World Data System
Webinar Series 2024

Enhancing Data Sharing: Insights and Initiatives

18 September 2024
19:00 UTC / 15:00 EDT

DOI: 10.5281/zenodo.13829739

WORLD DATA SYSTEM

Explore the Path to Proper Participation:
Visit Our Etiquette Page!

- Go to worlddatasystem.org/virtual-meeting-etiquette

  or scan the QR code

WDS Virtual Meeting Etiquette

- We welcome your questions in the Q&A section

- Please note that this webinar will be recorded

Speakers

Kerry Goetz

Andrea Guemez

Amber Reed

Agenda

5 Minutes – Welcome &Logistics - (Cameron)

10 Minutes – NIH & Data Sharing (Kerry)

20 Minutes – Yet2 data sharing study (Andrea)

10 Minutes – NIH data sharing new initiative (Amber)

10 Minutes – Q&A

5 Minutes – Closing

Please Welcome Kerry Goetz from NIH!

NIH Data Management and Sharing

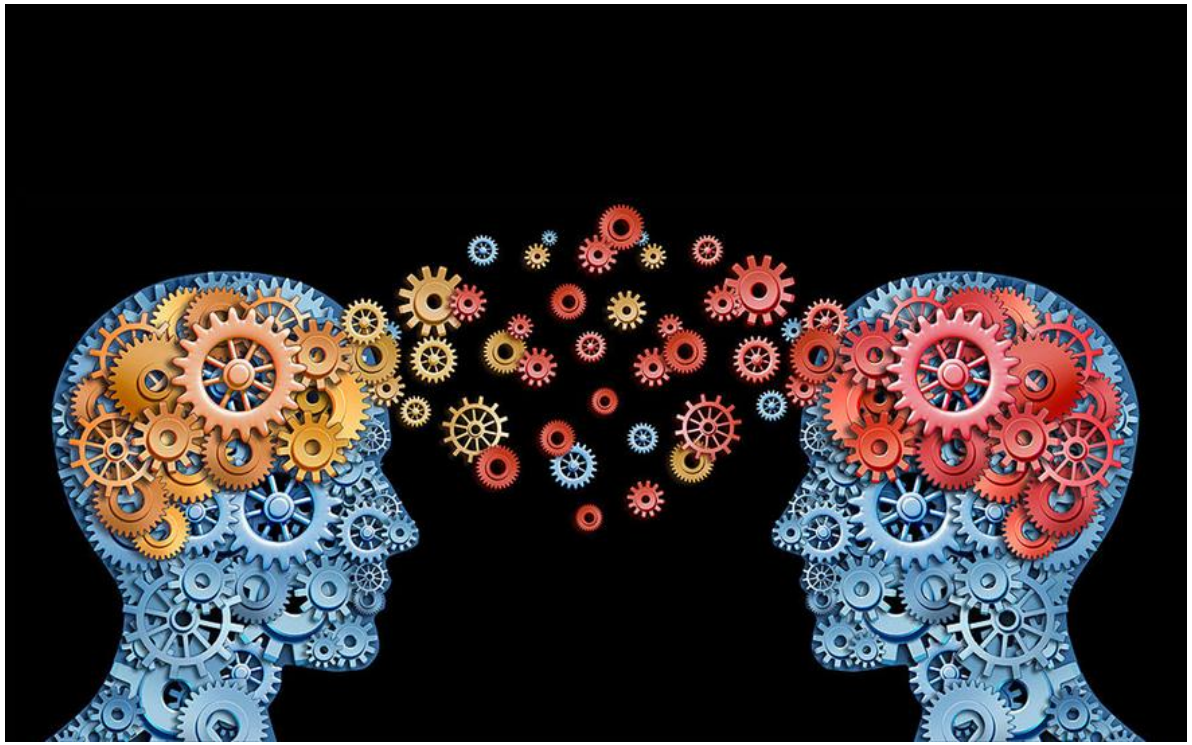
Kerry E Goetz

National Eye Institute, NIH

Associate Director, Office of Data Science and Health Informatics

World Data System

Sept. 18, 2024



Benefits to Data Sharing

- **Better Science** – reproducibility, collaboration, accelerate future science

- **Public Trust** – transparency, maximize contributions of participants, good stewardship

FAIR Principles

GO FAIR is committed to making data and services **findable**, **accessible**, **interoperable** and **reusable (FAIR)**.

