NSTC Conference

Building Bridges for Impacts across the S&T Enterprise

Balancing Openness and Risk at the NIH

Joni L. Rutter, PhD

Deputy Director

National Center for Advancing Translational Sciences (NCATS)

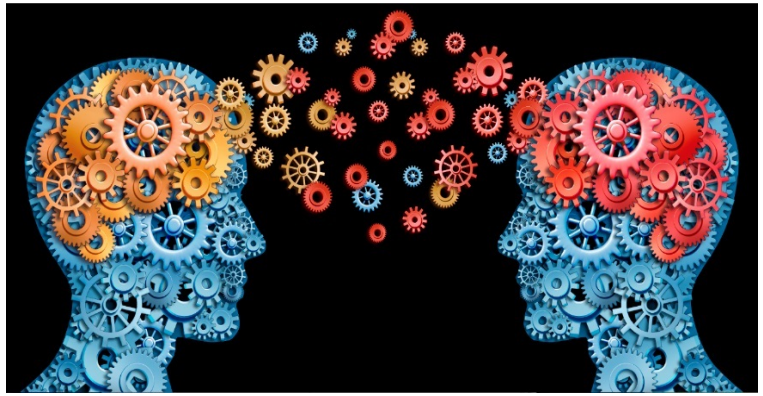
National Institutes of Health

June 13, 2019

National Center for Advancing Translational Sciences

How to build an enterprise level ecosystem to support data sharing?

- Findable, Accessible, Interoperable, and Reusable - (FAIR) data principles
- Highly structured data (via CDEs) – aggregation of data across studies
- Secure – A&A (NIST 800-53); FISMA low, moderate, high; HIPAA
  - Risk-based approach to date security
  - PIA – Privacy Impact Assessment
- Data workflows support
  (i.e. Private to Shared/Public datasets)
- Attribution for data sharing
- Reusable & scalable ecosystem
  (i.e. Cloud)



Making Data *FAIR*

Open Science is **a different way of doing science**, a paradigm in which the products and processes of research (research objects) are **broadly available & usable**

Best practices for open science abide by **FAIR principles**

Findable

- must have unique identifiers, effectively labeling it within **searchable** resources.

Accessible

- must be **easily retrievable** via open systems and effective and secure authentication and authorization procedures.

Interoperable

- should “use and speak the same language” via use of **standardized vocabularies**.

Reusable

- have clear information about data-usage licenses, and have a **traceable “owner’s manual,”** or provenance.

February 2013 - OSTP memo to federal agencies to increase public access to:

*Publications

*Data

NIH Strategic Plan for Data Science: 5 Overarching Goals

Data Infrastructure

- Optimize data storage and security
- Connect NIH data systems

Modernized Data Ecosystem

- Modernize data repository ecosystem
- Support storage and sharing of individual datasets
- Better integrate clinical and observational data into biomedical data science

Data Management, Analytics, and Tools

- Support useful, generalizable, and accessible tools and workflows
- Broaden utility of and access to specialized tools
- Improve discovery and cataloging resources

Workforce Development

- Enhance the NIH data-science workforce
- Expand the national research workforce
- Engage a broader community

Stewardship and Sustainability

- Develop policies for a FAIR data ecosystem
- Enhance stewardship

New: Office of Data Science Strategy

The NIH **Office of Data Science Strategy (ODSS)** in the Office of the Director:

- Provides leadership and coordination on the strategic plan for data science, in collaboration with the ICOs.
- Helps develop and implement NIH's vision for a **modernized** and **integrated** biomedical data ecosystem.
- Develops a diverse and talented data science workforce.
- Coordinates with trans-NIH governance committees.
- In coordination with the CIO, builds strategic partnerships to develop and disseminate advanced technologies and methods.

National Institutes of Health
Office of Data Science Strategy

Overview of Sharing Publication and Related Data

NIH strongly encourages
open access Data Sharing Repositories
as a first choice.

https://www.nlm.nih.gov/NIHbmic/nih_data_sharing_repositories.html

Options of scaled implementation for sharing datasets

Datasets up to **2 gigabytes**	Datasets up to **20*gigabytes**	High Priority Datasets **petabytes**
PubMed Central	**Use of commercial and non-profit repositories**	**STRIDES Cloud Partners**
• PMC stores publication-related supplemental materials and datasets directly associated publications. Up to 2 GB. • Generate Unique Identifiers for the stored supplementary materials and datasets.	• Assign Unique Identifiers to datasets associated with publications and link to PubMed. • Store and manage datasets associated with publication, up to 20* GB.	• Store and manage large scale, high priority NIH datasets. (Partnership with STRIDES) • Assign Unique Identifiers, implement authentication, authorization and access control.

