Guide to Choosing a Data Repository

It is important, and a requirement, to publish your data once you've completed your research. Recent federal guidance has changed data publication requirements for projects receiving federal assistance, such as your NOAA-funded work. This guide details what you need to know about the new guidance and can be used to help you select an appropriate repository for your data and instruments. Please know this document will continue to evolve as more clarity is given regarding this new guidance and how it specifically applies to NOAA-funded research.

Publishing Your Data: Why, When, and Where

In August 2022, the White House Office of Science and Technology Policy (OSTP) released new guidance for federal agencies to develop new policies for the sharing and publishing of federally-funded research data. In addition, the new guidance urges immediate access to research data as well as the use of persistent identifiers (e.g., DOIs) for researchers and research outputs.

In response, NOAA's Weather Program Office, in partnership with the Natural Hazards Center at the University of Colorado Boulder, convened a workshop on Open Data and Reuse in Social Science Weather Research in April 2023 to gather community input on best practices and considerations to keep in mind when implementing this guidance. We heard numerous concerns from our research community about publishing data, including what "immediate" means, how ethical considerations would be navigated, and where to even start with choosing where to publish. Despite these concerns, it is important to emphasize that publishing the results of federally-funded research ***is*** a government requirement and ***must*** be addressed. We recognize there are hundreds of different repositories to choose from and identifying an appropriate repository can be daunting. We offer this guide as a resource to assist you in navigating this challenge.

Generally speaking, NOAA does not endorse or require sharing data in any specific repository. In fact, NOAA encourages researchers to select the most appropriate repository for their research area(s) and data type(s).

In addition to meeting NOAA award requirements, there are a number of important reasons (and benefits for) why you should publish your data, including:

1. Increasing the research impact of your work (by getting a Digital Object Identifier like you get for your publications)
2. Promoting reproducibility and transparency in research and contributing to the scholarly enterprise
3. Preserving data for the long term
4. Strengthening the base of evidence for informed decision-making

Choosing a Suitable Data Repository

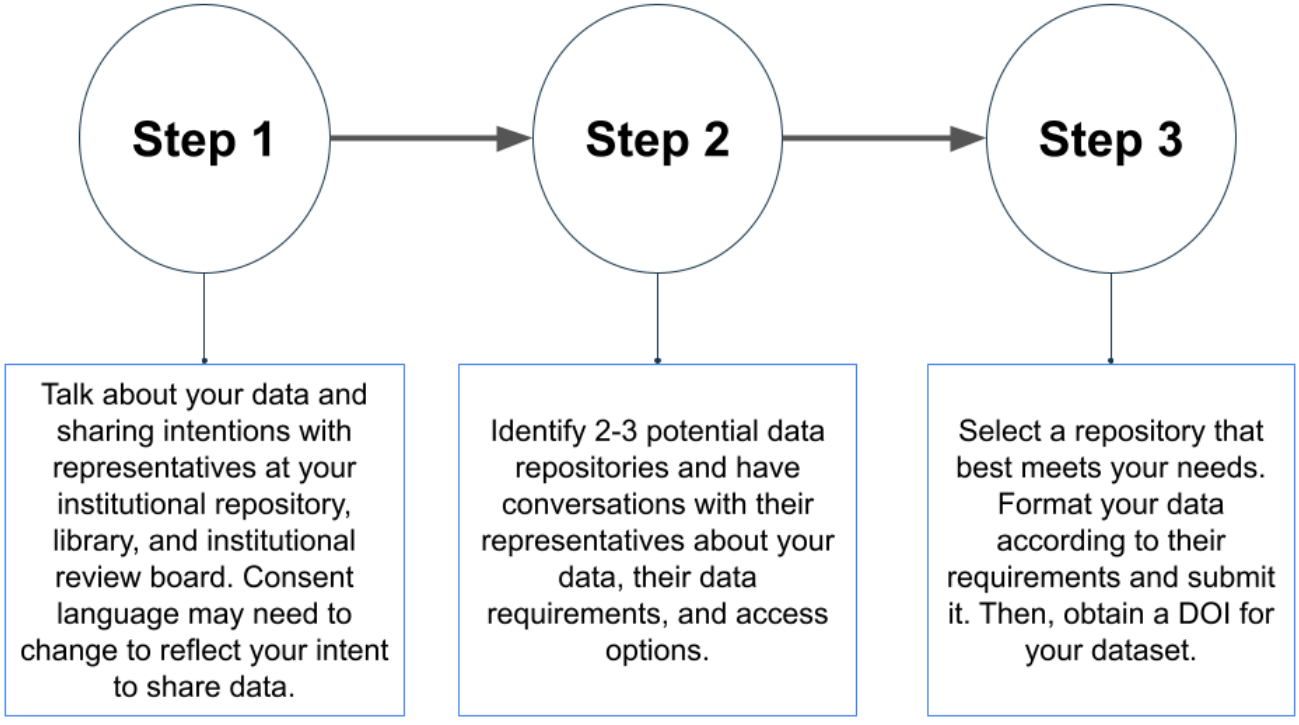
There are a myriad of things to consider when choosing an appropriate data repository, including:

- Security obligations
- Funder/publisher requirements
- Community conventions and norms (e.g., metadata schemes)
- Institutional policy

In addition, your specific data has a number of considerations:

- The type of data you are publishing (e.g., quantitative, qualitative, sensitive)
- The format and file size
- The potential reuse and reuse conditions (e.g., licenses)
- Version control of the data (e.g., tracking and managing changes to data over time)
- Storage location of the data
- Sensitivity of the data (e.g., personally identifiable information)
- Access conditions (e.g., restricted access, APIs)

With this in mind, the National Science and Technology Council has developed a comprehensive guide on desirable characteristics of data repositories for federally-funded research. We also suggest the following steps to support your process of choosing a suitable data repository for publishing your data:



At the beginning of your project, you should have a conversation with a representative from your institution's repository and/or library as well as with your Institutional Review Board (IRB) officials. Most universities have an institutional repository, such as the University of Oklahoma's SHAREOK or the University of Maryland System's MDSOAR. These representatives can help you think about your data sharing needs. In the absence of an institutional repository (or in addition to), a data librarian at your institution's library can also be a valuable resource. Having a conversation ***at the beginning of your project*** is important so you can identify an appropriate repository early and build any data requirements into your data collection and management, such as consent form language expressing your intent to share the data.

There are three primary types of data repositories to consider:

1. Institutional repositories (e.g., those associated with universities or organizations)
2. Subject- or domain-specific repositories (e.g., those limited to data from specific disciplines or domains, such as Social Science or Genomics)
3. Generalist repositories (e.g., those that accept data from all researchers regardless of discipline or data type)

Re3data provides a searchable database of repositories that meet the registry's minimal requirements and provides detailed information on the type, policies, and requirements for each repository included. Nature also provides a brief list of social science and generalist repositories, as well as some additional guidance in selecting a repository. In addition to the considerations listed above, we suggest three additional important things to look for in a potential data repository:

1. Does the repository assign digital object identifiers (DOIs)?
    a. What are DOIs and why is this important?
2. Does the repository adhere to FAIR and CARE data principles?
3. How trustworthy is the repository?
    a. CoreTrustSeal outlines 16 requirements and provides a listing of repositories that have been certified to meet these requirements.

The Next Steps

Contact your library and/or institutional repository to begin the process of deciding when, where, and how to publish your research data. For additional support in choosing a data repository, visit *Nature*'s data repository guidance page. Also know that our team will be discussing data publishing with you during our transition plan protocol co-development meetings.