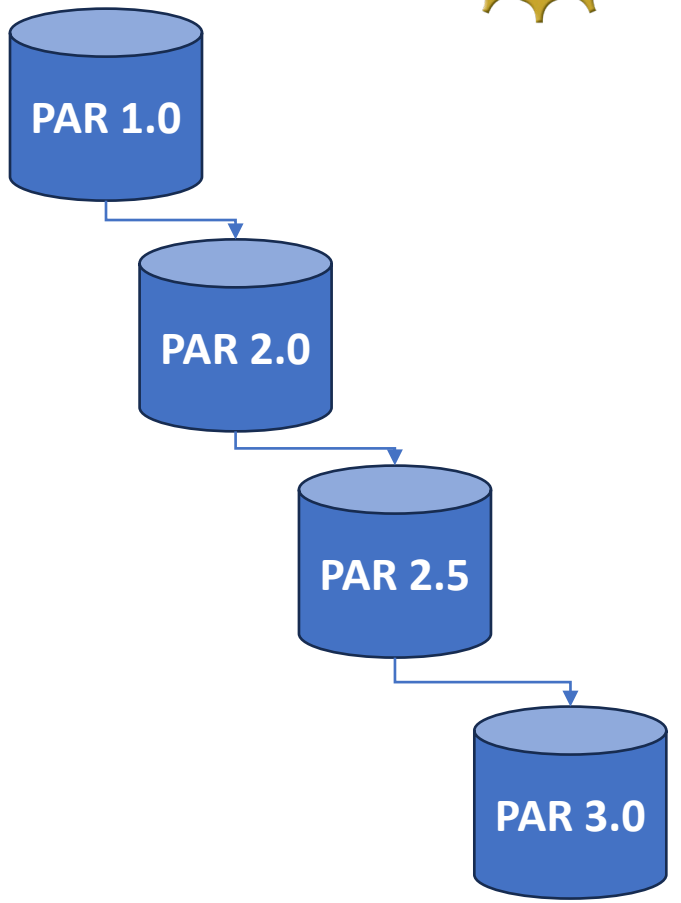
National Science Foundation

- Presentation by Martin Halbert, NSF Science Advisor for Public Access, mhalbert@nsf.gov
- The NSF Public Access Repository (NSF-PAR) is the formally designated agency repository used for compliance with federal public access deposit purposes by NSF awardees (URL: https://par.nsf.gov/)
- The NSF-PAR was developed and is maintained jointly as an ongoing infrastructure as a collaboration between the NSF and the Dept. of Energy
- The repository is a virtual system comprised of components within both the NSF Research.Gov and DOE Pages reporting systems
- Because this collaboratively maintained infrastructure leverages existing software and maintenance teams, it represents a significant annual savings in terms of federal investment in shared systems

DOI: 10.5281/zenodo.8208580

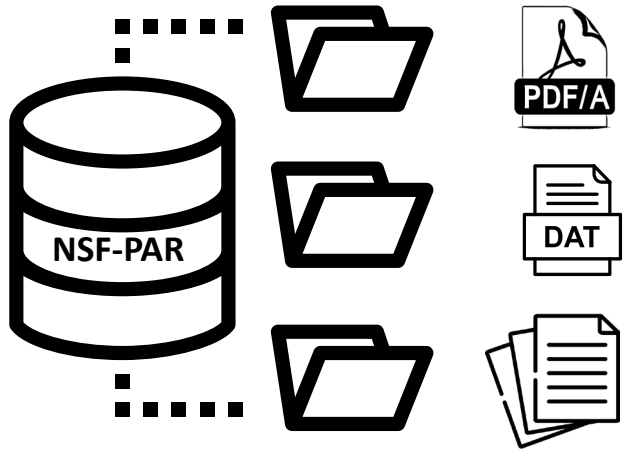
NSF PAR Development

- In 2013 the White House Office of Science and Technology Policy director John P. Holdren issued the memorandum entitled "Increasing Access to the Results of Federally Funded Scientific Research"
- PAR Version 1.0 focused on public access to peer-reviewed articles, and was brought online in 2015 as a part of the first NSF Public Access Plan (NSF 15-052) response to the Holdren Memorandum
- PAR 2.0 (implemented December 2021) added datasets as an optional public access format of research outputs
- PAR 2.5 will add conference proceedings and other formats in 2023
- PAR 3.0 will be implemented circa late 2024 to include requirements associated with the 2022 OSTP Nelson Memorandum



