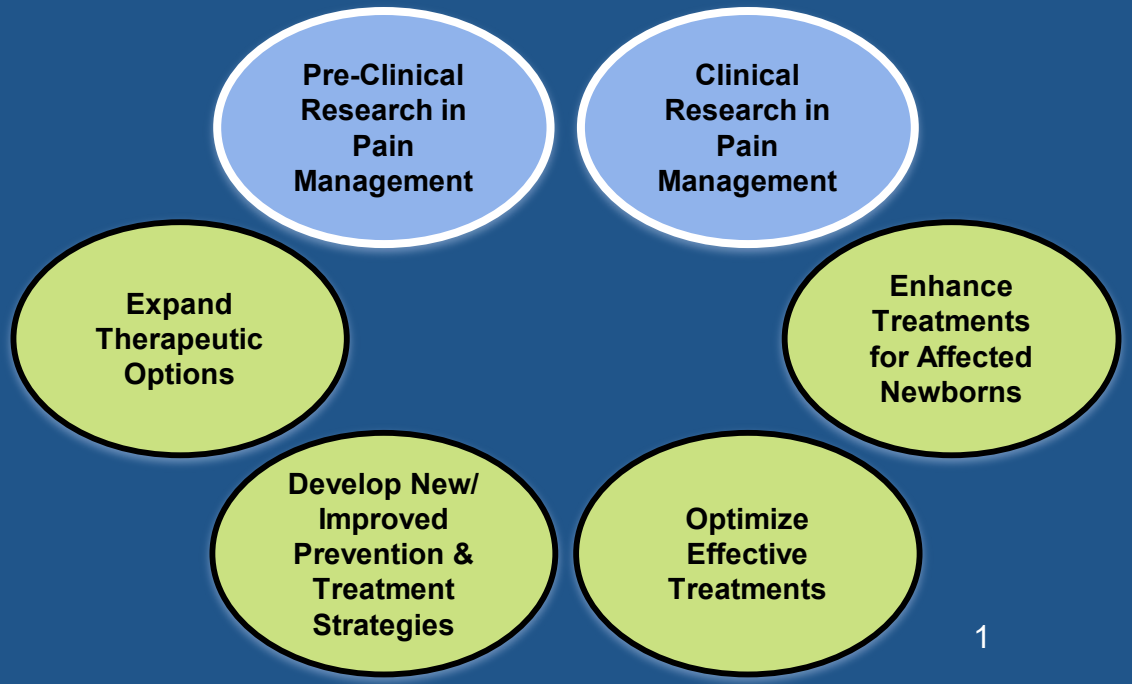
# NIH's Strategic Vision for Data Science: Enabling a FAIR-Data Ecosystem for HEAL

