NIH Common Data Elements (CDE) Workshop Executive Summary

September 30, 2015

BD2K, BMIC, NLM, NCI, NCATS, and NIEHS jointly organized a workshop for NIH staff to explore the role of CDEs in NIH Data Sharing. This workshop convened 40 representatives of the NIH community that were interested in CDEs with the goals of:

- Supporting NIH-wide understanding of current activities and opportunities related to CDEs.
- Identify current barriers/challenges for the adoption and use of CDEs by NIH-funded researchers, both intramural and extramural.
- Identify possible ways to modify development, implementation, and use of CDEs to increase adoption and value to research
- Identify incentives and opportunities for involvement of relevant communities in CDE development, use, and re-use.
- Develop evaluation plans for CDEs to test their assumed utility.
- Identify opportunities to improve coordination in the development of CDEs for research use and in infrastructure for developing and making them accessible.
- Determine how best to support CDE activities in the context of BD2K

Prior to the Workshop, a preparatory webinar was held on September 8, 2015. This webinar included presentations from several ongoing CDE programs by NIH staff, who are engaged in the BMIC CDE Working Group. This webinar and its presentations are archived on the NIH CDE Resources Portal.

The September 30 workshop opened with an introduction by the NIH Associate Director for Data Science, Dr. Phil Bourne to the "FAIR" principles, that data and other resources should be Findable, Accessible, Interoperable, and Re-usable, followed by an update on the status of the NIH Scientific Data Council. Drs. Dina Paltoo and JP Kim then gave a presentation and led a discussion on Data Sharing: How CDEs might fit into NIH's Data Management and Sharing Policies. Acting Director of the National Library of Medicine, Betsy Humphreys then gave an overview of the focus and charge of five break-out groups. Attendees split into groups, and at the end of the day each group reported back on their discussions and recommendations. This report includes the top recommendations arising from the break-out groups.

Recommendations:

1. **Establish NIH wide CDE-related governance.**

   NIH should ensure there is NIH wide governance and oversight of CDE-related policy and coordination in the context of broad data sharing.

   - It is recommended that this group report to the Scientific Data Council, who is responsible for data sharing oversight and policy.

- The NIH CDE Governance Group should address issues of NIH wide CDE policy and coordination among NIH ICs, with other federal agencies, the academic research community, and the private sector.
- The NIH CDE Governance Group should include representation from all NIH ICs to ensure NIH wide CDE coordination and policies enhance the work of all ICs.
- The NIH policies and practices should build on and incorporate the work and expertise of the existing BMIC CDE working group, which includes NIH staff with deep experience in multiple CDE programs.

2. **Ensure NIH wide forum for sharing and harmonization across NIH CDE programs.**

   The Trans-NIH Biomedical Informatics Committee (BMIC) established a subcommittee of technical and program leaders of NIH CDE programs to share and document their experience in developing, managing, and providing CDEs for their research communities. This group has enabled information sharing across broad and diverse sets of CDE programs leading to an appreciation and respect for the commonalities and differences of various CDEs. This forum should continue and these technical and program leaders should provide input and recommendations to the NIH CDE Governance Group concerning NIH wide policies.

   This program/technical forum should:

   - Develop shared common practices for CDE development, adoption, harmonization, modification, versioning, and use.
   - Ensure appropriate flexibility such that NIH CDE programs can enable all types of research and address the needs of particular IC use cases and research. communities; however, encourage use of shared CDEs as appropriate.
   - Develop a common definition/taxonomy of CDEs (with infrastructure leaders)
   - Develop (and implement) a proposal for common formats for CDEs (with infrastructure leaders)

3. **Coordinate and integrate NIH technical and software infrastructure(s) for supporting CDE development, discoverability, and access.**

   Several NIH groups have developed infrastructure for supporting CDE development, discoverability, and access, including NCI, CIT, NLM, and others. The leaders of these major infrastructure programs met at the September 30 workshop. Areas of overlap, differences, and opportunities for coordination and consolidation were evident. A group of the leaders of CDE infrastructure programs should begin immediately sharing information, tools, and coordinating activities.

   This infrastructure working group should:

   - Ensure **all** NIH required CDEs are easily discoverable and openly available. Inclusion of CDE availability in widely available registries/catalogues with links to de-centralized repositories where the individual CDEs are accessible could be one approach.

- Ensure NIH CDE infrastructures are not duplicative or siloed (unless well justified).
- Ensure NIH CDE infrastructures are user friendly for those looking to obtain information about NIH CDEs and to utilize NIH CDEs in their research; as well as meeting the needs of NIH programs and staff.
- Develop a common definition/taxonomy of CDEs (with technical/program leaders)
- Develop (and implement) a proposal for common formats for CDEs (with program/technical leaders)
- Develop (and implement) a proposal for common API for all NIH CDE-related resources.

4. **Obtain evidence of the impact and value of CDEs.**
   Evidence for the value and impact of CDEs on scientific advancement (positive or negative) is difficult to find. To obtain such evidence, NIH should support research to gather relevant data on impact of NIH CDE use on scientific research projects.

5. **Assess the interoperability of NIH CDEs.**
   The ease of exchange of information collected using various CDEs is unclear. This is a very relevant question when studies use different CDEs to collect data on the same characteristic or activity, e.g., smoking. With multiple CDEs in use for collecting data on smoking, it is unclear whether these data are comparable or not. Additionally, various CDEs collecting the same data may collect it in different units or formats. NIH should support collection of data on the comparability and interoperability of the NIH required CDEs.

Attachments:
CDE Workshop Agenda and Roster