Data Resources, Sustainability and NIH

Susan Gregurick
Director, Division of Biomedical technology, Bioinformatics and Computational Biology
NIGMS
Co-Chair NIH working group on Data Resource Sustainability

Data Resources in the Life Sciences Workshop ▶ HFSPO Secretariat Strasbourg, France ▶ November 18-19, 2016

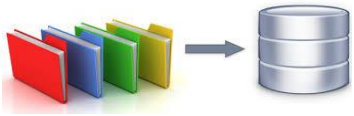
NIH National Institutes of Health
Office of the Director
Data Science at NIH

Role of Data Science at NIH

FAIR	Enable broad data sharing and reuse of data Findable, Accessible, Interoperable, and Re-usable (FAIR)
Commons	Support biomedical discovery by enabling the sharing of digital objects
Training	Enable an effective and diverse biomedical, data science workforce
Sustainability	Develop an NIH vision for economic, technical, and social stewardship of biomedical data repositories.

BD2K Investments

BD2K larger Centers & related Activities	Workforce	Targeted Software	Sustainability



- *Metadata tools and standards*
- *APIs for mapping, sharing, & integrating heterogeneous data*

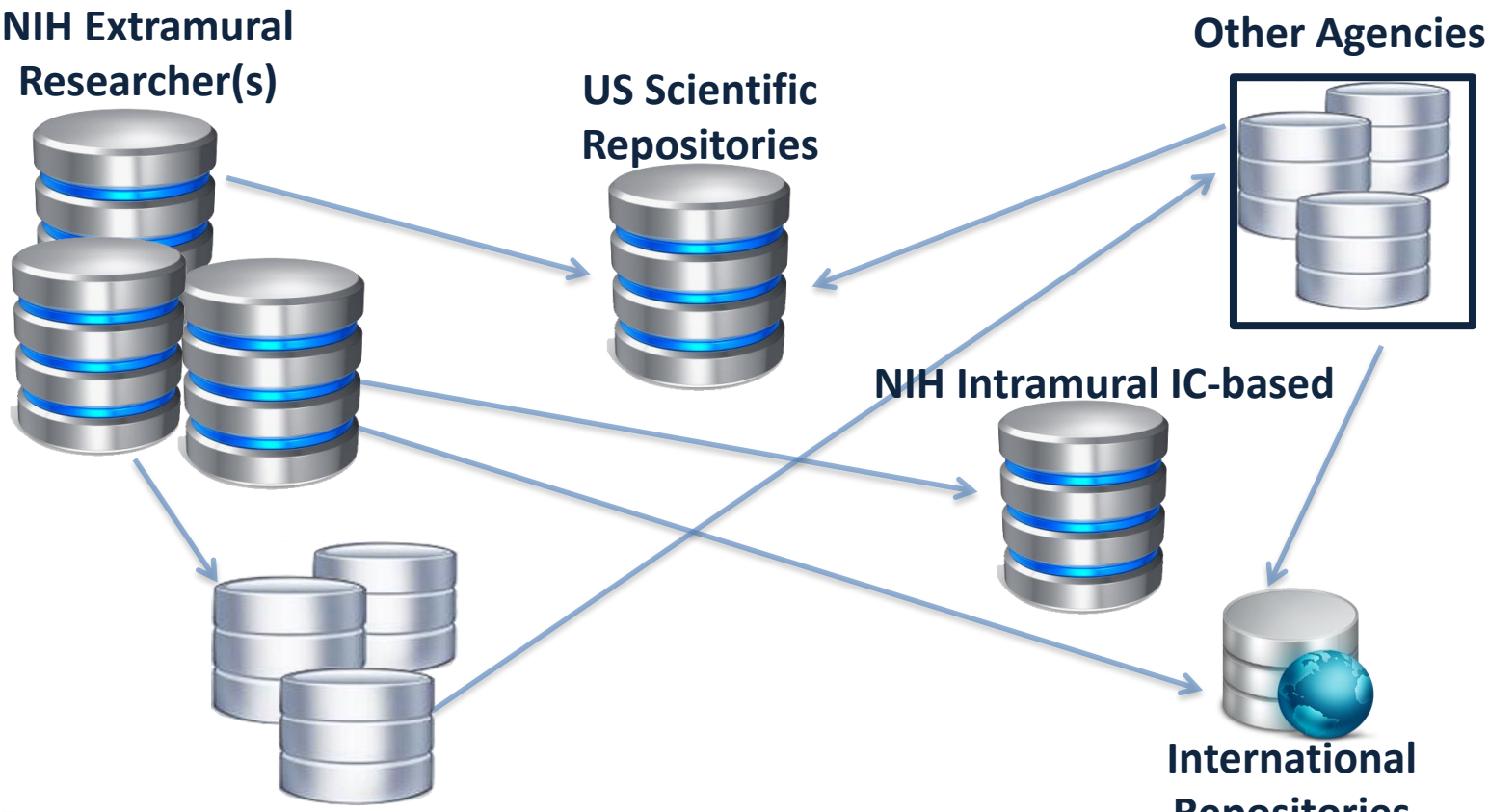
- *Open Educational Resources on data science*
- *Open Educational Resources on research data management*