2022 OSTP Nelson Memo & NSF-PAR

- For the NSF-PAR, the primary updates mandated by the 2022 OSTP Nelson Memorandum relate to process changes, not functionality
- Specifically, the Memo requires that, for PAR:
  - the 12-month automatic embargo stipulated by the 2013 OSTP Holdren Memo be replaced by zero-embargo immediate access
  - Datasets comprising evidence in PAR peer-reviewed articles also be made publicly accessible immediately (rather than optionally, as before in PAR)
- NSF plans to continue implementing new format types (such as software) for optional public access reporting purposes in PAR in coming versions of the system
- The system will continue to be jointly maintained and developed with DOE as a collaborative effort, leveraging expertise in both agencies



Sustainability through Responsible Stewardship

Ishwar Chandramouliswaran

ishwarc@nih.gov

Lead Program Director, FAIR Data & Resources Unit

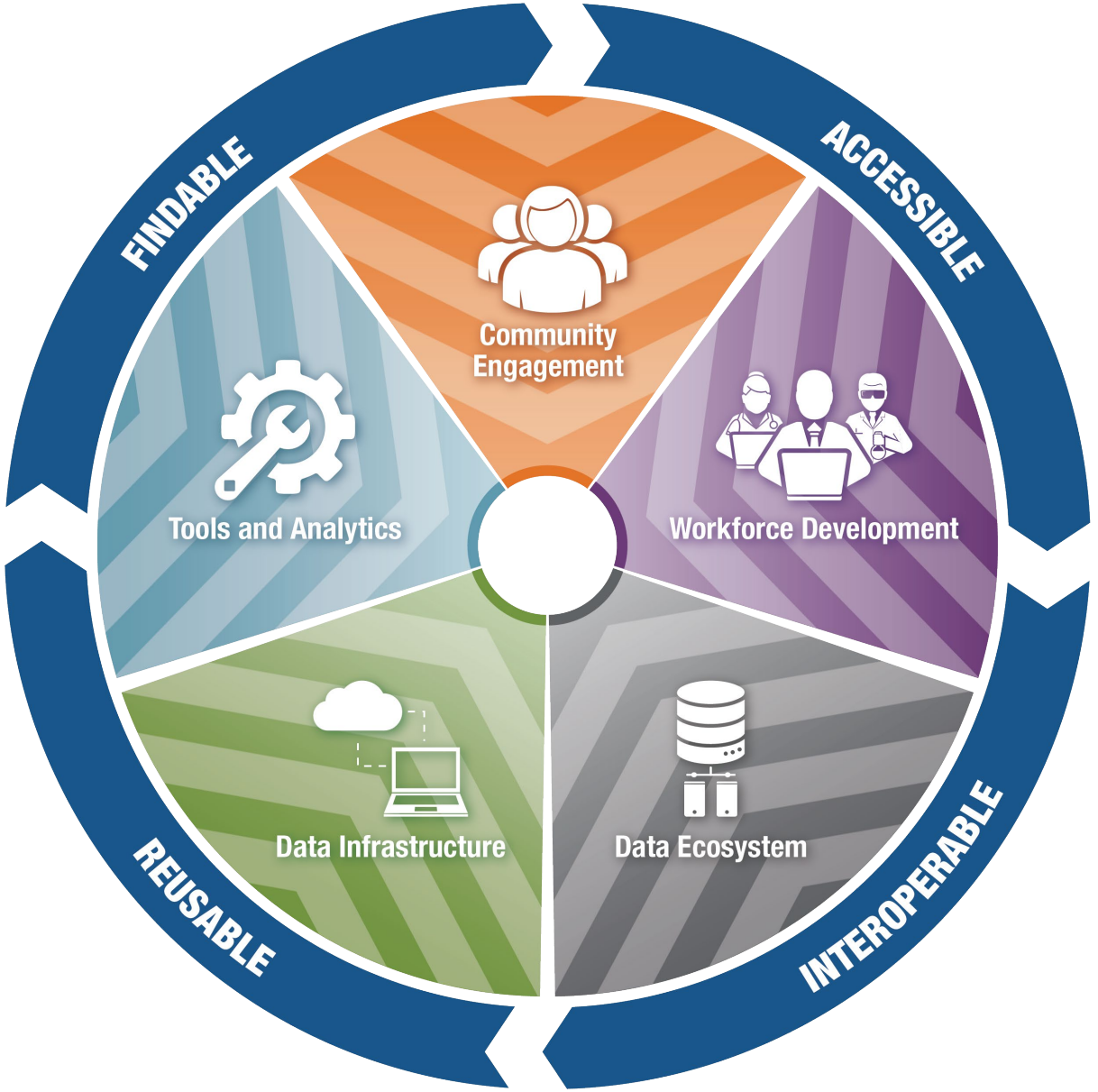
Office of Data Science Strategy, NIH

