

NATIONAL INSTITUTES OF HEALTH: Data Management, Open Access and Data Sharing Today and Beyond

Presentation to the ASEE
March 7, 2017

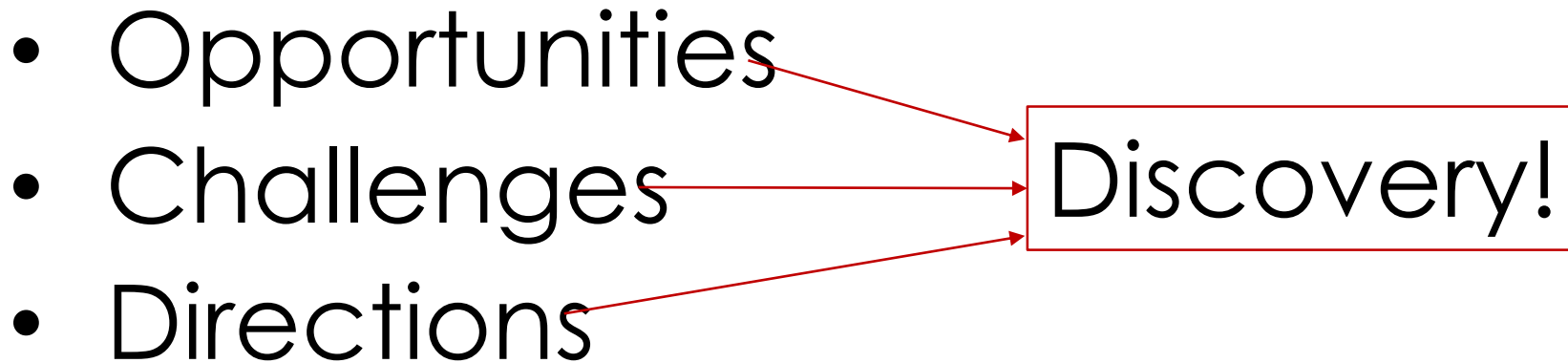
Patricia Flatley Brennan, RN, PhD
Director
National Library of Medicine

Enabling Biomedical
Scientists to capitalize more
fully on the Big Data being
generated by the
comm

Pivoting towards the future



DataScience@NIH

- Opportunities
 - Challenges
 - Directions
- Discovery!
- 
- A diagram illustrating the relationship between data science activities and discovery. On the left, a list of three bullet points: 'Opportunities', 'Challenges', and 'Directions'. On the right, the word 'Discovery!' is enclosed in a red rectangular box. Three red arrows originate from the right side of each bullet point and point towards the left side of the 'Discovery!' box, indicating that these three elements contribute to or lead to discovery.

