The New 2023 NIH Data Management and Sharing Policy

Wasila Dahdul, Data Curation Librarian

May 2, 2023

UCI Libraries

Topics for today

1. Policy Overview
2. Elements of the Data Management & Sharing Plan
3. Data Management Resources at UCI

Final NIH Policy for Data Management and Sharing

Notice Number:

NOT-OD-21-013

Key Dates

Release Date:	October 29, 2020
Effective Date:	January 25, 2023

NIH's Data Sharing Policy website: https://sharing.nih.gov

NIH Sharing Policies

- **2003: Data Sharing Policy** requires investigators seeking $500,000 or more in NIH funding to submit a data sharing plan (or rationale for not sharing). Superseded by the new policy in January 2023.
- **2008: Public Access Policy** requires NIH-funded scientists to submit final peer-reviewed journal manuscripts to PubMed Central (PMC) no later than 12 months after publication.
- **2014: Genomic Data Sharing Policy** requires investigators generating large-scale genomic data to submit a genomic data sharing plan.

NIH’s New Data Management & Sharing Policy (DMSP)

Effective January 25, 2023

- Requires researchers seeking NIH funding to prospectively submit a plan outlining how scientific data from their research will be managed and shared.
- Researchers should “maximize the appropriate sharing of scientific data.”
- Timeline: Data should be shared as soon as possible, and no later than the time of an associated publication or end of performance period (whichever comes first).

NIH’s New Data Management & Sharing Policy (DMSP)

- Applies to all research funded in whole or in part by NIH that generates scientific data:
  - Extramural grants
  - Contracts
  - Intramural research projects
  - All other NIH funding mechanisms

- Exception: funding that does not generate data (e.g., training grants).

NIH’s New Data Management & Sharing Policy (DMSP)

- The policy represents the minimum requirements. NIH ICOs may expect more specificity in their plans - *check funding announcements for additional information.*
- Plan is approved by NIH staff
- Report updates to Plan in progress reports

“Scientific data”

Defined as “the recorded factual material commonly accepted in the scientific community as of sufficient quality to validate and replicate research findings, regardless of whether the data are used to support scholarly publications.”

Does not include “laboratory notebooks, preliminary analyses, completed case report forms, drafts of scientific papers, plans for future research, peer reviews, communications with colleagues, or physical objects, such as laboratory specimens.”

Source: https://grants.nih.gov/grants/guide/notice-files/NOT-OD-21-013.html

Not all data needs to be shared

- informed consent limitations
- existing agreements prohibit sharing
- privacy or safety of research participants need protection
- explicit law or regulation prohibiting sharing
- data cannot be digitized with reasonable effort

Elements of the DMS Plan

Elements of the DMS Plan

1. Data type(s) and metadata (data description)
2. Related tools, software, and/or code
3. Standards for the data/metadata
4. Data preservation, access, and associated timelines
5. Access, distribution, or reuse considerations
6. Oversight of data management and sharing

Source: https://grants.nih.gov/grants/guide/notice-files/NOT-OD-21-014.html

NIH Sample Plans

Sample Plans

NIH has provided sample DMS Plans as examples of how a DMS Plan could be completed in different contexts, conforming to the elements described above. These sample DMS Plans are provided for educational purposes to assist applicants with developing Plans but are not intended to be used as templates and their use does not guarantee approval by NIH.

Note that the sample DMS Plans provided below may reflect additional expectations established by NIH or specific NIH Institutes, Centers, or Offices that go beyond the DMS Policy. Applicants will need to ensure that their Plan reflects any additional, applicable expectations (including from NIH policies and any ICO- or program-specific expectations as stated in the FOA).

Sample	Description	NIH Institute or Center
Sample Plan A	Clinical and/or MRI data from human research participants	NIMH
Sample Plan B	Genomic data from human research participants	NIMH
Sample Plan C	Genomic data from a non-human source	NIMH
Sample Plan D	Secondary data analysis	NIMH
Sample Plan E	Human genomic data	NHGRI
Sample Plan F	Technology development	NHGRI
Sample Plan G	Human clinical and genomics data	NICHD

https://sharing.nih.gov/data-management-and-sharing-policy/planning-and-budgeting-for-data-management-and-sharing/writing-a-data-management-and-sharing-plan#sample-plans

1. Data Type(s)

- Summarize the **types** and **amounts** of data you are collecting.
- Which scientific data from the project will be **preserved** and **shared**? Note: it doesn't have to be everything! You are expected to maximize the appropriate sharing of data; the plan should supply the **rationale**.
- What **metadata** or **documentation** (study protocols, data collection instruments) will be made accessible to facilitate interpretation of the data?

2. Related Tools, Software and/or Code

- Are **specialized tools** needed to access or manipulate your scientific data to support replication or reuse? If so, which ones?
- How can these tools be **accessed** (e.g., Github)? Are they open source or do they require a license?
- Are these tools likely to remain available for as long as the scientific data remain available?

3. Standards

- What standards will be applied to the scientific data and associated metadata (i.e., data formats, data dictionaries, data identifiers, definitions, unique identifiers, and other data documentation)?
- Some fields have community-developed standards while others do not. Indicate if no standards have been established.
- Example of standardized data: The OMOP Common Data Model

4. Data Preservation, Access & Timelines

- What research **data repositories** will you use for your data?
- How will your data be **findable** and **identifiable** (e.g., persistent unique identifier)?
- When will it be made **available** to others and for how long?