July 21, 2023

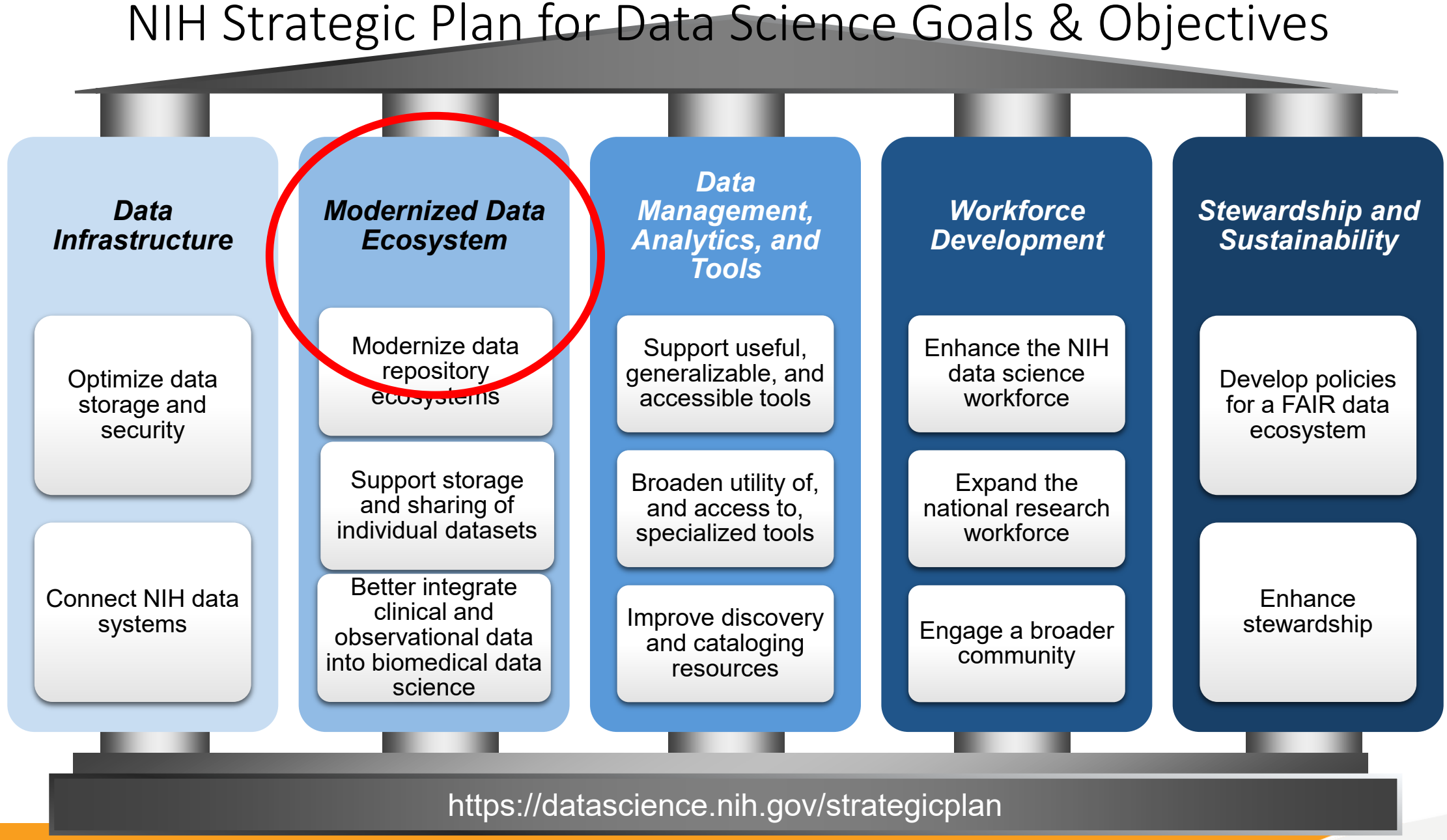
Office of Data Science Strategy

www.datascience.nih.gov

A modernized, integrated, FAIR biomedical data ecosystem



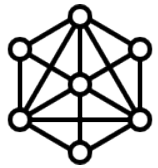
NIH Strategic Plan for Data Science Goals & Objectives



Reissue of Data Repositories and Knowledgebase Program to Supports Resource Life Cycle

NOT-OD-23-093 : Intent to fund early-stage biomedical data repositories and knowledgebases

- Align support with stage of the repository or knowledgebase
- Adopt best practices - NIH desirable characteristics of repositories
- Require open metrics – KPIs to measure scientific impact
- Emphasize domain-specific missions as a science driver



NOT-OD-23-099: Intent to fund established biomedical data repositories and knowledgebases

- Lower barriers for data sharing; reduce or eliminate silos
- Allow discovery and use of data (increase transparency)
- Optimize efficiency of operations and costs
- Disentangle data resources from research projects