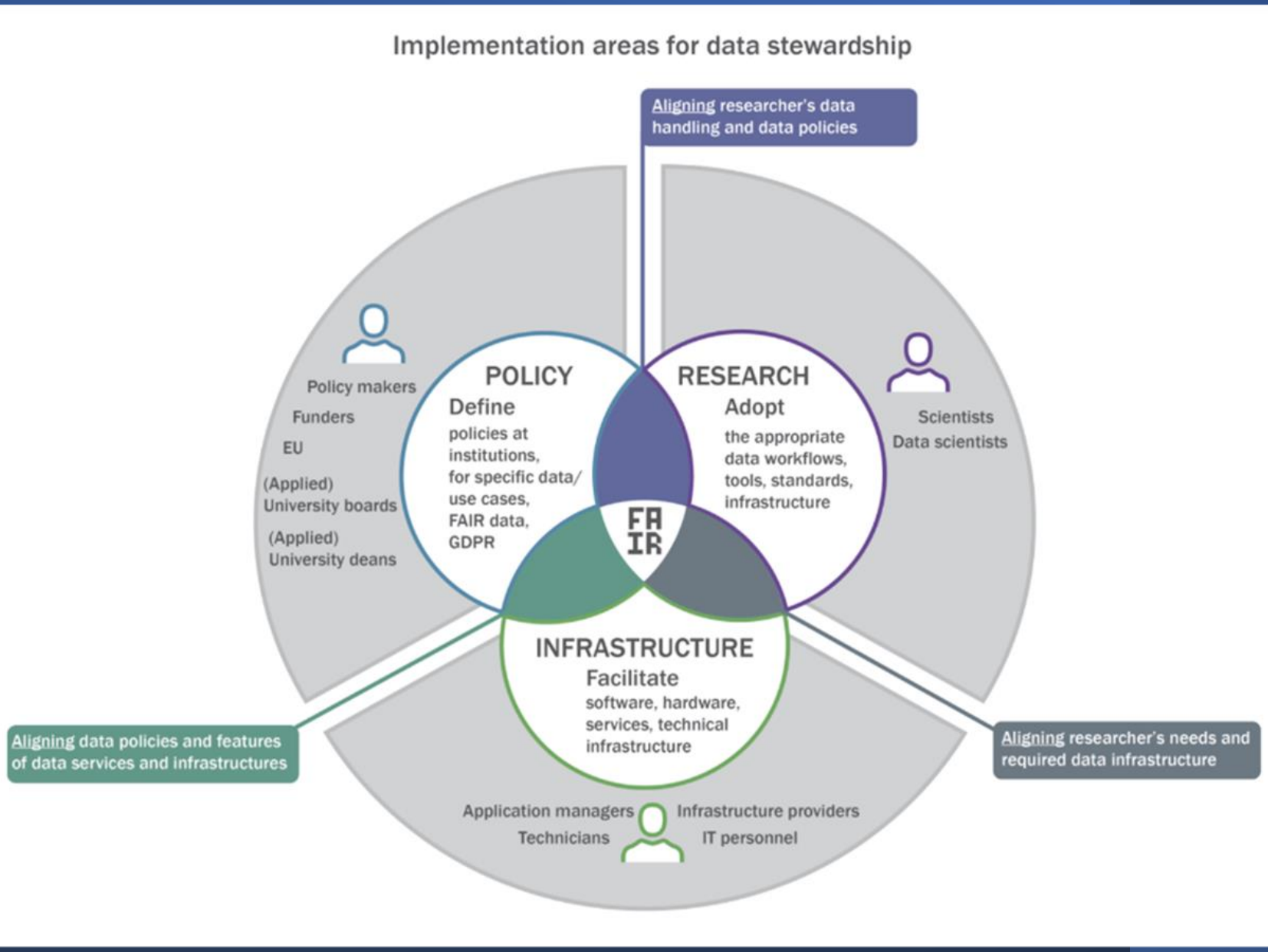
Findable: Metadata and data should be easy to find for both humans and computers.

Accessible: The exact conditions under which the data is accessible should be provided in such a way that humans and machines can understand them.

Interoperable: The (meta)data should be based on standardized vocabularies, ontologies, thesauri, etc. so that it integrates with existing applications or workflows.

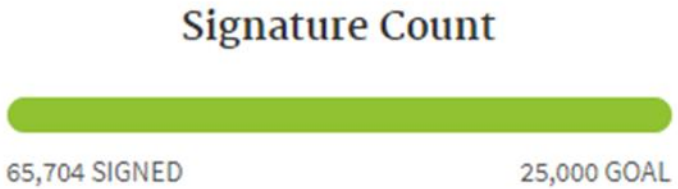
Reusable: Metadata and data should be well-described so that they can be replicated and/or combined in different research settings.

GO FAIR


Implementation areas for data stewardship
Aligning researcher's data handling and data policies
POLICY
Define
policies at institutions, for specific data/ use cases, FAIR data, GDPR
RESEARCH
Adopt
the appropriate data workflows, tools, standards, infrastructure
FA IR
INFRASTRUCTURE
Facilitate
software, hardware, services, technical infrastructure
Policy makers
Funders
EU
(Applied) University boards
(Applied) University deans
Scientists
Data scientists
Application managers
Technicians
Infrastructure providers
IT personnel
Aligning data policies and features of data services and infrastructures
Aligning researcher's needs and required data infrastructure

We the People…

…Require free access over the Internet to scientific journal articles arising from taxpayer-funded research.