  *Note: NIH encourages data to be made available **as soon as possible**, and no later than the time of an associated publication or end of the performance period, whichever comes first.*

Choosing a data repository

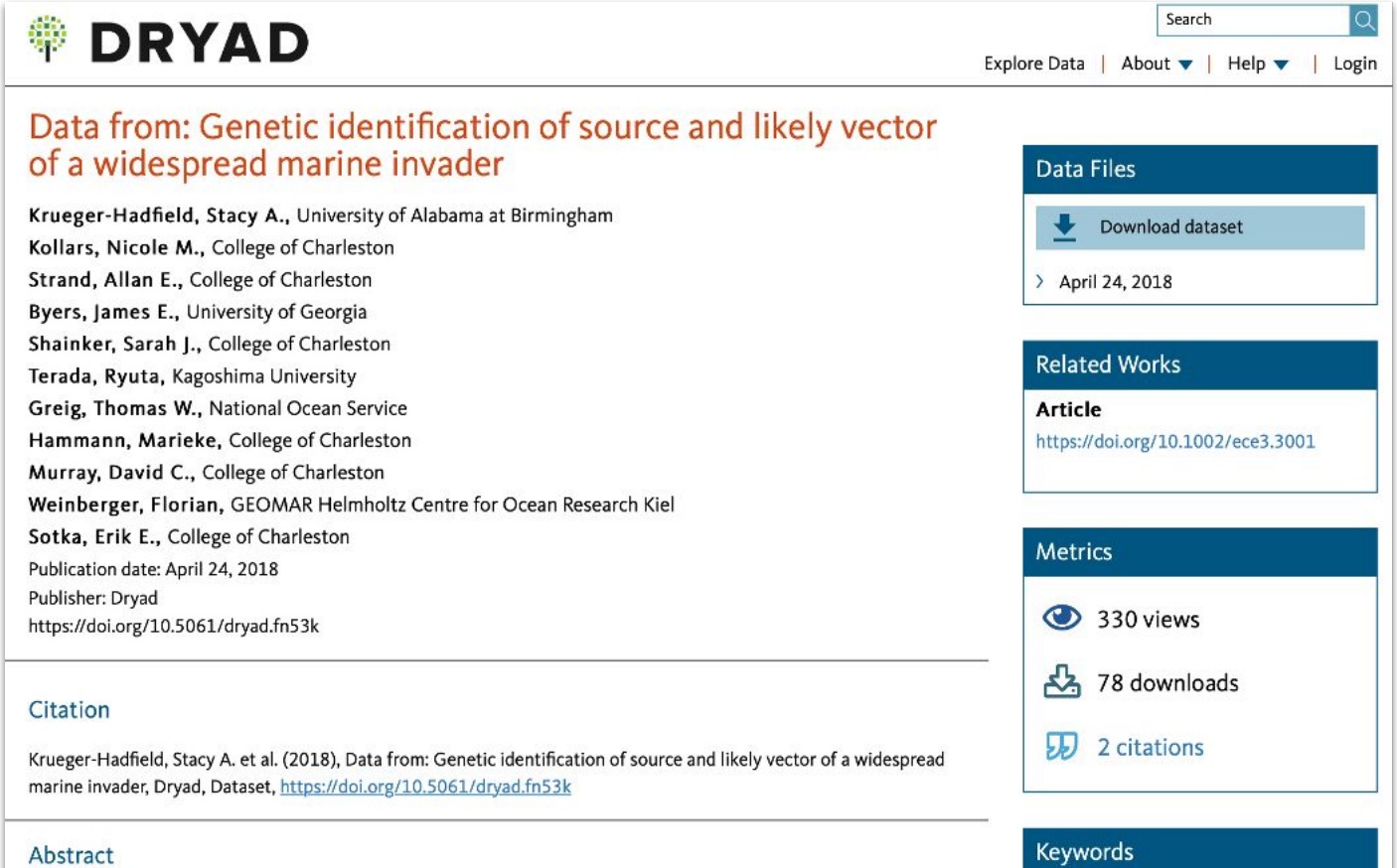
NIH strongly encourages the use of established repositories to the extent possible for preserving and sharing scientific data.

Desirable characteristics for all data repositories:

1. Unique Persistent Identifiers (e.g., https://doi.org/10.6075/J0J67H27, https://osf.io/uadxr, RRID:RGD_1566457)
2. Long-Term Sustainability
3. Metadata
4. Curation and QA (e.g., file checking, metadata enhancement, reproducibility)
5. Free and Easy Access
6. Broad and Measured Reuse (e.g., by providing easy-to-copy *citations* for your data)
7. Clear Use Guidance (e.g., *licenses* such as Creative Commons, MIT license, or CC0 public domain waiver)
8. Security and Integrity
9. Confidentiality
10. Common Format (i.e., widely-used, non-proprietary formats)
11. Provenance
12. Retention policy

Source:
https://sharing.nih.gov/data-management-and-sharing-policy/sharing-scientific-data/selecting-a-data-repository#desirable-characteristics-for-all-data-repositories

Data repository example: Dryad



What are some resources that are not considered repositories?

