2023 |

Final Report

Workshop on Open Data and Reuse in Social Science Weather Research

Final Report

Workshop on Open Data and Reuse in Social Science Weather Research

Workshop Organizers and Participants

This workshop was organized by the National Oceanic and Atmospheric Administration Weather Program Office and the Natural Hazards Center at the University of Colorado Boulder. The following people participated in the workshop and/or contributed to the development of this report:

> Alison M. Agather, Austin Becker, Ryan Berkheimer, Ann Bostrom, Jolie Breeden, Melissa H. Cantrell, Lauren Clay, Tim Cockerill, Joe Conran, Julie Demuth, Gustavo Durand, Chris Ellis, Gina Eosco, Maria Esteva, Jennifer Fagan-Fry, Joe Fillingham, Kim Fortun, Julian Gautier, Alan Gerard, Jen Henderson, Stephanie Hoekstra, John Ten Hoeve, David Hondula, Dessislava Kirilova, Jeremy Koster, Alison Krepp, Tony LaVoi, Ji Sun Lee, Victor Marchezini, Matt Mayernik, Kim Klockow McClain, Katie Mika, Rebecca E. Morss, Jonathon Mote, Alexandra Neal, Robert O'Connor, Mary Angelica Painter, Lori Peek, Amy Pienta, Patricia Pinto da Silva, Aditya Ranganath, Seann Regan, Joe Ripberger, Anaís Delilah Roque, Nathanael Rosenheim, Sameer H. Shah, Valerie Were, Cassandra Shivers-Williams, Sudhir Shrestha, Jennifer Tobin, Jamie Vickery, Castle Williamsberg, and Ariela Zycherman.

See the workshop webpage for a complete participant list with contact information for all contributors.

Suggested Report Citation

> Peek, Lori, Mary Angelica Painter, Gina Eosco, Alison M. Agather, Jonathon Mote, Cassandra Shivers-Williams, Jennifer Tobin, Jolie Breeden, and Open Data Workshop Participants. 2023. Final Report of the Workshop on Open Data and Reuse in Social Science Weather Research. Natural Hazards Center, University of Colorado Boulder. hazards.colorado.edu/research/weather-ready-research/open-data-workshop

Acknowledgements

This report is based on work supported through supplemental funding from the National Oceanic and Atmospheric Administration (NOAA) Weather Program Office to the National Science Foundation (NSF Award #1635593). Any opinions, findings, conclusions, or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of NOAA, NSF, or the Natural Hazards Center at the University of Colorado Boulder.

"Open Science is the principle and practice of making research products and processes available to all, while respecting diverse cultures, maintaining security and privacy, and fostering collaborations, reproducibility, and equity." –Science.gov

Executive Summary

This report summarizes the outcomes from the 2023 Workshop on Open Data and Reuse in Social Science Weather Research. The workshop brought together social science researchers, data experts from data repositories and libraries, and federal agency representatives.

The workshop was organized in response to the 2022 White House Office of Science and Technology Policy (OSTP) Memorandum on Ensuring Free, Immediate, and Equitable Access to Federally Funded Research. The memorandum delivered updated guidance to federal departments and agencies on public data access policies. This policy guidance helps ensure the results of publicly funded research—including data and publications—are made publicly accessible without delay or cost.

The overarching goal of the workshop was to gather insights and produce community-informed recommendations on the ethical publication and reuse of social science data in weather-related research. The table that follows summarizes nine key observations that emerged during the workshop, as well as recommended actions for NOAA and other federal agencies, data repositories, social science researchers, and the broader weather research community.

Key Observations	Recommended Actions
Observation 1: Knowledge of the open data policy change was varied.	✓ Federal agencies should clarify and widely share their plans for implementing the 2022 OSTP guidance. ✓ Federal agencies, data repositories, and social science researchers should partner to create open data and data reuse guidance, templates, and training materials.
Observation 2: Key terms such as "data," "timely," "immediate," "free," and "equitable" are not clearly defined in existing guidance.	✓ Federal agencies should define all key terminology in open data and reuse guidance. ✓ Terminology should be reviewed for accuracy, comprehensiveness, and clarity by those affected by new agency requirements. ✓ Federal agencies should provide flexible guidance on what constitutes data and consider data types relevant to social science researchers.

Observation 3: The emphasis on equity in the 2022 OSTP guidance represents an opportunity to address systemic inequities in research and advance ethical guiding principles for open data.	✓ Equity should be centered in all future guidance from federal agencies. ✓ Federal agencies should draw on existing community-driven guiding principles to advance equitable *and* ethical data publication and reuse strategies.
Observation 4: Stringent data publication requirements could potentially disincentivize the collection of sensitive data.	✓ Models for ethically publishing sensitive data should be developed collaboratively and shared. ✓ Social science researchers should receive additional training on best practices for protecting participant privacy while adhering to open data standards.
Observation 5: There are many certified data repositories where researchers can securely store and publish data and results from federally funded research.	✓ Federal agencies should recommend a select number of data repositories for social science data publication. ✓ Data repositories should receive additional funding to offer training on best practices for data publication, data citation, and data discovery, among other topics.
Observation 6: The 2022 OSTP guidance could positively influence interdisciplinary data integration and reuse efforts.	✓ Open data stakeholders should create metadata standards and data protocols that enable interdisciplinary integration for weather-related research. ✓ Federal agencies should track and promote published datasets to encourage data integration. ✓ Additional guidance and future convenings should focus explicitly on data reuse.
Observation 7: The 2022 OSTP guidance does not address requirements or incentives for publishing research protocols and instruments, independent of data.	✓ Federal agencies should provide guidance on how to make social science protocols, data collection instruments, and other research materials freely available to the public. ✓ Federal agencies should continue to incentivize the publication of research instruments and protocols.
Observation 8: Social science researchers lack sufficient funding to make all publications free and immediately available.	✓ Federal agencies should consult experts and tap available resources to ensure publications are made freely available without violating publishing rules or incurring elaborate fees.

