Promoting Public Access to Research Data

Beth A. Plale
Science Advisor for Public Access
Office of Advanced Cyberinfrastructure (OAC)
Computer and Information Science and Engineering (CISE)
National Science Foundation

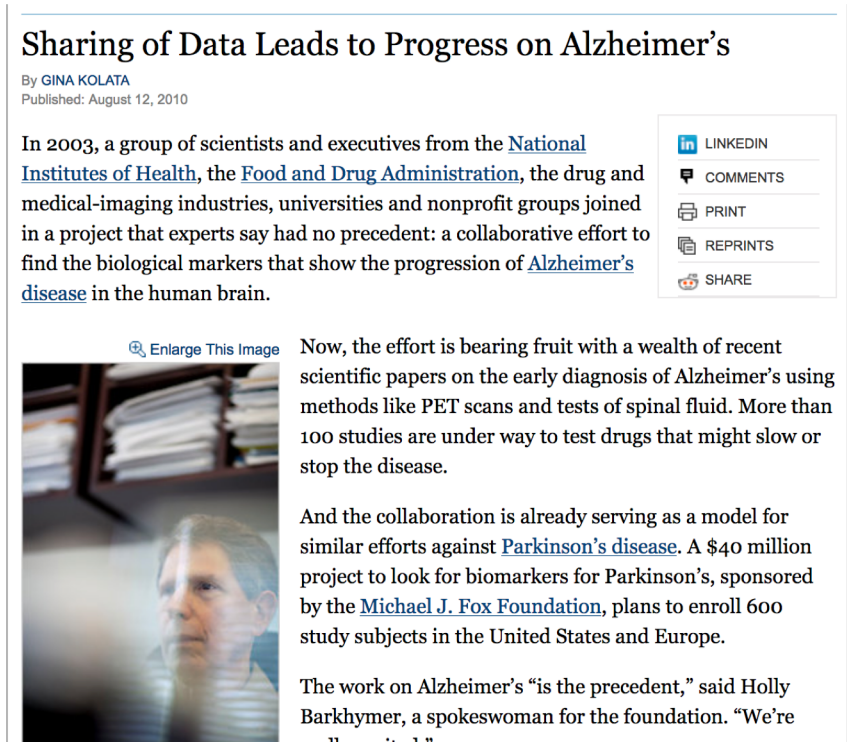
Disclaimer: the following slides convey personal observations on the current research data ecosystem landscape from the point of view of a US funding agency. It should not be construed as NSF policy.

APLU 2019 Joint CECE & COR Meeting

Why Focus on Research Data Reuse?

Environmental drivers of biodiversity: leveraging a history of NSF-funded research to test models of butterfly responses to global change (NSF 1839021)

Synchronization across terrestrial and aquatic ecosystems (NSF 1839011)



Sharing of Data Leads to Progress on Alzheimer's

By GINA KOLATA
Published: August 12, 2010

In 2003, a group of scientists and executives from the National Institutes of Health, the Food and Drug Administration, the drug and medical-imaging industries, universities and nonprofit groups joined in a project that experts say had no precedent: a collaborative effort to find the biological markers that show the progression of Alzheimer's disease in the human brain.

LINKEDIN
COMMENTS
PRINT
REPRINTS
SHARE

Enlarge This Image

Now, the effort is bearing fruit with a wealth of recent scientific papers on the early diagnosis of Alzheimer's using methods like PET scans and tests of spinal fluid. More than 100 studies are under way to test drugs that might slow or stop the disease.

