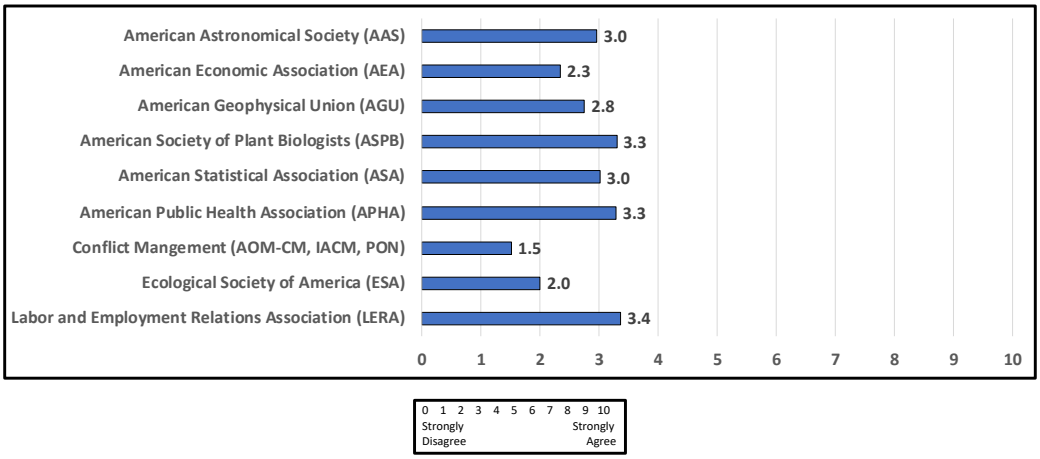
6.F. **Data Management Careers:** Actively develop and reward a pipeline of professionals in research data handling and management in support of research.

Survey Results

Statements	Strongly Agree
"The tenure, promotion, and rewards in my organization recognize and value researchers for sharing research data."	11%

Note: *Agree/Strongly Agree are 7-10 on a scale from 0-10 with 0=Strongly Disagree and 10=Strongly Agree.*

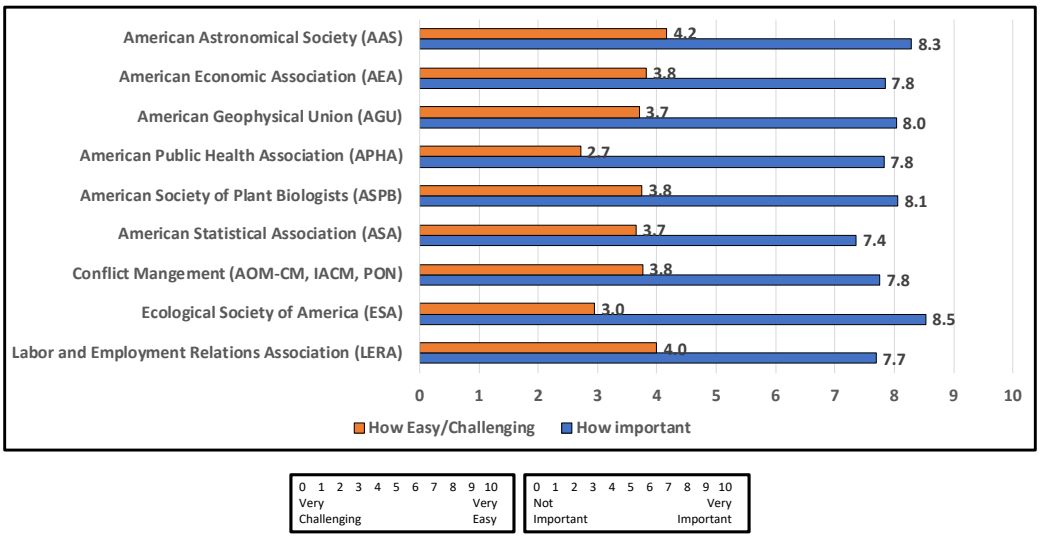
"The tenure, promotion, and rewards in my organization recognize and value researchers for sharing research data."



Indicator Issue	Important/ Very Important	Challenging/ Very Challenging
"Getting credit or recognition when your data is used by others in their research."	79%	54%

Note: *Important/Very Important responses are 7-10 on a scale from 0-10 with 0=Not Important and 10=Very Important; Challenging/Very Challenging are 0-3 on a scale form 0-10 with 0=Very Challenging and 10=Very Easy.*

"Getting credit or recognition when your data is used by others in their research."



4. DATA AS A FIRST-CLASS RESEARCH PRODUCT

Stakeholder Action Plan Notes

Stakeholder Alignment and Action: *A Living Document v2.0 (July 2021)*

Issue	Data are increasingly a primary output of research process. In some cases, key data sets are surrounded by communities of practice and associated articles, but these are exceptions. More often, data is just viewed as an input into research products (articles, presentations, etc.). What actions are needed to ensure that data is valued as a first-class product of research projects? Can we approach this with an input/output frame (borrowing from computer science programming)? Can we re-think the research enterprise as a data flow process? Can we re-frame data not as a competitive and/or commercial output, but as a common good? Can we think of data as having a life story, not just versions (is this data life cycle?)? How can peer review of data be part of the trust building process (i.e. related to data quality)? How can we incentivize sharing and give credit; brings you back to culture change [carrots and sticks!]? How can we overcome the clash of values related to data (e.g., control v sharing)? Can this be introduced in a more deliberate way, as a 'disruptive' impetus to drive change? How can we provide more training? Although submission to a repository can help with curation, how can this be funded? Can we have budgets that require a specific line item to cover data curation? Who is the champion to move all of this forward?
Vision	***Imagine a world where…*** • Producers of data are able to seamlessly share data. • Champions (individuals and technologies) in each discipline promote/support data sharing. • Adequate resources are provided for producers and data publishers (e.g., repositories). • Data means more than a collection of 'facts'; data is viewed as evolving, not static (e.g., having a life story). • Adequate training, depending on your role in the science enterprise. • Researchers prioritize sharing because of requirements and credit. • There are common understandings of minimum requirements and ways to enable interoperability. • Data is machine readable and machine actionable. • Talks at professional meetings start with sharing the metadata about where to get the data.

Stakeholders	• Researchers – academic (faculty and staff, graduate and undergraduate students) • Researchers – government or nonprofit • Researchers – private industry • Researchers – citizen scientists • Funders – government funding agencies • Funders – nonprofit/NGO funders • Funders – private entities • Community/culture builders – professional societies and consortia • Community/culture builders – academic departments, academic leadership • Publishers – society journals • Publishers – other nonprofit journals • Publishers – for-profit journals • Propellers – disciplinary research data service providers and repositories • Propellers – commercial research data service providers (AWS, and others) • Propellers – open-source repositories (GitHub, Kaggle, OSF, and others) • Propellers – open-source software/system providers (Docker, R, and others) • Propellers – other service providers (CHORUS, DataCite, and others) • Public interest – policy makers • Public interest – general public • Universities/schools/departments • Universities – development office • Universities – IP office
Interests	• Researchers: share versus not share (blanket non-sharing of confidential data, concerns over getting scooped). • Funders: money for curation as part of the science (not viewed as competing). • Publishers: peer review challenges, burden of responsibility. • Libraries: shifting primary mission from managers of information to libraries as central purveyors of data services and researchers recognize this. • Who gets to decide when the data is good enough to be a first-class object and when is it good enough? • How to make progress now; also realize it's a marathon, not a sprint?

Alignment	Misaligned Interests	Moderate Alignment	Well-aligned Interests
	• Asking faculty to do things they're not rewarded for • Current funding models don't support the necessary components to promote sharing over the full life cycle • Big universities versus smaller universities (equity) • Developed areas versus less developed areas (both domestic and international) • Entities within universities are not aligned on data sharing • Misalignment between professional societies (in the form of subscription journals) and open science	• Some alignment with data sharing already as reflected in policies	• Libraries moving towards embracing the data role • There are examples where the sharing and credit are well aligned, i.e., from a value perspective
Actions	• 2021 ◦ Create an interactive decision tree at the NSF program level for researchers submitting proposals to help with data management planning and to incorporate into their proposal. ◦ Clarify what kind of credit researchers would actually want for making their data available/clarify what might be meaningful metrics. ◦ Include credit for data as part of the results of prior research/rank proposals higher where PIs have a demonstrated track record of sharing. ◦ New grantees are given guides on how to do the data sharing stuff and require updates on data sharing in annual reports. ◦ Identify examples (practices) where these elements align, capturing that info, and making the examples available • 2022 ◦ Take what is done in 2021 and share with communities and see if it works for that community. ◦ Promote peer review of stand-alone data, i.e., not just data that is associated with a published article. ◦ Look at the CCDC as a case study and figure out if it could be generalized (process of validation, data certification by data analysts, get badged).		