	✓ Federal agencies that do not currently require investigators to share their publications should review policies from agencies with such requirements in place.
Observation 9: Fully understanding and implementing the 2022 OSTP guidance will take time, additional funding, and other forms of support.	✓ Data repositories and federal agencies should engage in regular dialogue about the resources needed to develop and sustain data infrastructure over time. ✓ Federal agencies should consider how they can facilitate additional time investigators will need to ensure compliance with new data publication requirements. ✓ Social science researchers should develop thorough data management and publication plans at the award proposal phase.

"Open data are generally defined as data that are freely and publicly available to any user." –Kassam-Adams and Olff, 2020

Introduction

In April 2023, the National Oceanic and Atmospheric Administration (NOAA) Weather Program Office and the Natural Hazards Center at the University of Colorado Boulder partnered to convene a two-day workshop on Open Data and Reuse in Social Science Weather Research. The workshop brought together social science researchers, data experts from data repositories and libraries, and federal agency representatives from NOAA and the National Science Foundation (NSF).

The workshop was organized in response to the 2022 White House Office of Science and Technology Policy (OSTP) Memorandum on Ensuring Free, Immediate, and Equitable Access to Federally Funded Research. The 2022 OSTP memorandum delivered updated guidance to federal departments and agencies on public data access policies. This policy guidance helps ensure that the results of publicly funded research—including data and publications—are made publicly accessible without delay or cost. According to the new guidance, all federal agencies will fully implement updated policies no later than December 31, 2025.

This most recent guidance is reflective of a more general trend toward open science and efforts to integrate the principle of equity throughout the federal government. For example, the 2022 OSTP memo builds upon other policies, such as the 2019 requirement that every federal agency have a chief data officer and the 2013 OSTP Memorandum that directs federal agencies to increase access to the results of federally funded research. Further, it augments the 2022 vision for equitable data, which provides recommendations for increasing data available for measuring equity and representing the diversity of the American people.

The 2022 OSTP guidance is significant because it accelerates the timeline for sharing data and publications, while also broadening data access for the public. Further, the guidance promotes equity in research by encouraging agencies to ensure that data and results are returned to historically underserved and underrepresented communities. It also seeks to increase public trust in taxpayer-funded research by enhancing scientific research integrity, accountability, and transparency.

Workshop Overview

The workshop described in this report focused on the publication of weather-related social science research data, challenges in ethically managing and sharing qualitative and quantitative data, and opportunities to support the infrastructure and culture needed for data publication and reuse. A total of 39 people attended the in-person meeting at the University of Colorado Boulder and an additional 12 people participated online. The participants represented three primary groups:

- **Social science researchers** with experience in multiple data types, sensitive data issues, publishing social science instruments and data, and multidisciplinary and interdisciplinary collaborations in weather-related research.

- **Data scientists and data librarians** with expertise in data governance and experience creating or maintaining the repositories necessary for securely managing, storing, and publishing social science data.

- **Federal agency representatives** from NOAA and NSF who evaluate and fund social scientific and interdisciplinary research.

The following four objectives guided the workshop: (1) **connect** researchers, data repository representatives, and federal funders; (2) **clarify** federal expectations of data publication and reuse; (3) **review** progress and pitfalls in social science data publication and reuse; and (4) **offer** actionable recommendations to NOAA and other federal agencies, the academic research community, and data repositories.

The workshop sessions were designed to be in dialogue with the broader open science movement, which strives to make research more inclusive, democratic, and transparent and allows for greater reproducibility, sharing, and collaboration. Open data—data that can be used and reused freely—is one cornerstone of this movement. At the workshop and in this report, the focus is primarily on open data. However, workshop presentations and discussions also focused on other open science aspects, such as open access publications and the process of publishing and sharing research instruments and protocols.

See the workshop webpage for a full description, the final agenda, and other supplementary materials.

Observations and Recommended Actions

Workshop participants identified nine key observations that underscore the challenges and promise of the recent open data guidance. The observations are presented below, along with corresponding recommendations that the group made to address each one.

Observation 1: **Knowledge of the open data policy change was varied.**

Federal agency and data repository representatives were keenly aware of recent and impending shifts in open data policy and had launched efforts in response. For example, both NOAA and NSF reported their agencies had already formed working groups or subcommittees in response to the 2022 OSTP memo and the guidance that preceded it. Representatives from the four data repositories who attended the meeting—DesignSafe-Cyberinfrastructure, Harvard Dataverse, Inter-university Consortium for Political and Social Research (ICPSR), and Qualitative Data Repository—were aware of changes in federal guidance and, in the case of one repository, had already hired additional staff in preparation for increased data publication activity. Data librarians from the University of Colorado Boulder Libraries were also aware of the changes and, in response, organized informational sessions to share policy updates with federally funded researchers at the university.