We believe in the power of the Internet to foster innovation, research, and education. Requiring the published results of taxpayer-funded research to be posted on the Internet in human and machine-readable form would provide access to patients and caregivers, students and their teachers, researchers, entrepreneurs, and other taxpayers who paid for the research. Expanding access would speed the research process and increase the return on our investment in scientific research

https://petitions.obamawhitehouse.archives.gov/petition/require-free-access-over-internet-scientific-journal-articles-arising-taxpayer-funded/

History

This is historical material "frozen in time". The website is no longer updated and links to external websites and some internal pages may not work.

the WHITE HOUSE
PRESIDENT BARACK OBAMA

BRIEFING ROOM | ISSUES | THE ADMINISTRATION | 1600 PENN

HOME · BLOG

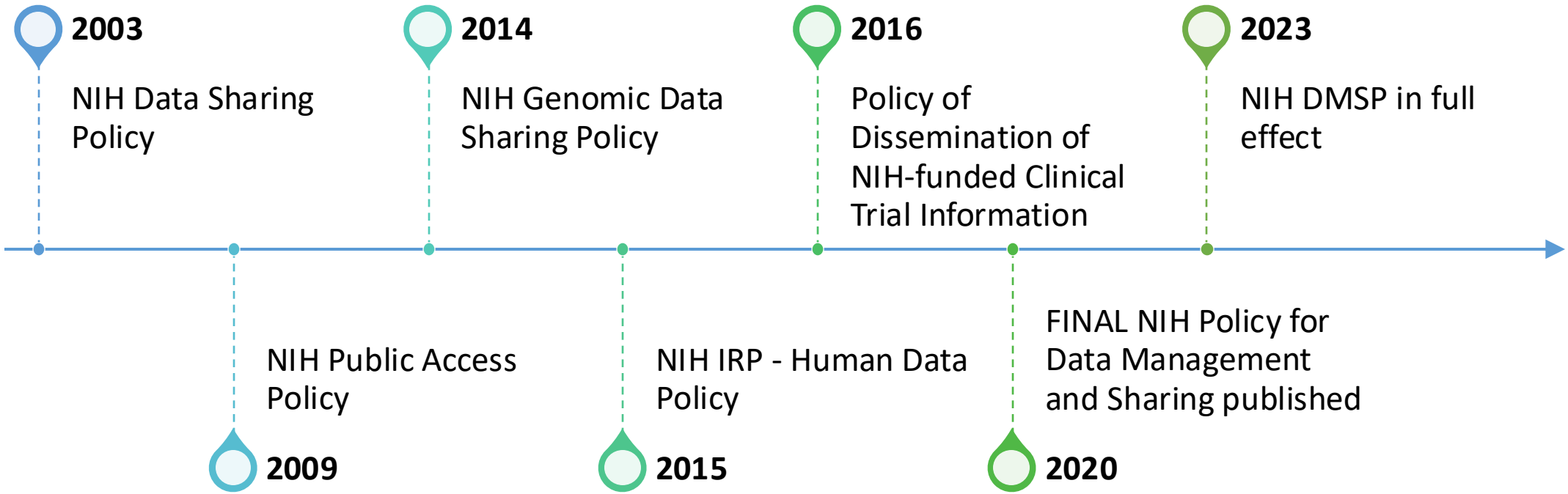
Expanding Public Access to the Results of Federally Funded Research

FEBRUARY 22, 2013 AT 12:04 PM ET BY MICHAEL STEBBINS

Summary: The Obama Administration is committed to the proposition that citizens deserve easy access to the results of research their tax dollars have paid for. That's why, in a policy memorandum released today, OSTP Director John Holdren has directed Federal agencies with more than $100M in R&D expenditures to

”directed Federal agencies with more than $100M in R&D expenditures to develop plans to make the published results of federally funded ***research freely available to the public*** within one year of publication and requiring ***researchers to better account for and manage the digital data*** resulting from federally funded scientific research”

NIH Data Policies



NIH ICOs

- 27 Institutes and Centers, Offices (ICOs)
  - NEI – National Eye Institute
  - NHLBI – National Heart, Lung, and Blood Institute
  - NCI – National Cancer Institute
  - NIEHS – National Institute of Environmental Health Sciences
  - CIT – Center for Information Technology
  - FIC – Fogerty International Center
  - Etc….

NIH Policy for Data Management and Sharing

- Data Management and Sharing Plan for all NIH funded research (Intramural, Grants, Contracts) generating **Scientific Data**
- Policy includes an example template and suggestions on data repositories
- **NIH ICOs to develop their own framework for compliance**
- Non-compliance may affect future funding

Scientific Data

- Scientific data is 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.
- Scientific data do 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.

- **Data sharing policies do not expect researchers to share all scientific data from a study**
- **Data does include negative results**

Key Elements

Outline of how scientific data are MANAGED and SHARED

Includes where there should be exceptions and limits (privacy and confidentiality, IP, Tribal sovereignty, etc.)

Timeline – no later than publication or end of award

Key Elements of Template

- Data type
- Related tools, software, code
- Standards
- Data preservation, access, timelines
- Access, distribution, reuse considerations
- Oversight of data management and sharing

Plan Submission and Review

Extramural Grant Awards*

Plan Submission

With application for funding in Budget Justification section

Plan Assessment

Peer reviewers only comment on (not score) budget

NIH program staff assess Plans

Plans can be updated

Plan Compliance

Incorporated into Terms and Conditions

Monitored at regular reporting intervals – mechanisms and tools to support oversight under development