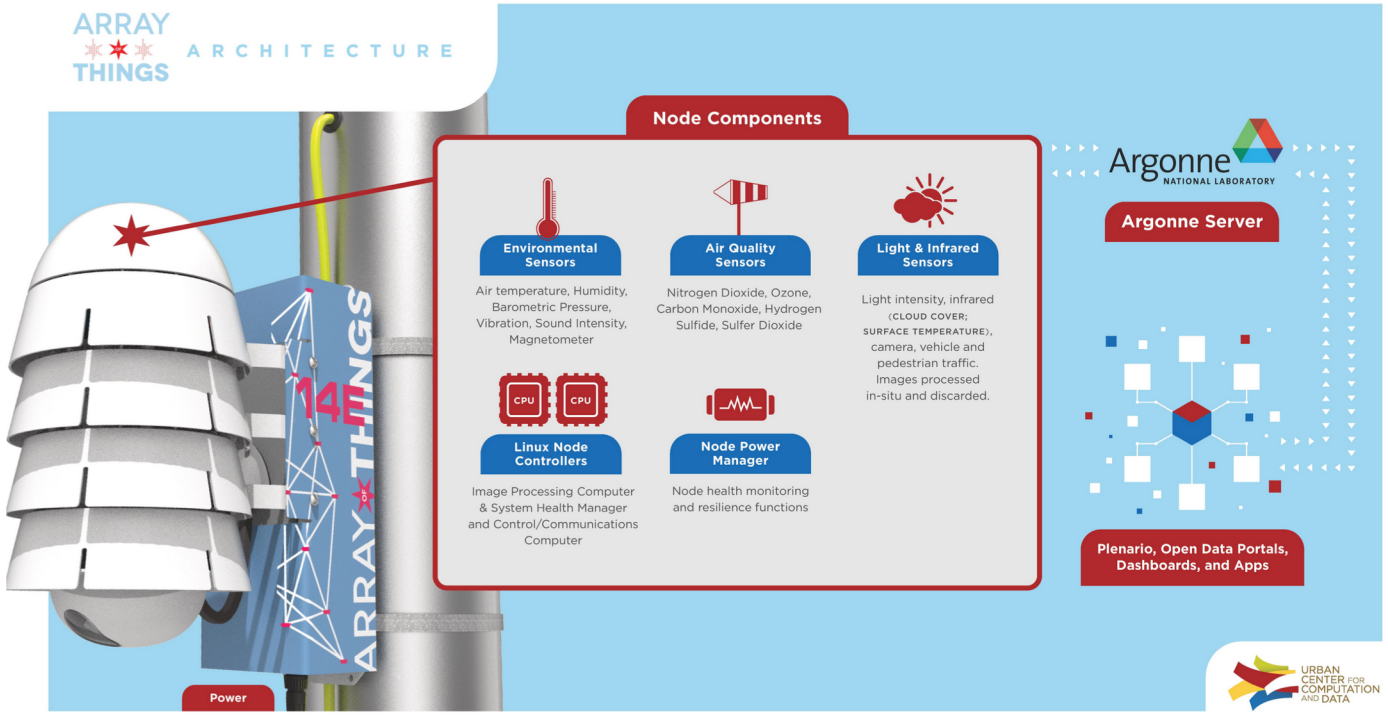
And the collaboration is already serving as a model for similar efforts against Parkinson's disease. A $40 million project to look for biomarkers for Parkinson's, sponsored by the Michael J. Fox Foundation, plans to enroll 600 study subjects in the United States and Europe.