Social scientists at the workshop were generally less familiar with the specifics outlined in the memo. Most indicated there was little to no dialogue in their departments or research units about the open data movement and related policy directives. As social science attendees learned more about the extent of the changes to federal policy, some expressed concern about the impacts on research and how agencies would interpret and potentially enforce them.

Recommendations:

Federal agencies should clarify and widely share their plans for implementing the 2022 OSTP guidance.

- Agencies should offer briefing sheets, informational web pages, and/or other materials that explain how they are responding to the guidance and expectations for funded investigators.

- Providing written direction about data publication expectations at both the proposal *and* annual grant reporting stages would help to socialize investigators regarding open data requirements while also setting expectations at the beginning and through the duration of a funded project.

- Agencies could hold online and in-person meetings that feature, for example, leaders from federal agencies who can explain how the guidance will impact the research community, representatives from data repositories who can highlight the tools and capabilities they offer, social science researchers who can share strategies for successful data publication and reuse, and journal editors who can discuss data publication requirements.

- Agencies should be encouraged to set clear and consistent expectations and metrics for compliance. They should also communicate the consequences of noncompliance.

Federal agencies, data repositories, and social science researchers should partner to create open data and data reuse guidance, templates, and training materials.

- Data repositories have already developed many materials that could help the social science research community and federal agencies respond to the 2022 OSTP guidance. For example, repositories offer clear-use guidance, curation assistance, quality assurance, interactive training sessions, and other services and resources.

- Social science researchers are encouraged to develop new undergraduate and graduate courses and assignments to make students aware of open data requirements for federally funded research. It is especially important to reach graduate students writing theses or dissertations with the support of federal funding.

- Federal agencies should support the development of online trainings that emphasize the importance of open data, the process for publishing data, and the ethical and practical aspects of reusing data. For example, a new CONVERGE Training Module could be created that specifically focuses on data publication and reuse in the social sciences.

- Materials should be widely shared in user-friendly and accessible formats, such as check sheets, case study examples, or brief videos. It is especially important to share materials with groups—such as students or academic social scientists—who may not yet have a record of data publication. Furthermore, partnering with organizations that have a wide reach in the weather-related research community—such as the American Geophysical Union and the American Meteorological Society—could help communicate information to more people.

Observation 2: Key terms such as "data," "timely," "immediate," "free," and "equitable" are not clearly defined in existing guidance.

Participants agreed the vision and intent of the 2022 OSTP guidance is laudable, but several questions were raised about operationalizing key terms used throughout the memorandum. Perhaps most fundamentally, social scientists questioned what constitutes data and, additionally, what data types and research materials will require publication. Social scientists collect and analyze widely varying data types, in different contexts,

and with different degrees of institutional and respondent protection needs. Data collection methods across the social sciences are also diverse, including approaches that generate quantitative and qualitative data. Further, social scientists create research protocols, instruments, data codebooks, and other materials that inform research design, data collection, and analysis. Participants wondered if products not traditionally considered data—such as field notes or observer/interviewer perceptions—would be part of data publication requirements, as this contextual information is often key to interpreting social science data.

Because data collection approaches vary widely in the social sciences, participants were unsure what making data "immediately available" or "available in a timely manner" would mean in practice. It was also unclear how researchers will make all data and publications "free and open to the public" when they may contain sensitive information or reside behind journal paywalls.

Representatives of the social science research community were especially concerned about conflicting guidance from funders and other institutions. For example, some Institutional Review Boards (IRBs) still recommend the destruction of data after data collection ends, which directly opposes federal guidance. Furthermore, what is considered "equitable" varies across institutions and communities. For example, some representatives of historically marginalized communities request that data and findings be shared with them before public release.

Recommendations:

Federal agencies should define all key terminology in open data and reuse guidance.

- Clear definitions of terms are essential for interpreting and enacting the 2022 OSTP guidance. For example, if an agency defines the "timely release of data" as a certain timeframe following the end of data collection, they should state this plainly, including their rationale for the definition and any exceptions.

Terminology should be reviewed for accuracy, comprehensiveness, and clarity by those affected by new agency requirements.

- Establishing working groups to review terminology and written guidance would likely encourage members of the weather research community to be engaged and to help set requirements that impact them. Working group members should be functionally diverse in terms of their skills, disciplines, career stages, and organizational affiliations. In addition, they should be composed of people who are demographically diverse to ensure that additional perspectives and ideas are represented.

- Working group members who review terminology should be compensated and given ample time to review and provide feedback on documentation.

- Future workshops could be organized to gather public comment on terminology and guidance materials. Such convenings might include groups similar to this workshop (i.e., federal agency representatives, representatives from data repositories, and social science researchers), as well as journal editors, researchers from other disciplines, and other potentially impacted groups. IRB representatives are also important partners to include to ensure that potentially conflicting terminology and guidance are addressed.

Federal agencies should provide flexible guidance on what constitutes data and consider data types relevant to social science researchers.

- Workshop participants generally recommended using a clear but broad and inclusive approach in defining what constitutes data. Participants believed this would both encourage researchers to publish their data and improve the potential for data reuse.

- Agencies should issue specific guidance for data types common to different disciplines. For example, in the social sciences, researchers often collect survey data, interview data, observational data, geospatial data, and focus group data. For such data, agencies might want to consider appointing experts to help craft guidance and answer questions regarding data publication and reuse for each data type.

- Any definition of data should be reviewed periodically, as artificial intelligence and other emerging technologies could change what constitutes data and data authorship.

