

Data Standards and the Public Value of Federally-Funded Research

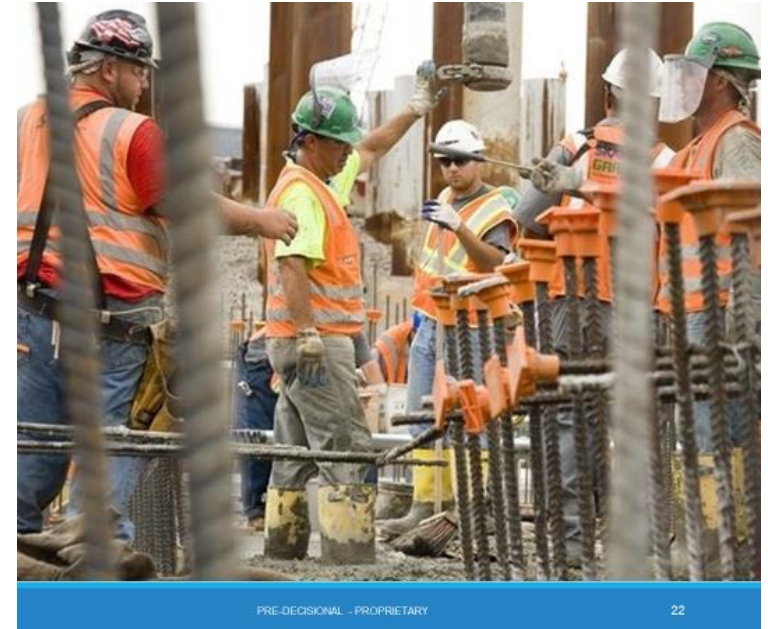
Jerry Sheehan, National Institutes of Health
Arthur Lupia, National Science Foundation
Co-Chairs, NSTC Subcommittee on Open Science

Thank you for your support of
research and researchers.

Together, we create...

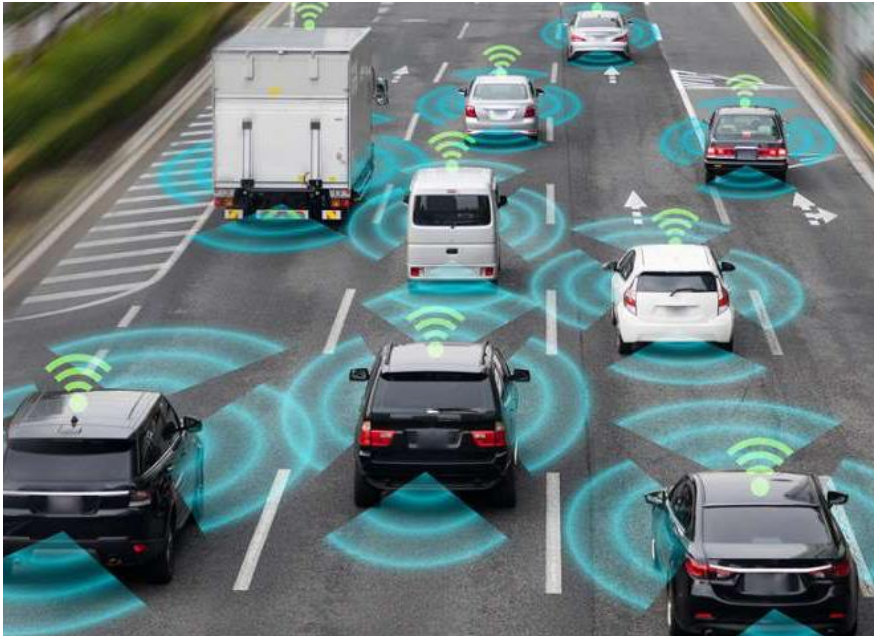


Science as service



PRE-DECISIONAL - PROPRIETARY

22



Science as a basis of innovation

Our shared endeavor

- Produce rigorous, precise research that empowers people and improves quality of life.





Today's Mission

- Increase the public value of federally-funded research activities.



Looming Questions for “Experts”

- “Why should we **pay** for what you do?”
- “Why should we **trust** what you do?”

In recent years, many researchers have...

- Strong incentives to discover and publish
- **Weak** incentives to explain how the discoveries emerged.
- **Weak** incentives to, and limited infrastructure for, sharing valuable research-related assets.

These practices threaten...

- the meaning of individual studies
- the meaning of scientific consensus
- the cultural authority of science.

Today's Mission

- Can we increase the public value of federally-funded research?
 - Publications
 - Data
 - Code
 - Related materials

Today's Mission

- Possible progress lanes...
 - Findability
 - Accessibility
 - Interoperability
 - Reusability

NSF Activities

- Public Access Repository (PAR)
 - 75k articles deposited
 - 93% available without embargo
 - Deposit rates up 230%/year for last four years.
- March '20 pilot: Make PAR 2.0 more responsive & easier to use.
- Dear Colleague Letter: Effective Practices for Data
 - Persistent IDs for Data
 - Machine-Readable DMPs



NIH Activities: Access to Literature

- Pub Med Central
 - Makes > 6.5M full-text articles available.
 - 1.5M link to associated data
 - 3.5M available for text mining
 - Coronavirus collection = 100K articles, downloaded 60M times
- Preprint Pilot
 - Launched June 2020
 - Initial focus: COVID-19, NIH-funded
 - More than 1000 preprints so far