- **Github** is a code repository for software, programs and scripts
- **REDCap** is a data collection tool for building and managing HIPAA- and IRB-compliant surveys and databases
- **Qualtrics** is a survey platform and data collection tool

Choosing a Data Repository

In order of priority…

1. Check the ICO and FOA for designated repositories
2. Choose a discipline or data-type specific repository
    a. NIH's list of supported data repositories
3. Choose a generalist repository if no appropriate specialist repository is available:
    a. UC's Dryad Digital Repository accepts data from any discipline
    b. NIH's list of generalist repositories

Choosing a repository, Priority #1: Meet any program-specific or FOA requirements

Home > Other Sharing Policies > NIH Institute and Center Data Sharing Policies

NIH Institute and Center Data Sharing Policies

Data sharing is a priority across NIH. To this end, many institutes, centers, and research programs have instituted specific data sharing policies in addition to the trans-NIH policies. These policies are listed in the table below. Note that individual funding opportunities may specify other requirements or expectations, so be sure to read all instructions carefully.

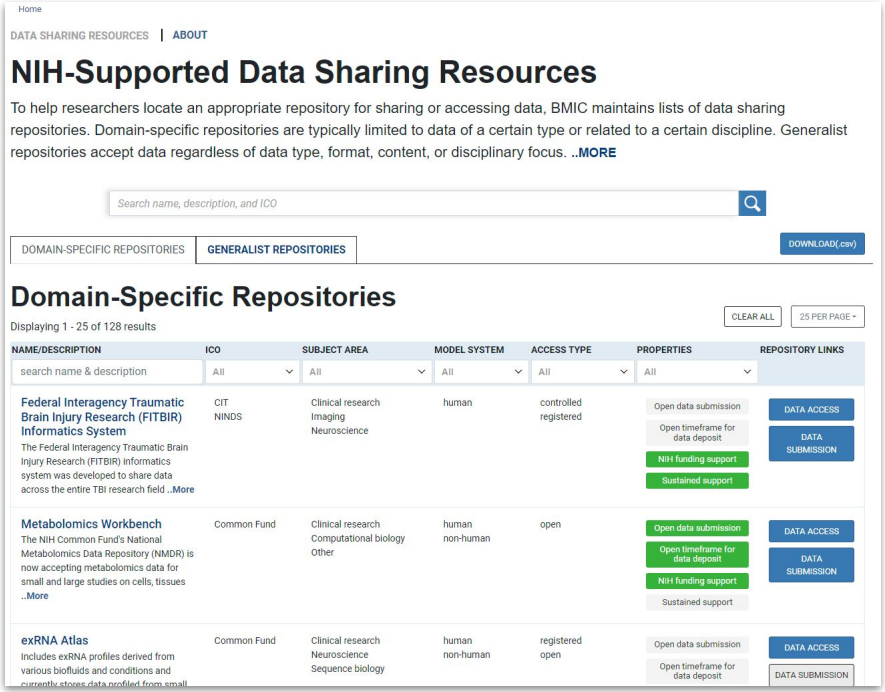
Institute or Center	HEAL
Data Sharing Policy Name	HEAL Public Access and Data Sharing
Description of Data Sharing Policy	Through the NIH HEAL Initiative Public Access and Data Sharing Policy (the Policy), NIH seeks to create an infrastructure that addresses the need for researchers, clinicians, and patients to collaborate on sharing their collective data and knowledge about opioid misuse and pain to provide scientific solutions to the opioid crisis. Under the Policy, applicants for extramural research funding (grants, cooperative agreements, contracts, and other transactions; "Applicants") for NIH HEAL Initiative Research Projects are required to submit a Public Access and Data Sharing Plan that (1) describes their proposed process for making resulting Publications and, to the extent possible, the Underlying Primary Data immediately and broadly available to the public or (2), if applicable, provides a justification to NIH if such sharing is not possible. Underlying Primary Data should be made as widely and freely available as possible while safeguarding the privacy of participants and protecting confidential and proprietary data.
Repositories	Various HEAL-Compliant repositories
Institute or Center	NCI
Data Sharing Policy Name	Cancer Moonshot℠ Public Access and Data Sharing Policy
Description of Data Sharing Policy	The primary goal of NCI's Cancer Moonshot℠ is to significantly accelerate cancer research discovery and meaningful implementation. The Cancer Moonshot Public Access and Data Sharing Policy addresses the recommendation of the Blue Ribbon Panel's Enhanced Data Sharing working group to the National Cancer Advisory Board that researchers, clinicians, and patients should collaborate in sharing their collective data and knowledge about cancer to accelerate progress towards improving cancer outcomes. Under this policy, applicants for Cancer Moonshot Research Projects are required to

NIH ICO Policies:
https://sharing.nih.gov/other-sharing-policies/nih-institute-and-center-data-sharing-policies

NIH supported repositories:
https://sharing.nih.gov/data-management-and-sharing-policy/sharing-scientific-data/repositories-for-sharing-scientific-data

Priority #2: Find an appropriate discipline-specific repository

- BMIC-maintained list of domain-specific repositories
- Nature's Data Repository Guidance
- The Registry of Research Data Repositories



Repositories for sharing human data

Source:
https://sharing.nih.gov/data-management-and-sharing-policy/sharing-scientific-data/selecting-a-data-repository#additional-considerations-for-human-data

Additional Considerations for Human Data

When working with human participant data, including de-identified human data, here are some additional characteristics to look for:

- **Fidelity to Consent:** Uses documented procedures to restrict dataset access and use to those that are consistent with participant consent and changes in consent.
- **Restricted Use Compliant:** Uses documented procedures to communicate and enforce data use restrictions, such as preventing reidentification or redistribution to unauthorized users.
- **Privacy:** Implements and provides documentation of measures (for example, tiered access, credentialing of data users, security safeguards against potential breaches) to protect human subjects' data from inappropriate access.
- **Plan for Breach:** Has security measures that include a response plan for detected data breaches.
- **Download Control:** Controls and audits access to and download of datasets (if download is permitted).
- **Violations:** Has procedures for addressing violations of terms-of-use by users and data mismanagement by the repository.
- **Request Review:** Makes use of an established and transparent process for reviewing data access requests.