Observation 3: The emphasis on equity in the 2022 OSTP guidance represents an opportunity to address systemic inequities in research and advance ethical guiding principles for open data.

Participants found the strong focus on equity in the 2022 OSTP guidance reassuring and beneficial. They viewed this emphasis as an important step to ensure that all members of the public have access to data and findings from federally funded studies. They also saw this as an opportunity for NOAA and other federal agencies to respond to systemic inequities in data collection practices and research.

Some participants were worried, however, that the policy—if not implemented using an equity lens—could have unintended consequences that exacerbate existing inequities. For example, minority-serving institutions such as Historically Black Colleges and Universities (HBCUs) that have been systematically underfunded and often operate with inadequate resources might be less likely to have access to the training, data infrastructure, and financial support necessary to ensure compliance with the policy guidance. If compliance is tied to future award eligibility, this, in turn, could worsen existing funding disparities by hindering funding access for institutions that are already resource constrained.

Researchers who conduct studies with marginalized groups and in historically marginalized communities expressed concern about how new rules could negatively impact these communities. Because Black, Hispanic, Asian, Indigenous, and other people of color have histories of being subjected to unethical, extractive data collection practices, new government rules and oversight may be viewed with mistrust and apprehension. Therefore, workshop participants agreed that it is imperative that any new directives be community-informed and grounded in a comprehensive ethical framework.

Recommendations:

Equity should be centered in all future guidance from federal agencies.

- Centering equity will require thoughtful and ongoing consultation with members of historically underfunded and oppressed groups and representatives of minority-serving institutions. Members of such groups should be involved in key leadership and decision-making roles and engagement activities.

- Federal agencies should reduce systemic barriers to applying for funding, collecting disaggregated data that can be broken down by key demographic variables, and meeting reporting expectations. For example, additional support could be offered to write a comprehensive data management plan or professional assistance could be made available to support data publication and reuse activities as part of research grants and other funding opportunities.

Federal agencies should draw on existing community-driven guiding principles to advance equitable *and* ethical data publication and reuse strategies.

- Communities with histories of unjust treatment in research may have a special interest in understanding and influencing the core guiding principles that underlie data publication or reuse rules. In response, federal agencies should use or build upon community-driven and widely accepted principles for equitable and ethical data publication and reuse. For example, the Vision for Equitable Data report offers priority uses for equitable data and identifies progress that could be made using existing data to answer equity-centered questions. Further, the TRUST Principles and FAIR Principles established frameworks to guide researchers who publish and reuse data, while maintaining the integrity of the open science process. The CARE Principles for Indigenous Data Governance specifically address data and innovation in Indigenous communities.

TRUST Principles	FAIR Principles	CARE Principles
Transparency, **R**esponsibility, **U**ser Focus, **S**ustainability, and **T**echnology	**F**indability, **A**ccessiblity, **I**nteroperability, **R**euse of Digital Assets	**C**ollective Benefit, **A**uthority to Control, **R**esponsibility, **E**thics
"... a common framework to facilitate discussion and implementation of best practice in digital preservation by all stakeholders."	"... specific emphasis on enhancing the ability of machines to automatically find and use the data, in addition to supporting its reuse by individuals."	"... people and purpose-oriented, reflecting the cruical role of data in advancing Indigenous innovation and self-determination."

Observation 4: Stringent data publication requirements could potentially disincentivize the collection of sensitive data.

Researchers who work with sensitive data were concerned that participant protections might conflict with the 2022 OSTP guidance. While the guidance states that agencies should follow other data policies that protect trade secrets, commercial information, personally identifiable information, and other data protected under law or policy, participants noted there may be other instances where funder requirements could conflict with best practices for ensuring respondent privacy. For example, a handful of workshop participants regularly collect data involving small samples of respondents who work in highly specialized fields and geographically identifiable regions, which can create difficulty in de-identifying data.

This is relevant to NOAA-funded weather research, especially when data are tied to a specific hazard or weather event. Examples include in-depth interview data collected from National Weather Service forecasters and broadcast meteorologists and observational data collected in local emergency operations centers after a major disaster. Without proper protocols in place, researchers at the workshop worried that open data requirements could unintentionally disincentivize the collection of qualitative data—which is difficult to de-identify without losing important context—or discourage research with focused population groups.

By extension, some participants pointed out that the guidance could lead researchers to avoid certain research, such as long-term ethnographic studies, that may not allow for the "immediate" release of data. This could further complicate already difficult-to-conduct research that has the potential to reduce harm among socially marginalized populations.

Recommendations:

Models for ethically publishing sensitive data should be developed collaboratively and shared.

- Privacy and security concerns are widely discussed in the open data literature. Annotated bibliographies and syntheses of available recommendations for navigating privacy and security concerns specific to social science data should be developed and widely shared.

- Data repositories have developed approaches for the ethical publication of sensitive data. For instance, ICPSR contains health data published with a permanent digital object identifier, but it is secured and inaccessible to the public due to privacy concerns. This example demonstrates that it is possible to partially implement open science standards while also maintaining high data privacy standards.

- It is important to remember that *not* sharing data can also be ethically problematic and lead to inequities, especially when study participants wish to access original, disaggregated data. Therefore, participants agreed that it will be important, moving forward, to develop case studies of successful collaborations that have led to open data publications—especially when sensitive data were involved.

- Researchers and representatives of data repositories should co-produce de-identification standards specific to social science datasets that involve, for example, easily identifiable or small sample populations.

