

# Exploring the Value of Open (EVO)

## Data Sharing: Why, How, and Measuring Impact

Kristen Ratan, Stratos

Maryrose Franko, HRA



November 10, 2022

# Welcome to the HRA EVO Program

- HRA's Open Science Community is sponsoring a one-year Exploring the Value of Open (EVO) Program with monthly webinar sessions, discussion groups, and opportunities to explore open science.
- The EVO program will cover not only the why, **but also the how**, and ways to measure impact of open research policies, practices, and incentives.



Kristen Ratan, Stratos



Salvatore La Rosa, CTF

# Why is Data Sharing Important?



## *Ensure reproducibility*

Proactively sharing data ensures that your work remains reproducible over the long term.



## *Inspire trust*

Sharing data demonstrates rigor and signals to the community that the work has integrity.



## *Receive credit*

Making data public opens opportunities to get academic credit for collecting and curating data during the research process.



## *Make a contribution*

Access to data accelerates progress. According to the 2019 State of Open Data report, more than 70% of researchers use open datasets to inform their future research.



## *Preserve the scientific record*

Posting datasets in a repository or uploading them as Supporting Information prevents data loss.

## Funders:

- Maximize Impact
- Improve equity

<https://plos.org/open-science/open-data/>

# Also, increasing pressure from policy

## NIH – effective Jan 25, 2023

The National Institutes of Health (NIH) is issuing this final NIH Policy for Data Management and Sharing (DMS Policy) to promote the management and sharing of scientific data generated from NIH-funded or conducted research. This Policy establishes the requirements of submission of Data Management and Sharing Plans (hereinafter Plans) and compliance with NIH Institute, Center, or Office (ICO)-approved Plans. It also emphasizes the importance of good data management practices and establishes the expectation for maximizing the appropriate sharing of scientific data generated from NIH-funded or conducted research, with justified limitations or exceptions. This Policy applies to research funded or conducted by NIH that results in the generation of scientific data.

## OSTP all Fed agencies – effective Dec 31, 2025

### b) **Scientific Data**

- i. Scientific data<sup>6</sup> underlying peer-reviewed scholarly publications resulting from federally funded research should be made freely available and publicly accessible by default at the time of publication, unless subject to limitations as described in Section 3(c)(i) and should be subject to federal agency guidelines for researcher responsibilities regarding data management and sharing plans, consistent with Section 3(c) of this memorandum.

# Open data and sensitive information

Privacy Act of 1974 (amended in January 2019).

Four key types of sensitive information are recognized:

1. Personally Identifiable Information (PII) (e.g., name, address, email address, etc.).
2. Sensitive PII (e.g., social security number, driver's license, etc.).
3. Protected Health Information (PHI) (e.g., patient's medical records, insurance claims, payment history, etc.).
4. Other information, such as financial records, grant applications, proprietary data, etc.

# Final NIH Policy for Data Management and Sharing - guidance

- Address data management and sharing in the **informed consent process**, such that prospective participants will understand what is expected to happen with their data
- Communicate any **limitations on subsequent use** of data to those individuals or entities preserving and sharing the scientific data
- Considering **controlling access** to scientific data derived from humans, even if de-identified and lacking explicit limitations on subsequent use

# Mechanics of Data Sharing

## What

### Per NIH Policy:

Scientific data includes any data needed to validate and replicate research findings.

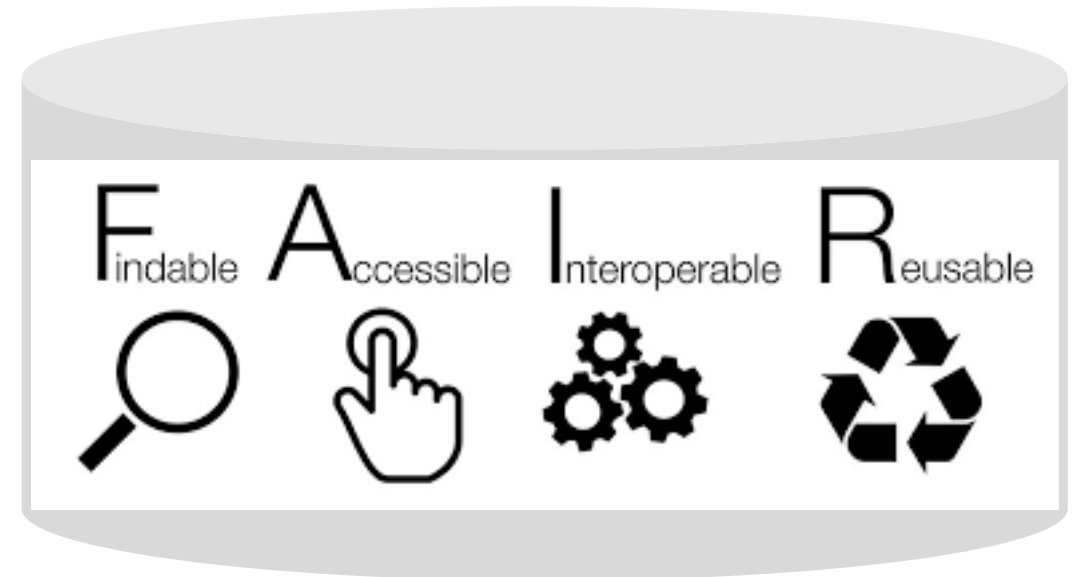
<https://tinyurl.com/562baatj>

## How

Data Prep and Curation:

- Clean
- Annotate
- Describe
- Complete metadata
- DOI
- License

## Where



<https://www.go-fair.org/fair-principles/>

# Stakeholder Opportunities

## Funders:

- Policy
- Guidance and training
- Curation services
- Preferred repositories
- Compliance monitoring
- Usage, citation, reuse metrics

## Publishers:

- Policy
- Interfaces and education
- Data services (curation, DOI)
- Facilitating reproducibility
- Connecting repos and pubs
- Usage, citation, reuse metrics



# Speakers

- **PLOS:** Publisher perspective on Data Policy and tracking
- **DataSeer:** Infrastructure and service provider to discover and track data sharing
- **CTF:** Funder use case to facilitate data sharing among grantees

# What's coming next at HRA

## EVO Program Schedule – Fall 2022

1. Nov 10 (3pm ET) - Data Sharing: why, how, and measuring impact
2. Dec 14 (1pm ET) – Government policies and plans (NIH and NSF)
3. Jan (TBD) – Birds of a Feature session options – look for a survey soon
  - Policy deep dive
  - Baseline analysis of OA and open data

## Future topics so far:

- Policy working group
- Preprints
- Software and Code sharing
- Protocols and tangible resources
- Data Management Plans and GMSs
- Career progression (see Value of ORCID)
- Compliance and reporting
- Evidence on the impact of open science
- Research Assessment and how it's evolving

[Suggest a topic!](#)

[Open Science Resources](#)

[Download the PDF here](#)



Thank you!