This is fundamental to creating a modernized biomedical data ecosystem

https://dpcpsi.nih.gov/council/january-27-28-2022-agenda

GREI Program –Include Generalist Repositories in the NIH Data Ecosystem & the Concept of "*Co-opetition*"

work together on Common Capabilities & Best-practices

- Implement consistent capabilities (NOT-OD-21-016)
- Create better access to & discovery of NIH funded data
- Conduct outreach & train on FAIR data practices
- Engage the research community

Expected Outcomes & Impact

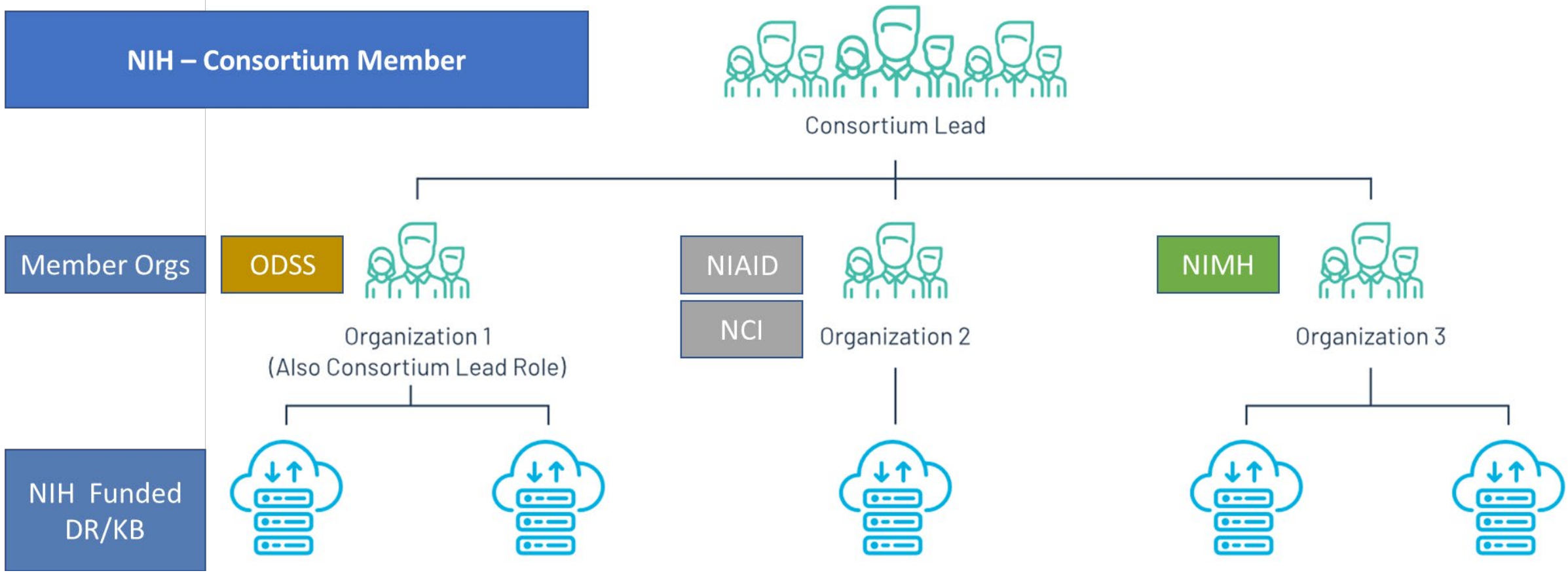
- Make data sharing easier
- Improve discoverability
- Increase reproducibility of research
- Encourage secondary use of data

https://datascience.nih.gov/data-ecosystem/exploring-a-generalist-repository-for-nih-funded-data

New Program to Build Sustainable Foundations for Open Software and Tools in Biomedical and Behavioral Science

- Support exceptional research software engineers (non-traditional PIs)
- Make sustainable impact on NIH research projects
- Enhance autonomy and career continuity for highly-skilled software engineers working in research
- Pilot solution to software engineering workforce challenge
- Support projects to develop robust, sharable, sustainable software and tools
- Bring software development best-practices & emerging technologies to NIH research projects
- Foster collaborations between biomedical and behavioral scientists and software engineers
- Extend impact of investment – for reuse; foster communities for open software development

***Most importantly**: Efforts to support open software and tools are critical to creating a modernized biomedical data ecosystem that will catalyze advances in science.*



Established NIH DataCite Consortium

ODSS established NIH as a consortium member in Aug 2022; will meet a critical need to mint PID's - digital object identifiers (DOIs) thereby supporting the implementation of FAIR principles for data generated from NIH funded and conducted research.