Compliance may factor into future funding decisions

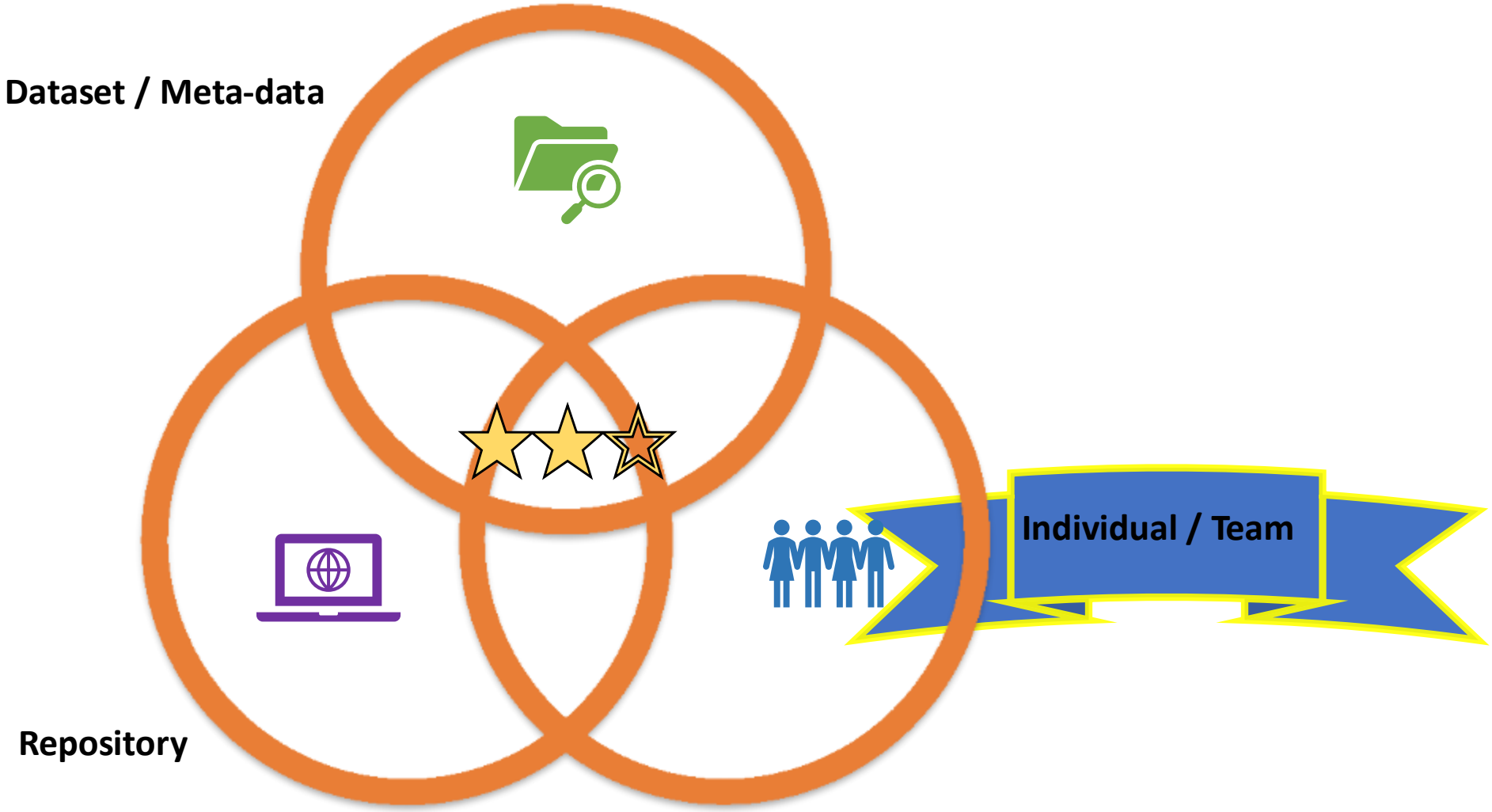
**Analogous requirements for contracts, OTAs, IRP*

Slide courtesy of the NIH OSP

Conclusion

- Responsible data management and sharing should be the default practice
- All scientific data should be managed; not all of it needs to be shared
- Prospective planning is essential for informed consent and data reuse

Rewarding Data Sharing



Next Welcome Andrea Guemez from Yet2!

IGNITING INNOVATION

NEI ODSHI Topic Specific Scouting

Enhancing Data Sharing: Insights and Initiatives

September 18th, 2024

Project Objective and Methodology

In conjunction with the NIH's Data Management and Sharing Policy, the NEI and ODSHI aim to identify incentives that encourage researchers to share data willingly, emphasizing motivation beyond mere obligation.

Project Objective

- **Identifying** metrics that can measure the impact of sharing primary data, including datasets, software tools and sequence data.
- **Identifying** alternative methods to encourage primary data sharing.

Results & Methodology

- *yet2* has evaluated potential factors that may have influenced the h-index's rapid adoption
- *yet2* has conducted a global search effort to identify novel and nascent author-level indexes with the potential to encourage and incentivize data sharing.
- *yet2* has interviewed over 40 experts across different scientific domains, including life, computer and social sciences, to better understand the potential of an index to incentivize data sharing.
- *yet2* has identified gaps within the data sharing ecosystem that can potentially hinder the success of an index.
- *yet2* has included potential alternatives to an index to incentivize data sharing.