Repositories with access controls:

- Harvard Dataverse
- Qualitative Data Repository (QDR)
- Vivli
- ICPSR

Helpful resource:
https://qdr.syr.edu/guidance/human-participants

Vivli: a generalist repository for clinical data

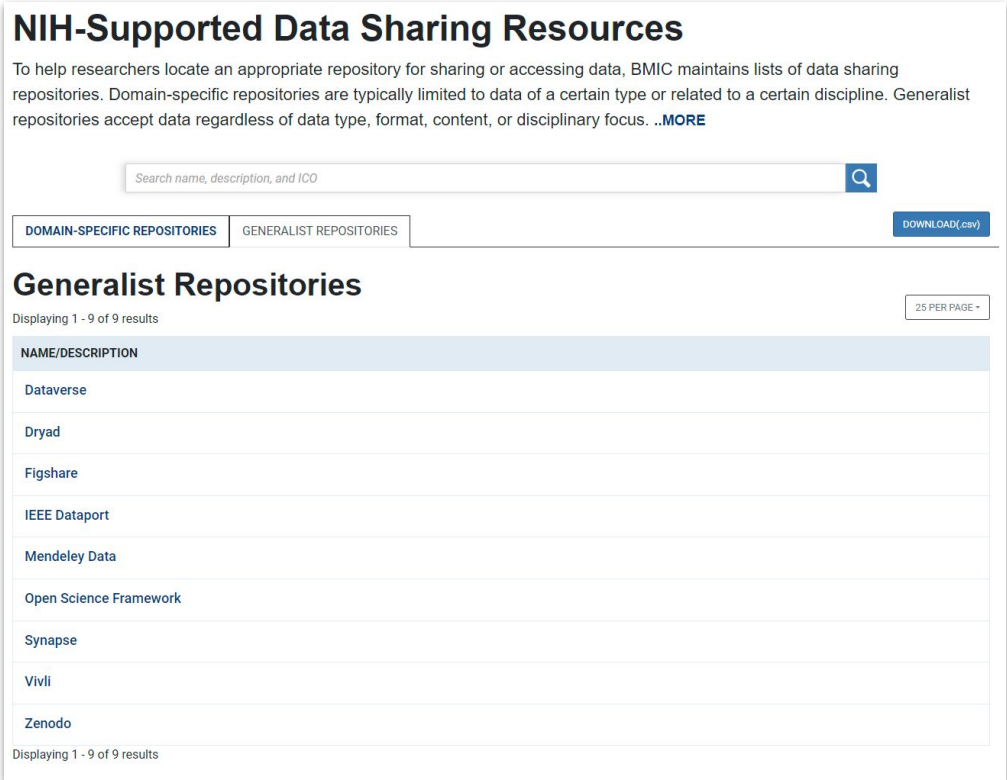
Service	Item Type	Cost
One-time service fee • includes initial consult • metadata curation • archiving 10 yrs securely in the cloud • DOI for publications and grants	Clinical trial dataset (<500GB)	$4,000 USD
Anonymization fee per dataset	Optional fee	$10,000 USD
Larger datasets	Clinical trial datasets (>500GB).	$10,000 USD

https://vivli.org/resources/sharedata/

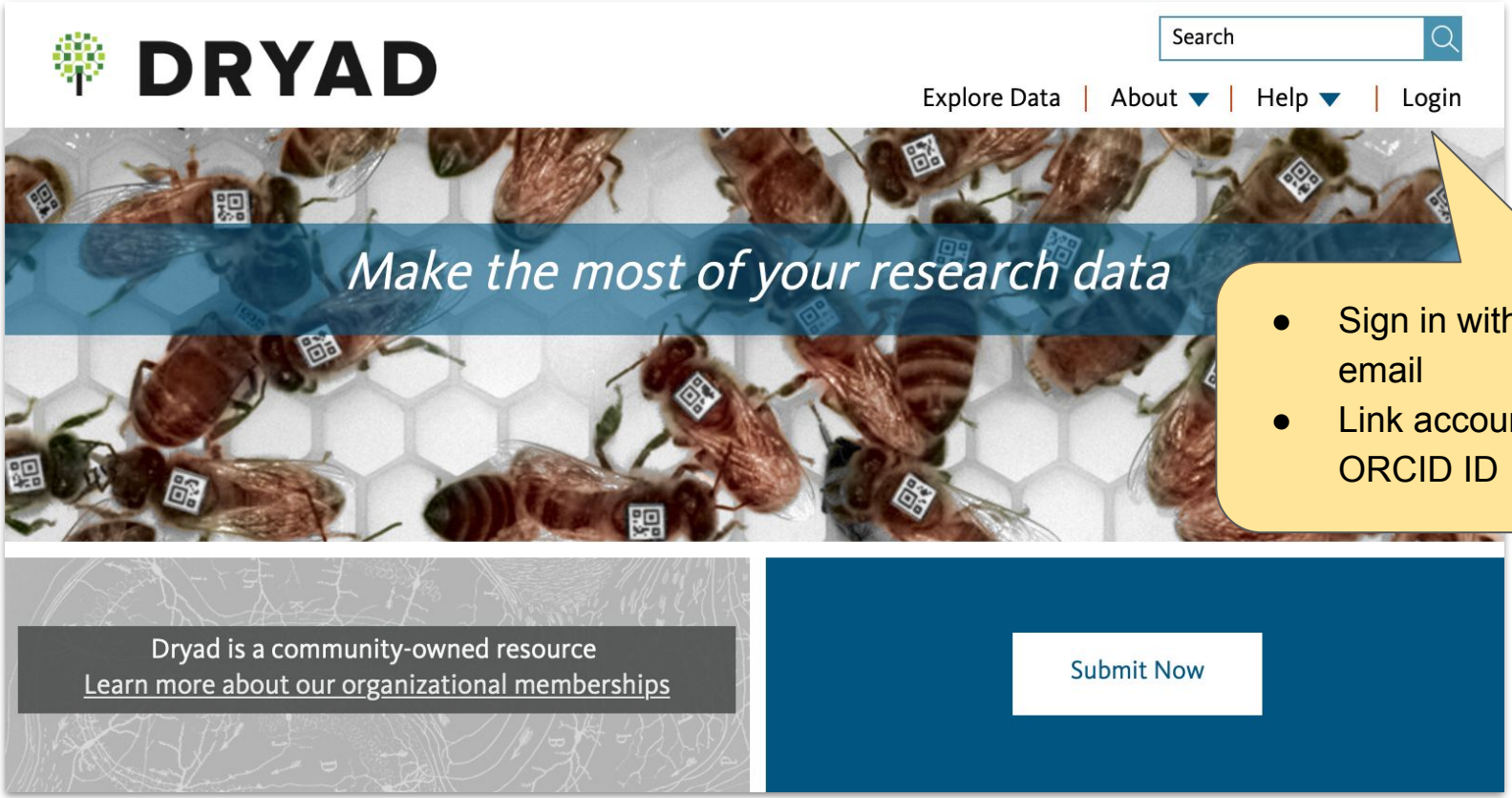
Vivli
CENTER FOR GLOBAL CLINICAL RESEARCH DATA

