



Forecasting Costs for Preserving, Archiving & Promoting Access to Biomedical Data

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



U.S. National Library of Medicine



Charge to the Committee

Develop and demonstrate a
framework for forecasting long-term costs
for preserving, archiving, and accessing various
types of biomedical data
and estimating potential future benefits...

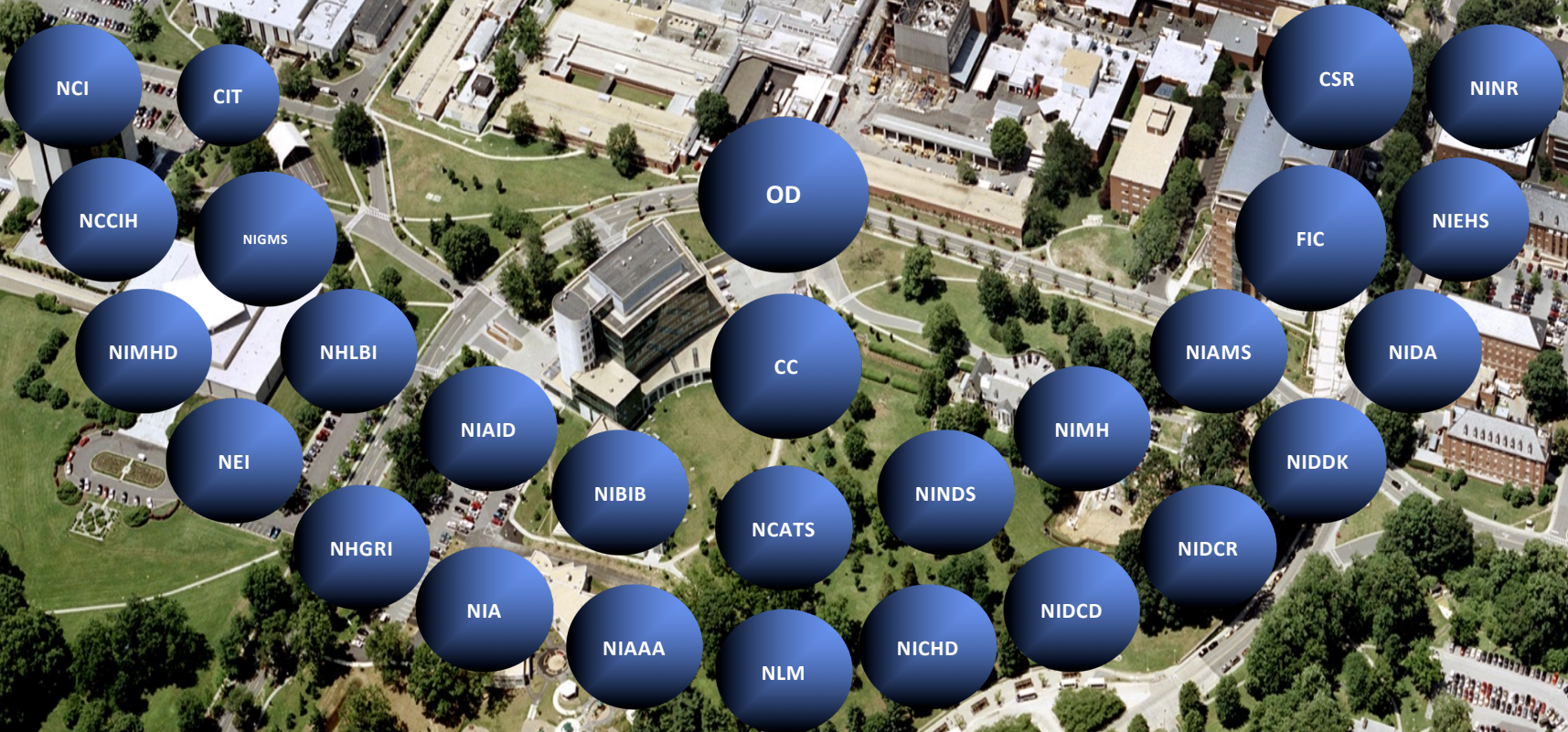


U.S. National Library of Medicine



U.S. National Library of Medicine

The National Institutes of Health

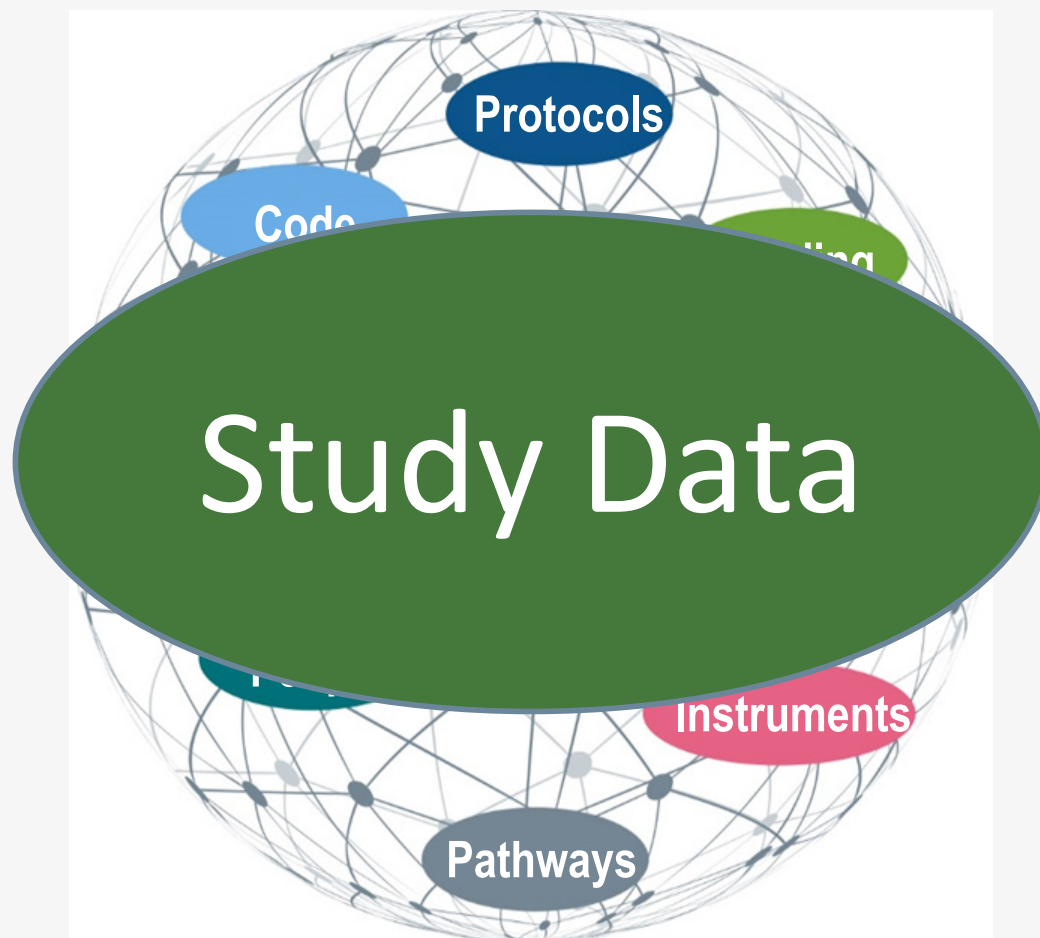


U.S. National Library of Medicine



Fostering a ecosphere of discovery

digital research objects



U.S. National Library of Medicine

NIH DATA DEPOSIT 2019

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

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

Options for scaled implementation for datasets

Datasets up to
2 gigabytes

PubMed Central

- 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.

Datasets up to
20* gigabytes

Dryad and FigShare, or other

- Assign Unique Identifiers (DOI) to datasets associated with publications and link to PubMed
- Store and manage datasets associated with publication, up to 20* GB.