Metrics

Potential Targets identified/reviewed by *yet2*	**107**
Interviews carried out by *yet2*	40
Targets presented to Client	40
Life Sciences	36
Data Sharing / Open Science	35
Social Sciences	14
Computer Science / AI-ML	11
Multiple	11
Introductory Calls request	**0**
Countries Represented	16

Data Indexes Identified

- T-index (2015)
- U-Index (2018)
- OmicsDI Index (2019)
- Data Index (2021)
- SCIENCE-Index (2023)

Alternative efforts have been implemented to encourage data sharing

Paper-like-format Data

IUPHAR BPS crafts concise summaries in a 'paper-like format' from their data.

This approach could enable data to be cited similarly to papers, acknowledging both the creators of the dataset and those who contributed to its development

GABA$_A$ receptors in GtoPdb v.2021.3

Delia Belelli,[1] Tim G. Hales,[1] Jeremy J. Lambert,[1] Bernhard Luscher,[2] Richard Olsen,[3] John A. Peters,[1] Uwe Rudolph,[4] and Werner Sieghart[5]

Author information · Copyright and License information PMC Disclaimer

Contents Go to:

This is a citation summary for GABA$_A$ receptors in the Guide to Pharmacology database (GtoPdb). It exists purely as an adjunct to the database to facilitate the recognition of citations to and from the database by citation analyzers. Readers will almost certainly want to visit the relevant sections of the database which are given here under database links.

Database links Go to:

GABA$_A$ receptors

https://www.guidetopharmacology.org/GRAC/FamilyDisplayForward?familyId=72

References Go to:

1. Albaugh PA, Marshall L, Gregory J, White G, Hutchison A, Ross PC, Gallagher DW, Tallman JF, Crago M and Cassella JV. (2002) Synthesis and biological evaluation of 7,8,9,10-tetrahydroimidazo[1,2-c]pyrido[3,4-e]pyrimdin-5(6H)-ones as functionally selective ligands of the benzodiazepine receptor site on the GABA(A) receptor. *J Med Chem* 45: 5043–51 [PubMed] [Google Scholar]

Source: GABA$_A$ receptors in GtoPdb v.2021.3

Network Graphs

Network graphs could be used to measure how centralized data is to surrounding research.

DataCite and CrossRef have been developing PID Graphs to create connections between PIDs (Persistent Identifiers) and other resources Source

Data Sharing Rewards

Showcasing "Gold Stars" for good sharing practices like the Data Sharing Passports developed by ICPSR and the Center for Open Science's Data Sharing Badges

What led to the rapid adoption of the h-index?

***yet2* has identified 4 factors that could have potentially influenced the rapid adoption of the h-index:**

1. **Simplicity and Inclusivity:**

   The h-index offered a straightforward alternative to established indexes, such as the Science Citation Index, as it disregarded the journals where the papers were published and omitted considering author order.

2. **Early Adopters:**

   Following Dr Hirsch's publication, there was enthusiasm among colleagues, leading to the index being included in Stanford Physics Information Retrieval System (SPIRES).

3. **Education and Knowledge Diffusion:**

   Distinguished journals like Nature and Proceedings of the National Academy of Sciences published papers on the index.

4. **Accessibility:**

   Distinguished search engines, such as Scopus, Web of Science (WoS), and Google Scholar, integrated bibliometric tools for the automated calculation of the index.

Timeline Source

2005

- H-index is created by J.E. Hirsch
- SPIRES implemented the algorithm on its dataset
- Nature presented the index in an article and PNAS published a paper confirming interest in this metric.

2006

- Ann-Will Harzing developed a bibliometric tool to allow a quick & accessible calculation.

2007

- Elsevier included the h-index in Scopus

2008

- Thomson-Reuters integrated the h-index into the WoS.

2011

- Google Scholar Citations was opened. Any academic could create and control their profile page on Google Scholar.

Index - Key Insights

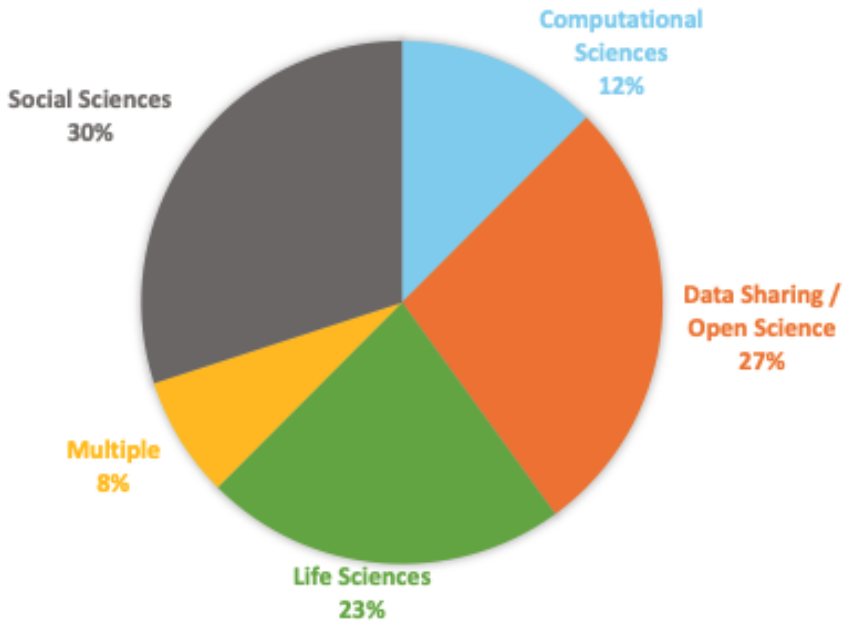
- Indexes **reward data sharing** by recognizing the work of researchers.
- Biggest challenges across indexes include **biases towards popular fields, lack of data citation, and potential misuse.**
- Indexes focus on scientific domains that share data, and don't allow a correction for geographical regions with fewer resources.
- There is **lack of evidence to suggest use of the indexes**. However, the OmicsDI and T-index have been previously used.