Related Findings

Research Practice Recommendations

1.D. **Awareness and Motivation to Value All Data:** Mechanisms to increase understanding of the value of all research data, including null results.

3.B. **Data Citation Expectations:** Data providers should specify appropriate citation format and other aspects of appropriate use (e.g., DUA, license).

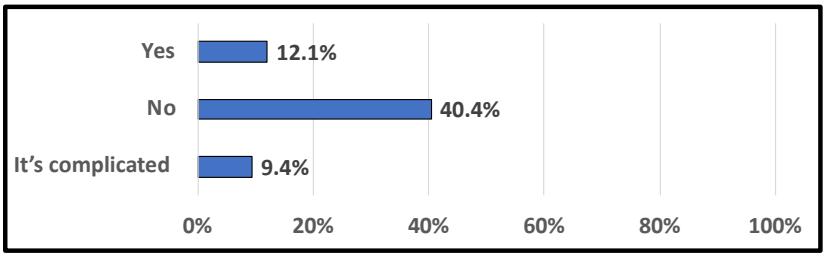
Survey Results

"What, if any, factors motivate you to share or reuse publicly accessible research data?"

Categories into which a sample of responses were sorted and their percentages:

Advancement of science	32%
Public good/Personal responsibility/Public policy	16%
Reproducibility/Transparency	12%
Efficiency/Effectiveness	11%
Availability/Quality of data	10%
Required	9%
Teaching/research needs	6%
Recognition	1%
Collaboration	1%

"When you make data publicly accessible, do you generally hold it privately for an embargo period (such as two years)?"



Note: *Percentages are for all respondents (those not making data available were not asked this question, but constitute the additional percentages to equal 100%).*

5. DATA PRODUCER TRAINING MODULES

Stakeholder Action Plan Notes

Stakeholder Alignment and Action: *A Living Document v2.0 (July 2021)*

Issue	Data producers need a way to build skills that reduce burden in their research process and increase the impact of their shared data. This is true for researchers at all stages of their careers. Interview and survey results say that this is important but hard to do. Single-point modules (30–45 minutes) can be particularly impactful. What are examples of modules that should be developed for data producers? How can we bring Reproducibility in earlier in the curriculum (for current researchers that would be "a change" in practice)? Here are examples of possible modules that could be used as a part of curriculum and also be used to enhance existing knowledge: 1. Techniques for discovery of useful datasets (e.g., making datasets more discoverable in a Google dataset search—the "Findable" in FAIR) 2. Understandable data—metadata 3. Data reuse—licensing 4. Citation and credit 5. Protection/security 6. Data dictionaries/vocabularies/ontologies 7. File naming/organization/storage structure (e.g., Tidy data) 8. Scripting is your friend.
Vision	***Imagine a world where…*** • Researchers are aware of all past and current research and are able to be more innovative. • Researchers spend less time on data discovery and making it interoperable and more time on research and innovation. • Researchers share their data openly from the start (assuming it is well documented/validated), and collaborate in the open (not realistic for sensitive data, and possibly limited sharing). • Researchers are motivated to be open and share because that is valuable to the research and they no longer are "required" but instead "motivated." • Data production best practices were so well understood and internalized that ad hoc specialized training was not necessary.

Stakeholders	**Primary** • Researchers – academic (faculty and staff, graduate and undergraduate students) • Researchers – government or nonprofit • Researchers – private industry • Researchers – citizen scientists **Secondary** • Funders – government funding agencies • Funders – nonprofit/NGO funders • Funders – private entities • Community/culture builders – societies and consortia • Community/culture builders – academic departments, academic leadership • Publishers – society journals • Publishers – other nonprofit journals • Publishers – for-profit journals • Public interest – policy makers • Public interest – general public • Libraries **Remaining** • Propellers – disciplinary research data service providers and repositories • Propellers – commercial research data service providers (AWS, and others) • Propellers – open-source repositories (e.g., GitHub, Kaggle, OSF, and others) • Propellers – open-source software/system providers (e.g., Docker, R, and others) • Propellers – other service providers (CHORUS, DataCite, and others)
Interests	• Scientific progress • Transparency • Reproducibility • Accountability • Raising the value of digital outputs (data, software)—Return on investment • Educational progress (how people learn from the research) • Career advancement • Personal researcher motivation (drive) • Priority given to original data as opposed to reuse (journals, funders?) • Use of research • Researcher incentive methods • Discovery of relevant training material to the takeholder

Alignment	Misaligned Interests	Moderate Alignment	Well-aligned Interests
	• Career Advancement • Personal Researcher Motivation (Drive) • Priority given to original data as opposed to reuse (journals, funders?) • Educational Progress (how people learn from the research) • Researcher Incentive Methods • Discovery of Relevant Training Material	• Accountability • Raising the value of digital outputs (data, software)—Return on investment	• Scientific Progress • Transparency • Reproducibility
Actions	• 3Q 2021 ◦ Identify experts on training development ◦ Module design—Prioritization of stakeholder need ◦ Module development ◦ Instructor guides ◦ Train the trainer ◦ Schedule events • 4Q 2021 ◦ Training partner summit		

Potential partners/sponsors for implementation on this topic:

- Carpentries
- Openscapes
- FSCI
- Data Management Training Clearing House
- Society training
- Institutions
- AAU/APLU
- Research Data Framework effort (NIST)
- RDA
- Funder community that commits to initial development and ongoing management
- Librarians
- American Statistical Association?

Individual commitments to support implementation on this topic:

- Societies are willing to work more on this topic. Need training experts to support.
  - AGU
  - FABBS

Related Findings

Research Practice Recommendations

1.C. **Training and Toolkits to Provide Replicable Frameworks:** Develop training and toolkits that enable data producers to develop documented data sources (e.g., data, study goals, methodology, metadata, software code) with replicable frameworks, particularly if not in an R1 institution.

3.C. **Consistent Identifiers:** Encourage and expect a consistent set of identifiers and their related metadata (e.g., DOIs and PIDs for data, researcher, institution, funder).

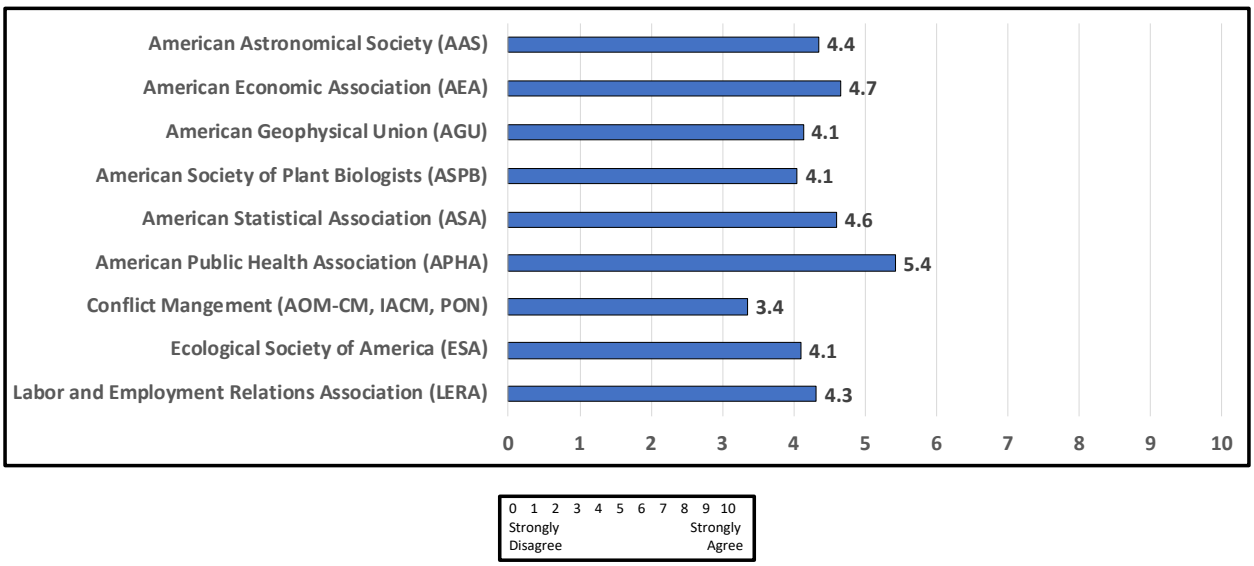
5.C **Training and Resources:** Provide institutional support for training in data sharing and reuse practice (e.g., leading practices and tools, study planning for data sharing, ethics training) for researchers of all career stages and with an understanding that a mix of formats will be needed (e.g., components of courses, regular "booster shots," modular lessons, asynchronous, just-in-time modules) .