High priority Datasets
T-Pbyte range

NIH managed repositories

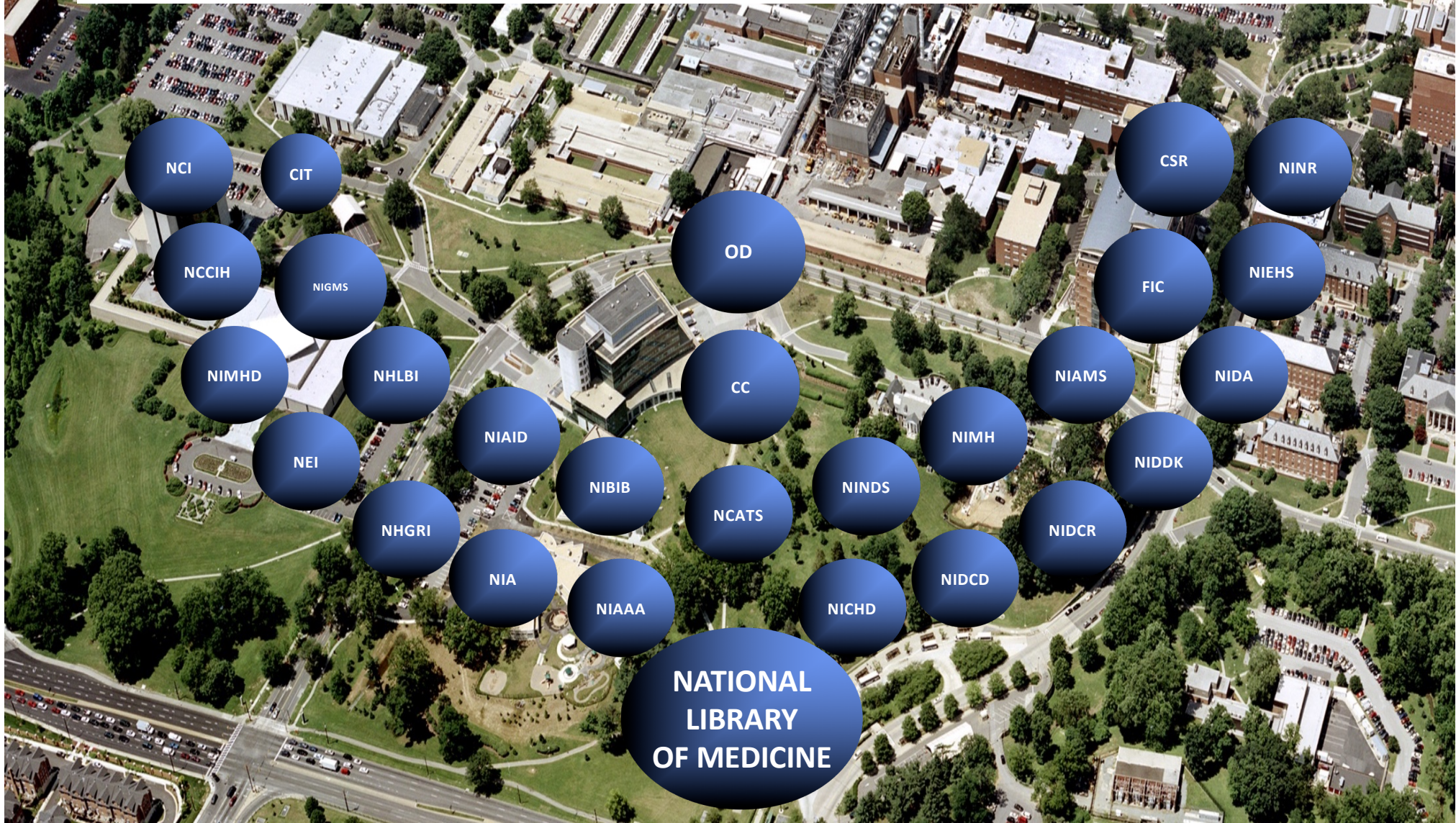
- Store and manage large scale, high priority NIH datasets (Partnership with STRIDES)
- Assign Unique Identifiers (dbGap), implement authentication, authorization & access control (eRA Commons)

Data Preservation, Storage and Reuse is essential for Discovery

- Annually, NIH spends
 - \$30B generating data
 - >\$1B managing data resources (intramural)
 - ~.25B supporting data (extramural)
- Changing role of researcher
 - Data generator → data contributor → Data user



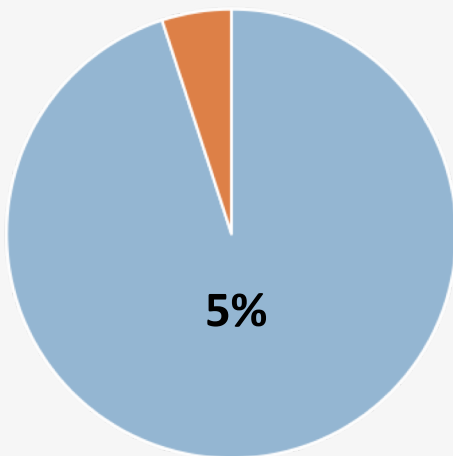
The National Institutes of Health



U.S. National Library of Medicine

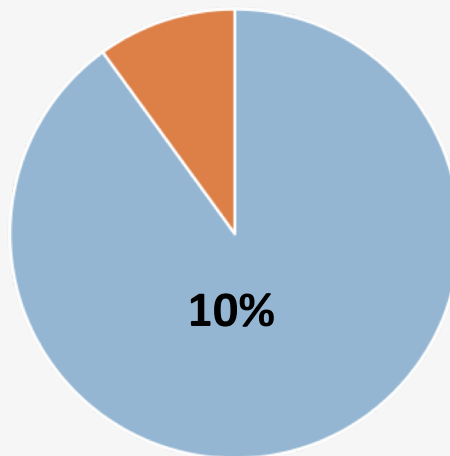
Out of every \$100 invested in research, how much should be invested in data sustainability?