Index	Description	Advantages	Challenges	Inclusivity	Acceptability
data-index	Author-level metric measuring dataset usage. **Authors**: Hood, et al.	• Promotes generating and sharing data • Reward data sharing and **data usage**	• Potential misuse, like h-index. • Does not account for the varying amounts of data created across different disciplines • Requires robust citation architecture	• Can be used across different scientific disciplines & with varying data types	No evidence of use
OmicsDI-index	Composite index weighing data sharing through views, download, citations, and other metrics. **Authors**: Perez-Riverol et al.	• Considers more than a single aspect of data sharing. • Provides connections between different provider's datasets.	• Lack of data citation and ID assignation • Biases towards popular fields - Data volume influences index weights • Dependence on Provider's data and cooperation	• Omics data • Mostly US and EU datasets.	Datasets in OmicsDI platform.
u-index	Single metric index based on usage / awareness data citations form a citing universe. **Authors**: Callahan, et al.	• Can be modified to exclude self citations. • Anyone may define their own citing universe for a given resource.	Citation-based: • Biases towards popular fields • Lack of Data Citation • Positive/negative view not distinguished • Collected articles inaccuracy	Biological, medical, bioinformatical and medical informatically data	No evidence of use. *Authors appear to show no interest.
SCIENCE-index	Blockchain-based metric measuring a researcher's scientific contributions. **Authors**: Adams, et al.	• Takes data quality into consideration. • Transparency and decentralized control through Blockchain	• Require large, accurate data sets for model training • Computational and cost constraints of blockchain technology	• Claim geographical inclusivity/fairness. • Only target fields with large amounts of data.	No evidence of use – Recently published (2023)
T-index	How to measure the quality of financial Tweets (Twitter) **Authors**: Paola Cerchiello, et al.	• T-Index takes into account the confidence intervals; includes a variation mechanism	• Not specifically designed for research datasets	• All media contexts providing there is a count frequency and an impact.	• European Central Bank • Deutsche Bundesbank

yet2 Interviewed 40 Experts Globally

EXPERTS BY COUNTRY

EXPERTS BY SCIENTIFIC FIELD



EXPERTS BY ORGANIZATION

- Academia – 53%
- Non-profit – 15%
- For-profit – 7.5%
- Government – 7.5%
- Other – 17%

Data Sharing Simplified Workflow

Pre-Sharing		Sharing			Post-Sharing
Data Creation	**Data Curation**	**Repository Storage**	**ID Designation**	**Journal Publication**	**Data Citation and Implementation**
Areas of opportunity					
Early implementation of best practices in data creation (*curating as you go*)	Identify levels of data curation based on re-use potential and value - Creating awareness for researchers on curation services available (Dryad)	Creating awareness on bona fide repositories best to store data - Dedicated repositories for curated and un-curated data (i.e., controlled access vs generalist)	Get journals, repositories and institutes on board early - Engage Data Cite for support	Get journals and institutions on board early - Best Practices for journal data citation (including typesetters)	Journals need to enforce data citation prior to publishing papers - Institutes to value the data sharing index as much as the h-index

Researchers Have Suggested Alternative Efforts To Incentivize Data Sharing

Elevating the importance of data to the same level as papers might spark a cultural shift towards data sharing. This would require a Stick + Carrot approach.

- Collaboration among policymakers, universities, journals, and repositories is key to elevate data importance.
- Streamlining and simplifying the data sharing process could serve as an incentive for researchers to share more, as it is currently a difficult and time-consuming task.
- E.g., The **Royal Society uses a Stick + Carrot approach**:
  - Stick - firm open data mandate: Whenever a script is submitted, the author is required to make available all the code, research materials and datasets in the article. If not made available, the process will not move forward.
  - Carrot - The journals provide authors with data sharing support through data curation services provided by Dryad and encouraging proper data storage through a list of different available repositories.

People will only share data if they receive a personal benefit:

- Recognition/career progression
  - Could we start ranking schools through the amount of data they share?
  - Get employers and schools to use data sharing as a measure of success.
- Money – future funding relies on sharing data from precious grants.

Collecting the data and curating it need to have equal or greater value than the papers you have written about it. But don't devalue publications.

- Create more opportunities for data to be published
- The uptake of citation here is slow
- Behavior needs to be changed and the number of outlets that accept this needs to be increased

Concluding Remarks

There is a strong need for a data sharing incentive scheme

- Five indexes have been identified to incentivize data sharing, but none have been widely adopted.

Simplicity is key

- An index's acceptance potentially relies on an easily understandable and calculable formula, as well as a structured framework (I.e., automatic WoS/Google Scholar calculation).
- Support from journals, universities, search engines, and repositories can potentially enhance usage, as these control key elements of academic recognition (tenure, promotion, grants, visibility).