Social science researchers should receive additional training on best practices for protecting participant privacy while adhering to open data standards.

- Representatives of data repositories and university data librarians should be supported in training, mentoring, and communicating ethical compliance with participant and community-level requests for privacy, while also adhering to federal open data guidance.

- Researchers who work with sensitive data may require additional training on, for example, new privacy tools that help reduce the risk of participant re-identification in published datasets. Researchers should also follow recommendations for making detailed, disaggregated data that is reflective of the diversity of the American population available while protecting participant privacy.

Observation 5: There are many certified data repositories where researchers can securely store and publish data and results from federally funded research.

The data repositories represented at the workshop included one generalist repository (Harvard Dataverse) and three domain-specific repositories (ICPSR for the social science community, the Qualitative Data Repository for qualitative researchers, and DesignSafe-Cyberinfrastructure for the natural hazards research community). In addition to these four respected repositories, there are many more repositories where researchers can securely store, manage, and publish their data. For example, one registry of repositories lists thousands of data repositories. Data librarians at the workshop noted that while the number of repositories might seem overwhelming, a much more limited number are CoreTrustSeal certified, meaning they have met standards that independently validate the integrity of the repository.

Even still, workshop participants noted problems could emerge if NOAA or other federal agencies require researchers to publish their data without offering guidance on which repositories they recommend or how to find the data once published. Given the number of options, researchers might unintentionally select a repository with less rigorous standards or one that is unsuitable for their data type or research, resulting in a lower quality final product. If data are spread across many repositories or improperly cited, it could impede reuse because social science weather data will be difficult if not impossible to find. Consider, for example, that current estimates suggest less than 1% of published data are properly cited. Representatives of data repositories face an uphill battle, as they struggle with user engagement and a lack of human resources to educate the research community. Further, research shows that early career researchers and social scientists who collect qualitative data involving human participants are among those most in need of training support.

Recommendations:

Federal agencies should recommend a select number of data repositories for social science data publication.

- Federal agencies should consult data librarians and other experts to develop a catalog of preferred repositories for social science weather-related data publication. As a starting point, agencies can draw on other federal efforts to identify sound repositories, such as National Science and Technology Council guidance on Desirable Characteristics of Data Repositories for Federally Funded Research and National Institutes of Health advice on Selecting a Data Repository.

- An agency-specific catalog of recommended data repositories could help researchers more easily identify trusted repositories suitable for their research. Such a catalog could inform researchers about domain-specific and generalist repositories, as well as how the metadata processes and search functions of each work.

Data repositories should receive additional funding to offer training on best practices for data publication, data citation, and data discovery, among other topics.

- Publishing data can encourage data integration and open new lines of inquiry. For published data to reach its full potential, data repositories should be supported to offer training on how to properly cite, share, and reuse published data. Because efforts to reuse data hinge on the ability to locate data, data librarians also have an important role to play in educating researchers regarding best practices for finding and using open data.

- Training sessions should be open to the entire research community, but also should be targeted toward those with less experience with open data such as early career and qualitative researchers.

"Data integration is a deliberate and iterative process of connecting diverse raw data and results to develop new frameworks, datasets, and analyses capable of generating public impacts beyond the application of single or multiple data." -Shah et al., 2023

Observation 6: The 2022 OSTP guidance could positively influence interdisciplinary data integration and reuse efforts.

Interdisciplinary research requires data integration across disciplinary lines and temporal and geographic scales. For example, weather research often involves the use of climatological, engineering, Earth, and social science data. However, working in interdisciplinary teams to integrate such datasets can pose unique challenges. Workshop participants shared stories of anonymity concerns that emerged when physical science observations were combined with qualitative data. Others described challenges with the spatial scale of physical and social science data, noting incompatibility for data integration if data collection is not coordinated at the research design phase.

Workshop participants agreed these hurdles to interdisciplinary data integration are significant, but not insurmountable. In fact, the implementation of the 2022 OSTP guidance may represent an opportunity for agencies, researchers, and data repositories to co-develop data publication and data reuse standards that could support interdisciplinary research and the integration of diverse datasets. Further, participants agreed that it is important to prioritize the development of common metadata documentation. Metadata helps researchers from different disciplines understand, combine, and use data. At present, there are no common metadata standards across the interdisciplinary weather research enterprise. Creating such standards will be critical moving forward to allow data to be combined and analyzed together. If successful, these efforts could help make data more accessible and usable across disciplines and scales.

Recommendations:

Open data stakeholders should create metadata standards and data protocols that enable interdisciplinary integration for weather-related research.

- Key open data communities should work together to develop a basic set of metadata standards that can help researchers working across disciplines to combine and analyze data.

- Data publication and reuse protocols, including authorship standards, should be developed by key open data communities. With its strong physical and social science mission, NOAA is positioned to convene representatives of data repositories and researchers from many disciplines to create such standards and protocols.

Federal agencies should track and promote published datasets to encourage data integration.

- Federal agencies should promote published datasets to encourage discovery of datasets across disciplines and to foster interdisciplinary collaborations. Such efforts will be necessary as it is likely that social scientists and physical scientists will continue to publish data in different repositories. Such promotion could also help increase the impact of published data.

Additional guidance and future convenings should focus explicitly on data reuse.

- There is currently little available guidance in the weather research community on how to find and reuse open data. Members of the research community could create a list of keywords repositories could use uniformly, such as a published controlled vocabulary to standardize searchable metadata fields. The creation of clear terms and definitions would simplify researchers' efforts to search, find, and use open data across disciplines.