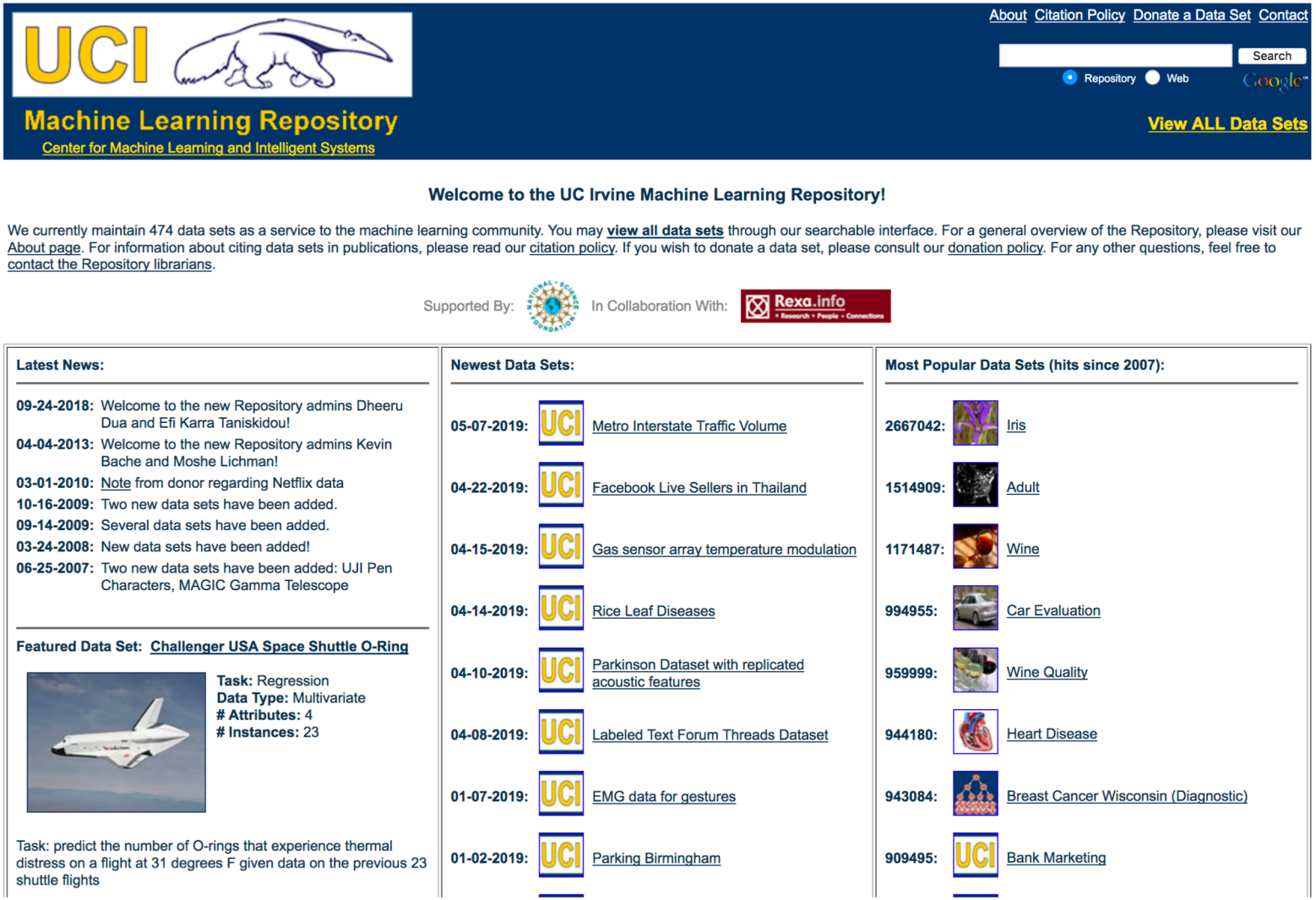
The work on Alzheimer's "is the precedent," said Holly Barkhymer, a spokeswoman for the foundation. "We're

Beth Plale, NSF
bplale@nsf.gov

Aggregating data for broad use: Smart and connected communities

MRI:
Development of an Urban-Scale Instrument for Interdisciplinary Research
(NSF 1532133)





Curated and growing community asset that fuels much of machine learning research

(2007—current)

Larger landscape of public access

Investigators are expected to share with other researchers, at no more than incremental cost and within a reasonable time, the primary data, samples, physical collections and other supporting materials created or gathered in the course of work under NSF grants.