**Susan Gregurick, Ph.D.**

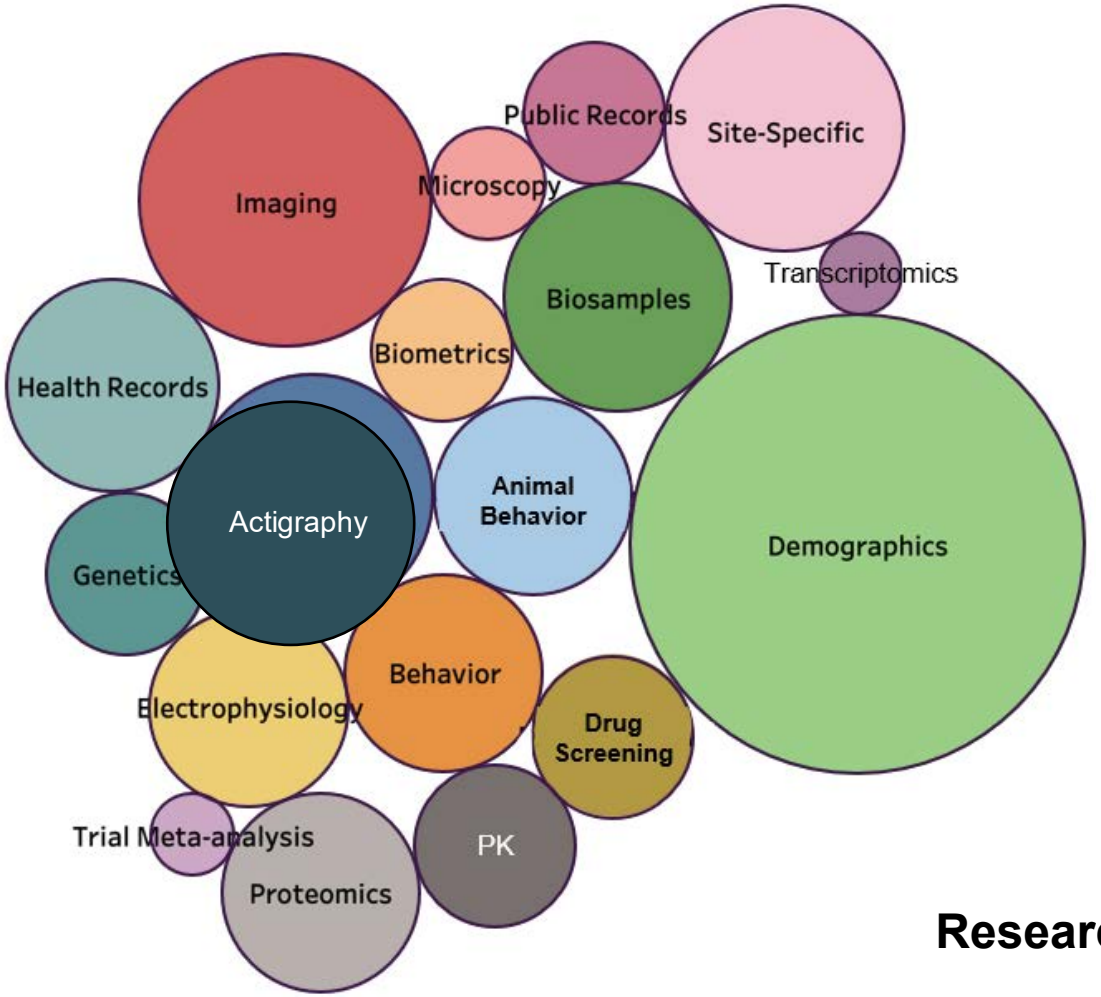
**Associate Director for Data Science and Director, Office of Data Science Strategy**

*January 16, 2020*




National Institutes of Health
*Office of Data Science Strategy*

## IMAGINE… the ability to link data in the HEALing Communities Study with data on opioid prescribing practices and measures of opioid use in other HEAL studies.




**Clinical Researchers Need Access to Standardized, Cross-study Data**

**Clinical Researchers Need to Know About HEAL Research in Their Region**

**Clinical Researchers Want to Build Synthetic Cohorts Across Studies**

**Clinical Researchers Want to Study Comorbidity**

**Researchers Want All HEAL Data to be Searchable and Discoverable**

# Making Data *FAIR*

| | |
|---|---|
| **F**indable | • must have unique identifiers, effectively labeling it within searchable resources. |
| **A**ccessible | • must be easily retrievable via open systems and effective and secure authentication and authorization procedures. |
| **I**nteroperable | • should “use and speak the same language” via use of standardized vocabularies. |
| **R**eusable | • must be adequately described to a new user, have clear information about data-usage licenses, and have a traceable “owner’s manual,” or provenance. |


# Generalist Repository Pilot: NIH Figshare

## Make your research FAIR in a few easy steps

1. Create an account at nih.figshare.com/ .
2. Create a new item.
3. Assign important metadata to your dataset to help provide context for reuse, link to relevant funding information or associated publications, and make your research more discoverable.
4. Publish! Once your content is published, it'll go into a review queue to be checked for metadata completeness and ensure all submitted content adheres to NIH policies.
5. Once live, Figshare will track all attention and potential impact around your research. All published research receives a DOI, which will help with data citation.