- More guidance and examples are needed of social science data being effectively (or inappropriately) reused. Future convenings could center on a range of such use cases.

> "Data reuse can be seen as honoring research participants' contributions by maximizing the scientific value of the data they have provided rather than treating data as an asset 'owned' by the researchers who happened to collect it, from which they alone can extract value." –Sadeh et al., 2023

Observation 7: The 2022 OSTP guidance does not address requirements or incentives for publishing research protocols and instruments, independent of data.

The 2022 OSTP guidance focuses on providing open access to data and publications, with no specific discussion of the independent publication of research protocols and instruments. Yet, workshop participants agreed publishing protocols and instruments is important because these materials provide the context for published data. Further, because some social science researchers remain reticent or are unable to release data publicly, publishing research instruments can shed light on aspects of their research design and data collection strategies. By publishing tools and instruments—even in the absence of data—future cohorts of researchers can potentially reproduce or update studies in the future.

Recommendations:

Federal agencies should provide guidance on how to make social science protocols, data collection instruments, and other research materials freely available to the public.

- Sharing guidance on how to publish research instruments and other materials, independent of data, could encourage a wider range of researchers to participate in the open science movement.

- Federal agencies are encouraged to use instrument publication training materials developed by data repositories and members of the research community. For example, in 2020, the NSF-funded CONVERGE facility established a novel Data Ambassadors program where social science hazards and disaster researchers receive funding to attend a three-hour training that culminates with publishing a research instrument, data set, or report on DesignSafe.

Federal agencies should continue to incentivize the publication of research instruments and protocols.

- As presented during the workshop, the social science community has made considerable progress in the areas of data *and* instrument publication. With support from NOAA and NSF, the Natural Hazards Center has administered multiple special funding calls through the Weather Ready Research Award Program to incentivize data and research instrument publication. As a component of this program, researchers receive funding to participate in training activities and ultimately publish data or research instruments.

Observation 8: Social science researchers lack sufficient funding to make all publications free and immediately available.

The 2022 OSTP guidance encourages agencies to ensure investigators make findings—largely in the form of publications—free and immediately available. Many researchers, however, cannot afford to pay article processing charges or open access fees, which can cost thousands of dollars for a single article. Further, workshop participants were unsure how publications that are included in print books or edited volumes would be shared publicly.

Recommendations:

Federal agencies should consult experts and tap available resources to ensure publications are made freely available without violating publishing rules or incurring elaborate fees.

- In preparing publication requirements for researchers, federal agencies should center equity concerns and be aware that many researchers cannot afford to pay to publish their work.

- Researchers should not violate protocols for the journals, books, and other outlets where they publish their work. They should work with federal funders to ensure they can comply with both journal publication standards and agency requirements.

- Libraries can assist federal agencies and researchers in finding information to support open access initiatives. For example, many institutions have open access policies that secure the rights of faculty and other affiliated researchers to publish pre-prints openly in institutional repositories. Additionally, many journals offer to publish pre-prints during the review phase of an article. Libraries are at the forefront of this movement—they have services, initiatives, and knowledge to help researchers publish openly and ethically.

Federal agencies that do not currently require investigators to share their publications should review policies from agencies with such requirements in place.

- NSF established the Public Access Repository (NSF-PAR) to provide public access to its funded research. At the time investigators submit their annual reports, they are required to deposit their peer-reviewed, published journal articles and juried conference papers in the NSF-PAR. They also can upload citations of other publications such as book chapters, data publications, theses, and dissertations. This and other models should be considered by other federal agencies.

Observation 9: Fully understanding and implementing the 2022 OSTP guidance will take time, additional funding, and other forms of support.

Once fully implemented, the 2022 OSTP guidance will add layers of responsibility and reporting requirements to funded investigators. Workshop participants expressed concerns that these additional requirements would not be met with necessary resources to support them. Research participants questioned whether there would be funding or other support to help researchers and data repositories comply with upcoming changes.

Data repositories—which often operate with limited staff and funding support—are already feeling the impacts from journals requiring researchers to publish their data. Representatives expressed concern about the ramifications of increased data publication requirements without similar increases in support for both the short-term and long-term to build the necessary infrastructure to address increased demand.

Workshop participants also raised questions about who will be responsible for verifying the integrity of data and instruments as researchers seek to publish these materials in response to the new requirements. As time passes and repositories potentially cease to exist, participants also wondered how data will be maintained over time to ensure its future reuse.

Recommendations:

Data repositories and federal agencies should engage in regular dialogue about the resources needed to develop and sustain data infrastructure over time.

- Regular convenings between federal agencies and data repositories should be held to discuss available resources and what is needed to ensure data infrastructure can support a major infusion of new projects.

- As technology advances and funding sources change, it will be vital to sustain data repositories longer than typical three- to five-year funding cycles currently allow.

Federal agencies should consider how they can facilitate additional time investigators will need to ensure compliance with new data publication requirements.

- Federal agencies could consider increasing current award amounts by 10%, for example, to ensure investigators have the proper resources to manage, store, and publish their data and make it openly available in a timely manner.

- Federal agencies could also decrease other reporting requirements to save investigators time that could then ostensibly be dedicated to publishing their data.

Social science researchers should develop thorough data management and publication plans at the award proposal phase.