National Science Foundation

NSF Public Access Repository

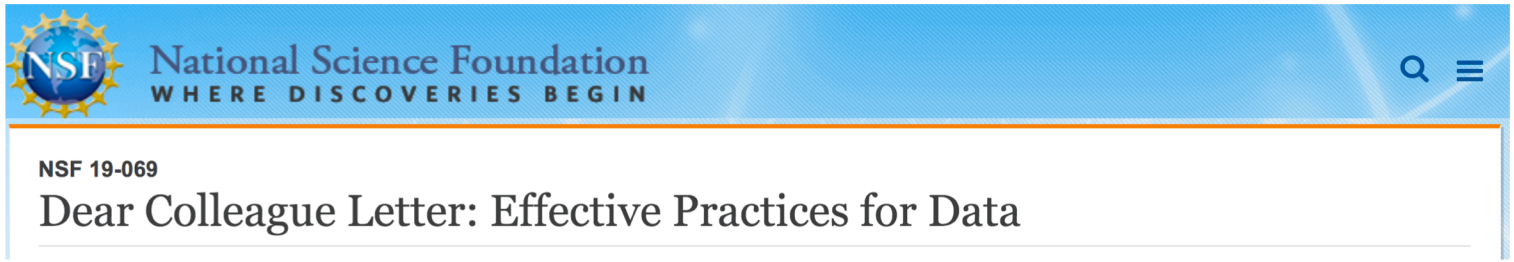
NSF Public Access Repository (NSF PAR)

- Provides public access to journal and juried conference papers
- Open access after 12 month embargo
- Award recipients deposit author copy of publication as part of reporting process
- Partnership with US Dept. of Energy. DOE PAGES system provides back end storage

https://par.nsf.gov/faq

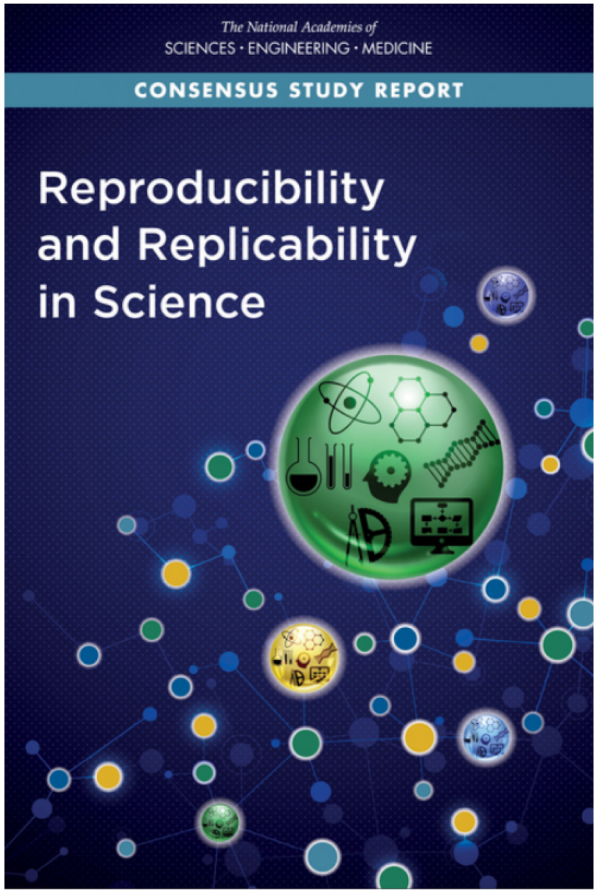
NSF Dear Colleague Letter (2019): Effective Practices for Data

- Encourages scientific community to learn effective practices for data, and implement them in proposals to NSF
- Focused on data that accompanies a publication
- Encourages use of
  - Globally Persistent IDs (PIDs) for data
    - Citation of data
  - Use Data Management Plan tool
    - To produce machine readable DMPs