Survey Results

Statements	Strongly Agree
"There is currently sufficient knowledge and training in my primary field or discipline on software and tools that can reduce burden in producing and documenting research data."	23%

Note: *Agree/Strongly Agree are 7-10 on a scale from 0-10 with 0=Strongly Disagree and 10=Strongly Agree.*

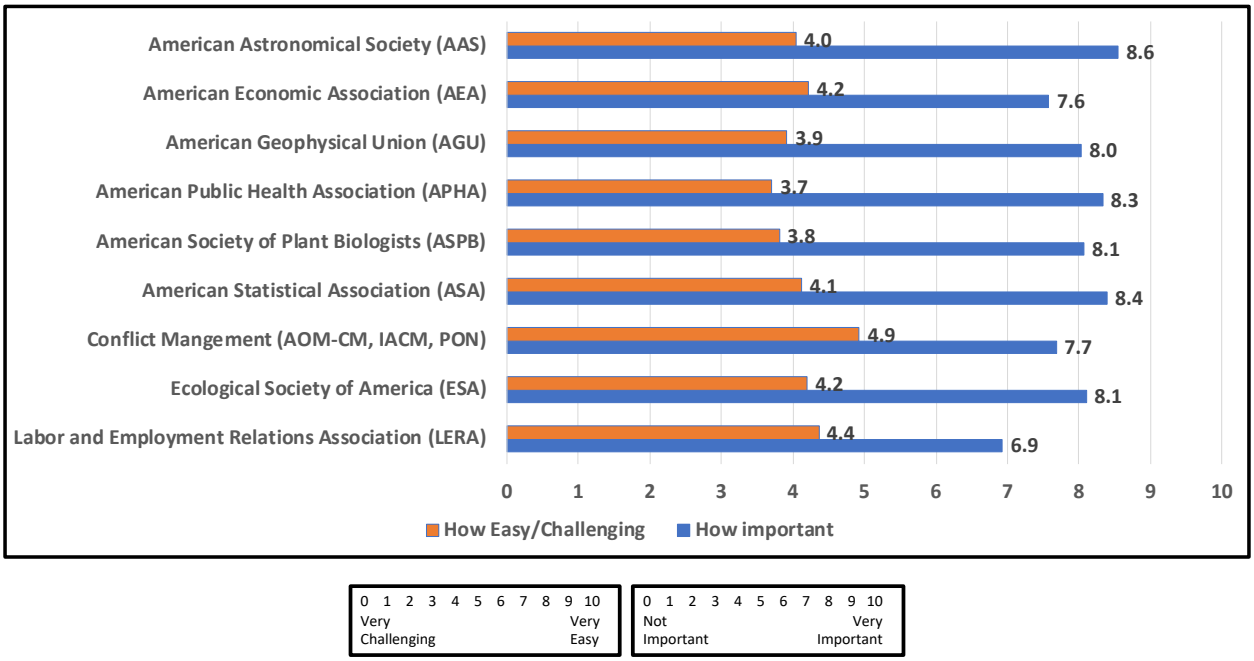
"There is currently sufficient knowledge and training in my primary field or discipline on software and tools that can reduce burden in producing and documenting research data."



Indicator Issue	Important/ Very Important	Challenging/ Very Challenging
"Learning the skills and tools needed to prepare and document data for reuse."	83%	48%

Note: *Important/Very Important responses are 7-10 on a scale from 0-10 with 0=Not Important and 10=Very Important; Challenging/Very Challenging are 0-3 on a scale form 0-10 with 0=Very Challenging and 10=Very Easy.*

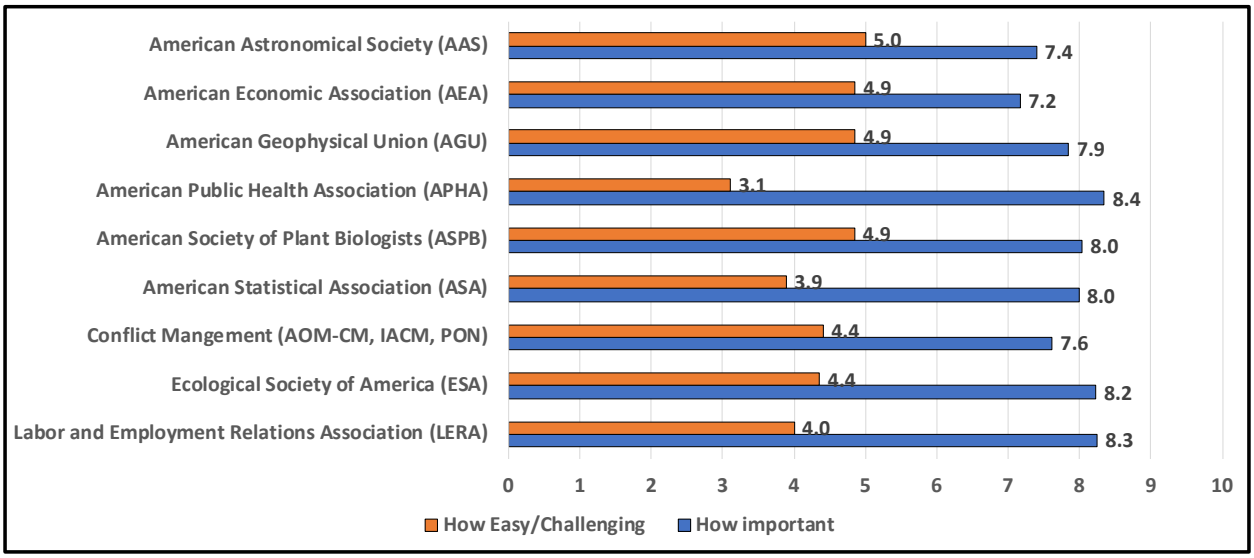
"Learning the skills and tools needed to prepare and document data for reuse:"



Indicator Issue	Important/ Very Important	Challenging/ Very Challenging
"Knowing which repository or repositories are trusted and appropriate for your data."	75%	42%

Note: *Important/Very Important responses are 7-10 on a scale from 0-10 with 0=Not Important and 10=Very Important; Challenging/Very Challenging are 0-3 on a scale form 0-10 with 0=Very Challenging and 10=Very Easy.*

"Knowing which repository or repositories are trusted and appropriate for your data."



6. DATA REUSER TRAINING MODULES

Stakeholder Action Plan Notes

Stakeholder Alignment and Action: *A Living Document v2.0 (July 2021)*

Issue	Data reusers need a way to build skills to ensure appropriate and effective reuse of research data. This is true for researchers at all stages of their careers. Interview and survey results say that this is important but hard to do. Single-point modules (30–45 minutes) can be particularly impactful. What are examples of modules that should be developed for data reusers? Here are examples of possible data wrangling and tidying modules that could be in multiple platforms to support a variety of audiences (e.g. live webinars, recorded webinars, text books, cookbooks): 1. Data import 2. Data manipulation 3. Statistics 4. Graphing data 5. Ethical responsibilities of data generation and reuse 6. How to find and access data 7. Data quality 8. Discipline-specific considerations 9. IRB issues 10. Terms of reuse (licenses)
Vision	***Imagine a world where…*** Researchers have supported time to learn data access and analysis tools such as R and Python (e.g., Summer salary support). Modules/books/training material and workshops are vetted and easily available for different fields (biologists, data scientists, social scientists, geoscientists) and different levels from beginner to more advanced. Menu of options for different learning styles. Online help is available by email (or in-person during workshops), but independent problem-solving is encouraged. You can easily find reusable data that is relevant to your research question. Websites exist where vetted data are available (NCBI, NASA Data Search: https://search.earthdata.nasa.gov/search) Communities of practice exist to support one another with shared tools and data analysis resources (e.g., PANGEO, https://pangeo.io/) Tools and training for combining data sets are available, support cross-disciplinary analysis (e.g., data fusion of selected parameters from disparate sources)

Stakeholders	• Researchers – academic (faculty and staff, graduate and undergraduate students, administrators) • Researchers – government or nonprofit • Researchers – private industry • Researchers – citizen scientists • Educators – graduate, undergraduate, K-12 • Funders – Institutional Support • Funders – government funding agencies • Funders – nonprofit/NGO funders • Funders – private entities • Community/culture builders – professional societies and consortia • Community/culture builders – academic departments, academic leadership • Publishers – society journals • Publishers – other nonprofit journals • Publishers – for-profit journals • Publishers – newspapers (NY Times) • Propellers – disciplinary research data service providers and repositories • Propellers – commercial research data service providers (AWS, and others) • Propellers – open-source repositories (GitHub, Kaggle, OSF, and others) • Propellers – open-source software/system providers (Docker, R, and others) • Propellers – other service providers (CHORUS, DataCite, and others) • Public interest – policy makers • Public interest – general public
Interests	• Facilitate broad ease of data reuse. • Replication and rigor, greater care in data collection. • Knowledge advancement is accelerated. • Computational tool pedagogy, leading practices defined. • Clear training, defined, small-bite modules that reach a broad audience. • Open access to educational materials, some resources listed below. • Consortium/grant for developing educational materials and assessing effectiveness. • Broader sharing of data analysis tools throughout community.