Priority #3: Find a generalist or institutional repository that meets desirable characteristics

Source: https://www.nlm.nih.gov/NIHbmic/generalist_repositories.html



- General-purpose data repository that makes data and code discoverable, freely reusable, and citable
- UC campuses partner with CDL to co-develop Dryad
- Publisher and researcher workflow integrations
- Curation and quality assurance workflow during submission
- Free for UCI researchers



- Sign in with UCI email
- Link account to ORCID ID

https://datadryad.org/stash

Data from: Genetic identification of source and likely vector of a widespread marine invader

Krueger-Hadfield, Stacy A., University of Alabama at Birmingham
Kollars, Nicole M., College of Charleston
Strand, Allan E., College of Charleston
Byers, James E., University of Georgia
Shainker, Sarah J., College of Charleston
Terada, Ryuta, Kagoshima University
Greig, Thomas W., National Ocean Service
Hammann, Marieke, College of Charleston
Murray, David C., College of Charleston
Weinberger, Florian, GEOMAR Helmholtz Centre for Ocean Research Kiel
Sotka, Erik E., College of Charleston

Publication date: April 24, 2018
Publisher: Dryad
https://doi.org/10.5061/dryad.fn53k

Citation

Krueger-Hadfield, Stacy A. et al. (2018), Data from: Genetic identification of source and likely vector of a widespread marine invader, Dryad, Dataset, https://doi.org/10.5061/dryad.fn53k

Abstract

The identification of native sources and vectors of introduced species informs its ecological and evolutionary history and may guide policies that seek to prevent future introductions. Population genetics represents a powerful set of tools to identify origins and vectors, but can mislead when the native range is poorly sampled or few molecular markers are used. Here, we traced the introduction of the Asian seaweed Gracilaria vermiculophylla (Rhodophyta)

Data Files

Download dataset

› April 24, 2018

Related Works

Article
https://doi.org/10.1002/ece3.3001

Metrics

330 views

78 downloads

2 citations

Keywords

Gracilaria vermiculophylla

Oysters

Dryad is appropriate for sharing open and re-usable data

- Dryad does not have access control features for restricted data
- Human subjects data must be properly anonymized to remove personally identifiable human subject information
- Licensing terms must be compatible with the Creative Commons Zero waiver

5. Access, Distribution, Reuse

Describe any factors affecting access, distribution or reuse including:

- Informed consent - will you get consent to share? *Note: NIH recommends addressing data management and sharing during the informed consent process.*
- How are you ensuring privacy and confidentiality (de-identification, etc)?
- Will access to the data be controlled?
- Is your data subject to any restrictions on access (HIPAA, Tribal or state laws, etc.)?

6. Oversight of Data Management and Sharing

- How will you monitor compliance with this plan?
- Who will monitor compliance and how often?
- A list of possible data management roles: https://tinyurl.com/DMProles

How do I pay for this?

Allowable costs may be included in NIH budget requests, including costs associated with:

1. **Curating data and developing supporting documentation,** including formatting data according to accepted community standards; de-identifying data; preparing metadata to foster discoverability, interpretation, and reuse; and formatting data for transmission to and storage at a selected repository for long-term preservation and access.