For more information about the NIH Figshare pilot or to share your questions, ideas, or suggestions, please email datascience@nih.gov. For technical support, please email nihsupport@figshare.com .

https://nih.figshare.com/f/faq

### Share

- Self-publish any data type and file format
- Link grant information
- Bulk-upload with API
- 100GB storage per user

### Discover

- Access open, de-identified data
- Search and filter on metadata
- Indexed in Google
- Track usage metrics

### Cite

- Assign a DOI
- Attach a license
- Ability to embargo
- Secure storage on FedRAMP AWS S3

# FHIR® Standard and Application Program Interface

**F**ast

**H**ealthcare

**I**nteroperability

**R**esources

- Developed by Health Level Seven International (HL7), a non-profit organization
- Designed specifically for exchanging electronic health care record data
- For patients and providers, it can be applied to mobile devices, web-based applications, and cloud services
- FHIR is already widely used in hundreds of applications across the globe for the benefit of providers, patients and payers

HL7 FHIR

# Common Characteristics of Our Large-Scale Platforms

- Access to high value biomedical data spanning multi-data domains and disease areas harmonized by domain-specific, extensible data models and dictionaries
- Rich suites of computational resources and tools to explore, analyze, and visualize data
- Individual and group workspaces to enable researcher to upload or access data, create experiments and conduct analysis, and store or share results
- Common approaches to assure only right people access data for right purposes and that data remain safe, secure and private

# NIH Has Significant Investments in Data and Analysis Ecosystems

| | |
|---|---|
| **Cancer Research Data Commons** | NCI's largest collection of cancer data, tools, and computational workspaces for analysis in support of the Cancer Precision Medicine and Cancer Moonshot |
| **National Data Archive** | NIMH human subjects' data collected from hundreds of research projects and analysis tools and methods for collaborative science |
| **BioData Catalyst** | NHLBI TOPMed datasets and tools for analysis including phenotype, genomics, omics, and imaging data |
| **Analysis and Visualization Informatics Lab Space (AnVIL)** | NHGRI genomics, phenotype, EHR data, and analysis tools |

# NIH Researcher Auth Service (RAS): Toward Single 'Sign-on' Across NIH Data Resources

- Streamline login for authorization of controlled-access data
- Make use of industry standard technology (web tokens)
- Enforce multi-factor authentication for security
- Keep flexible for different NIH needs: 'do no harm to existing systems'

**End goal:**

NIH-wide system for a consistent method to access data across NIH data resources

# Leveraging NIH Data Science Opportunities for HEAL

| | | |
|---|---|---|
| HEAL Data | → | Support community standards, leverage FHIR |
| Storage of HEAL data | → | STRIDES program |
| HEAL data ***not*** in the HEAL repositories | → | Figshare, general repositories |
| “Protected” HEAL data | → | Platform, can take advantage of single sign-on system |

# HEAL Data Ecosystem

HEAL Consortium

Therapeutics Development

Basic Science of Pain & OUD

Phase 2 & Effectiveness Trials

Implementation & Sustainability Trials

Data Generators

Data Submission

DMC

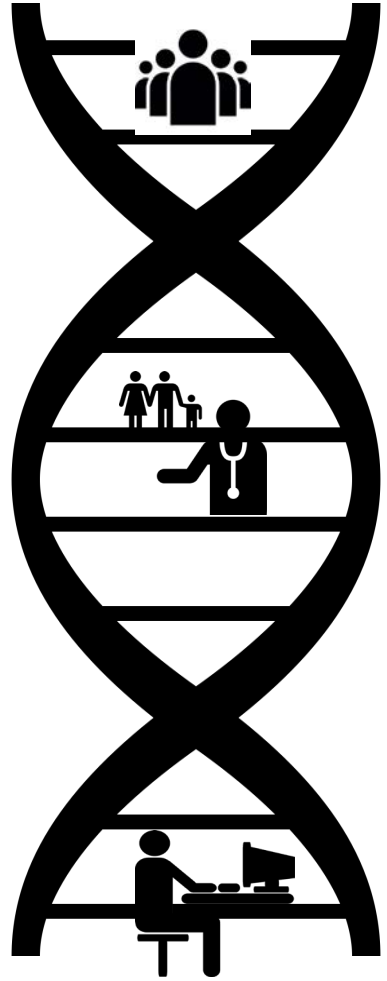
DMC

DMC

Data Management Centers
- Storage
- Harmonization by SMEs
- Security

HEAL Platform
- Access, Query, Share
- Tools, Computational Services
- Metrics

Larger Community



Data Consumers

# Stay Connected

@NIHDataScience

/NIH.DataScience

www.datascience.nih.gov

National Institutes of Health
*Office of Data Science Strategy*