

Exploring the Value of Open (EVO)

Birds of a Feature on Data Sharing

Kristen Ratan, Stratos
Maryrose Franko, HRA
Salvatore La Rosa, CTF



March 7, 2023

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

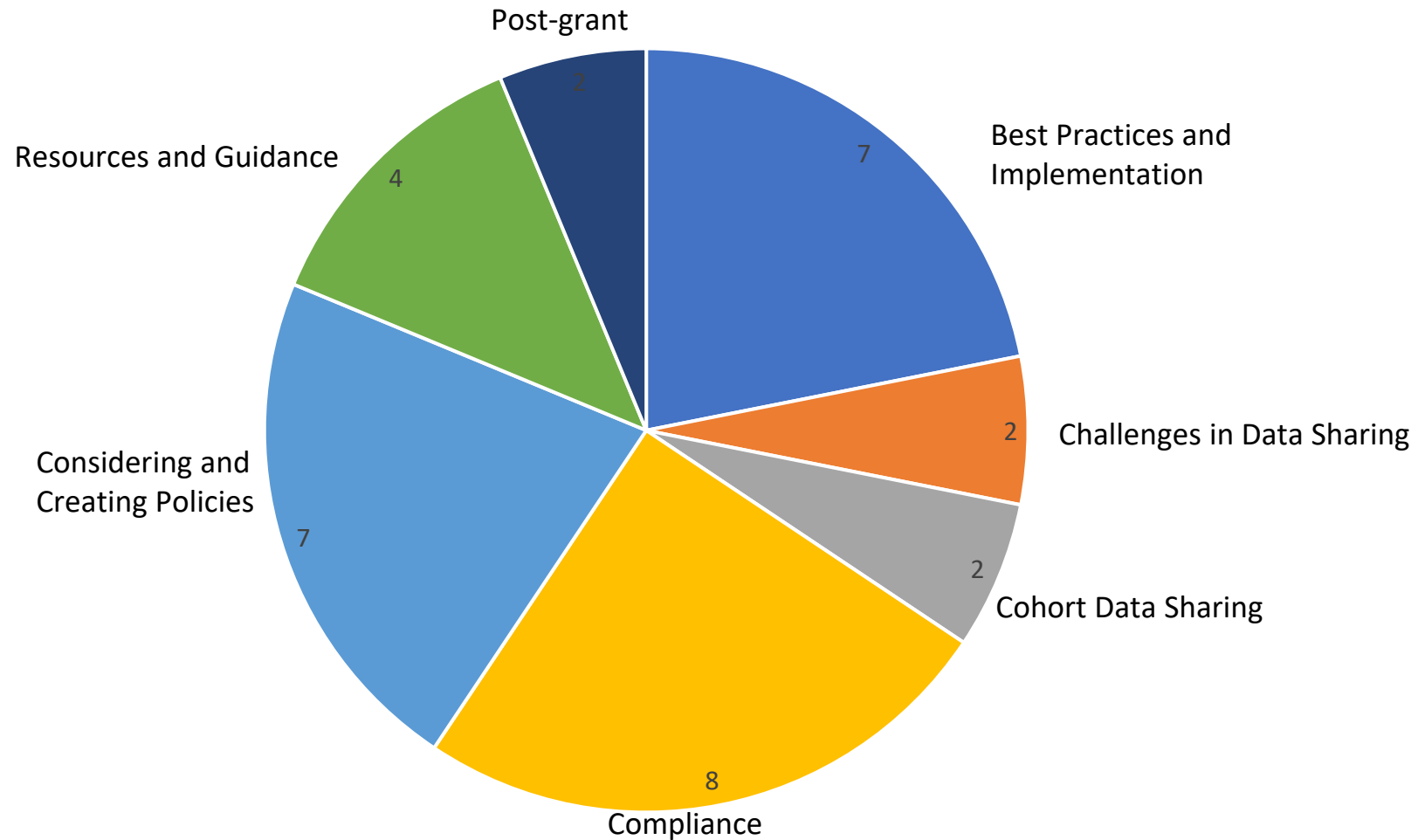
b) **Scientific Data**

- i. Scientific data⁶ 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.