Total Research \$\$



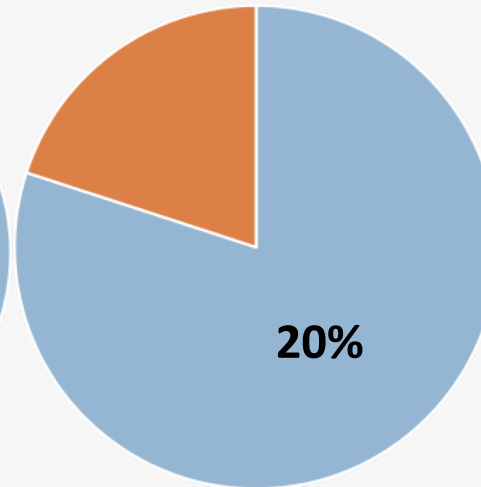
■ Direct cost ■ Data Sustainability

Total Research \$\$



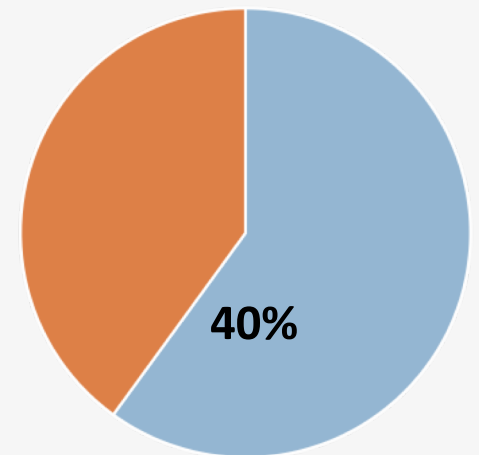
■ Direct cost ■ Data Sustainability

Total Research \$\$



■ Direct cost ■ Data Sustainability

Total Research \$\$



■ Direct cost ■ Data Sustainability



U.S. National Library of Medicine

Who will use the forecasting framework?

- Researchers
- Program Officers
- Funders

What do we need to learn from this workshop?

From Researchers:

- Incentives and barriers to deposit
- Barriers to subsequent use
- Potential market

The National Academies of
SCIENCES • ENGINEERING • MEDICINE
Forecasting Costs for Preserving, Archiving, and Promoting
Access to Biomedical Data: A Workshop

July 11 - 12, 2019

National Academy of Sciences Building
Room NAS 125
2101 Constitution Ave NW
Washington, DC

Thursday, July 11, 2019

8:30 AM

Welcome and Introductory Remarks
David Chu (Institute for Defense Analyses, study committee chair)
Tyler Kloefkorn (NASEM)
Samantha Magsino (NASEM)

8:45

Sponsor Expectations
Patricia Flatley Brennan (National Library of Medicine)

9:00

The Burdens and Benefits of "Long Tail" Data Sharing
Adam Ferguson (University of California, San Francisco)

10:00

Break

10:20

Panel Discussion: Researchers' Perspectives—Managing Risks and Forecasting Costs for Long-Term Data Preservation
Moderator: Margaret Levenstein (University of Michigan, study committee member)

Panelists:

Nuno Bandeira (University of California, San Diego)
Jessie Tenenbaum (Duke University, North Carolina Department of Health and Human Services)
Gina Tourassi (Oak Ridge National Laboratory)
Robert Williams (University of Tennessee Health Science Center)

11:40

Panel Discussion: Addressing Data Risks and Their Costs
Moderator: Michelle Meyer (Geisinger, study committee member)

Panelists:

Amy O'Hara (Georgetown University)
Brad Malin (Vanderbilt University)
Trevor Owens (United States Library of Congress)

Lunch



U.S. National

Other considerations that may emerge over the next two days

- Who pays?
 - Investigator?
 - Funder?
 - Future user?
 - Society?
- Who benefits?
- Cost Sharing model
- What is the true value of the data?



Questions?



U.S. National Library of Medicine

Reaching NLM



@NLM_news
@NLMdirector



patti.brennan@nih.gov



U.S. National Library of Medicine