



Introduction to bioCADDIE

George Alter, Ph.D.

University of Michigan
bioCADDIE Executive Committee

Presented to “UTMB Health Policy Dialogues”
University of Texas Medical Branch
April 8, 2015

National Institutes of Health

Data and Informatics Working Group

Draft Report to

The Advisory Committee to the Director

June 15, 2012

“The NIH should act decisively to enable a comprehensive, long-term effort to support the creation, dissemination, integration, and analysis of the many types of data relevant to biomedical research.”

National Institutes of Health

Data and Informatics Working Group

Draft Report to

The Advisory Committee to the Director

June 15, 2012

Recommendation 1: Promote Data Sharing through Central and Federated Repositories

1a. Establish a Minimal Metadata Framework for Data Sharing

1b. Create Catalogues and Tools to Facilitate Data Sharing

1c. Enhance and Incentivize a Data Sharing Policy for NIH-Funded Data



Enabling Data Utilization

- New policies that better encourage data and software sharing
- A catalog of research datasets that will enable researchers to find and cite
- Community-based data and metadata standards



Analysis Methods and Software:

- Development and hardening of software to meet needs of the biomedical research community
- Access to large-scale computing to enable data analysis on Big Data
- Dynamic community engagement of users and developers

Enhancing Training:

- Increase number of computationally and quantitatively skilled biomedical trainees
- Strengthen the computational and quantitative skills of all biomedical researchers
- Make training available to NIH staff to enhance NIH review and program oversight

Centers of Excellence:

- Investigator-initiated centers
- NIH-specified centers

Main goals of bioCADDIE

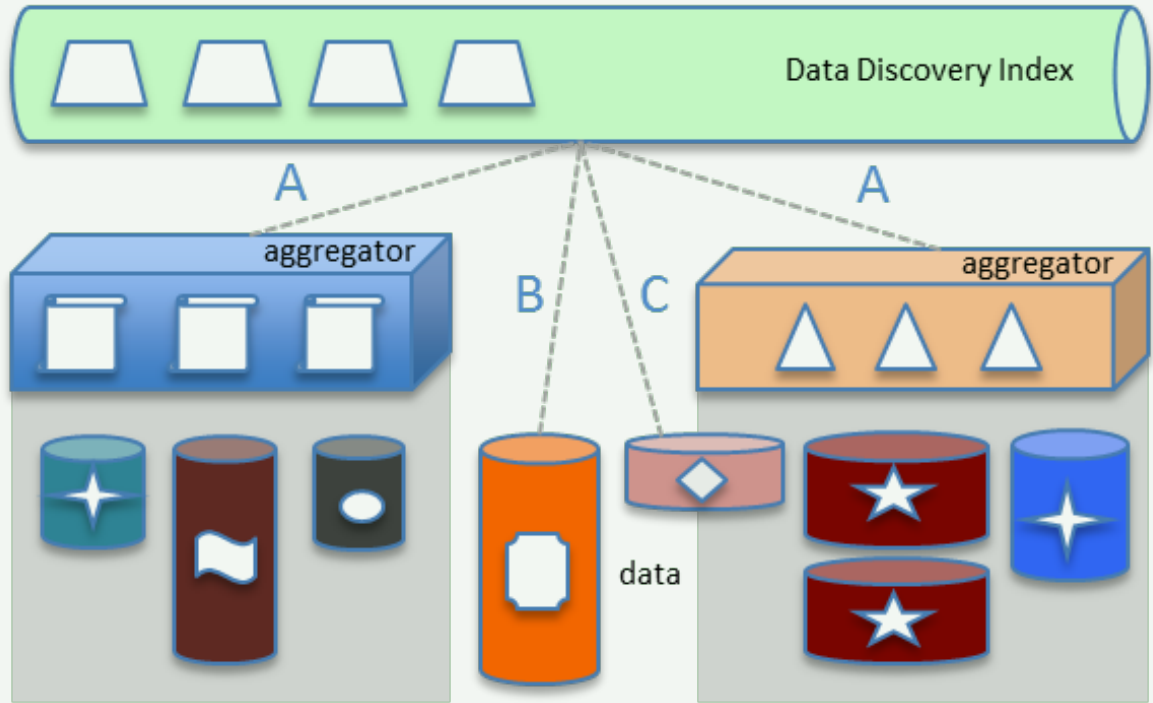
- Build a searchable index of data objects in biomedical data repositories
- Help users find data and conditions for access
- Facilitate, monitor, and reward
 - Data sharing
 - Data curation and annotation
 - Data reuse
 - Data citation

Expected outcomes

- Prototype Data Discovery Index
- Pilot applications that “dock” with the prototype
- Serve as an incubator and clearinghouse for
 - Innovative designs
 - Partnerships
 - Standards for data curation, metadata, and repositories
- Contribute to design of the NIH Commons, a digital environment for storage, manipulation, and sharing of research objects

Use cases

- **Disease-based search across scales:** Find all datasets from Alzheimer's patients that have RNA seq, behavioral, and imaging data available.
- **Molecular-based search across organisms and scales:** What proteomics and metabolomics datasets are related to the same biological process?
- **Molecular data/phenotype associations:** What datasets are available that have genome data about IDH1 and IDH2 in humans or other species for a particular phenotype of interest (e.g., glioma)?
- **Behavioral and environmental data:** What is the effect of stress on health? Could different components (family, work, neighborhood) have stronger associations with health?



(A) Aggregator services: Repositories and indices with a common metadata format

(B) Databases not covered by aggregators

(C) Annotation of databases only partially annotated by aggregators

White shapes represent metadata in various formats.

bioCADDIE Working Groups

1. BD2K Centers of Excellence Collaboration
2. Data Identifiers Recommendation
3. Core Metadata Specifications
4. Use Cases and Testing Benchmarks
5. Dataset Citation Metrics
6. Criteria for Being Included in the DDI
7. Machine Actionable Licenses
8. Ranking Algorithm
9. End User Evaluation Criteria
10. Repository Collaboration
11. Outreach Meeting: Repository Operators
12. Standard-driven Curation Best Practices
13. Evaluation of Harvesting and NLP Pilot Projects

All this by
August 2017!

bioCADDIE Executive Committee

- Lucila Ohno-Machado (PI) - University of California, San Diego
- George Alter - ICPSR, University of Michigan
- Ian Fore - NIH
- Jeffrey Grethe - University of California, San Diego
- Susanna-Assunta Sansone - University of Oxford and Nature Publishing Group
- Hua Xu - University of Texas Houston