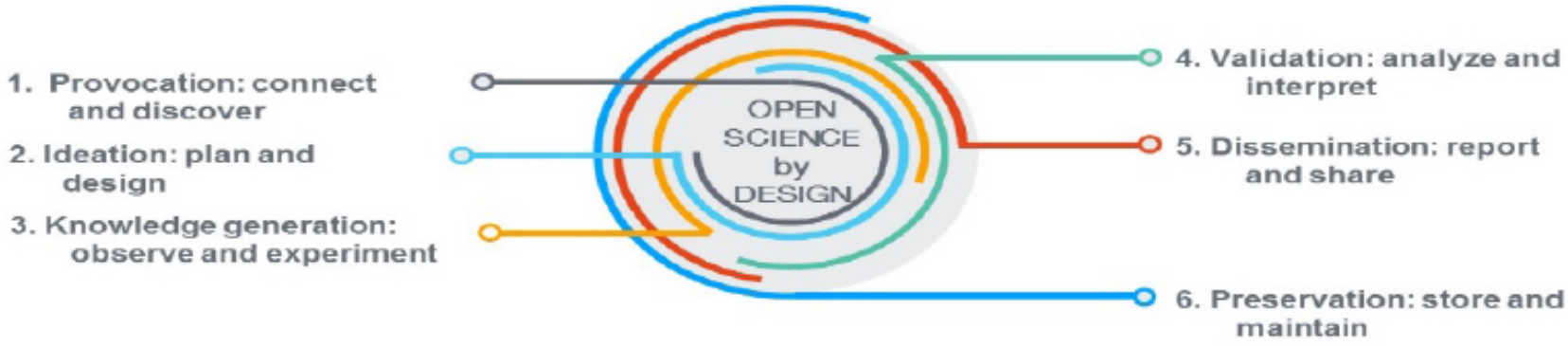


Reproducibility and Replicability in Science (2019)

- National Science Foundation, Directed by Congress, tasked an Academies committee to define what it means to reproduce or replicate a study, explore issues across science, and assess impact of these issues on public trust in science
- Distinguishes reproducibility from replicability



NASEM Open Science by Design: Realizing a Vision for 21st Century Research (2018)



November 29, 2017

AAU-APLU Public Access Working Group Report and Recommendations

PUBLIC/OPEN ACCESS | REPORTS | INTELLECTUAL PROPERTY

Fall 2017

The Association of American Universities (AAU) and Association of Public and Land-grant Universities (APLU) today released a report that details actions universities and federal agencies can take to ensure public access to federally-sponsored research data.

"Ensuring that research data are more accessible clearly has tremendous potential to fuel scientific analysis and discovery by making data more open to scrutiny, re-analysis, and extension," the report says. "By committing to a set of shared principles and minimal levels of standardization across institutions and agencies, we can help minimize costs, enhance interoperability between institutions and disciplines, and maximize the control institutions can exert over how they ensure access to publicly funded scholarship."