- Researchers are more likely to publish their instruments and data when they start with an explicit plan to do so. Such planning takes time and training, and researchers are encouraged to take advantage of available opportunities to support their efforts at providing open data.

- Social science researchers could also consider developing informal or formal peer review networks for their data and instruments to ensure the final published information is clear, comprehensive, complete, and useful to others.

Conclusion

The 2023 Workshop on Open Data and Reuse in Social Science Weather Research reviewed current policy guidance, created a space to share challenges and concerns, and provided the opportunity to give feedback about how to best approach open data and reuse practices. Workshop participants were enthusiastic about making data, research protocols and instruments, and publications available in a timely, open, and equitable manner. Every workshop participant expressed willingness to take on new responsibilities to foster open science.

In recognition of the capacity and motivation of the attendees, our closing session was organized in response to the following two prompts:

- What is one thing you would be willing to do to advance open data and publications?
- What is one collective action that you think our social science weather research community should take?

Workshop participants generated ideas that ranged from developing new assignments and academic courses (social science researchers), to updating the metadata standards to be more inclusive of the social sciences (data repository representatives), to changing reporting requirements to ensure investigators publish data and instruments as well as results from their research (federal agency representatives). Participants also considered how we can work together and with other partners to continue to shift the culture toward equitable open data as called for in the 2022 OSTP guidance.

Following the workshop, participants began to act on their personal commitments. For example, CONVERGE and the Qualitative Data Repository partnered to host a webinar for the social science community on publishing qualitative data. Other participants completed the CONVERGE Data Ambassadors training and published data on DesignSafe. Social science researchers are now working with data scientists at DesignSafe to create a de-identification protocol for qualitative social science datasets with small sample sizes. Faculty members are integrating the content of this workshop into their undergraduate and graduate research methods courses.

While this workshop generated many important insights and recommended actions, we acknowledge this meeting was simply one step in a much longer journey. Leadership, commitment on the part of many communities, and resources will be needed to institute the recommendations in this document. Time is of the essence, and the time to advance equitable and open data practices in the social science weather research community is now.

References

Borycz, Joshua, Robert Olendorf, Alison Specht, Bruce Grant, Kevin Crowston, Carol Tenopir, Suzie Allard, Natalie M. Rice, Rachael Hu, and Robert J. Sandusky. 2023. “Perceived Benefits of Open Data are Improving but Scientists Still Lack Resources, Skills, and Rewards.” *Humanities and Social Sciences Communication* 10, 339. https://www.nature.com/articles/s41599-023-01831-7

Kassam-Adams, Nancy and Miranda Olff. 2020. “Embracing Data Preservation, Sharing, and Re-Use in Traumatic Stress Research.” *European Journal of Psychotraumatology* 11, 1739885. https://doi.org/10.1080/20008198.2020.1739885

Khan, Nushrat, Mike Thelwall, and Kayvan Kousha. 2022. “Are Data Repositories Fettered? A Survey of Current Practices, Challenges, and Future Technologies.” *Online Information Review* 46(3): 483-502. https://doi.org/10.1108/OIR-04-2021-0204

Khan, Nushrat, Mike Thelwall, and Kayvan Kousha. 2023. “Data Sharing and Reuse Practices: Disciplinary Differences and Improvements Needed.” *Online Information Review* 47(6): 1036-1064. https://doi.org/10.1108/OIR-08-2021-0423

Lin, Dawei, Jonathan Crabtree, Ingrid Dillo, Robert R. Downs, Rorie Edmunds, David Giaretta, Marisa De Giusti, Hervé L’Hours, Wim Hugo, Reyna Jenkyns, Varsha Khodiyar, Maryann E. Martone, Mustapha Mokrane, Vivek Navale, Jonathan Petters, Barbara Sierman, Dina V. Sokolova, Martina Stockhause, and John Westbrook. 2020. “The TRUST Principles for Digital Repositories.” *Scientific Data* 7, 144. https://doi.org/10.1038/s41597-020-0486-7

Morss, Rebecca E., Heather Lazrus, and Julie L. Demuth. 2018. “The ‘Inter’ Within Interdisciplinary Research: Strategies for Building Integration Across Fields.” *Risk Analysis: An International Journal* 41(7): 1152-1161. https://doi.org/10.1111/risa.13246

Peek, Lori and Seth Guikema. 2021. “Interdisciplinary Theory, Methods, and Approaches for Hazards and Disaster Research: An Introduction to the Special Issue.” *Risk Analysis: An International Journal* 41(7): 1047-1058. https://doi.org/10.1111/risa.13777

Peters, Isabella. 2022. “Meaningful Indicators for Research Data.” *DataCite.* https://doi.org/10.5281/zenodo.6581155

Rathje, Ellen M., Clint Dawson, Jamie E. Padgett, Jean-Paul Pinelli, Dan Stanzione, Pedro Arduino, Scott J. Brandenburg, Tim Cockerill, Maria Esteva, Fred L. Haan, Jr., Ahsan Kareem, Laura Lowes, and Gilberto Mosqueda. 2020. “Enhancing Research in Natural Hazards Engineering Through the DesignSafe Cyberinfrastructure.” *Frontiers in Built Environment* 6. https://www.frontiersin.org/articles/10.3389/fbuil.2020.547706/full.

