

NIDA Strategic Planning – Big Data Workgroup
Co-Chairs: Roger Little and Massoud Vahabzadeh
SPB Coordinator: Ericka Boone

Workgroup Webinar
Wednesday, April 22, 2015
2:00 p.m.

Attendees

Ericka Boone, Roger Little, Maryann Martone, Michael Neale, Udi Ghitza, Steve Gust, Maureen Boyle, Philip Bourne, Tom Radman, Tisha Wiley, Michael Milham, Michele Rankin, Christopher Chute

Technical Note

Due to difficulties with WebEx, the slide presentation could not be viewed by participants. In addition, some workgroup members were unsuccessful at dialing in to the program, so the meeting discussion was limited to only a few participants, and the duration was very brief.

Overview of Big Data Challenges/Discussion

Dr. Roger Little provided an overview of written comments received from workgroup members and then opened the floor to discussion.

- In general, the primary concerns included reproducibility and harmonization of the data. While it may be theoretically impossible to align all data points, it will be helpful to include annotation on how the data was generated and why it is important.
- Dr. Michael Neale agreed with the need for annotating methods and explaining any discrepant findings; he said the keys to quality data included good curation and the use of open software.
- Dr. Little and Dr. Michael Milham discussed the challenge of developing metrics for assessing the quality and usability of complex versus simple data.
- Dr. Philip Bourne questioned the layout of the priorities document, suggesting that harmonization of data was related more to standardization than to reproducibility. He brought up the FAIR concept (Find, Access, Interoperability, Reuse), pointing out that reproducibility has more to do with access to similar data for future studies.
- Dr. Bourne voiced concern over the idea of data sharing. He and Dr. Neale discussed the issues of policy, security, infrastructure, logistics, and resources related to sharing data, particularly big data. Storage, sustainability, and preservation are also points of concern, and the research field will need to develop solutions to meet the mandate for data sharing on all publicly funded grants beginning in 2016.
- Dr. Little spoke to the common notion of creating a shared space for data, narrative, and identification of contents for each research item measured, and stated that a potential solution is currently being tested. He said the goal is to develop big data storage using optimal formats. We have a large repository of genetics data, and BD2K will be addressing our other issues.

- Dr. Milham mentioned several storage possibilities for shared data but had concerns about privacy and security. He and Dr. Bourne discussed the concept of centralized vs. federated/discoverable databases and usage considerations. Patterns of expected use and scope of the data all depend on the architecture that supports the database. Structured data, unstructured data, metadata, common data elements all come into play—these issues make it clear that there is no one optimal solution.

Data Storage Models

- Workgroup members discussed some of the research activities involving development of data repositories and dictionaries. Groups within and outside of NIH are currently working on data capture and storage initiatives. Several communities are already data savvy, including the clinical communities at the National Library of Medicine, and data scientists in the physics and astronomy fields. It would be helpful to examine which models have been adapted by others and why.
- Dr. Maryann Martone suggested that NIDA could better align with an existing platform if we first identify characteristics of the NIDA portfolio and its constituents. Dr. Little stated that the results of the Addictome project might help explain how NIDA would go about adopting a compatible model.

Workgroup Charge

Dr. Little and Dr. Bourne revisited the workgroup's charge. The Big Data Workgroup was asked to develop a recommendations document for NIDA executives that identifies key issues surrounding big data and offer guidance on how to manage those issues. The recommendations from each workgroup should revolve around how NIDA science can benefit in their particular areas, and they are expected to be considered for inclusion in NIDA's overall Strategic Plan.

Next Steps

Dr. Little proposed a process for the next several meetings of the workgroup. Referring to the priority spreadsheet distributed to all members (Homework document, April 22 meeting), he suggested they discuss Issue 5 - Data Sharing at the next meeting (April 29th), followed by Issue 1 – Capture, and from there discuss issues related to storage and visualization, with a final meeting to wrap on June 17. From there, Dr. Little proposed that he, Massoud and an extramural workgroup member work together to help draft up the recommendations document based on all discussions.

Public Comment Period

- No comments were submitted to the workgroup.

Action Items

- There were no action items.

Next Meeting

The next webinar is scheduled for Wednesday, April 29 at 2 p.m.