NIH-OD | NCI | NEI | NHLBI | NIA | NIAID | NIBIB | NICHD | NIDDK | NIDA | NIMH

Established a Data Management Center of Excellence

The purpose of this project is to support FAIR and DMS objectives through establishing a multi-functional Data Management Center of Excellence at NIH (DMCOE) which will:

- Research, collect, and annotate FAIR and DMS resources and best practices
- Knowledge transfer on FAIR and DMS from community led efforts to key stakeholders
- Coordination and communication of DMS and FAIR amongst ODSS, OER and ICs
- Provide additional relevant expertise

Deliverable

- **Landscape/Gap Analysis** - Universities' Efforts in NIH DMS Policy and FAIR Guidelines; repositories review
- **FAIR and DMS Best Practice Knowledge Guides (**teaching/learning content) - Incorporating FAIR throughout the Research Journey; Content (OER Program Officer Workshop: Data Management & Sharing Plan Assessment, held June 22)
- **FAIR, DMS Events** – Data Management & Sharing Seminar Series; Getting Practical with FAIR; joint events with community led efforts such as DCN, ELIXIR, GO FAIR – Train-the-Trainer

DATAWorks! Prize

2022 Prize
2023 Prize

Entries Open: June 1, 2023
Entries Close: August 15, 2023

$500,000 Prize Purse

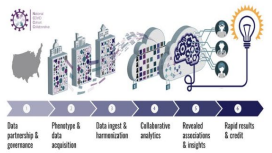
Up to 16 monetary prizes recognizing team developing in data sharing or reuse recipes

Highlighting the Power of Data Sharing and Reuse in the Biological & Biomedical Sciences

Learn More & Enter

www.herox.com/dataworks

Grand Prize, Data Sharing

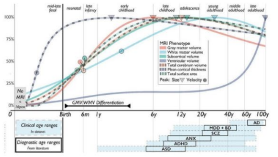


National Covid Cohort Collaborative

Democratizing Access to Clinical Data

The National Covid Cohort Collaborative (N3C) is a diverse multidisciplinary community dedicated to collaborative analytics at scale.

Grand Prize, Data Reuse



BrainChart

Brain Charts for the Human Lifespan

BrainChart combines MRI data from over 100,000 people to create and share growth charts of the human brain across the lifespan.

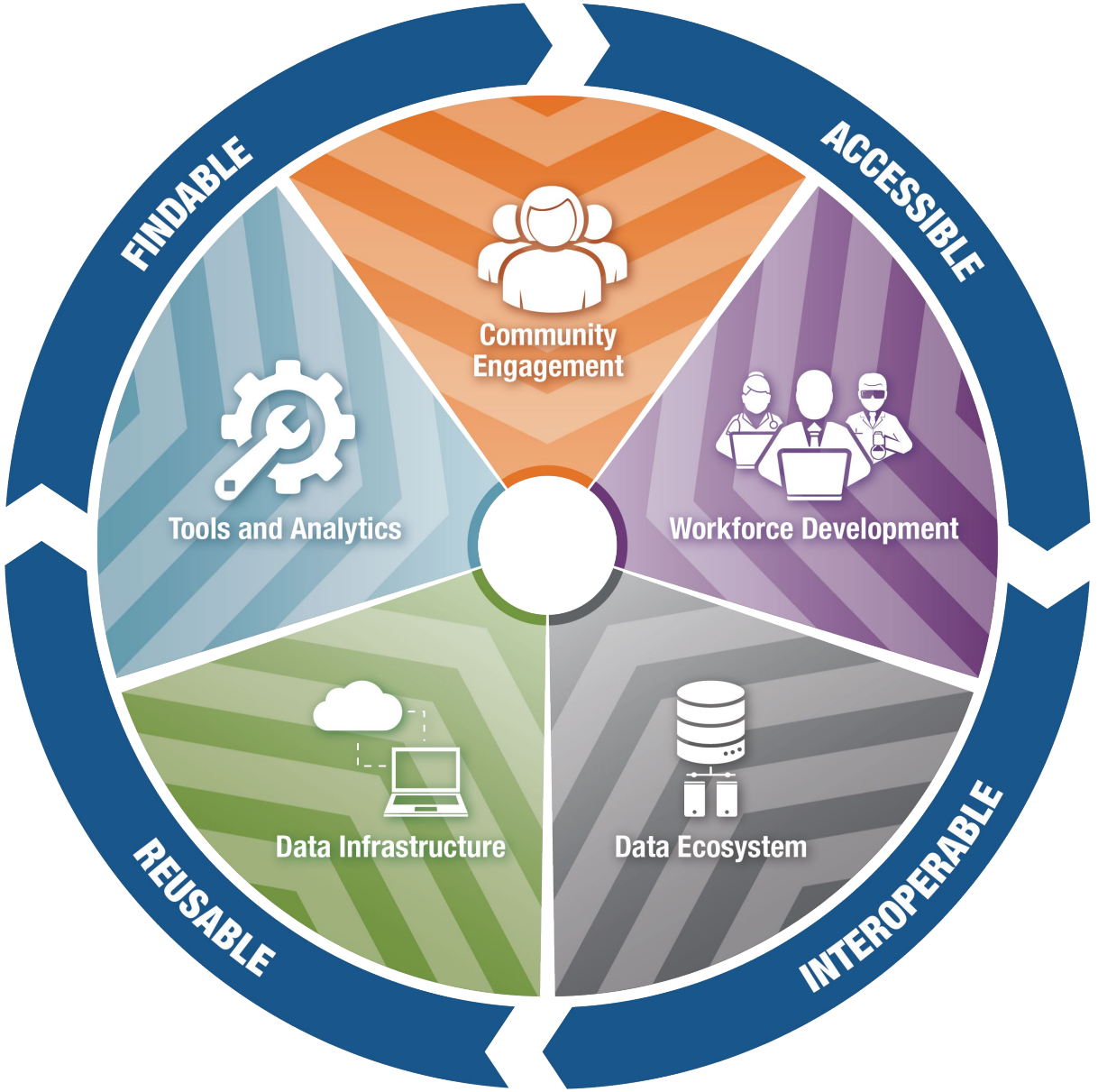
National Institutes of Health
Office of Data Science Strategy