Alignment	Misaligned Interests	Moderate Alignment	Well-aligned Interests
		• Facilitate broad ease of data reuse (proprietary and academic) • Replication and rigor, greater care in data collection • Knowledge advancement is accelerated	• Clear training, defined, small-bite modules that reach a broad audience • Open access to educational materials, some resources listed below • Consortium/grant for developing educational materials and assessing effectiveness • Broader sharing of data analysis tools throughout community • Computational tool pedagogy, leading practices defined
Actions	• 3Q 2021 ◦ Inventory of existing modules ◦ Compose outline of first modules to develop ◦ Platform: identify outlets for modules (webinars, conference workshops to get training started, asynchronous, cross-Institution, pilot) ◦ Identify content curators, vetters, teachers ◦ Outreach to journals and publishers that support shared data • 4Q 2021 ◦ Identify and invite authors for modules that need to be developed ◦ Assess existing modules ◦ Identify funding sources and start writing proposals ◦ Evaluate different platforms ◦ Establish website with resources and guided training outlines for different disciplines, webinars, podcases • 1Q 2022 ◦ Assess new modules ◦ View and evaluate assessment data ◦ Start defining leading practices • 2Q 2022 ◦ Submit proposals for funding ◦ Pilot modules continue		

Potential partners/sponsors for implementation on this topic:

- Societies
- NSF
- Nonprofits (e.g., Sloan foundation)

Individual commitments to support implementation on this topic:

- Outreach to journals and academic associations in breakout group member's own fields to gauge interest
- Serve on a team of people committed to seeing modules developed
- Compiling list of existing modules
- Seek venues for presenting webinars (e.g., Doug Schuster, American Met Society Board on Data Stewardship Webinar series—highlight available resources).

Good training books or websites mentioned in discussion:

- Getting Started with R: An Introduction for Biologists by Andrew P. Beckerman, Dylan Z. Childs, and Owen L. Petchey
- Nasa Data Recipes: https://earthdata.nasa.gov/earth-observation-data/data-recipes
- NCAR Python Xdev: https://ncar.github.io/xdev/
- Pangeo project: https://pangeo.io/
- Examples on data quality, fitness for reuse: https://climatedataguide.ucar.edu/
- Data Help Desk: https://www.esipfed.org/data-help-desk

Related Findings

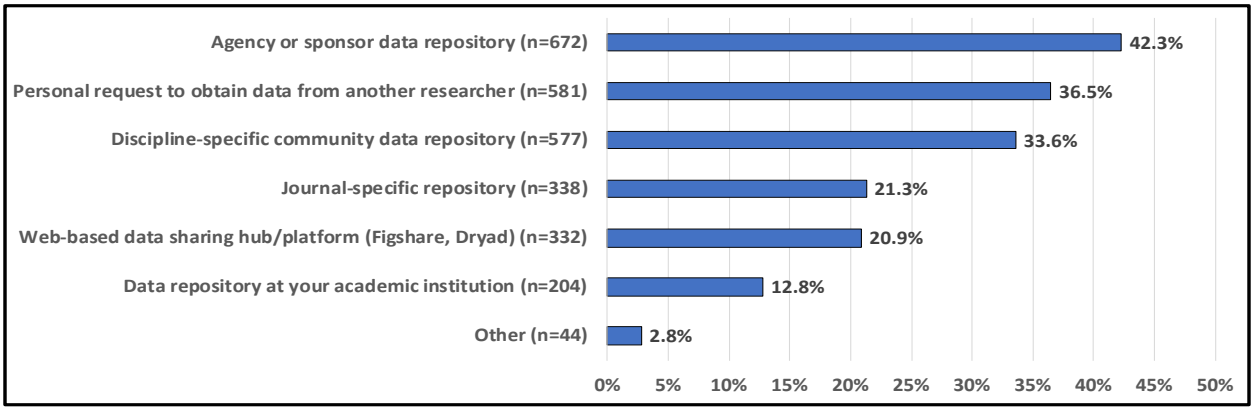
Research Practice Recommendations

2.E. **Training on Data Usage:** Ensure researchers of all career stages have access to professional development on all aspects of data usage.

6.E. **Curriculum for Data Ethics and Handling:** Teach researchers (faculty, staff, students, scientists) about ethical and other considerations around data and open-source sharing, which promotes increased data literacy in broader society.

Survey Results

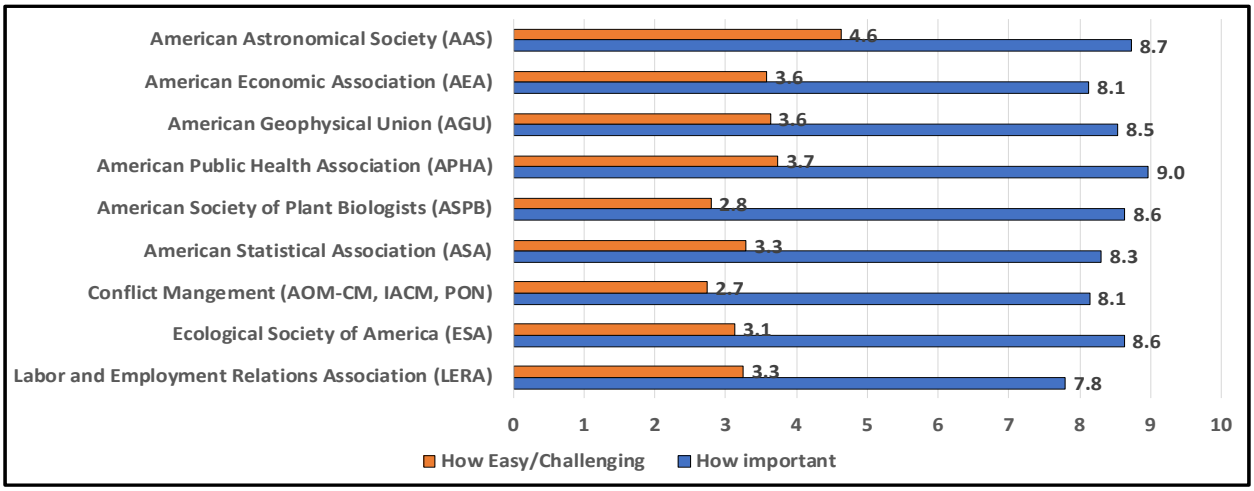
"Where do you obtain data?" (Select all that apply)

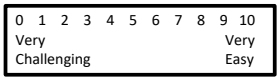


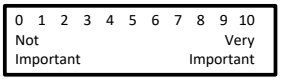
Indicator Issue	Important/ Very Important	Challenging/ Very Challenging
"Finding and getting access to data that has been made available for reuse by others."	89%	58%

Note: *Important/Very Important responses are 7-10 on a scale from 0-10 with 0=Not Important and 10=Very Important; Challenging/Very Challenging are 0-3 on a scale form 0-10 with 0=Very Challenging and 10=Very Easy.*

"Finding and getting access to data that has been made available for reuse by others."







7. DATA SHARING AND REUSE RESEARCH ETHICS

Stakeholder Action Plan Notes

Stakeholder Alignment and Action: *A Living Document v2.0 (July 2021)*

Issue	Data sharing involves ethical considerations around data sharing and reuse. This includes, for example, protecting confidentiality, ensuring appropriate reuse of data, and appropriate citations to give credit to data sharer. What actions are needed to codify ethical guidelines, ensure awareness, and track ethical compliance in data sharing and reuse? How to address variation in ethical guidelines based by countries? It is important to ensure ethical guidelines that relate to human and animal subjects—ethical guidelines for protection. Note that ethics are not only related to sharing and reuse. Also important is the dignity of participants as part of considering ethics of reuse. It is important not to have difference about data sharing for faculty members at different career stage—exceptions result in more and more exceptions. Also, it is important to support early career in the spirit of data sharing. How do we address ethical disputes related to data sharing (FORCE—Future of Research Communication and E-scholarship 11 (2011 started, open science issues)? How to include the committee on Publication Ethics?