An Early Example…NCATS TRANSLATOR

https://ncats.nih.gov/translator

Before Translator

From age 4 years
Multiple vomiting episodes daily
No effective treatments
Weight at 19 years = 78lbs

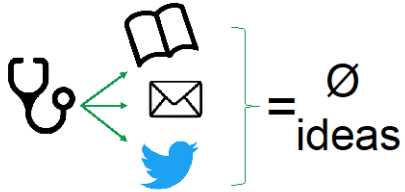
After Translator

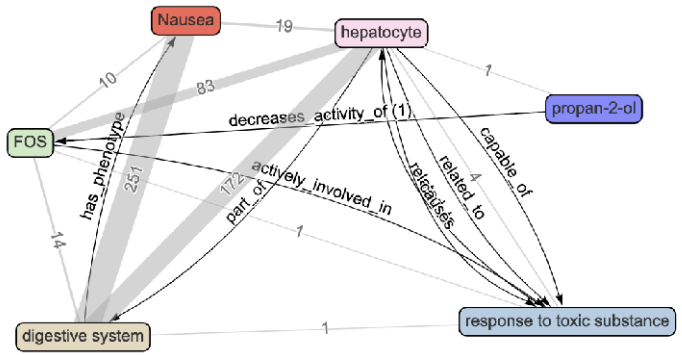
Translator

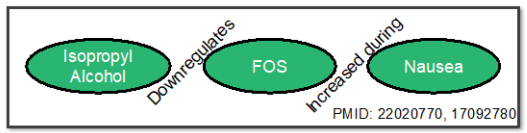
Hidden alternatives for refractory cyclic vomiting

Does it make sense?

By what mechanism would inhaled isopropyl alcohol treat nausea?





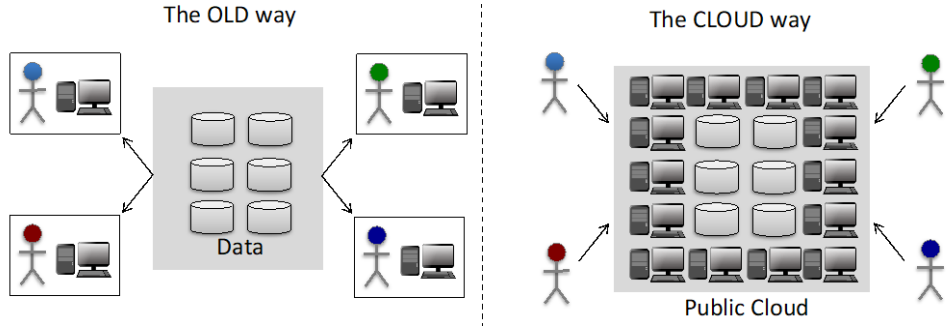


29.1M Pubmed Abstracts in seconds

National Center for Advancing Translational Sciences

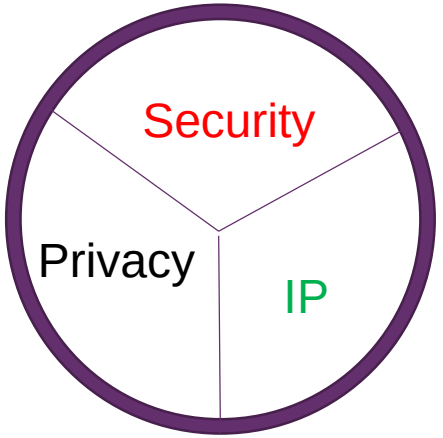
Key Takeaways

Data science & open science promise to accelerate **innovation & progress** in biomedical research



NIH has extensive **plans** to realize this promise, relying on **FAIR principles**

Implementation of these plans is underway, carefully **balancing** and always **assessing data, technology, policy, and…**



Special Thanks

- **Office of Data Science Strategy: Susan Gregurick**
- **STRIDES:** Andrea Norris, Nick Weber and NMDS team
- **Connecting NIH Data Resources:** Vivien Bonazzi, Regina Bures, Ishwar Chandramouliswaran, Tanja Davidsen, Valentine Di Francesco, Jeff Erickson, Tram Huyen, Rebecca Rosen, Steve Sherry, Alastair Thomson, Nick Weber, and BioTeam
- **Linking Publications to Datasets:** Jim Ostell and NCBI Implementation Team
- **Data Repository and Knowledgebase Resources**: Valentina di Francesco, Ajay Pillai, Qi Duan, Dawei Lin, **Christine Colvis,** and James Coulombe
- **Trustworthy Data Repositories**: Dawei Lin, Kim Pruitt, **Jennie Larkin**, Elaine Collier, Christine Melchior, Minghong Ward, and Matthew McAuliffe
- **Criteria for Open Access Data Sharing Repositories: Jerry Sheehan, Mike Huerta**, Dawei Lin**,** Maryam Zaringhalam, Lisa Federer and BMIC Team
- **Pilot for Scaled Implementation for Sharing Datasets:** Ishwar Chandramouliswaran and **Jennie Larkin**
- **Coding-it-Forward Fellows Summer Program:** Jess Mazerik
- **Data Science Training**: Valerie Florance, Jon Lorsch, Kay Lund, Kenny Gibbs, Shoshana Kahana, Erica Rosemond, Carol Shreffler
- **Diversity in Biomedical Data Science**: Valerie Florance, Jon Lorsch, Hanna Valantine, Roger Stanton, Charlene Le Fauve, Ravi Ravichandran, Zeynep Erim, Derrick Tabor, Rick Ikeda