NIH Activities

Access to Data

- Data Management and Sharing Policy
 - Draft for Public comment – November 2019
 - Supplemental guidance - Elements of data management & sharing plan; Allowable costs;
- STRIDES Initiative
 - Commercial clouds for big data
 - Discount costs, training
 - Sequence Read Archive – 35PB



NIH Activities

Data Standards

- Clinical terminologies (NLM)
 - SNOMED, LOINC, RxNORM
- Research use of clinical data standards
 - Fast Healthcare Interoperability Resource (Guide Notice + Tools)
 - US Core Data for Interoperability
- Common Data Elements
 - NIH CDE Repository
 - Disease-specific CDEs (incl. COVID-19)



Subcommittee on Open Science



- Principals
 - Kelvin Droegemeier, Director of the Office of Science and Technology Policy
 - Lisa Nichols, OSTP Assistant Director for Academic Engagement

Subcommittee on Open Science



- Steering Committee
 - Lisa Nichols, OSTP
 - Arthur Lupia, NSF
 - Jerry Sheehan, NIH
 - Brian Hitson, DOE
 - Gary Matlock, NOAA
 - Christopher Thomas, DOD

Subcommittee on Open Science



- Goal
 - Advance Federal efforts to support open science by increasing access to, and use of, the results of federally funded research and development, including but not limited to scholarly publications and digital data.

Subcommittee on Open Science



- Recent Activities
 - Request for Comment on Desirable Characteristics of Data Repositories
 - Request for Information on Public Access to Publications, Data, and Code resulting from Federally-Funded research

Agencies are Working to Coordinate on...

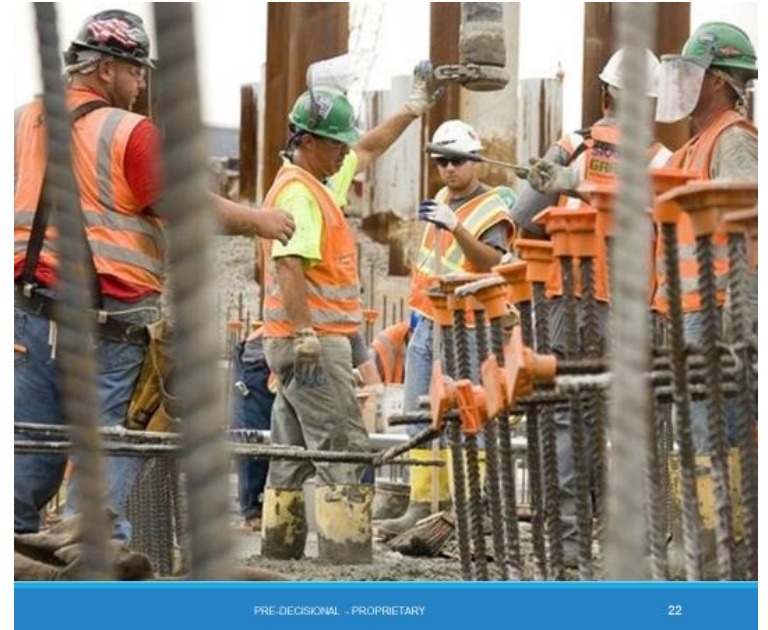
- Desired Repository Standards
- Desired Metadata Characteristics
- Desired Data Management Plan Characteristics
- Desired Attributes of Digital Identifiers
- Shared principles for data access

We want to hear and learn from you.

- For what data-related topics do your communities need more guidance from government agencies?
- For what data-related topics do government agencies need more guidance from your communities?

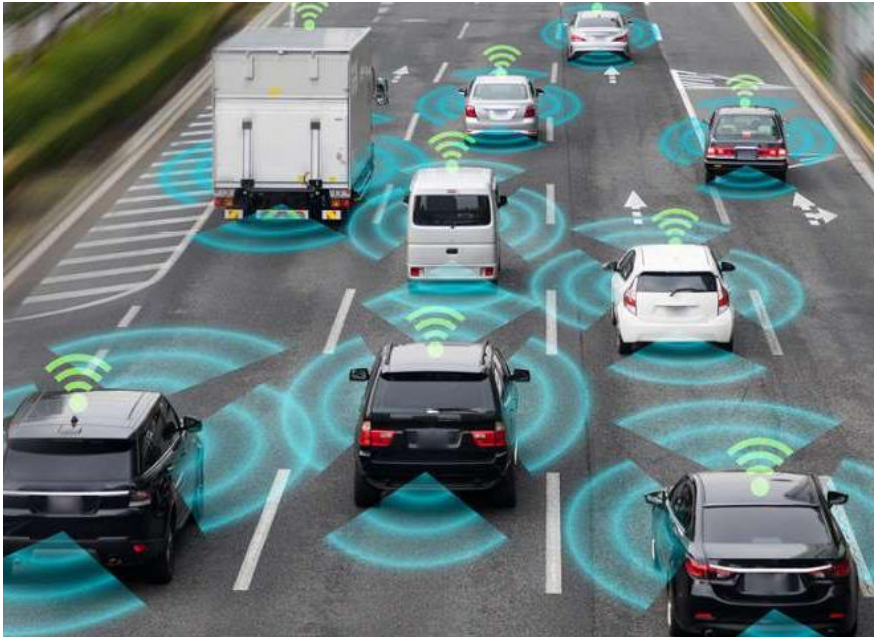


Science as service



PRE-DECISIONAL - PROPRIETARY

22



Science as a basis of innovation

Our shared endeavor

- Produce rigorous, precise research that empowers people and improves quality of life.

Data Standards and the Public Value of Federally-Funded Research



Thank you!