Association of Public & Land-grant Universities

Association of American Universities
Inquiry · Innovation · Impact

2018 WORKSHOP ON ACCELERATING PUBLIC ACCESS TO RESEARCH DATA

October 29-30, 2018 – FHI 360 Conference Center in Washington, DC

EUROPEAN OPEN SCIENCE CLOUD

European Commission

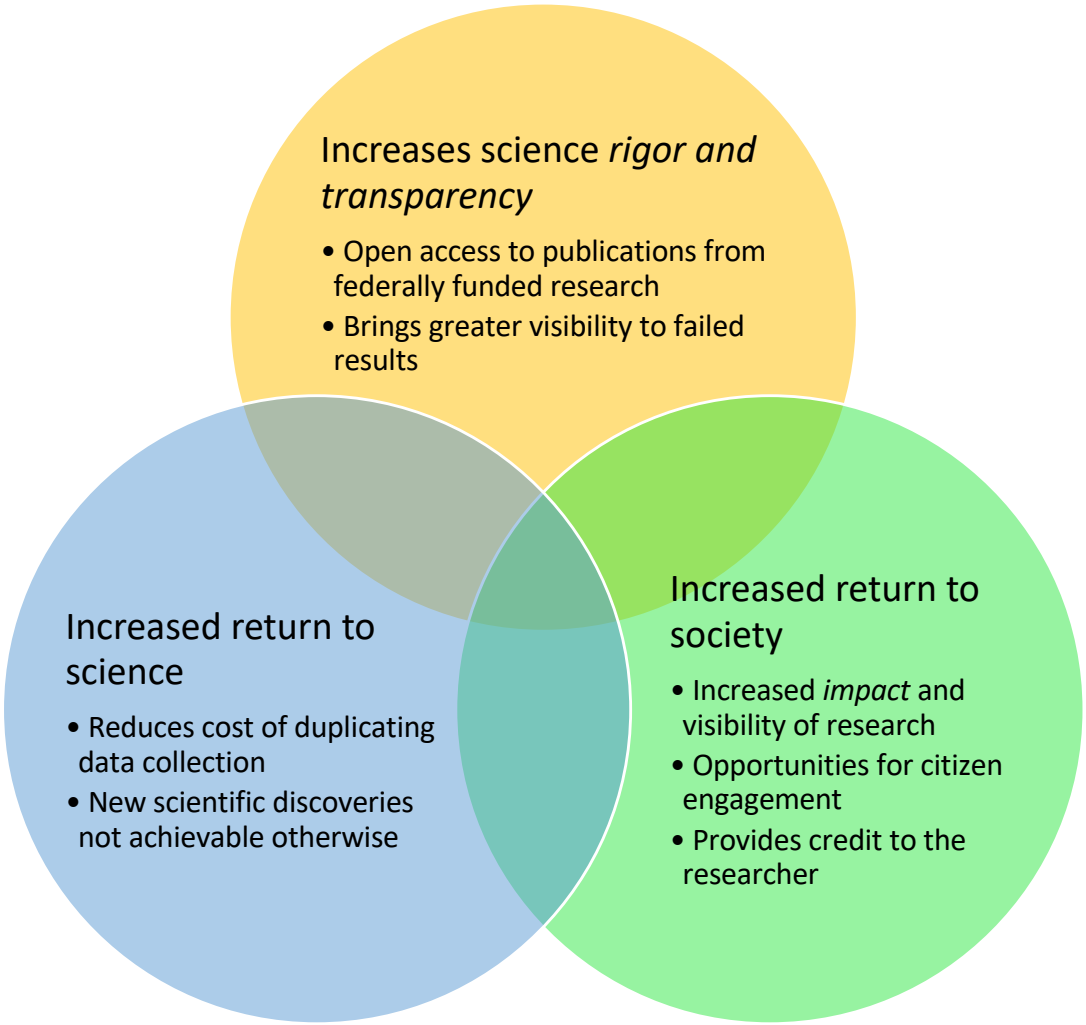
NEWS | 23 November 2018 | Vienna, Austria | Research and Innovation

Commission launches European Open Science Cloud

Following a major effort by the European Commission, the Member States and the scientific community, the European Open Science Cloud (EOSC) was launched today to provide a safe environment for researchers to store, analyse and re-use data for research, innovation and educational purposes. The Commission presented the governance structure and the portal to EU science ministers and future users at an Austrian EU Presidency conference in Vienna.

zenodo

Benefits of Public Access



FAIR Principles (2015)

- Research data objects are
  - **F**indable
  - **A**ccessible
  - **I**nteroperable
  - **R**eusable

FORCE11
The Future of Research Communications and e-Scholarship

Findable:

"Easy to find by both humans and computer systems and based on mandatory description of the metadata that allow the discovery of interesting datasets"

- e.g. Able to locate data by individual patient, patient segment, intervention, outcome metric

Accessible:

"Stored for long term such that they can be easily accessed and / or downloaded with well-defined license and access conditions (Open Access when possible), whether at the level of metadata, or at the level of the actual data content"

- e.g. Patients should be able to access parts of their own data via a patient controlled record

Interoperable:

"Ready to be combined with other datasets by humans as well as computer systems"