Shift attention away from impact

- Measuring impact is not trivial and goes beyond the scope of citations and downloads. It is important to avoid the intricacies of impact to promote index development.

A poorly designed index may result in more negative consequences than positive outcomes

- Investment in gating mechanisms, similar to a Data Peer-Review process, to ensure a standardized data and metadata quality.
- The data sharing ecosystem, particularly data citation, requires strengthening.

IGNITING INNOVATION

Thank You

www.yet2.com

Welcome Amber Reed from NIH!

National Eye Institute
Research Today...Vision Tomorrow

Advancing Data Sharing Through a New Initiative: The NIH Data Sharing (S-index) Challenge

Amber Reed, MPH
Assistant Director, Office of Data Science and Health Informatics
ICO Challenge Manager
National Eye Institute, National Institutes of Health

World Data System Webinar
September 18, 2024
No Financial Disclosures

Challenges: A Tool in NIH's Biomedical Innovation Toolbox

NIH Data Sharing Index (S-index) Challenge

Goal: The S-index Challenge aims to advance the NIH mission by actively promoting data sharing and developing a robust metric to reward exemplary data sharers. This initiative seeks to incentivize and recognize researchers who excel in making their data accessible, ultimately fostering a culture of openness and collaboration within the scientific community.

Mandatory Registration
[March 2025]

Phase 1 Deadline
[June 2025]

Phase 2 Deadline
[February 2026]

Phase 1 — Inspiration & Ideation

- Host virtual convening to build energy and engage participants
- Applicants compete in first round and finalists are chosen
- Finalists are awarded an interim prize, which is encouraged to be used for travel expenses to NIH Innovation Event for Phase 2

Phase 2 — Implementation

- Finalists compete for Grand Challenge to identify preferred S-metric at NIH Innovation Event
- Highlight each of the implementers and collect sentiment around sharing
- Winner(s) are announced & prizes are awarded

NIH
National Eye Institute

S-index Challenge Partners

Office of
Data Science Strategy

National Institutes of Health
Office of Extramural Research

National Institutes of Health
The BRAIN Initiative®

National Eye Institute
Research Today...Vision Tomorrow

National Institute of
Biomedical Imaging
and Bioengineering
Creating Biomedical Technologies to Improve Health

National Institute
on Drug Abuse

Eunice Kennedy Shriver National Institute
of Child Health and Human Development
Healthy pregnancies. Healthy children. Healthy and optimal lives.

National Institute of
Environmental Health Sciences

National Institute
of Mental Health

National Institute of
Neurological Disorders
and Stroke

National Center for
Biotechnology Information

How Can I Participate?

This Challenge is open to the public, including researchers, healthcare professions, data scientists, informaticians, and anyone interested in expanding data sharing in research. The competition aims to **stimulate innovation** and **advance open science principles**.

Registration Categories		
Individual	Registering on behalf of yourself.	Prize is awarded to individual.
Team	Registering as a group of individuals competing together but not on behalf of an established organization, institution, or corporation.	Prize is awarded to designated team lead.
Entity	Registering as a group of individuals competing together on behalf of a legally established organization, institution, or corporation.	Prize is awarded to entity.

To be eligible for a prize award, individual participants and team captains must be U.S. citizens or permanent residents. Entities must be incorporated and have their primary place of business in the U.S.

NIH
National Eye Institute

S-index Key Dates

Mandatory Registration (intent to participate): **March 3, 2025**

Phase 1 open: **April 21, 2025**

Phase 1 deadline: **June 2, 2025** (submit via challenge.gov)

Phase 1 Technical & Federal Judging: **July 7 – August 15, 2025**

Phase 1 Announcement of Finalists: **September 16, 2025**

Phase 2 open: **September 17, 2025**

Phase 2 deadline: **February 13, 2026** (submit via challenge.gov)

NIH Innovation Event: **Q1, CY26** (date TBD)

Phase 2 Technical & Federal Judging: **March 2 -April 10, 2026**

Announce Winner(s): **May 3-7, 2026**

*Tentative dates subject to change

Thank You!

Join us in promoting data sharing by developing a robust metric to reward exemplary data sharers:

https://bit.ly/47fcM5a

neiodshi@nei.nih.gov

NIH National Institutes of Health
Turning Discovery Into Health

NIH National Eye Institute

Q&A

Thank you!

WORLD DATA SYSTEM

October WDS Webinar

World Data System
Webinar Series 2024

Open Climate Knowledge & Data with Creative Commons

Taylor Campbell

Monica Granados

22 October 2024 @ 14:00 EDT/18:00 UTC

WDS ITO Workshop in October

In this workshop, participants will be given an introduction to the ESIP AI Data Readiness Checklist which encapsulates considerations for data quality, documentation, access and preparation. Following rounds of review and discussion for each of these checklist categories, participants will be invited to take part in a pilot exercise to apply this checklist to datasets in their own repositories. A brief synopsis of other AI-readiness related developments will be presented towards the end.

The workshop will be delivered in English and offered at two different dates/times to support a wider spectrum of geographical time zones.

Stay Tuned for Our Upcoming Dates with the Webinar Series!