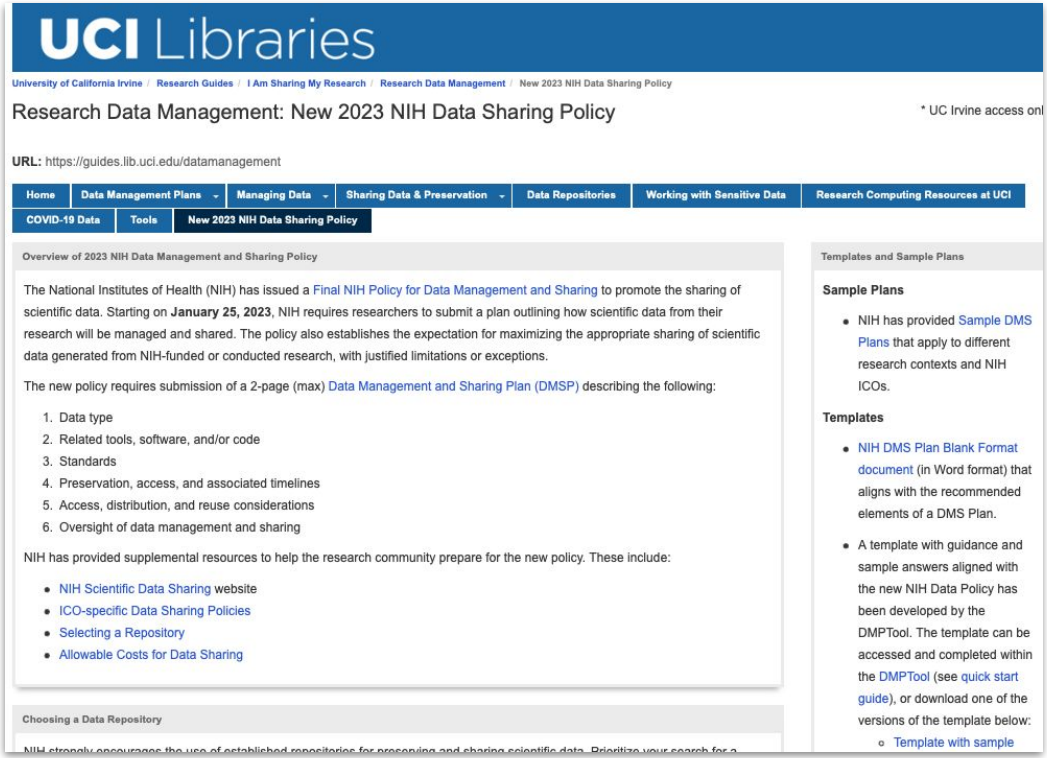
2. **Local data management considerations,** such as unique and specialized information infrastructure necessary to provide local management and preservation (e.g., before deposit into an established repository).

3. **Preserving and sharing data through established repositories,** such as data deposit fees necessary for making data available and accessible. For example, if a Data Management and Sharing Plan proposes preserving and sharing scientific data for 10 years in an established repository with a deposition fee, the cost for the entire 10-year period must be paid prior to the end of the period of performance. If the Plan proposes deposition to multiple repositories, costs associated with each proposed repository may be included.