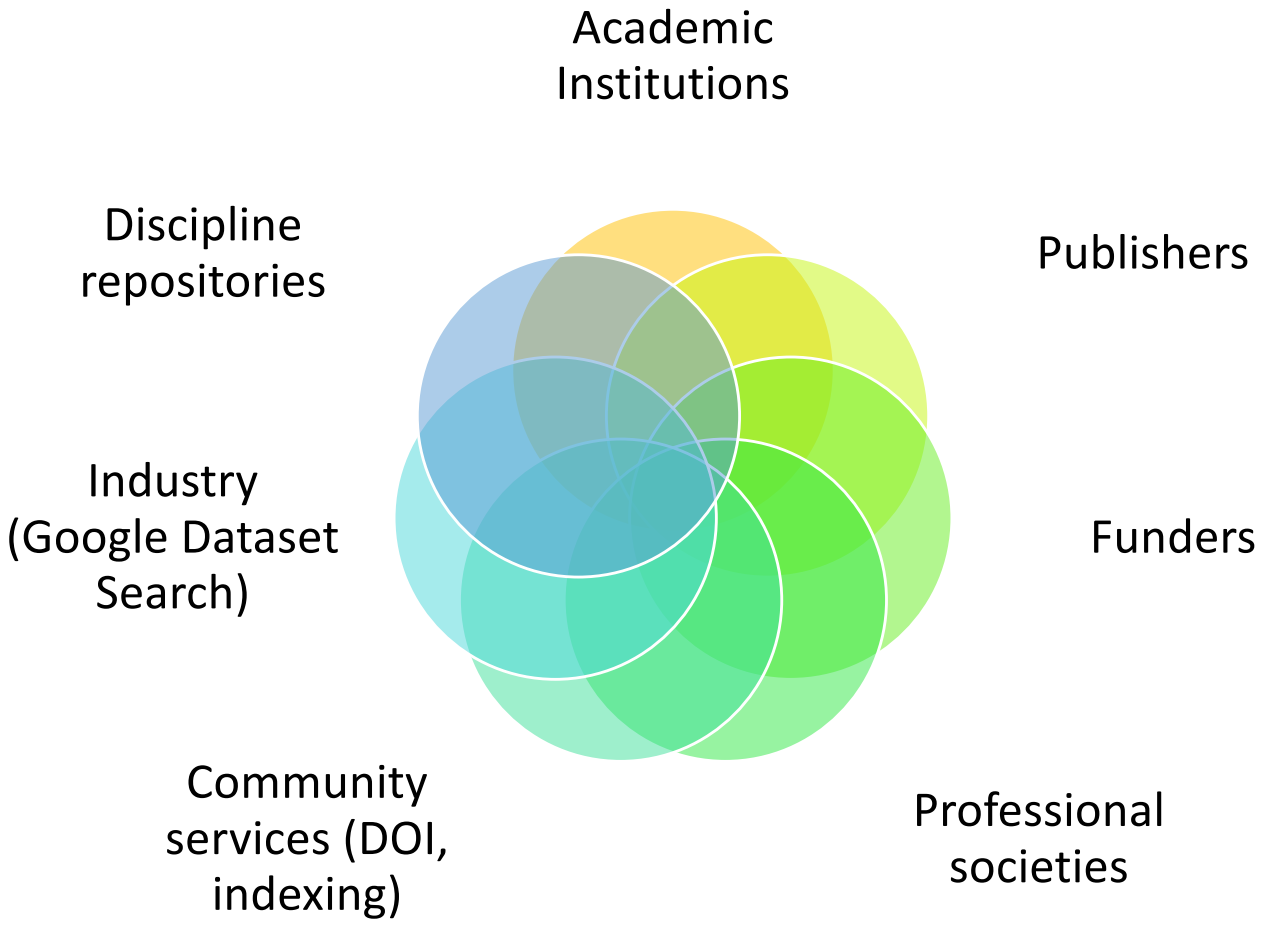
- Semantic interoperability: mapped data taxonomies across diseases and population groups e.g. consistent methodology & scale for measuring pain / quality of life
- Technical interoperability: specifications to allow different systems to communicate with each other

Reusable:

"Ready to be used for future research and to be processed further using computational methods"

- e.g. Outcomes data should be available for the long-term for systematic analysis or clinical research (with permission from data owner)

In conclusion



Research Data Ecosystem is Multi-stakeholder

Research Data Ecosystem Takeaways

- Data has both value and lifecycle
  - Curate what is truly valuable
  - Save for as long as is useful (and no longer)
- Innovative sustainability thinking being done across the board
- Academic institution involvement is currently coalition of the willing
  - Many institutions in formative stages
  - Fortunately, more mature exemplars from which to draw

Beth Plale
bplale@nsf.gov