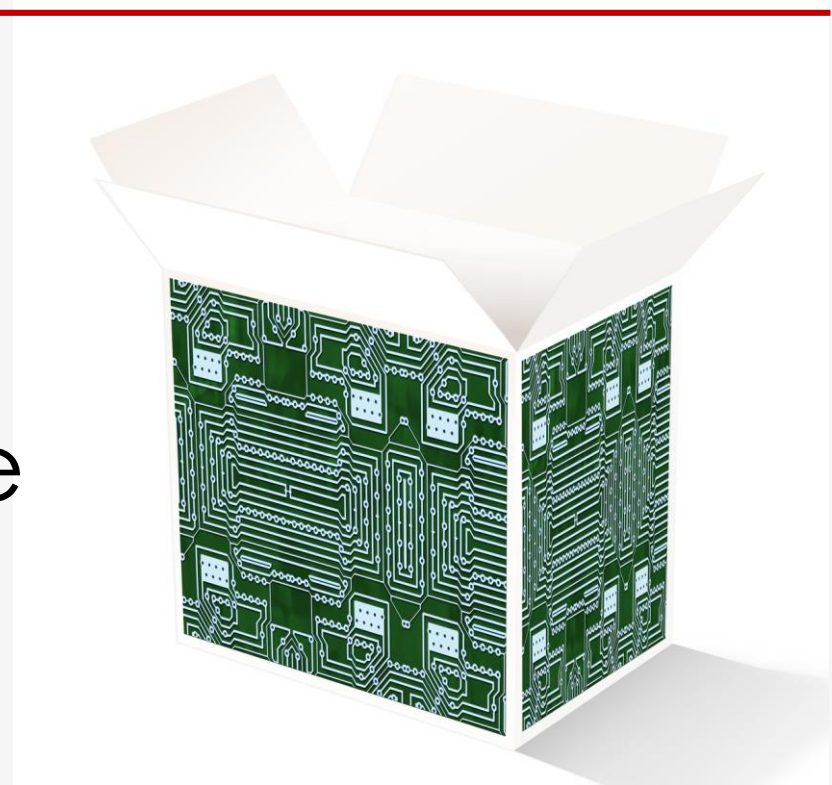
Data

Findable

Accessible

Interoperable

Reusable





FINDABLE

- Persistent identifier
- Metadata
- Curation
- Indexing
- Catalogs
- Source on the fly



ACCESSIBLE

- Security
- Authentication
- Authorization
- Metadata persists
- Public-private solutions



INTEROPERABLE

- Formal, accessible, shared and broadly applicable language
- Vocabularies follow FAIR principles

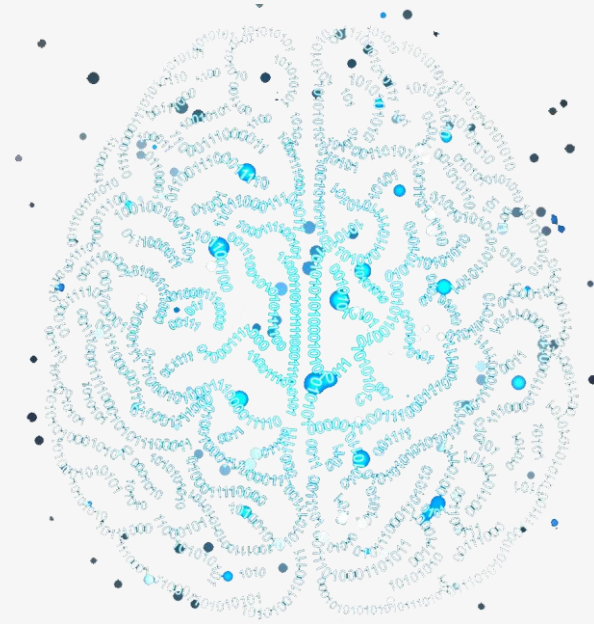


REUSABLE

- Meet domain-relevant community standards
- Provenance
- Sustainable
- Shift to Discovery

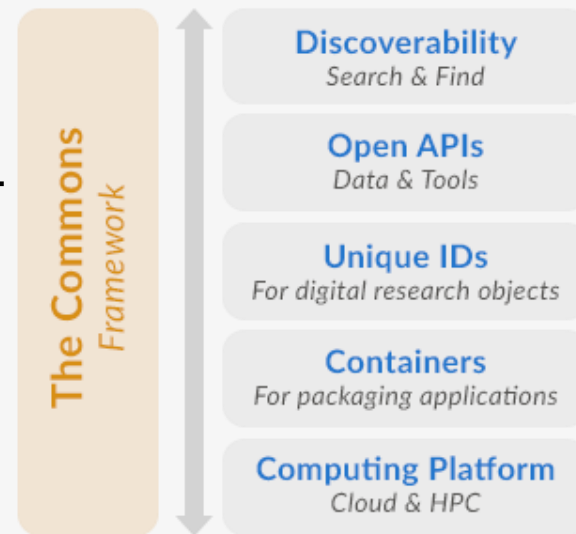
Computation

- Analytics
 - Biostatistics
 - Statistics
 - Distributed analytics
 - Machine learning
 - Optimization
- Visualization
 - Visual Analytics
 - Depicting results
- Management
 - Business process
 - Preserving provenance of analytical strategies
 - Maintaining version control