DATAWorks! FASEB
Advancing Data Sharing and Reuse

Office of Data Science Strategy

www.datascience.nih.gov

A modernized, integrated, FAIR biomedical data ecosystem



@NIHDataScience /NIH.DataScience datascience@nih.gov

Michael Cooke, Department of Energy

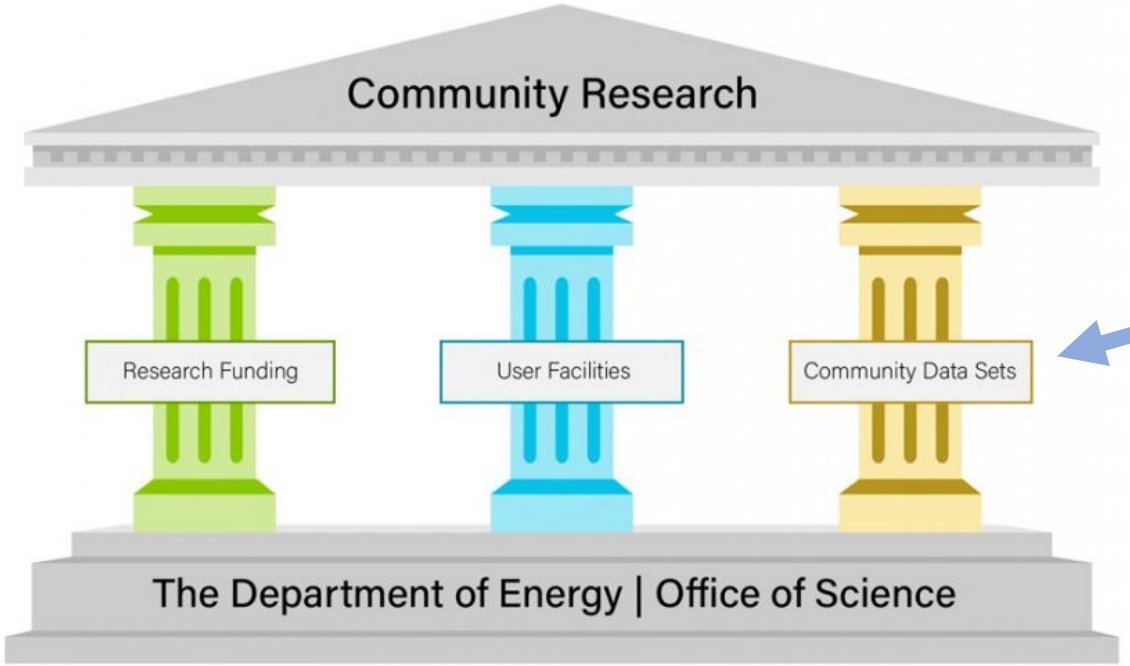
Senior Technical Advisor, Office of the Deputy Director for Science Programs

michael.cooke@science.doe.gov

- 2023 DOE Public Access Plan provides Data Management Principles:
  - Accelerate discovery through equitable data sharing
  - Share and preserve data to validate and replicate research
  - Maximize appropriate data sharing
- When new plan goes into effect, scientists will need to work with data experts to develop responsive Data Management and Sharing Plans
  - Share data underlying publications at time of publication
  - Provide timeline for sharing data not underlying publications
  - Use repositories that align with Desirable Characteristics of Data Repositories