Sadeh, Yaara, Anna Denejkina, Eirini Karyotaki, Lonneke I. M. Lenferink, and Nancy Kassam-Adams. 2023. “Opportunities for Improving Data Sharing and FAIR Data Practices to Advance Global Mental Health.” *Cambridge Prisms: Global Mental Health* 10, E14. https://doi.org/10.1017/gmh.2023.7

Shah, Sameer H., Cassandra R. O'Lenick, Jessica S. Wan, Alexandra Ramos-Valle, Kevin D. Ash, Olga V. Wilhelmi, Catrin M. Edgeley, Maria J. Molina, Jessica Moulite, Carlo Andre Chunga Pizarro, Kelsey Emard, Olivia Z Cameron, James M Done, Cleo Wölfle Hazard, Thomas M Hopso2, Mikah Jones, Forrest Lacey, Michée Arnold Lachaud, Danica Lombardozzi, Michael Méndez, Rebecca E. Morss, Katharine Ricke, Fernando Tormos-Aponte, William R. Wieder, and Christopher L. Williams. 2023. "Connecting Physical and Social Science Datasets: Challenges and Pathways Forward." *Environmental Research Communications* 5, 095007. https://iopscience.iop.org/article/10.1088/2515-7620/acf6b4/pdf

Wilkinson, Mark D., Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, Jan-Willem Boiten, Luiz Bonino da Silva Santos, Philip E. Bourne, Jildau Bouwman, Anthony J. Brookes, Tim Clark, Mercè Crosas, Ingrid Dillo, Olivier Dumon, Scott Edmunds, Chris T. Evelo, Richard Finkers, Alejandra Gonzalez-Beltran, Alasdair J.G. Gray, Paul Groth, Carole Goble, Jeffrey S. Grethe, Jaap Heringa, Peter A.C 't Hoen, Rob Hooft, Tobias Kuhn, Ruben Kok, Joost Kok, Scott J. Lusher, Maryann E. Martone, Albert Mons, Abel L. Packer, Bengt Persson, Philippe Rocca-Serra, Marco Roos, Rene van Schaik, Susanna-Assunta Sansone, Erik Schultes, Thierry Sengstag, Ted Slater, George Strawn, Morris A. Swertz, Mark Thompson, Johan van der Lei, Erik van Mulligen, Jan Velterop, Andra Waagmeester, Peter Wittenburg, Katherine Wolstencroft, Jun Zhao, and Barend Mons. 2016. "The FAIR Guiding Principles for Scientific Data Management and Stewardship." *Scientific Data* 3, 160018. https://doi.org/10.1038/sdata.2016.18

Web Links

CARE Principles for Indigenous Data Governance: https://www.gida-global.org/care

CONVERGE Data Ambassadors: https://converge.colorado.edu/data/data-ambassadors/

CONVERGE Organizing and Archiving Qualitative Data with the Qualitative Data Repository Webinar: https://converge.colorado.edu/social-science-fridays/organizing-and-archiving-qualitative-data-with-the-qualitative-data-repository/

CONVERGE Training Modules: https://converge.colorado.edu/resources/training-modules/

CoreTrustSeal for Core Trustworthy Data Repositories: https://www.coretrustseal.org/

DesignSafe-Cyberinfrastructure: https://www.designsafe-ci.org/

Foundations for Evidence-Based Policymaking Act, 2018: https://www.congress.gov/bill/115th-congress/house-bill/4174/text

Harvard Dataverse: https://dataverse.harvard.edu/

Inter-university Consortium for Political and Social Research (ICPSR): https://www.icpsr.umich.edu/web/pages/

National Institutes of Health, Selecting a Data Repository: https://sharing.nih.gov/data-management-and-sharing-policy/sharing-scientific-data/selecting-a-data-repository

National Science Foundation Award #1635593: A Clearinghouse on Natural Hazards Applications: https://www.nsf.gov/awardsearch/showAward?AWD_ID=1635593

National Science Foundation Public Access Repository (NSF-PAR): https://par.nsf.gov/

National Science and Technology Council Guidance on Desirable Characteristics of Data Repositories for Federally Funded Research: https://www.whitehouse.gov/wp-content/uploads/2022/05/05-2022-Desirable-Characteristics-of-Data-Repositories.pdf

Natural Hazards Center, University of Colorado Boulder: https://hazards.colorado.edu/

OpenDP: https://opendp.org/about

Qualitative Data Repository: https://data.qdr.syr.edu/

Registry of Research Data Repositories (re3data): https://www.re3data.org/

United States Department of Education Press Release on Secretaries of Education, Agriculture Call on Governors to Equitably Fund Land-Grant HCBUs: https://www.ed.gov/news/press-releases/secretaries-education-agriculture-call-governors-equitably-fund-land-grant-hbcus

A Vision for Equitable Data: Recommendations from the Equitable Data Working Group: https://www.whitehouse.gov/wp-content/uploads/2022/04/eo13985-vision-for-equitable-data.pdf

Weather Ready Research Award Program: https://hazards.colorado.edu/research/weather-ready-research/index

White House Office of Science and Technology Policy (OSTP) Memorandum on Ensuring Free, Immediate, and Equitable Access to Federally Funded Research, 2022: https://www.whitehouse.gov/wp-content/uploads/2022/08/08-2022-OSTP-Public-Access-Memo.pdf

White House Office of Science and Technology Policy (OSTP) Memorandum on Increasing Access to the Results of Federally Funded Scientific Research, 2013: https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/ostp_public_access_memo_2013.pdf

Workshop on Open Data and Reuse in Social Science Weather Research: https://hazards.colorado.edu/research/weather-ready-research/open-data-workshop