- *Data Provenance & Data Wrangling, Data Privacy*

- *Inventory of NIH Data Repositories and Costs*
- *NIH Commons*
- *Interoperability projects*
- *Sustainability FOAs*
- *Policy Recommendations*

NIH National Institutes of Health
Office of the Director
Data Science at NIH

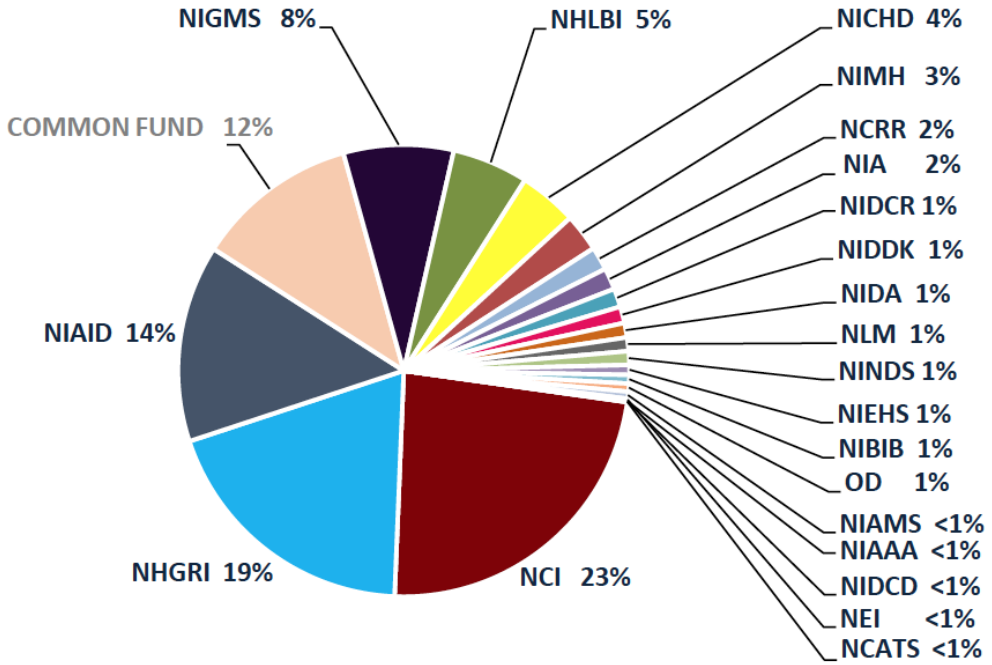
Current Ecosystem of Biomedical Research Data Resources



Standards, data formats, accessibility, utility, completeness vary

National Institutes of Health
Office of the Director
Data Science at NIH

Estimated Extramural Data Repository Spending Across NIH FY2007-FY2014, Gross Totals*

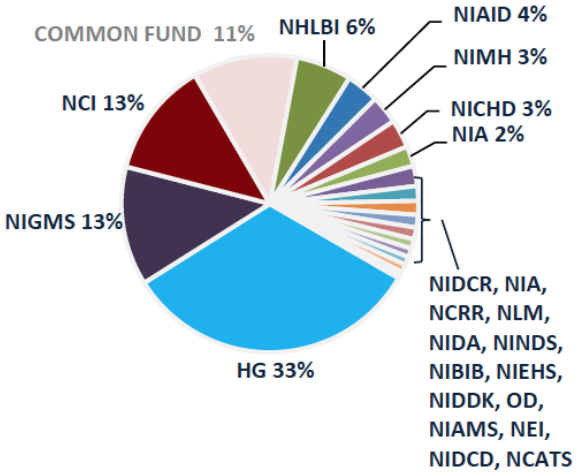


Number of Extramural Data Repositories Sampled	**131**
Gross Total Spent (Unadjusted)	**$1.34 Billion**
Gross Total Spent (Inflation-Adjusted)	**$1.27 Billion (2014 USD)**

*Please note that some repositories are funded in a hybrid scheme involving multiple institutes or funding mechanisms, which confounds subtotals. Sample data last updated 10-2016.