Infrastructure

- Commons
- Identity management and authentication
- Planning and forecasting tools
- Business analytics



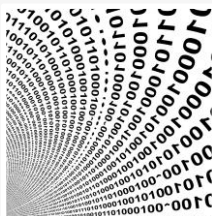
Courtesy Warren Kibbe

National Library of Medicine (NLM)

- **Mission:** To acquire, organize, disseminate, and preserve the biomedical knowledge of the world for the benefit of the public's health
- Open science is key to NLM's mission of supporting scientific discovery
- NLM products highlighted:
 - PubMed Central
 - ClinicalTrials.gov
 - PubChem



Our Future Builds on Our Past



Data
2016 -

Network
1984-2015



Digitize
~1970-1983



Index
1838-~1960





U.S. National Library of Medicine



Assess Value

Mine
high-value
datasets

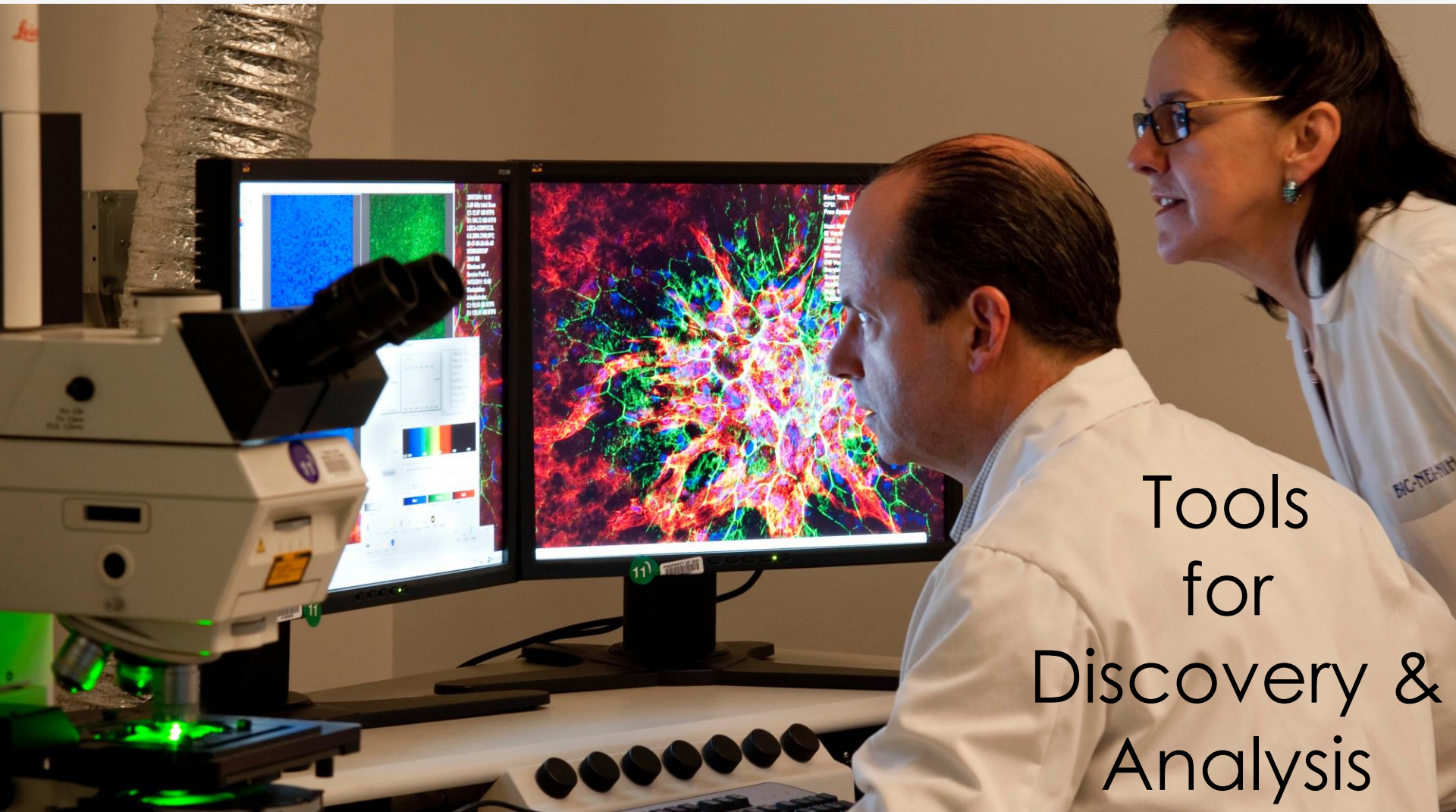
Preserve
with a
purpose



Promote Standards

LOINC SNOMED CT UMLS CDE RxNorm VSAC





Tools for Discovery & Analysis



U.S. National Library of Medicine



Promote Training



Engage with Others





NLM Collection

National Library of Medicine Collection

- ~17,000 serial titles
- All Serials – journals, annuals, statistics, etc.
- Discoverable in NLM Catalog & LocatorPlus
- NLM provides ILL and ensures archiving

PubMed

- 26.5 M records
- Only bibliographic records

MEDLINE

- ~5600 selected journal titles
- Only bibliographic records
- 93% of PubMed

PMC = PubMed Central Archive

- ~2,000 full participation journals
- 4 million full text articles
- ~1M federally funded public access articles
- Bibliographic records display in PubMed

2008: NIH
Public Access
Policy & PMC

International
collaboration:
Europe PMC &
PMC Canada



Enhancing Access

- Research results more readily accessible to the public, providers, educators, and the scientific community
- Permanent preservation of research findings
- Common format used to store data from diverse sources
- Collections available for bulk download and text mining



PubMed Central

Direct Deposit of Data
in support of any
PMC article
as a supplement
by October, 2017

al and life sciences