U.S. DEPARTMENT OF ENERGY | Office of Science

Office of Science PuRe Data Resources



PuRe Data Resources

Public Reusable Research (PuRe) Data Resources aim to make data publicly available in order to advance scientific or technical knowledge

ATcT | JGI Joint Genome Institute | NNDC | PDG particle data group | The Materials Project | ARM | KBase Predictive Biology – DOE Systems Biology Knowledgebase

Characteristics of PuRe Data Resources

- Authoritative provider of public data/capabilities in its area
- Uses Persistent Identifiers (PIDs) to facilitate discovery, reuse, and citation
- Strives to advance the FAIR Data Principles, demonstrate the "Desirable Characteristics of Data Repositories" guidance by the NSTC SOS

Relevant Desirable Characteristics for PuRe Data Resources:

Organizational Infrastructure	Digital Object Management	Technology
• Free and Easy Access • Clear Use Guidance • Risk Management • Retention Policy • Long-term Organizational Sustainability	• Unique Persistent Identifiers • Metadata • Curation and Quality Assurance • Broad and Measured Reuse • Common Format • Provenance	• Authentication • Long-term Technical Sustainability • Security and Integrity

U.S. DEPARTMENT OF ENERGY | Office of Science

2023: The Year of Open Science

"The principle and practice of making research products and processes available to all, while respecting diverse cultures, maintaining security and privacy, and fostering collaborations, reproducibility, and equity."

https://open.science.gov/

PIDs@ OSTI.GOV

https://www.osti.gov/pids/

PuRe Data Resources

https://science.osti.gov/Initiatives/PuRe-Data

DOE CODE

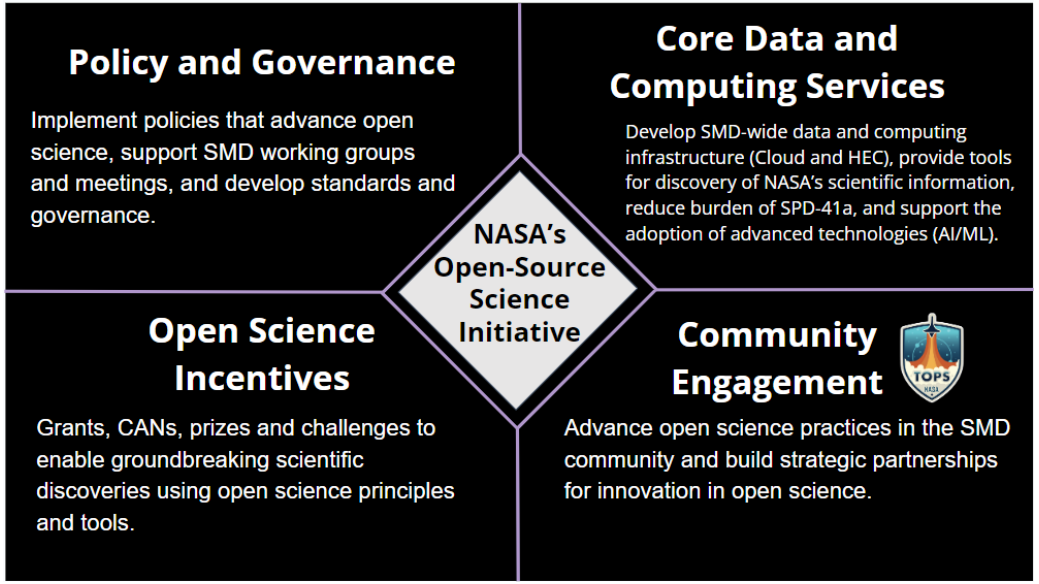
https://www.osti.gov/doecode/

Cerese Albers, Ph.D. NASA

Lead Program Executive, NASA Earth Science Data Systems program
cerese.m.albers@nasa.gov

- The Earth Science Data Systems (ESDS) program falls within the purview of the Earth Science Division (ESD), under the Science Mission Directorate.
- The ESDS program oversees the life cycle of NASA's Earth science data from acquisition through processing and distribution
  - Providing free and open data to all users for any purpose since 1994.
  - Since 2015 all data systems software developed through research and technology awards have been made available to the public as Open Source Software.
- NASA has a dedicated Open-Source Science Initiative (OSSI), led by the Chief Science Data Office at the SMD level. ESDS represents data repositories and the interests of the Earth Science Division on its council.
  - Policy, infrastructure, funding, and community