Source: https://grants.nih.gov/grants/guide/notice-files/NOT-OD-21-015.html

Data management resources at UCI

UCI Libraries' NIH webpage with templates and sample plans

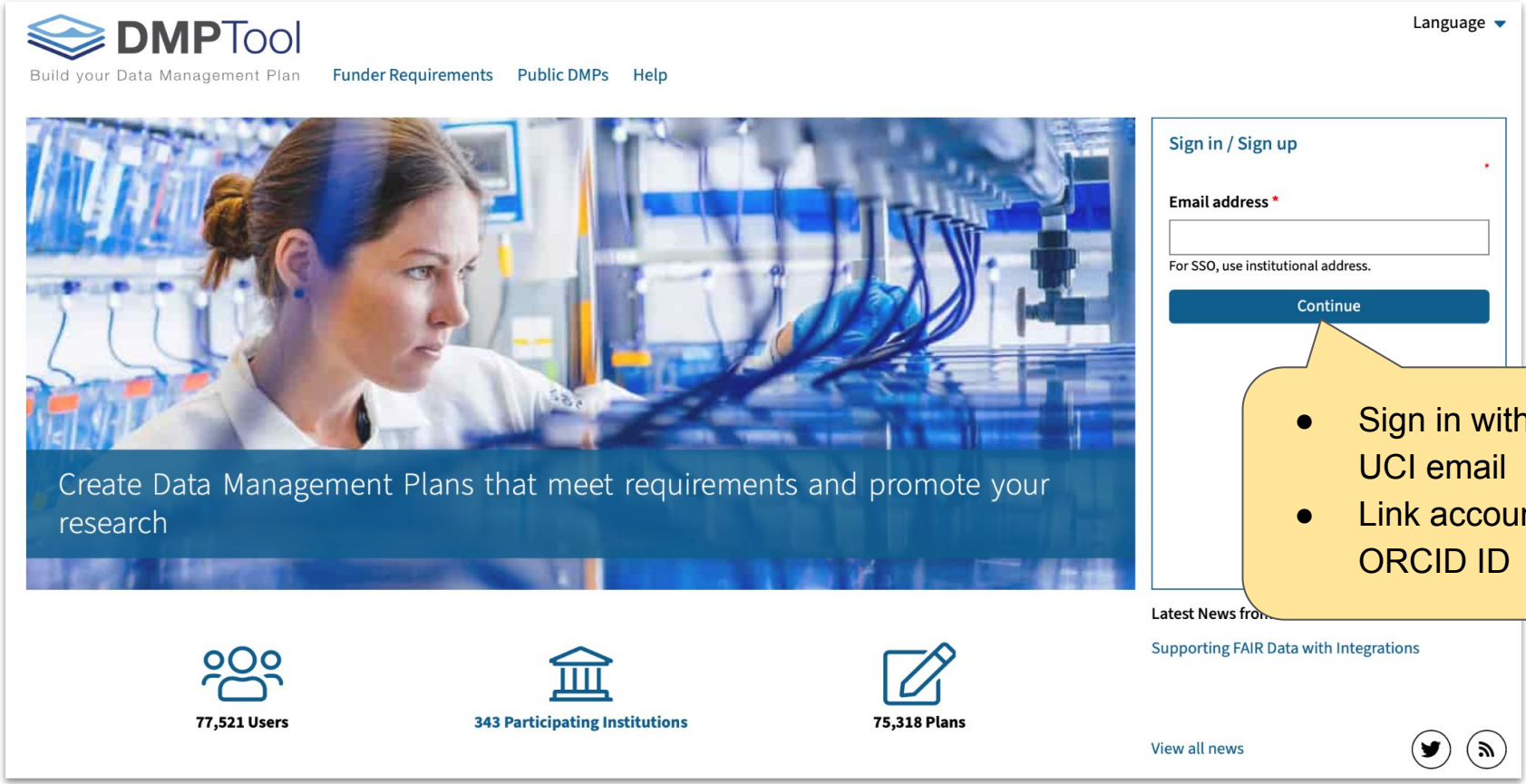


https://guides.lib.uci.edu/datamanagement/NIH_2023_data_sharing_policy

Templates also available in DMPTool

- Features templates for writing data management plans for 22 federal and private funders, including **NIH**
- UC campuses partner with CDL to offer DMPTool
- Free for UCI researchers

https://dmptool.org



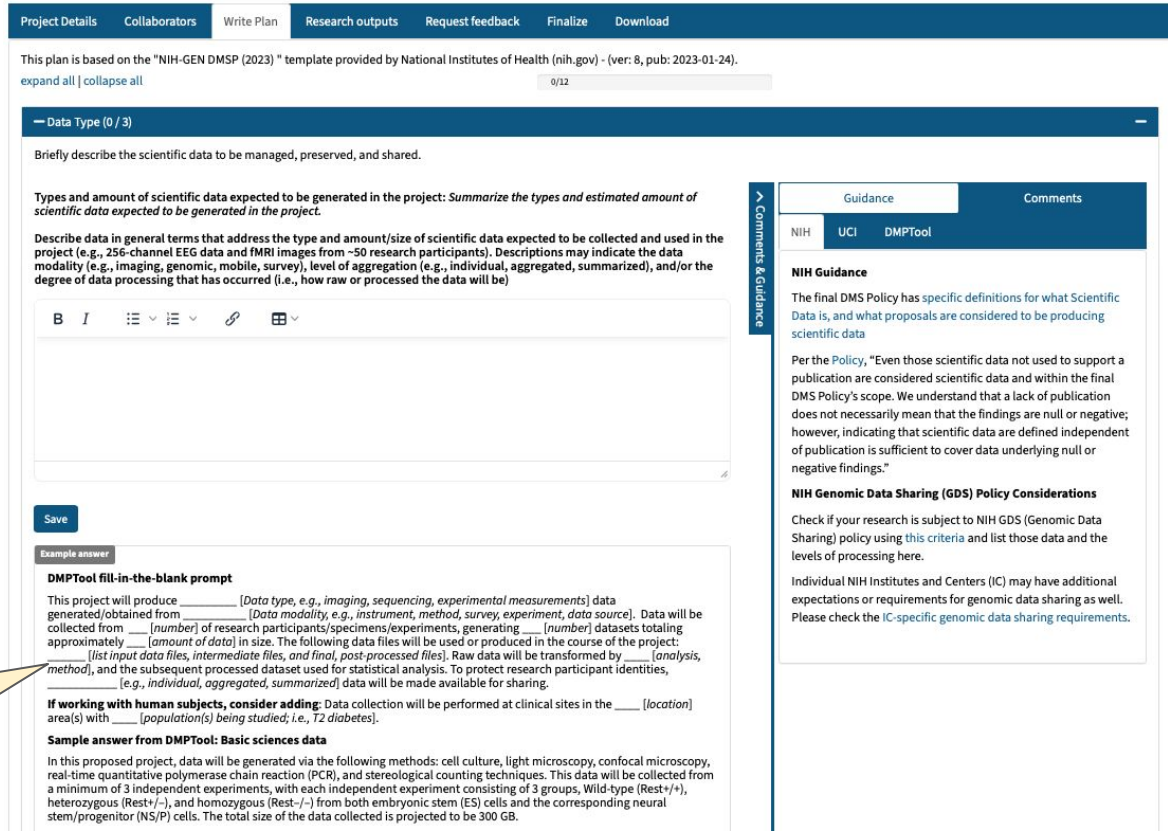
- Sign in with UCI email
- Link account to ORCID ID