nal Library of

PMC Articles

archived in PMC.

Content provided in part by:

1949

Full Participation
Journals

323

NIH Portfolio
Journals

4015

Selective Deposit
Journals



ClinicalTrials.gov

A service of the U.S. National Institutes of Health

- A registry and results database of publicly and privately supported clinical studies of human participants conducted around the world
- Sponsor or researcher registers study when it begins; updates information throughout; reports summary results upon completion
- Over 233,000 studies in all 50 States and 195 countries; nearly 24,000 studies with posted summary results

Now Available: Final Rule for FDAAA 801 and NIH Policy on Clinical Trial Reporting

[Find Studies](#) ▾

[About Clinical Studies](#) ▾

[Submit Studies](#) ▾

[Resources](#)

[About This Site](#) ▾

ClinicalTrials.gov currently lists **228,054 studies** with locations in all 50 States and in **191 countries**.

Text Size ▾

Search for Studies

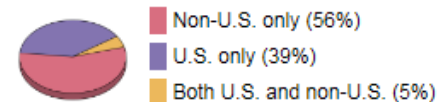
Example: "Heart attack" AND "Los Angeles"

[Advanced Search](#) | [See Studies by Topic](#)
[See Studies on Map](#)

Search Help

- [How to search](#)
- [How to find results of studies](#)
- [How to read a study record](#)

Locations of Recruiting Studies



Total N = 40,060 studies
(Data as of October 17, 2016)

- [See more trends, charts, and maps](#)

For Patients and Families

- [How to find studies](#)
- [See studies by topic](#)
- [Learn about clinical studies](#)
- [Learn more](#)

For Researchers

- [How to submit studies](#)
- [Download content for analysis](#)
- [About the results database](#)
- [Learn more](#)

For Study Record Managers

- [Why register?](#)
- [How to register your study](#)
- [FDAAA 801 requirements](#)
- [Learn more](#)

Learn More

- [Final Rule Webinar Series](#)
- [Tutorials for using ClinicalTrials.gov](#)
- [Glossary of common site terms](#)
- [For the press](#)
- [Using our RSS feeds](#)

[HOME](#)

[RSS FEEDS](#)

[SITE MAP](#)