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

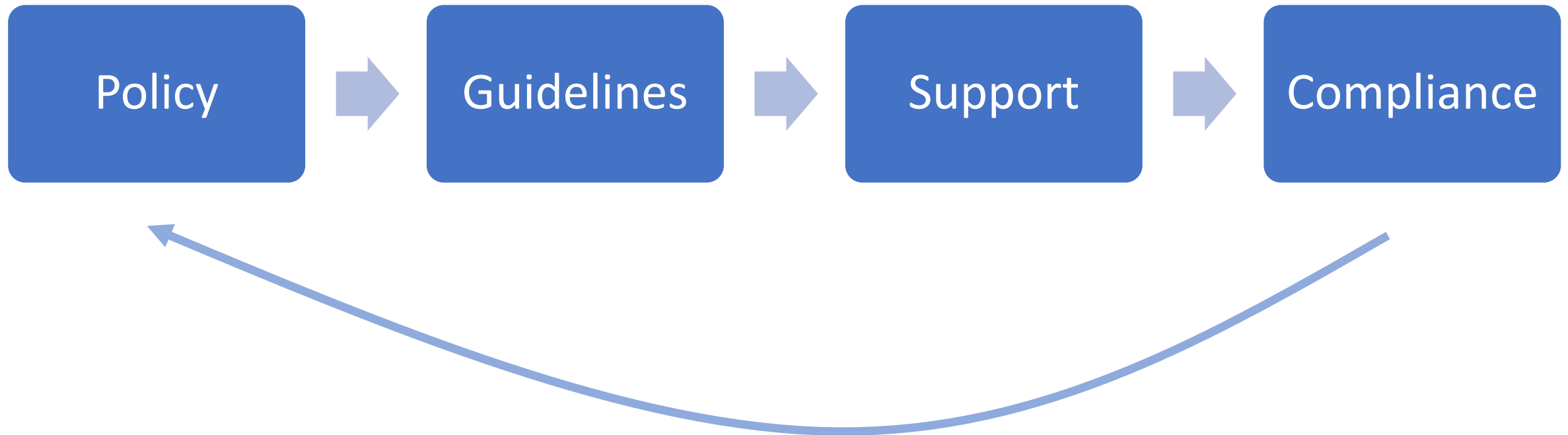
Member Topics of Interest




Three Broad Categories

1. Policy-making
 - What to prioritize, consequences of policy choices
 - How to create and communicate policies
2. Best Practices and Implementation
 - Grantee guidance and support
 - Infrastructures, tools, and partners
 - Overcoming major challenges
 - Special cases like cohort sharing, post-grant sharing
3. Compliance and Impact Tracking