Vision	***Imagine a world where…***
	• Researchers share data enthusiastically using leading practices that are well established (working with repository, guidelines (e.g., GDPR), credit to participants)
	  ◦ What are the leading practices?
	  ◦ Level of implementation all over the place, do IRB offices on campuses know about the national and international groups.
	  ◦ Branch out of the academy then IRB does not exist in the same way.
	• Entities which the data is about is excited to have their data shared. Development of trust between participant and researcher.
	• Participants wishes about sharing data is respected, time dedicated is respected because of reuse.
	• Unified understanding of what is required to respect privacy.
	• Who has access and for how long and what can be done with the data?
	• Data is identifiable even when you think it is not.
	• Ways are available to make data anonymous.
	• Participant data is important, but one part of data reuse (e.g., weather predication).

Stakeholders	• Researchers – academic (faculty and staff, graduate and undergraduate students) Note the need to consider the various higher education sectors and various colleges/ universities infrastructures. This is not equitable. • Researchers – government or nonprofit • Researchers – private industry • Researchers – citizen scientists • Funders – government funding agencies • Funders – nonprofit/NGO funders • Funders – private entities • Community/culture builders – professional societies and consortia • Community/culture builders – academic departments, academic leadership • Publishers – society journals • Publishers – other nonprofit journals • Publishers – for-profit journals • Propellers – disciplinary research data service providers and repositories • Propellers – commercial research data service providers (AWS, and others) • Propellers – open-source repositories (GitHub (Note that this is not a repository, more of an infrastructure provider, more of a development platform/communication platform), Kaggle, OSF, Databray and others) • Propellers – open-source software/system providers (Docker, R, and others) • Propellers – other service providers (CHORUS, DataCite, and others) • Public interest – policy makers • Public interest – general public • Participants in research – humans (e.g., minors (for example, patients, organizations) and animal
Interests	• Privacy • Confidentiality • Accessibility • Integrity—attention paid to data being changed/cyber security around data being shared. Connected to being curated and stored . • Integration into workflows. • Propriety views of data. • Time and resources this work will required to make the data accessible and the flip side is also true for participants (even life of animals). • Funding used collect the same data – if data is shared then it would reduce costs (people, animals, environment, time and resources etc.). • Ethical cost of failing to share. • Wasted animals because sample size is not large enough.

Alignment	Misaligned Interests	Moderate Alignment	Well-aligned Interests
	• Security • Privacy • Ethical costs because people want to hoard data • Infrastructure to achieve these goals • Leading practices are created in some spaces but not being implemented in practice		• Agreement in need for security, privacy, credit etc., but misalignment to actualize/achieve/do it
Actions	• Time in the process is too late. Needs to be done earlier in the process not when it gets to publication. Needs to be at the beginning as part of the proposal before data is even collected or reaching publication. • Threat of publishers to not publish paper incentivizes researcher to make data public. • Ethically not all data should be made public. Open data sharing (anyone can get data), but accessibility can also be limited. • Authors should not be deciding who gets access (e.g., agree, are a friend) • Restricted data—who holds the power about sharing it. Develop guidelines. Role of the institution (e.g., grants and contracts offices). • What governs the people that are not within institutions? • Equity in the access to the infrastructures to support this type of data sharing		

Related Findings

Research Practice Recommendations

4.D. **IRB Standards:** facilitate greater standardization of IRB practices and expectations across institutions and other organizations.

5.D. **Modernize IRB Approaches:** Research offices should increase awareness of and work with IRBs to appreciate the need for compliance with data sharing mandates and balanced share requirements against risk (e.g., evaluate templates in use by IRBs, such as provisions to destroy data after five years).

6.A. **Templates to Simplify Sharing:** Provide guidelines, templates, and checklists to help data producers evaluate compliance with requirements (e.g., DMP, confidentiality) and best practices in sharing data (e.g., open-source options, selecting repository).

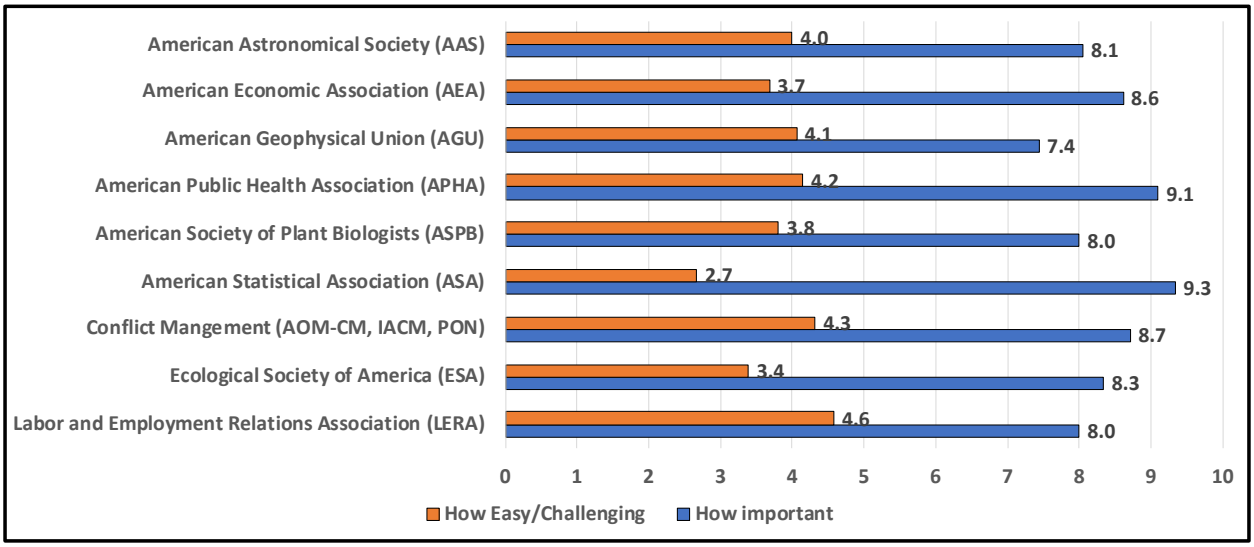
6.D. **Review DMPs:** Review DMPs to ensure compliance with commitments stated in DMPs.

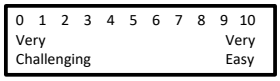
Survey Results

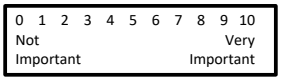
Indicator Issue	Important/ Very Important	Challenging/ Very Challenging
"Knowing how to protect sensitive data before sharing."	82%	55%

Note: *Important/Very Important responses are 7-10 on a scale from 0-10 with 0=Not Important and 10=Very Important; Challenging/Very Challenging are 0-3 on a scale form 0-10 with 0=Very Challenging and 10=Very Easy.*

"Knowing how to protect sensitive data before sharing."







8. EQUITY AND INCLUSIVE DATA SHARING AND REUSE

Stakeholder Action Plan Notes

Stakeholder Alignment and Action: *A Living Document v2.0 (July 2021)*

Issue	Issues of equity and inclusion are woven throughout data sharing and reuse. For example, some research institutions lack resources to provide full support for data sharing, some repositories require extra resources to access (e.g., proprietary repositories, secure data centers such as the Federal Statistical Data Research Centers), and some populations are studied without access to their own data. What initiatives will advance equitable access to data, support for inclusive data sharing and reuse, and propel broader impacts in society through data sharing and reuse?
Vision	***Imagine a world where…*** • Anyone can access any metadata from anywhere (to maximize data findability). • Everyone has the tools/skills to make use of the data they access. • Everyone is trained to be empathetic/compassionate in sharing their data. • All metadata is machine accessible, enabling translation of keywords into a suitable local human language. • There is no inequality baked into the data and its usage.
Stakeholders	• Researchers – academic (faculty and staff, graduate and undergraduate students) • Researchers – government or nonprofit • Researchers – private industry • Researchers – citizen scientists • Funders – government funding agencies • Funders – nonprofit/NGO funders • Funders – private entities • Community/culture builders – professional societies and consortia • Community/culture builders – academic departments, academic leadership • Community/culture builders – academic libraries • Publishers – society journals • Publishers – other nonprofit journals • Publishers – for-profit journals

	• Propellers – disciplinary research data service providers and repositories • Propellers – commercial research data service providers (AWS, and others) • Propellers – open-source repositories (GitHub, Kaggle, OSF, and others) • Propellers – open-source software/system providers (Docker, R, and others) • Propellers – other service providers (CHORUS, DataCite, and others) • Public interest – policy makers • Public interest – general public • Public interest – participants in research studies, populations who are the subjects of the data • Public interest – Underrepresented groups • Note: All stakeholders are responsible for ensuring equity and inclusion in all aspects of their work.
Interests	• Autonomy over data about oneself/the right to be forgotten/the right to privacy. • Funders of research to obtain maximum value from their investment in data generation. • Data to be used in a way which is equitable to all stakeholders. • Data to be collected in an equitable way. • Inclusion of underrepresented voices in decision making processes. • Underpopulated pathways for minorities into careers with research data and computing.