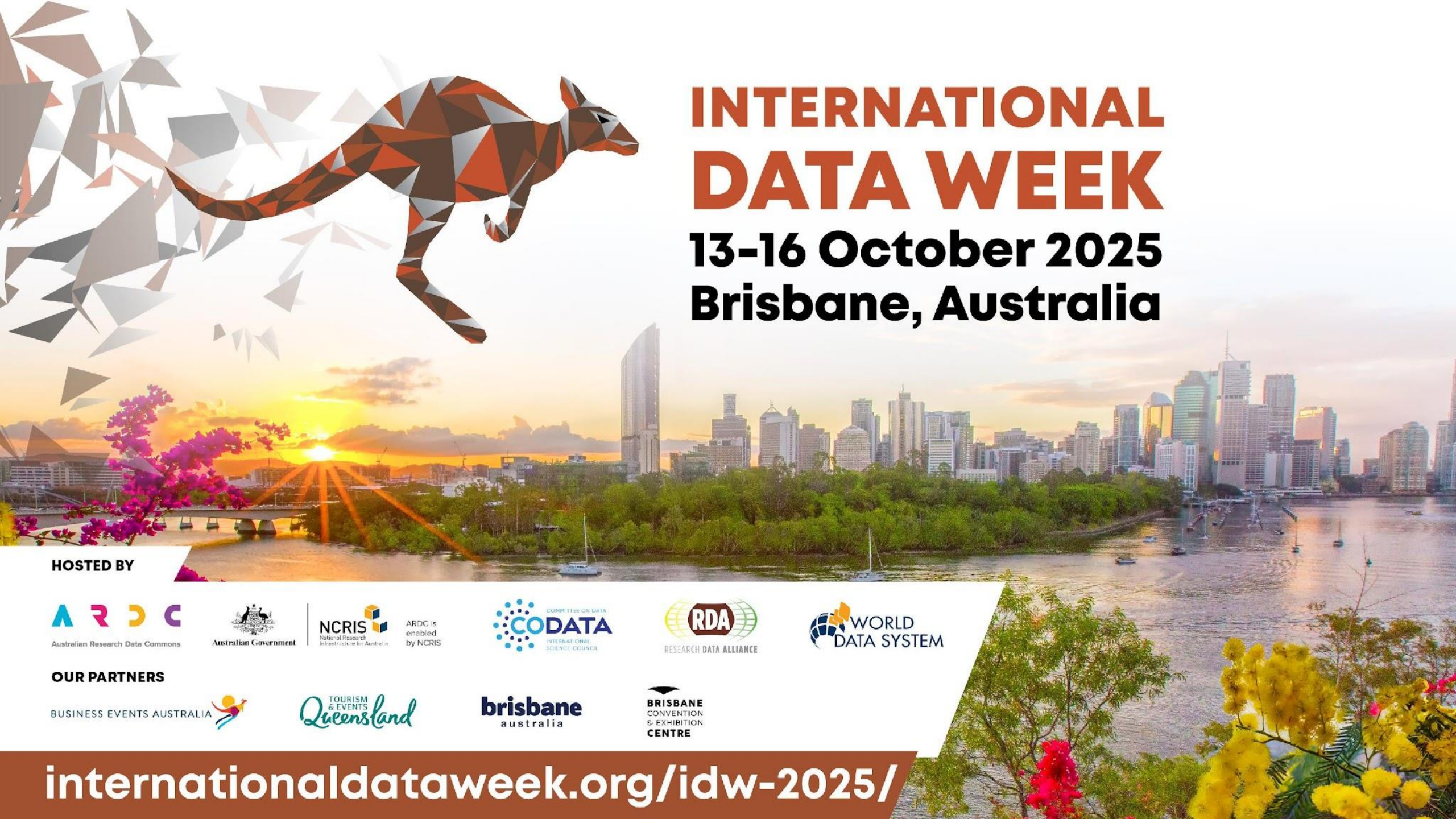
https://worlddatasystem.org/webinars/

Mark your calendars for our future webinars

- 19 November WDS Webinar with CTS

Rewind and Rediscover: Watch Our Past Webinar Sessions

WORLD DATA SYSTEM


INTERNATIONAL
DATA WEEK
13-16 October 2025
Brisbane, Australia
HOSTED BY
Australian Research Data Commons
Australian Government
NCRIS
National Research Infrastructure for Australia
ARDC is enabled by NCRIS
CODATA
RDA
RESEARCH DATA ALLIANCE
WORLD DATA SYSTEM
OUR PARTNERS
BUSINESS EVENTS AUSTRALIA
TOURISM & EVENTS Queensland
brisbane australia
BRISBANE CONVENTION & EXHIBITION CENTRE
internationaldataweek.org/idw-2025/

Participation Survey

- Reminder to participate in a short survey after the webinar concludes.
- Your feedback is key to refining our future webinars to cater more effectively to your interests and needs.
- We appreciate your input!

WORLD DATA SYSTEM

WDS and ECR Newsletters

- Next WDS Newsletter Release: October 3rd
- Next ECR Newsletter Release: September 19th

Read and subscribe for future updates around the WDS community!

Join a WDS Community!

- WDS Membership

- WDS-Early Career Researcher Network

For more info:

Email us at wds-ipo@utk.edu

Visit our website at worlddatasystem.org

Instagram: @wds_ipo

LinkedIn: @company/world-data-system

X (formerly Twitter): @ISC_WDS

Vimeo: @worlddatasystem