National Institutes of Health
Office of the Director
Data Science at NIH

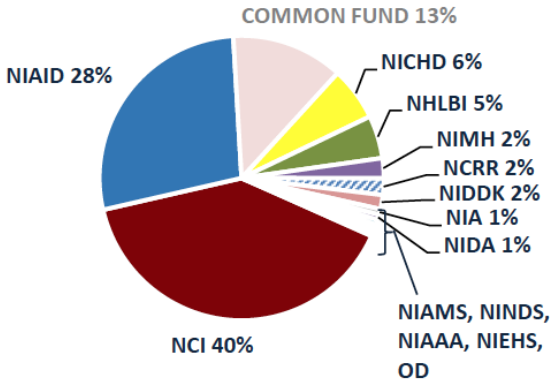
Estimated Extramural Data Repository Spending Across NIH FY2007-FY2014, by Funding Type*



Grant Funding

86 Repositories
(Approx. 65% of Sample)

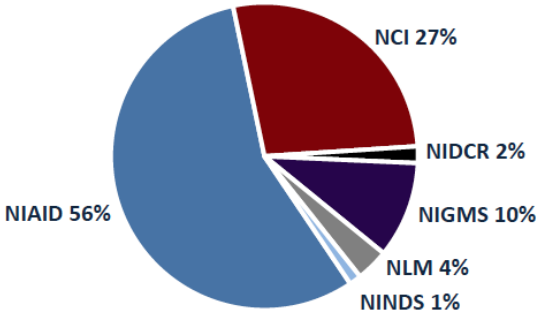
Gross Total Spent (Inflation-Adjusted)
>$795 Million (2014 USD)



Contract Funding

41 Repositories
(Approx. 31% of Sample)

Gross Total Spent (Inflation-Adjusted)
>$487 Million (2014 USD)



Interagency Funding

8 Repositories
(Approx. 6% of Sample)

Gross Total Spent (Inflation-Adjusted)
>$30 Million (2014 USD)

*Please note that some repositories are funded in a hybrid scheme involving multiple institutes or funding mechanisms, which confounds subtotals. Sample data last updated 10-2016.

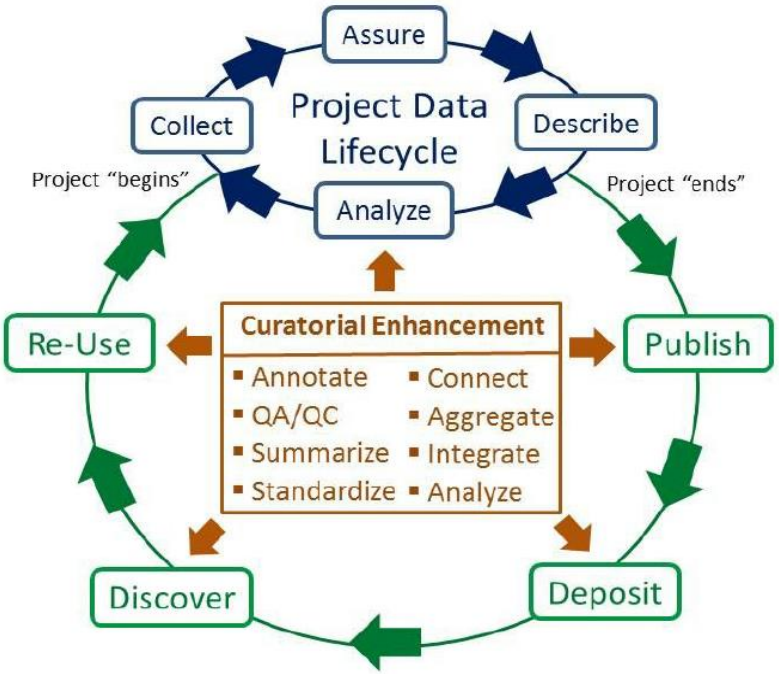
National Institutes of Health
Office of the Director
Data Science at NIH

Sustainability at NIH

Function	Develop an NIH vision for economic, technical, and social stewardship of biomedical data repositories.
Goals	**Goal 1**: Define Metrics for Evaluation of Biomedical Data Repositories and Assess Value. **Goal 2:** Develop a Sustainable Lifecycle and Coherent Funding Plan in support of Biomedical Research Data.

Funding Opportunity Announcement (FOA):
Improving Efficiencies and Innovations in Curation

- Tools to facilitate the use of community data standards
- Automated or semi-automated approaches
- Improve speed and accuracy and harmonization
- Support data annotation at points throughout the research lifecycle
- Distributed, Crowdsource approaches to curation