Alignment	**Misaligned Interests**	**Moderate Alignment**	**Well-aligned Interests**
	• Underpopulated pathways for minorities into careers with research data and computing	• Autonomy over data about oneself/the right to be forgotten/the right to privacy	• Funders of research to obtain maximum value from their investment in data generation • Data to be used in a way which is equitable to all stakeholders • Data to be collected in an equitable way • Inclusion of underrepresented voices in decision making processes

Actions	

	• 3Q 2021
	  ◦ Audit of current data sharing processes and their impacts on equity
	  ◦ Identify underrepresented voices, invite to participate
	• 4Q 2021
	  ◦ Draft action plan based on audit
	  ◦ Ask stakeholders to comment on draft action plan
	• 1Q 2022
	  ◦ Revise draft action plan based on stakeholder feedback
	  ◦ Policy development to include stated consequences for non-compliance
	• 2Q 2022
	  ◦ Refine and revise policy after stakeholder feedback in an agile method
	  ◦ Share learnings from this project with equivalent funding agencies (1 domestic; 1 international)

Potential partners/sponsors for implementation on this topic:

- NSF
- NIH
- Gordon and Betty Moore Foundation

Related Findings

Research Practice Recommendations

4.A. **Equity in Resources for Data Sharing:** Increased equity among large and small institutions by considering consortia for developing and supporting infrastructure and resources needed for quality data sharing.

Survey Results

"If you could use one phrase or metaphor to summarize your vision of success for creating, sharing and reusing publicly accessible research data, what would it be?"

- Eliminate the barriers to use of data by non-academic researchers.
- A world in which I have the same access to administrative data as those from top universities.
- Research data as a public good—one use of research data does not limit other uses of the same data.
- Our science is better if we raise one another up.
- Data sharing for sustainable democracy—now and for our future
- The house we all build together

9. JOURNAL DATA SHARING POLICIES

Stakeholder Action Plan Notes

Stakeholder Alignment and Action: *A Living Document v2.0 (July 2021)*

Issue	Journals vary in how they handle data associated with articles (e.g., paywall repositories, allowing authors to make data available upon request, no policy, etc.). Journals may or may not support researchers in depositing research data in disciplinary or other repositories. Note that journal policies often implicitly assume that the data is generated for the publication. What models will support public access to data associated with journal publications while minimizing data sharing costs for journals? Who is accountable? Do we (editors, publishers) need a reviewer team and/or dedicated staff for data? Question of code versus data? How much data that you share? Just data that is in the figure or all of the data sets? Need nuanced approach to data sharing that takes into account larger data sets, MOUs, etc. Some disciplines may have high level of expected document/metadata associated with data, and that can be a barrier. Data is also found embedded in-line in manuscripts and/or supplemental/supporting Information docs (often PDF in final form). How to transition "traditional analog" methods to prepare "sharable" data vs. accessible datasets?
Vision	***Imagine a world where…*** • Researchers can check the data analysis, replicate, and build on any published research. • Shared data leads to novel findings, especially in ways that cross disciplines. • We are able to use data 20 years from now and see the roots of our discipline, using new data analytical tools on foundational data sets. • All data that comes into the journal are FAIR and equitable (not just R1s but 1890s schools as well). • Sharing data does not increase administrative or cost burdens. • Grants include the cost for doing this work. • Clear for researchers how to share their data and easy to find a repository. • Data is citable with persistent identifier and trackable. • Easy to find and identify what data are open in various disciplines. • Data are curated, easy to replicate, and build on the research in a paper—rated as first-class output (to support access, archive/preservation if warranted).

Stakeholders	• Researchers – academic (faculty and staff, graduate and undergraduate students) • Researchers – government (especially survey/stats divisions in agencies) or nonprofit • Researchers – private industry (note: Consider industry as first-class stakeholder for some fields/disciplines, such as Chemistry (user & contributors) and Agriculture (contributors and users)) • Researchers – citizen scientists • Funders – government funding agencies • Funders – nonprofit/NGO funders • Funders – private entities • Community/culture builders – professional societies and consortia, standards societies/ unions • Community/culture builders – academic departments, academic leadership • Publishers – society journals • Publishers – other nonprofit journals • Publishers – for-profit journals • Propellers – disciplinary research data service providers and repositories • Propellers – commercial research data service providers (AWS, and others) • Propellers - commercial service providers (institutional impact) • Propellers – open-source repositories (GitHub, Kaggle, OSF, and others) • Propellers – open-source software/system providers (Docker, R, and others) • Propellers – other service providers (CHORUS, DataCite, and others) • Propellers – for-profit vendors that create data capture/collection and analysis (e.g. speech recognition) • Public interest – policy makers and decision-makers • Public interest – general public
Interests	• Time • Money • Having consensus/buy-in • Control and influence • Recognition and credit • Risk mitigation • Public interest, public good, public impact (from local to national to global) • Global standing (who leads in data sharing) • Market share • Openness • Transparency • Ethics (sensitive information) • Reproducibility • Security and economic impact

<table>
<tr><td></td><td colspan="3">

- Return on Investment of the sharing
- Quality of the data
- Accountability (no one wants to own/pay for this challenge)
- Embargo and timeliness and exclusivity (Fear of being scooped)
- Fear of being harassed (climate change, feminist scholarship)
- Data equity
- Easy to access and share data
- Intellectual property

</td></tr>
<tr><td>Alignment</td><td>Misaligned Interests</td><td>Moderate Alignment</td><td>Well-aligned Interests</td></tr>
<tr><td></td><td>

- Control and Influence
- Profitability for publishers (data analytics and journals)
- Time
- Global standing
- Market share
- Accountability
- One-stop shopping

</td><td>

- Ethical use/sharing of data (sensitive data)
- Security and economic impact
- Equity
- Embargo and exclusivity
- Intellectual property
- Recognition
- Openness
- Transparency
- ROI of data sharing
- Fear of harassment

</td><td>

- Efficiency
- Reducing cost
- Higher quality data
- Reproducible
- Public interest
- Risk mitigation
- Credit and tracking
- Ease of access

</td></tr>
<tr><td>Actions</td><td colspan="3">

- 3Q 2021
  - Sign on to COPDESS/Make Data FAIR agreement.
- 4Q 2021
  - Develop a COPDESS-like process for other disciplines.
  - Develop community-based discipline-specific practices in data sharing—building toward data standards across journals in a discipline. Maybe work with librarians with disciplinary expertise, identify other stakeholders to collaborate (e.g., standards bodies).
  - Plan for a special issue that shares the outcomes from this work?
- 1Q 2022
  - Develop tools and resources for researchers, e.g., template for readme file to accompany data archive for publication. Reduces number of conversations needed with the researcher.
  - This could happen earlier as some templates exist that could be built from
- 2Q 2022
  - General parameters for data editor role?
  - Develop a book to share success of data sharing.

