

BD2K Concept: Innovations in Biomedical Digital Curation

BD2K Sustainability Workgroup

Allen Dearry & Susan Gregurick, Cochairs

BD2K MCWG ► Bethesda, MD ► April 25, 2016





Sustainability Working Group

Developing an NIH vision for economic, technical, and social stewardship of biomedical repositories.

Goals

1. Assessing Value and Defining Metrics for Evaluation of Data Repositories
2. Develop a Sustainable Lifecycle and Coherent Funding Plan in Support of Research Data.



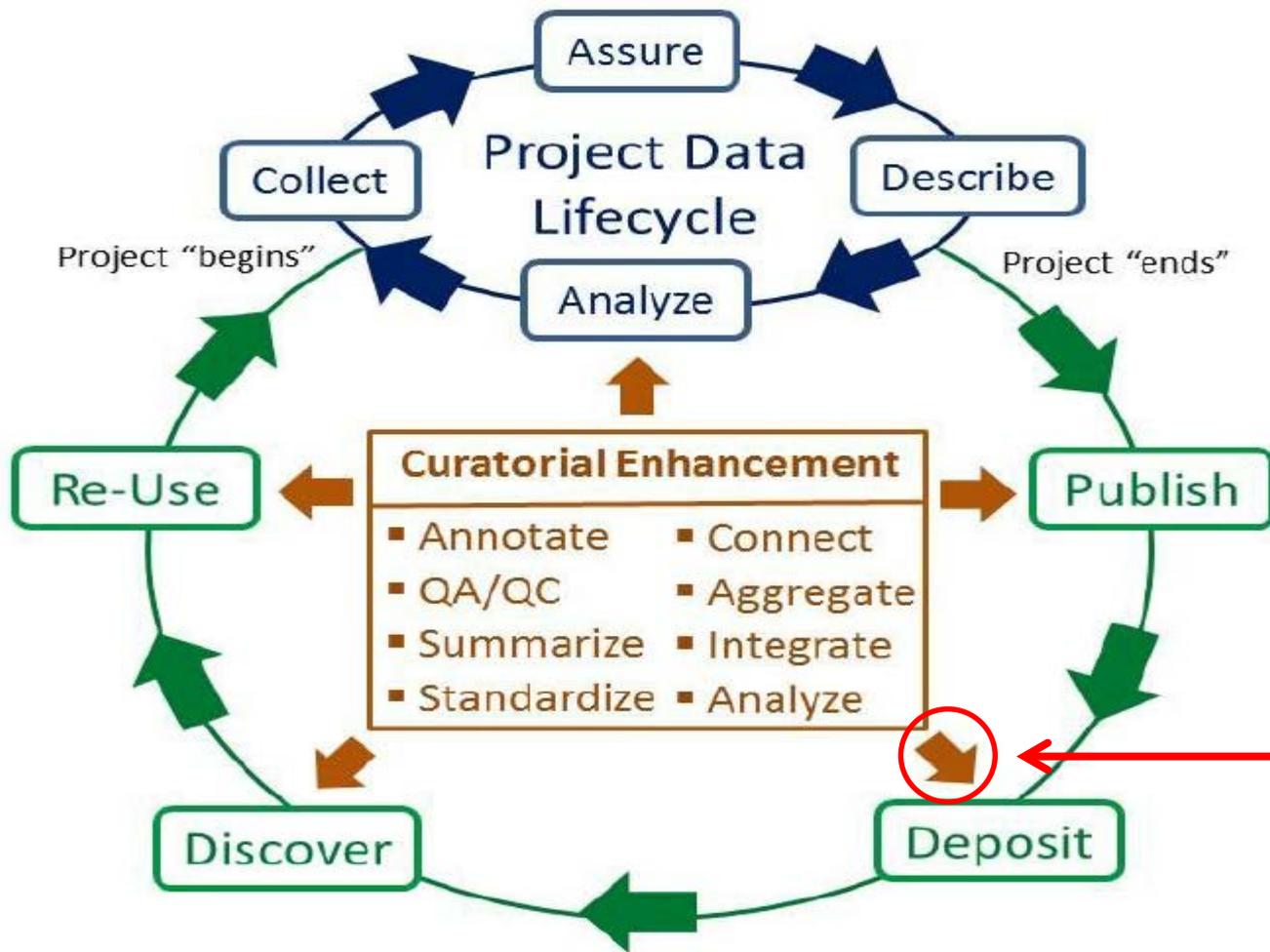


Why focus on data curation?

- Provides significant value-add to data resources.
 - Good curation makes research more FAIR and reproducible
- Current approaches are expensive, inefficient, and not scalable.
 - ~40% of cost of some biomedical data resources goes to curation
- Need to integrate data curation throughout the research lifecycle.
 - Capture knowledge of researchers more readily at time of data generation
- Enhance the professional trajectory of those who conduct curation.



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



Presently, curation is mostly conducted "after the fact," i.e., after publication (and deposition), by professional bio-curators who extract information from tables, figures, text, and authors.

Can curation occur elsewhere in the lifecycle?



Curation Workshops

- **Ninth International Biocuration Conference**
 - Geneva, April 10-14, 2016
 - Costs of information stewardship often considered an IT expense *rather than a value-added, essential component of biological research in the 21st century.*
- **FORCE2016**
 - Portland, April 17-19, 2016
 - Scholars, librarians, archivists, publishers, funders.
 - Focus on incorporating curation throughout the data lifecycle



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



Proposed Curation FOA

- Support development of methods to enhance efficiency and effectiveness of curation
 - Support more effective and consistent curation, throughout the research lifecycle.
 - Elevate status and recognition of professionals who conduct curation.





Curation FOA Potential Topics

- Tools to facilitate implementation, harmonization, and consistency of standards and protocols.
- Mechanisms and systems to make curation interoperable and scalable.
- Techniques to improve accuracy of extracting information from text or other digital sources.
- Approaches to organize and integrate curation throughout the research lifecycle.
- Methods that employ (semi)automation or crowdsourcing.





Curation FOA & other BD2K activities

Targets an area not currently addressed

- CEDAR Center, Stanford (PI: Mark Musen)
 - Developing metadata templates, but does not address wider array of curation challenges this FOA would support.
- Targeted software RFA-CA-15-017
 - Similarly, developing software for metadata application.
- BioCADDIE and Standards Coordinating Center
 - Seek to help community find data, resources, and standards; will benefit from improved curation.





Curation FOA Logistics

- Cooperative Agreement mechanism: to ensure coordination among grantees and other BD2K and IC efforts.
- Encourage those who conduct curation to serve as PIs or co-PIs.





Sustainability WG Members

- Vivien Bonazzi, OD
- Allen Dearry, NIEHS
- Valentina di Francesco, NHGRI
- Valerie Florance, NLM
- Susan Gregurick, NIGMS
- 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, OD
- Linda Ogu, NIGMS
- Vinay Pai, NIBIB
- Dina Paltoo, OD
- Kim Pruitt, NCBI
- Thomas Radman, NIDA
- Taner Sen, OD
- Maria Shatz, NIEHS

Data Science at NIH

- ▶ <https://datascience.nih.gov/adds>
- ▶ bd2k@nih.gov
- ▶ [@NIH_BD2K](https://twitter.com/NIH_BD2K)
- ▶ [#BD2K](https://twitter.com/BD2K), [#BigData](https://twitter.com/BigData)