NASA SPD-41a Science Information Policy

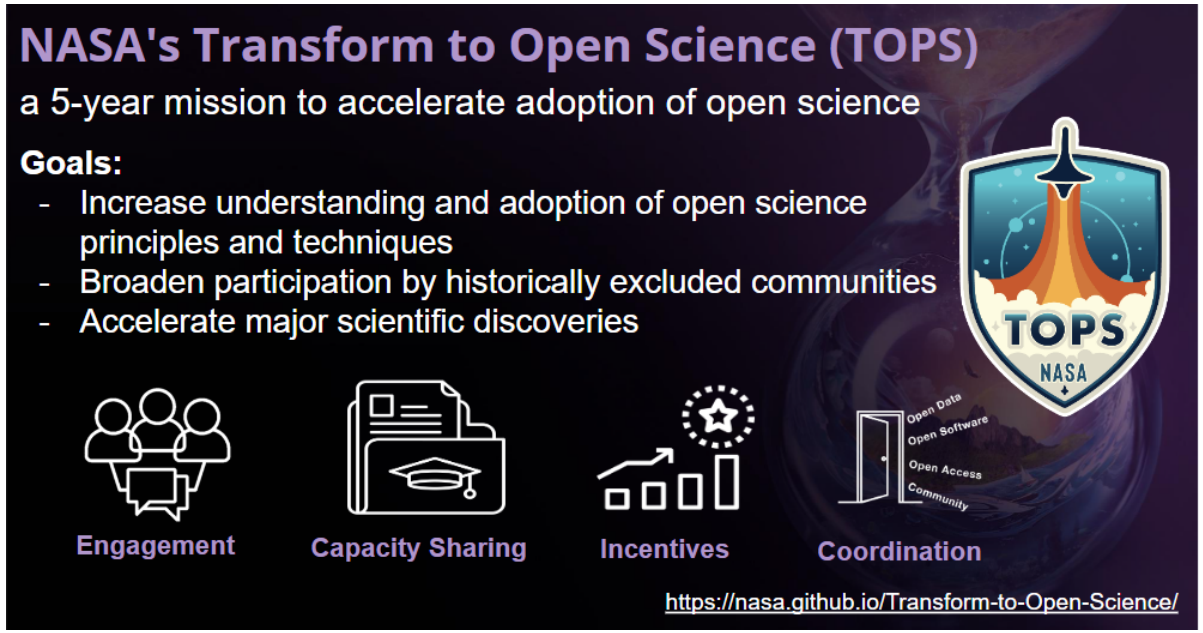


Open Source Science in Practice at NASA

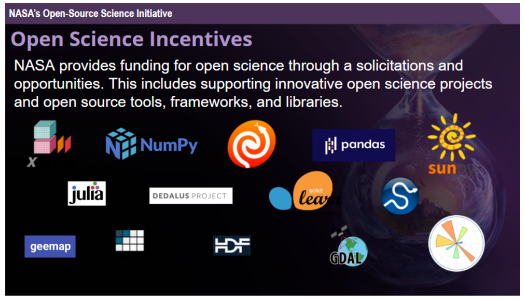
- The NASA is putting Open Source Science into practice:
  - ***Open*** the entirety of the scientific process from start to finish
  - ***Broaden*** community involvement in the scientific process
  - ***Increase*** accessibility of data, software, & publications
  - ***Facilitate*** inclusion, transparency, and reproducibility of science

- Data repositories play a critical role in implementing:
  - Data and Metadata standards and guides, DOIs, working Data Management Plans with data producers and ensuring data requirements are met for missions, increasing support and community engagement in open software and accessibility of science, preparing users to work in the cloud, policy compliance, reaching new communities, etc.

- Science Mission Directorate support for community through TOPS



How is NASA prioritizing open data access now and preparing for the future?



- ***Changing*** policy: NASA SMD policy SPD-41a and NASA Procedural Requirement
  - ESDS has updated its data and information policies, standards, templates and guidelines.
  - We created new requirements for upcoming mission data and software.
- ***Investigating*** potential new architectures that can meet upcoming mission science processing objectives.
- ***Studying*** new enterprise approaches to common services that enable increasing user-oriented focus.
- ***Prioritizing*** dataset cloud migration.
  - More than 4PB of data have already been migrated.
  - Already have some entire missions in the cloud from start-to-finish.
  - Two of our data repositories are already completely in the cloud, more on the way.
  - Co-locating datasets and observations of past missions, current missions, and future missions in analysis-ready cloud optimized (ARCO) formats to enable science in the cloud.
- ***Investing*** in cost-saving mechanisms with commercial vendors and collaborating amongst agencies.
- ***Funding*** open science tools, open science community support, open engagement and collaboration, and open software initiatives.

Funding Opportunities and Announcements

- ***Transforming*** out data systems to prepare for huge volumes of data and new ways of conducting science

NASA's Open-Source Science Initiative

NASA's approach for putting Open Science into practice.

https://science.nasa.gov/open-science-overview