</td></tr>
</table>

Potential partners/sponsors for implementation on this topic:

- NASEM RT on Incentives for Open Science
- NSF/NIH funding to support disciplinary convenings.
- CODATA, ISC (International Science Council)
- Publisher groups? (STM)

Individual commitments to support implementation on this topic:

- American Society of Plant Biologists

Related Findings

Research Practice Recommendations

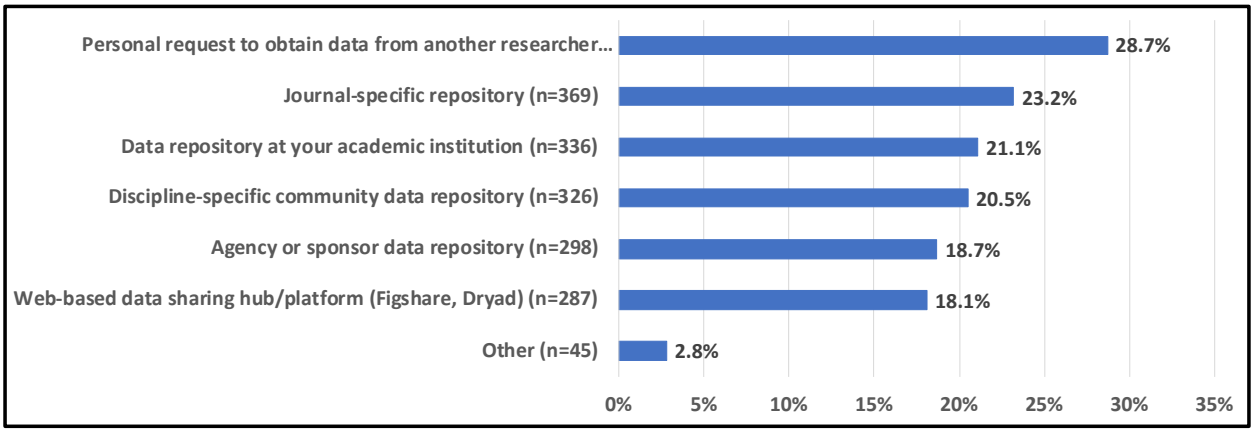
3.D. **Journal Requirements:** Work with journals to develop consistent approaches and require authors to give cite data and credit data producers.

5.A. **Organizational Leadership:** Publishers, societies, and funding organizations should support data publication with guidance on domain standards and via incentives (e.g., teeth in DMPs, consideration in tenure and promotion).

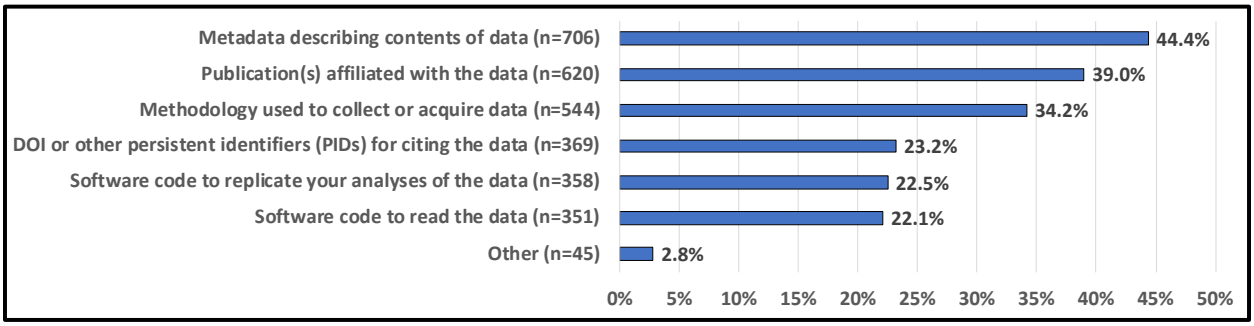
6.G. **Support FAIR Data and Reproducibility/Replicable Practices:** Establish a closer dialog between funders, publishers, and repositories to ensure publication of data and materials for reuse, reproducibility, and replicability are required and support producers/authors in meeting these requirements (e.g., with code checking, removing limits on methods section).

Survey Results

"How do you make your data accessible to the public (select all that apply)?"



"When you make data accessible, what documentation do you typically share with your data?"



10. PLANNING TO SHARE AND REUSE DATA

Stakeholder Action Plan Notes

Stakeholder Alignment and Action: *A Living Document v2.0 (July 2021)*

Issue	Producers experience challenges in preparing data, with proper documentation, when they have not planned ahead. Too often, they do not know what materials need to be shared with data to support its effective and appropriate reuse. Interview and survey results say that this is important but hard to do. What actions can be taken to promote advance planning for the sharing and reuse of data? How are privacy issues and confidentiality assured for sensitive data (human subjects, PII, informed consent, etc.)? Absent planning things won't change! What are reasonable embargo periods? Need to think like a data user to make sure that all necessary information is collected up front. Cultural issue is to convince data producers of the value in sharing. Before one can "plan to share and reuse data", one needs to be convinced of the value of doing so.
Vision	***Imagine a world where…*** • Data capture that facilitates sharing is easy and embedded in the research practice because researchers have been convinced of the value of openly sharing data. • Data are born FAIR. • Good data management practices and skills are intrinsic and part of the training of research practitioners. • Data can be effectively utilized by a broad range of literate people. • Data are collected with broader utility, beyond the current research question and field, borne in mind. • It is easy to discover, access, and assess data to be reused. • Data sharing earns proper credit and recognition. • Researchers value data sharing before they begin (don't need to be convinced). • The cost of data management is accommodated in research funding. • Planning includes end-of-life decision-making criteria.
Stakeholders	• Researchers – academic (faculty and staff, graduate and undergraduate students) • Researchers – government or nonprofit • Researchers – private industry • Researchers – citizen scientists • Researchers – community-based partners • Researchers – Data contributors • Funders – government funding agencies • Funders – nonprofit/NGO funders • Funders – private entities

	• Community/culture builders – professional societies and consortia • Community/culture builders – academic departments, academic leadership • Publishers – society journals • Publishers – other nonprofit journals • Publishers – for-profit journals • Propellers – disciplinary research data service providers and repositories • Propellers – commercial research data service providers (AWS, and others) • Propellers – open-source repositories (GitHub, Kaggle, OSF, and others) • Propellers – open-source software/system providers (Docker, R, and others) • Propellers – other service providers (CHORUS, DataCite, and others) • Public interest – policy makers, media • Public interest – general public
Interests	• Reproducibility, reliability, rigor, transparency • Public trust in research results • Priority of discovery and associated consequences of missing out • Sustainability, curation, preservation, managing cost • Liability, confidentiality, and re-identification possibilities • Embarrassment • Recognition • At this point it is extra work to plan and prepare data for sharing; unfunded mandate • Data management plans that are carefully prepared, maintained, and followed through; how is adherence checked and what are the ramifications for failing? • Huge diversity of data types, including proprietary data • Discipline-based standard formats, minimum metadata conventions • Dynamic data • Retention

Alignment	Misaligned Interests	Moderate Alignment	Well-aligned Interests
	• Embarrassment • Sustainability, curation, preservation, managing cost • Priority of discovery and associated consequences of missing out • Retention • At this point it is extra work to plan and prepare data for sharing; unfunded mandate	• Liability, confidentiality, and re-identification possibilities • Data management plans that are carefully prepared, maintained, and followed through; how is adherence checked and what are the ramifications for failing? • Recognition • Huge diversity of data types, including proprietary data • Discipline-based standard formats, minimum metadata conventions • Dynamic data	• Reproducibility, reliability, rigor, transparency • Public trust in research results
Actions	• 4Q 2021 ◦ Inventory and survey of resources related to above misaligned interests. Discuss with NAS Roundtable on Aligning Incentives for Open Science. • 2Q 2022 ◦ Reconvene: what progress have we made since the June 2021 workshop?		

Potential partners/sponsors for implementation on this topic:

- NIST, Research Data Framework (RDaF)
- AAU, APLU, ARL, CHORUS, DataCite, ORCID
- RDA, CODATA, GO-FAIR, WDS (Data Together)
- Foundations (Sloan, etc.)
- NSF/XSEDE
- ICPSR
- GitHub, Zenodo, Figshare, Dataverse, Dryad, Project Jupyter, re3data, Knowledge Network for Biocomplexity
- OHRP
- Professional societies

Individual commitments to support implementation on this topic:

- NIST RDaF Initiative
- AMS Board on Data Stewardship
- Helmsley Charitable Trust on implementation of data sharing in clinical and observational health studies

Related Findings

Research Practice Recommendations

4.B. **Advance Repository Consultations:** Consult with institutional or domain repository before launching into a proposal and DMP to better anticipate what will be needed for curating and depositing data that will maximize reuse by others.

4.C. **Data Governance:** Include the Institutional Review Board (IRB) and other offices in advance planning and education on the types of data that will be generated and used to ensure appropriate use and confidentiality protections (e.g., other forms of restricted data such as proprietary data, issues of data governance, DUAs, licensing).

4.E. **Understand Potential Data Reusers:** Talk to potential data reusers in advance of beginning the project to anticipate what they will need.