From Policy to Practice



Policy Worksheet

 Stratos	Updated 23 Feb 2023				
Action Area	Encourage	Require	Support	Rules	Options and best practices
ORCIDs	Encourage that all grantees and their trainees have ORCIDs	ORCIDs at the time of grant proposal submission	Enable and support ORCID signup through proposal process	Use ORCIDs with publications and other outputs. In ORCID record, add affiliations, set funders as "trusted"	Give permission for "auto-update" tying ORCIDs to publications
Articles	Encourage gold/diamond OA journals or green OA in repositories without embargo	Require OA per CoalitionS-type policies	Require and support OA through guidelines, how-tos, and staff help	DOI, CC-BY license	make freely available through HRA Open, a repository, with CC-BY licence and ensure RRS
Preregistration	Encourage preregistration that specifies a research plan in advance and submits it to a registry	Require preregistration be open upon submission	Facilitate and help preregistration	Discoverable (DOI preferred)	Make publicly open immediately, ensure DOI
Preprints	Encourage preprints upon or before journal submission	Require preprints in an open and nonprofit service	Guidelines, how-tos, and staff help	DOI, CC-BY license	Post to recognized preprint service that reserves author rights
Datasets	Encourage a data availability statement on accessing datasets underpinning publications	Require dataset sharing through FAIR repositories	Require and facilitate data sharing through FAIR repositories	DOI, open license, cited in publications	Curate datasets with accurate and complete metadata, contributorship, and usage notes
Code	Encourage a code availability statement on accessing software underpinning publications	Require code sharing through open repositories such as Github	Facilitate code sharing and persistent storage	DOI, ORCID or other real name, cited in publications	Curate code with accurate and complete metadata, contributorship, and usage notes and DOI assigned. Consider a repository record pointing to code
Resources/materials	Encourage an availability statement explaining how to access tangible materials underpinning publications	Require registering and sharing through Addgene, Scicrunch or other open repository	Facilitate registration and sharing through Addgene, Scicrunch or other open repository	RRID	Post detailed records of real world resources and materials into a repository associated with SciCrunch
Protocols	Encourage an availability statement explaining how to access detailed protocols underpinning publications	Require open sharing through Protocols.io, Zenodo, or other open repository	Facilitate and support sharing through Protocols.io, Zenodo, or other open repository	DOI, CC-BY license	Post protocols in full in a repository, preferably one that specializes in protocols and assigns DOIs

Mechanics of Data Sharing

What

Per NIH Policy:

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

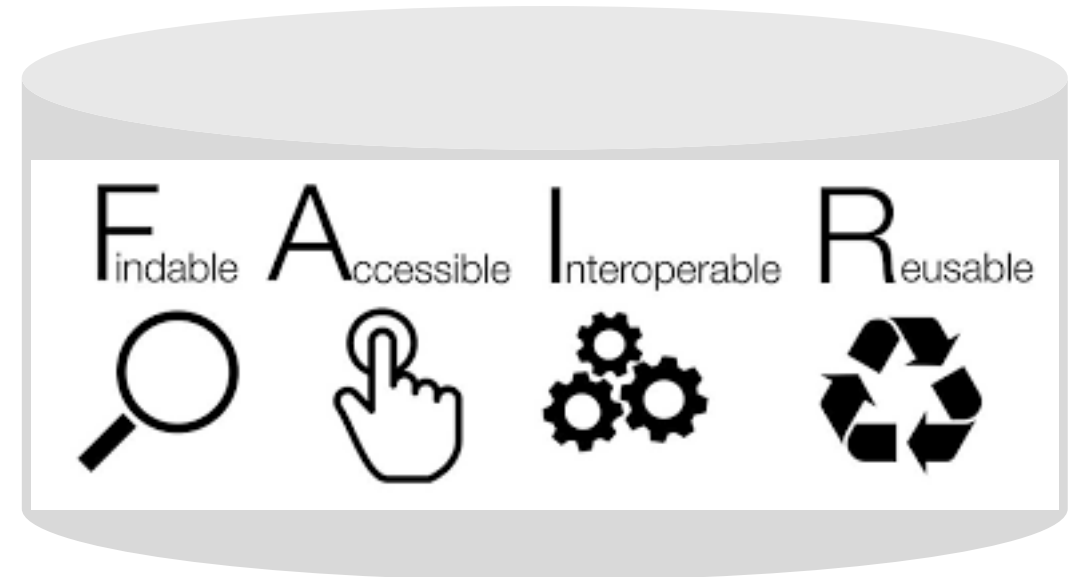
<https://tinyurl.com/562baatj>

How

Data Prep and Curation:

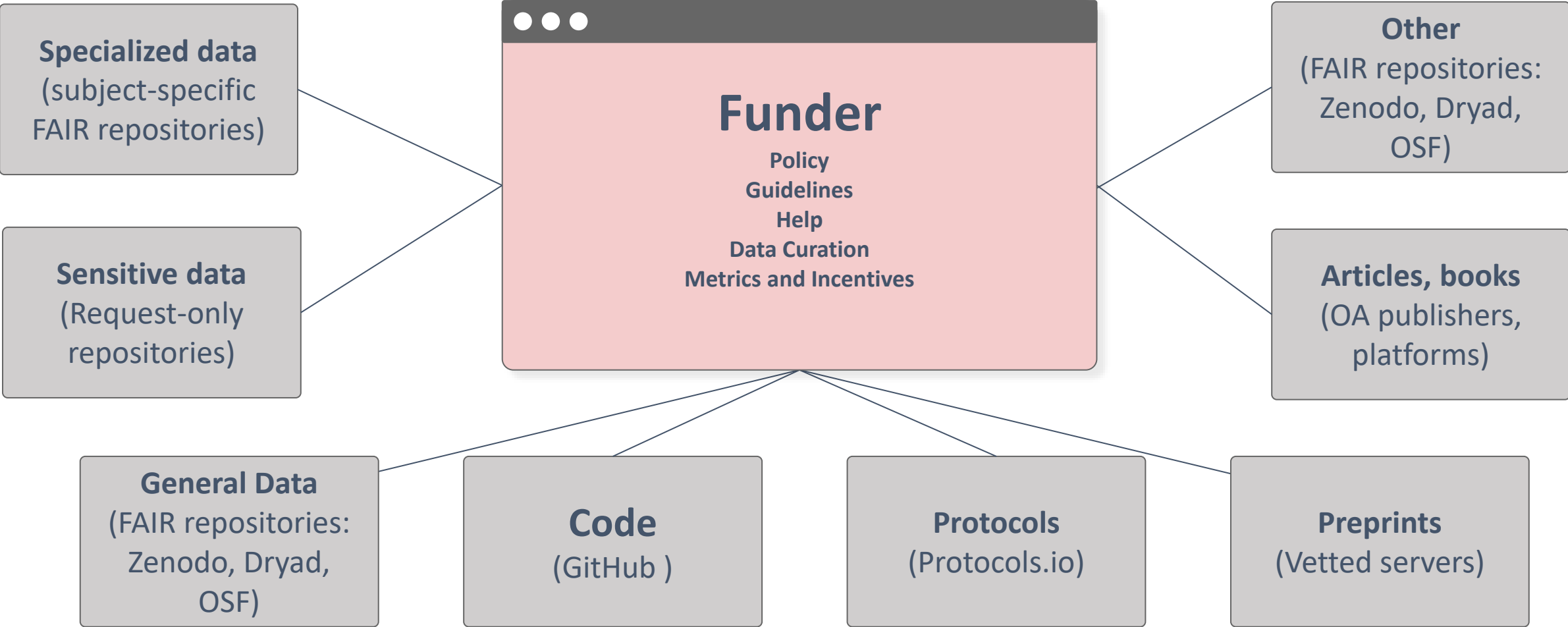
- Clean
- Annotate
- Describe
- Complete metadata
- DOI
- License

Where



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

Preferred Tools Network



Compliance

- Emerging tool chain
- Look for the more transparent or open source partners
- Support efforts to establish standards
- Leading Services:
 - OA.Works – open access and preprint monitoring and reporting
 - OpenAlex – open access, preprint, and person-data reporting
 - DataSeer – data sharing monitoring and reporting
 - DataCite – persistent identifier organization for datasets and other non-article outputs

Open Mic

Please share a sentence or two about a data sharing policy, question, experiment or other activity that you think might interest the group or that you want to see covered

Coming Next: Consultative Sessions

- Sign up for one or more deeper dive small group sessions that will cover:
 - Policy-making
 - Best Practices and Implementation
 - Compliance and Impact Tracking

What's coming next at HRA

EVO Program Schedule – Spring 2023

1. April – May: Data Sharing Consultative Sessions (sign up)
2. May – Setting up Open Research programs, first steps and best practices
3. June- July BOF - Facilitating Open Research with grantees

Future topics:

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

[Suggest a topic!](#)

[Open Science Resources](#)

[Download the PDF here](#)



Health
Research
Alliance

Thank you!