[TERMS AND CONDITIONS](#)

[DISCLAIMER](#)

[CONTACT NLM HELP DESK](#)

ClinicalTrials.gov: API & Data Use

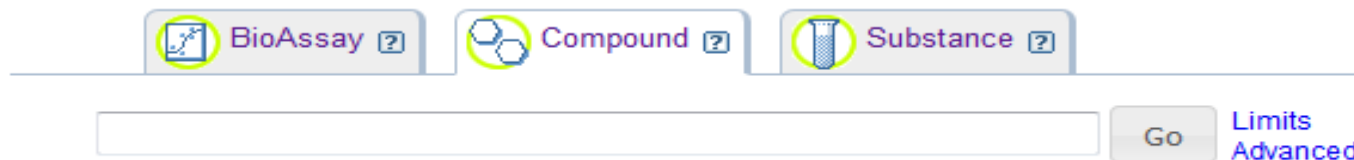
- **API allows third parties to submit queries for specified trials and display records:** e.g., BreastCancerTrials.org, Foundation Fighting Blindness, Colon Cancer Alliance

The screenshot shows the BreastCancerTrials.org website. The header includes the logo and navigation links: Tour, Privacy, Donate, Contact Us, Log-In, and Share. Below the header are main navigation tabs: Find a Trial, Why Participate, Who We Are, Resources, and FAQs. A breadcrumb trail reads: Home / Find a Trial / See All Trials. On the right, there are links for HELP, AYUDA, 求助, help-desk@bctrials.org, and a phone number (415) 476-5777 with a Print icon.

The left sidebar contains a menu with the following items: How BCT Matching Works, What You'll Need, Get Started, Making Your Decision, and See All Trials (highlighted). Below the menu, it states "Last updated: December 21, 2016". Under "See All Trials", there are links for "All Trials (586)", "All Metastatic Trials (276)", and "NEW Trials (0)". The "QuickViews" section includes "Match by Tumor Type:" followed by "DCIS (Ductal Carcinoma In Situ) Newly Diagnosed", "Stage I-III", and "Inflammatory Breast Cancer".

The main content area features a large banner with the text "see All Trials" and a "Sort by zip code:" field with a "go" button. To the right of the banner are social media share icons for Facebook, Twitter, and Email. Below the banner is a "Search ALL trials:" field with an "OR" dropdown and a "search" button.

The trial listing section is titled "All Trials" and features a trial entry: "AC Chemo Followed by Taxol with or without Paraplatin for Triple-Negative Breast Cancer". The description for this trial is: "Doxorubicin Hydrochloride and Cyclophosphamide Followed by Paclitaxel With or Without Carboplatin in Treating Patients With Triple-Negative Breast Cancer". To the right of the trial entry, there is a box indicating "Weekly visits for 5 months".



- Comprehensive source of chemical structures of small organic molecules and their biological activities
- Organized as three linked databases
 - PubChem Substance, PubChem Compound, PubChem BioAssay
- Records link to other databases, including scientific literature in PubMed



- Provides guidance on citing the original generators of the data sets
 - Corresponding scientists get credit
 - Readers can locate data source
- Includes approximately 93 million compounds, 225 million substances

NLM & the Future of Data Science



- Evolving data storage, communications, and computer security technologies
- Methods for generation, formalization, management, and sharing of knowledge resources
- Training for data scientists, data-informed investigators, data librarians
- Partnership with other NIH components and agencies promoting best practices for data storage, access, discovery and analysis.
- Strategic planning underway

DataScience@NIH pivots to NLM

- Create mechanisms to determine high value data sets, locate them, forecast their cost and utilization
- Implement efficient, secure preservation strategies that facilitate access and reuse
- Re-engage and stimulate intramural and extramural efforts in standards
- Develop new methods for for data management and data-driven discovery
- Grow a talented workforce
- Engage with government, national, and global collaboratives
- Foster open science
- Ensure integration of lessons learned from BD2K and Cloud Pilot

Reaching me



@NLMdirector



patti.brennan@nih.gov