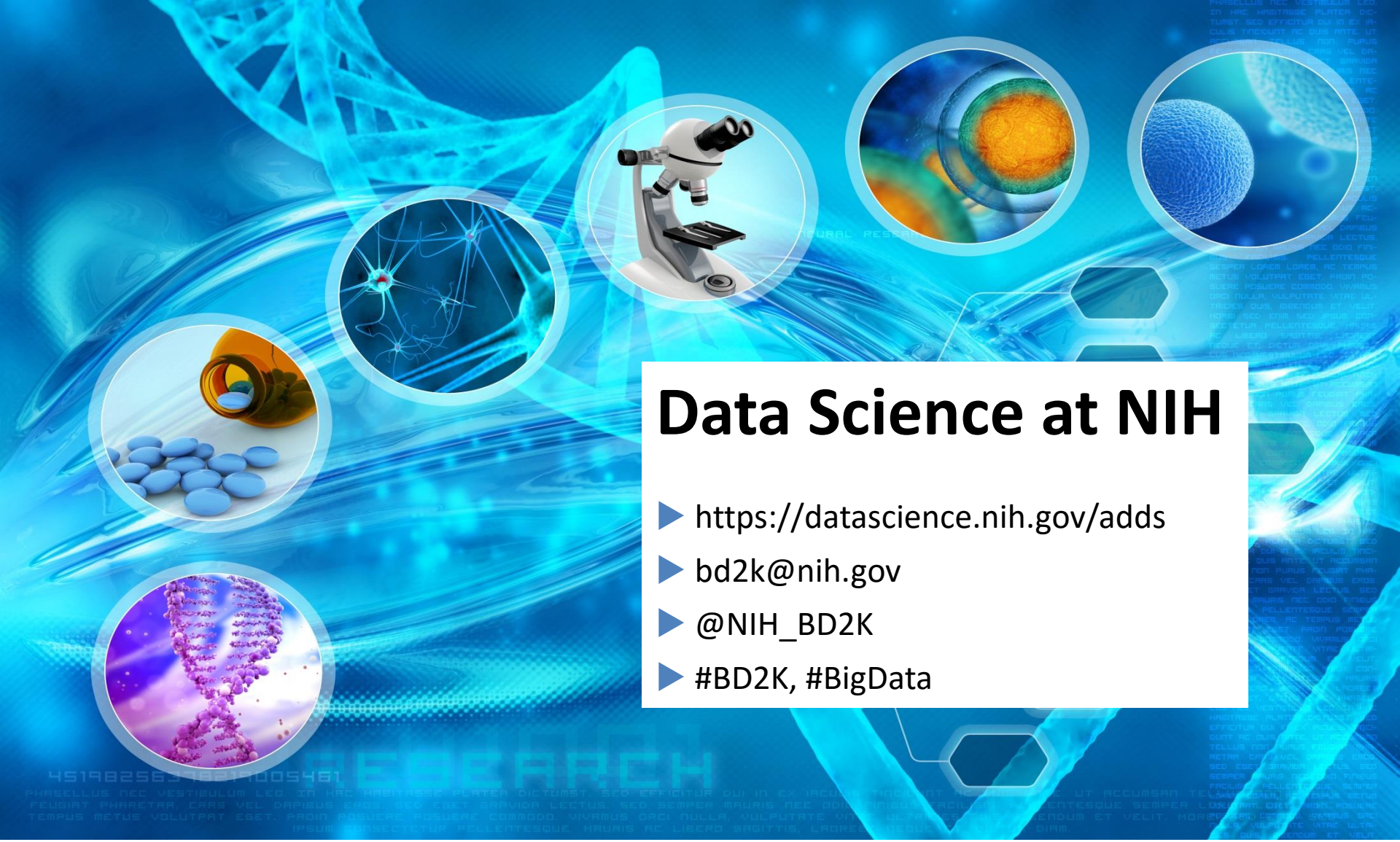
National Institutes of Health
Office of the Director
Data Science at NIH

Other solutions explored

- Request For Information (RFI): Metrics to Assess Value of Biomedical Digital Repositories
  - Responses mainly from US and UK academic stakeholders (98 total)
- Recommendations for NIH repository support
  - Mechanism(s), review criteria, implementation
- International collaboration on business and funding models
- Pilots: MODs reorganization and the Commons

Sustainability WG Members

- Vivien Bonazzi, OD
- Allen Dearry, NIEHS
- Valentina Di Francesco, NHGRI
- Valerie Florance, NLM
- Susan Gregurick, NIGMS
- Izumi Hinkson, OD
- Betsy Hsu, NCI
- Jonathan Horsford, NIDCR
- Mike Huerta, NLM
- Juli Klemm, NCI
- Malgorzata Klosek, OD
- Jennie Larkin, OD
- Dawei Lin, NIAID
- Erin Luetkemeier, OD
- Peter Lyster, NIGMS
- Andrew Miklos, NIGMS
- Linda Ogu, NIGMS
- Vinay Pai, NIBIB
- Dina Paltoo, OD
- Kim Pruitt, NCBI
- Agnes Rooke, NIDDK
- Taner Sen, OD
- Maria Shatz, NIEHS



Data Science at NIH

- https://datascience.nih.gov/adds
- bd2k@nih.gov
- @NIH_BD2K
- #BD2K, #BigData

National Institutes of Health
Office of the Director
Data Science at NIH