Jerry Sheehan
Deputy Director, NLM
International Council for Scientific & Technical Information
London, 11 September 2018
NLM provides—
Literate Science
Data Science
NLM sends—

> 100 TB of data
to > 4 M users
daily
NLM receives—

> 10 TB of data from

> 3,000 users daily
Strategic Planning Context

2012 NIH Data & Informatics Working Group
2016 New NLM Director
2015 Working Group on NLM’s Future
2016 NLM Strategic Planning
Strategic Planning Process

Input

Analysis

Writing

approved
Feb 13, 2018
NLM Supports NIH Objectives

- Advance opportunities for biomedical research
- Foster innovation
- Enhance scientific stewardship
- Manage for results
Accelerate discovery & advance health through data-driven research

Reach more people in more ways through enhanced dissemination & engagement

Build a workforce for data-driven research & health
Goal 1

1.1 Connect the resources of a digital research enterprise

1.2 Advance research and development in biomedical informatics and data science

1.3 Foster open science policies and practices

1.4 Create a sustainable institutional, physical, and computational infrastructure
Fostering a ecosphere of discovery
digital research objects
The 21st Century Collection

• Innovative attribution
• Automated indexing
• Personalized presentation & delivery

COLLECT | CONNECT | KNOW
Reach more people in more ways through enhanced dissemination and engagement

Goal 2

2.1 Know NLM users and engage with persistence

2.2 Foster distinctiveness of NLM as a reliable, trustable source of health information and biomedical data

2.3 Support research in biomedical and health information access methods and information dissemination strategies

2.4 Enhance information delivery
New users, New ways
Biomedical & health information access methods & information dissemination strategies
Enhance information delivery

PubMed labs

NIH NLM inside
Foster distinctiveness of NLM as a reliable, trustable source of health information & biomedical data
Goal 3

3.1 Expand and enhance research training for biomedical informatics and data science

3.2 Assure data science and open science proficiency

3.3 Increase workforce diversity

3.4 Engage the next generation and promote data literacy
Expand & enhance research training for biomedical informatics & data science
Training across society
NLM Promotes Outreach through the 6000+ Members of the National Network of Libraries of Medicine
What is PubMed Labs?

PubMed Labs is a test site where we are experimenting with new features and tools that eventually may be incorporated in PubMed, in their current or a revised form based on the input we receive. Please try the site and let us know what you think.
Journal Article as a Nexus for Discovery

PubMed Central (PMC) contains full text records of publisher’s versions and author manuscripts.

PubMed contains citations, abstracts, and associated metadata.

https://www.crossref.org/blog/the-article-nexus-linking-publications-to-associated-research-outputs/
Linking Datasets to Articles in PubMed Central

**Supplementary Material**
- PMC policy requires all referenced supplementary data files to be deposited with the article or manuscript
- May be supplied by journal/publisher under a PMC participation agreement or a funded author via NIHMS

**Data Availability Statements**
- Required by select journals
- Provide a statement about where data supporting the results reported in a published article can be found

**Data Citations**
- NLM encourages authors and journals to make datasets available in a public repository and include the relevant machine-readable data citation(s) in the paper

https://www.ncbi.nlm.nih.gov/pmc/about/submission-methods/
https://www.ncbi.nlm.nih.gov/pmc/about/guidelines/#supplementary
<table>
<thead>
<tr>
<th>Curation at Scale</th>
<th>Expanded Metadata</th>
<th>Efficient and Connected</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 million or more citations indexed per year, to keep pace with publishers</td>
<td>Optimized metadata to support access to ClinicalTrials.gov and other trial registries</td>
<td>Publishers supply metadata and allow metadata extraction from full text</td>
</tr>
<tr>
<td>Citations indexed within 24 hours, to meet user needs for discovery</td>
<td>Comprehensive genetic metadata to support the scientific research community</td>
<td>MeSH vocabulary and workflow enhancements support user needs</td>
</tr>
<tr>
<td>Automated indexing applied, to ensure efficient processing</td>
<td>Improved chemical metadata spanning pharmacological/toxicological research areas</td>
<td>Authoritative vocabularies in addition to MeSH increase access</td>
</tr>
<tr>
<td>Expert human curation applied, to ensure indexing accuracy and quality</td>
<td>Enhanced funding metadata to support research portfolio analysis</td>
<td>Links to datasets support NLM and NIH data science goals</td>
</tr>
</tbody>
</table>
Assessing NLM’s Systems and Services
Wash & Learn
Engaging in NIH’s *All of Us* Program

- Increase Awareness of the *All of Us* Research Program
- Create educational materials around health and wellness
- Develop health and wellness “programs in a box” for libraries
- Promote ‘trusted and reliable’ consumer health resources to the general public
- Increasing capacity for libraries providing health programming
- Facilitate partnerships with *All of Us* Research Program partners
Blue Ribbon Panel Review of NLM’s Intramural Research Program

- Review strengths and weaknesses of NLM intramural research & training
- Consider optimal balance among research, development, and services
- Identify priority areas in biomedical informatics & data science research
- Find ways to support training in biomedical informatics & data science
- Recommend (as warranted) changes to NLM’s organizational structure, budget, staffing, internal and external partnerships
- Suggest ways to assess outcomes and impact of research and training
- Align NLM’s intramural research with NLM and NIH Strategic Plans
Training in Biomedical Informatics and Data Science

- **Research training** for biomedical informatics and data science
- Strengthening connections to **Minority-Serving Institutions**
- Strengthening connections to **iSchools**
- NLM-wide and NIH-wide training in **data science**