https://dmptool.org

DMPTool

Build your Data Management Plan

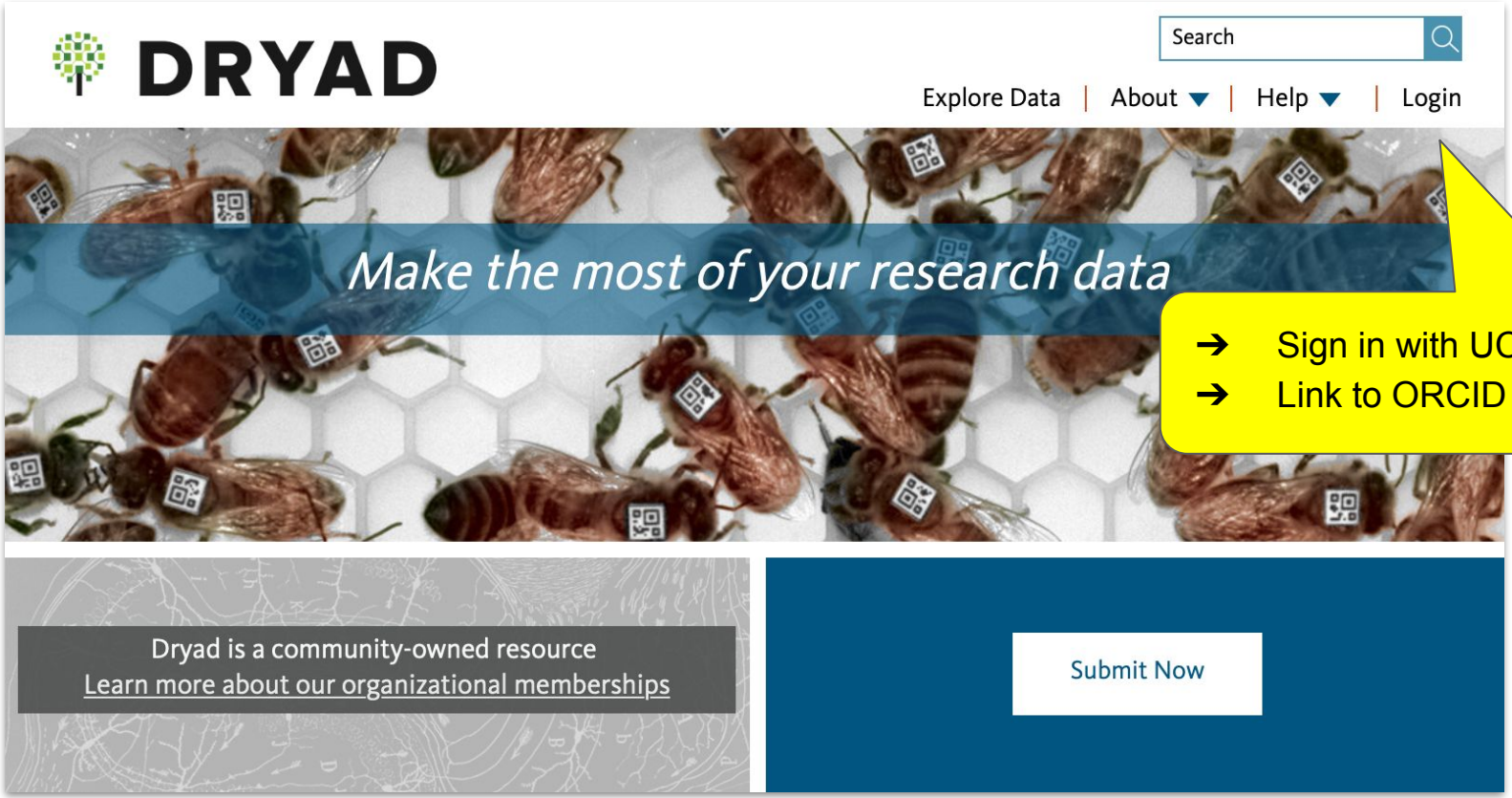
- Select the NIH template when creating your plan



Example answers available in template

- General-purpose data repository that makes data and code discoverable, freely reusable, and citable
- Publisher and researcher workflow integrations
- Curation and quality assurance workflow during submission
- UC campuses partner with CDL to co-develop Dryad
- Free for UCI researchers

https://datadryad.org/stash


DRYAD
Search
Explore Data | About | Help | Login
Make the most of your research data
➔ Sign in with UCI email
➔ Link to ORCID ID
Dryad is a community-owned resource
Learn more about our organizational memberships
Submit Now

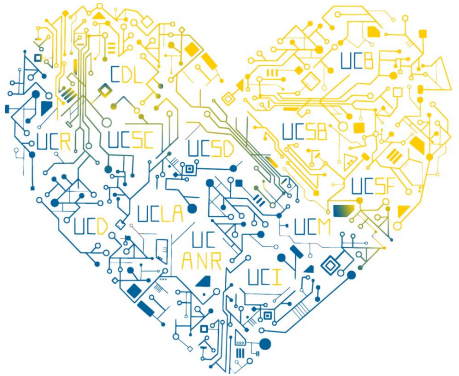
Consultations on Data Management

- Topics:
  - Data repository selection
  - Data standards
  - Data management planning
  - Development of draft DMS Plans
- Email: wdahdul@uci.edu
- Schedule an appointment: https://spaces.lib.uci.edu/appointments/wasila

Recorded webinars on NIH Data Sharing Policy

- **Writing a Data Management and Sharing Plan for NIH,** UC Love Data Week, February 2023
  Presented by Wasila Dahdul (UCI) and Ariel Deardorff (UCSF)
  View recording and materials via Box
- **Managing and Sharing Data for NIH Projects,** UC Love Data Week, February 2023
  Presented by Wasila Dahdul (UCI), Derek Devnich (UCM), Ho Jung Yoo (UCSD), Reid Otsuji (UCSD)
  View on Vimeo
- **The New 2023 NIH Data Management and Sharing Policy.** November 2022, hosted by the School of Medicine, Office of Research.

  Webinar recording, Slides



Questions?

Wasila Dahdul, wdahdul@uci.edu

Digital Scholarship Services: http://lib.uci.edu/dss

UCI Libraries