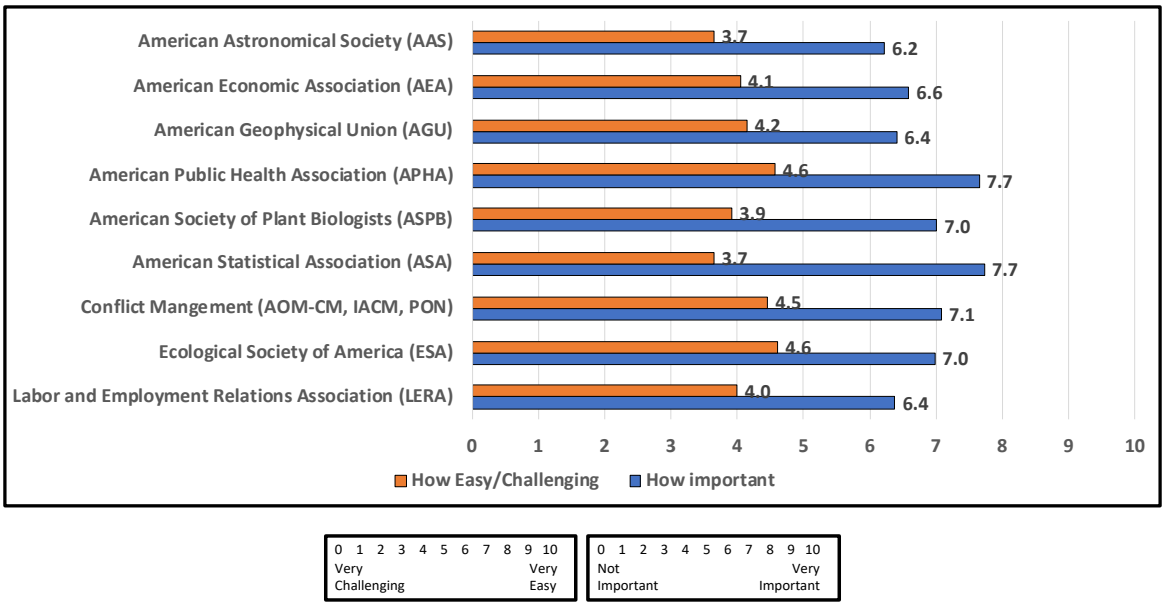
6.B. **Lowering Barriers:** Develop templates to assist data producers (in addition to what comes from IRBs).

Survey Results

Indicator Issue	Important/ Very Important	Challenging/ Very Challenging
"Planning for how the data will be documented and shared before you start your research study."	62%	49%

Note: *Important/Very Important responses are 7-10 on a scale from 0-10 with 0=Not Important and 10=Very Important; Challenging/Very Challenging are 0-3 on a scale form 0-10 with 0=Very Challenging and 10=Very Easy.*

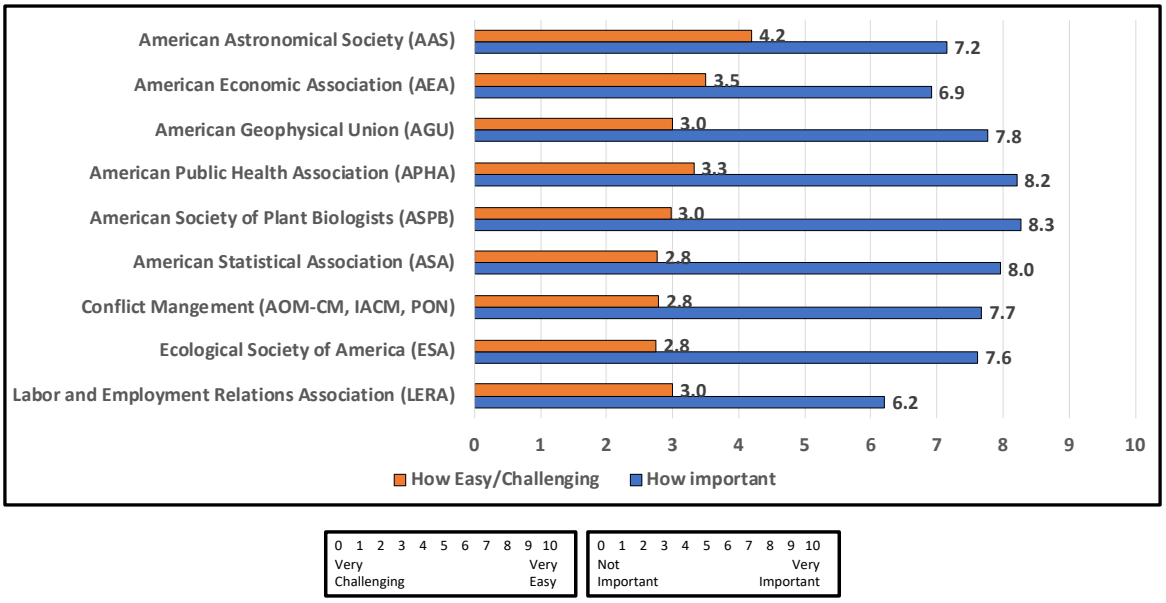
"Planning for how the data will be documented and shared before you start your research study."



Indicator Issue	Important/ Very Important	Challenging/ Very Challenging
"The availability of standards and controlled vocabularies to facilitate effective sharing and reuse of your data"	74%	64%

Note: *Important/Very Important responses are 7-10 on a scale from 0-10 with 0=Not Important and 10=Very Important; Challenging/Very Challenging are 0-3 on a scale form 0-10 with 0=Very Challenging and 10=Very Easy.*

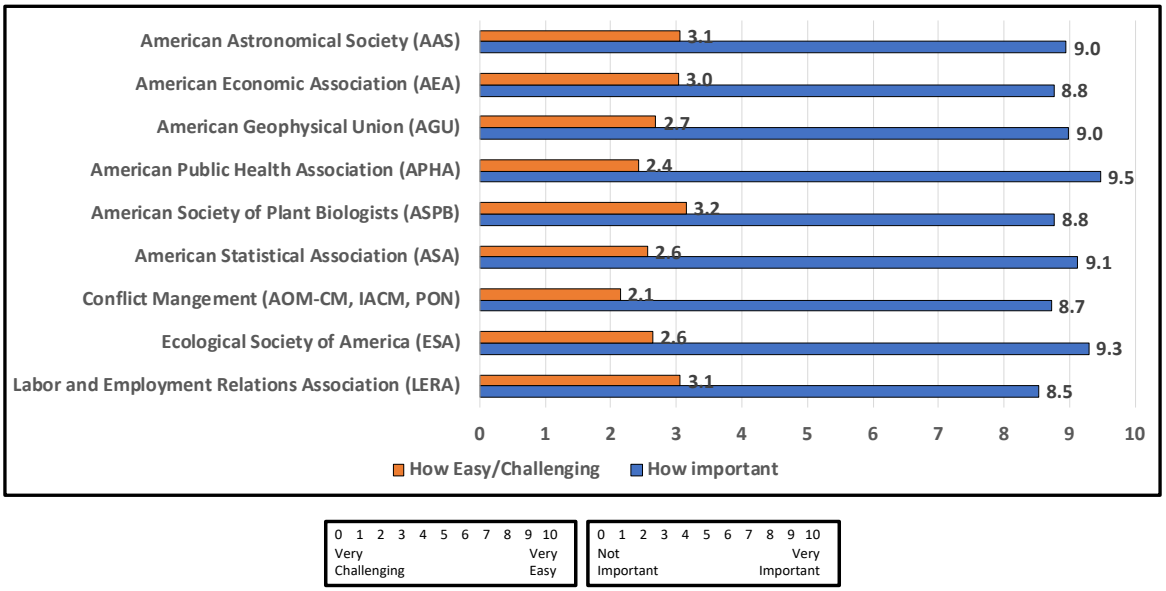
"The availability of standards and controlled vocabularies to facilitate effective sharing and reuse of your data"



Indicator Issue	Important/ Very Important	Challenging/ Very Challenging
"Ensuring that there is clear, accurate and complete metadata and documentation associated with shared data."	93%	69%

Note: *Important/Very Important responses are 7-10 on a scale from 0-10 with 0=Not Important and 10=Very Important; Challenging/Very Challenging are 0-3 on a scale form 0-10 with 0=Very Challenging and 10=Very Easy.*

"Ensuring that there is clear, accurate and complete metadata and documentation associated with shared data."



Indicator Issue	Important/ Very Important	Challenging/ Very Challenging
"Having confidence in the quality of data made available for reuse by others."	95%	66%

Note: *Important/Very Important responses are 7-10 on a scale from 0-10 with 0=Not Important and 10=Very Important; Challenging/Very Challenging are 0-3 on a scale form 0-10 with 0=Very Challenging and 10=Very Easy.*

"Having confidence in the quality of data made available for reuse by others."