

**U.S. Department of Health and Human Services (HHS)
National Institutes of Health (NIH)
Office of the Director
Division of Program Coordination, Planning, and Strategic Initiatives
Office of Data Science Strategy (ODSS)**

**Advancing the Use and Development of
Common Data Elements (CDEs) in Research**

**March 6–7, 2024
Auditorium Balcony C, Natcher Building
NIH Main Campus
Bethesda, MD
and Virtual**

Workshop Executive Summary

Overview and Summary

This workshop focused on the use of CDEs—standardized, defined questions paired with specific allowable responses that—if used systematically across sites, studies, or clinical trials—ensure the data collected are consistent. The goals of the meeting were to assess, enhance, and broaden the development, adoption, and use of CDEs for research across diseases and conditions; demonstrate successes and discuss strategies to encourage the adoption and use of CDEs; and engage participants from diverse professional backgrounds.

Dr. Susan Gregurick, Associate Director for Data Science and Director of ODSS, provided opening remarks, and Dr. Monica Bertagnolli, Director of NIH, emphasized the importance of CDEs in making science valuable across the biomedical research community. To set the stage, Ms. Denise Warzel from the National Cancer Institute explained that using CDEs to make data interoperable allows researchers to compute meaning by assessing the concepts associated with the question and allowable responses and emphasized that adopting CDEs will improve data quality and consistency, support data harmonization, enhance knowledge acquisition, simplify collaboration, and reduce project startup time.

Session I: The Value of Common Data Elements

In the first session, Dr. Richard Scheuermann from the National Library of Medicine (NLM) outlined the value of ontologies—formal, explicit specifications of a shared conceptualization—for structuring the content of data and facilitating interoperability, allowing users to make inferences within the semantic knowledge. Ms. Helena Sviglin of the U.S. Food and Drug Administration (FDA) outlined the FDA policy framework for sharing data standards, which uses an NIH-maintained vocabulary and supports improved public health in the United States. Mr. Avinash Shanbhag of the HHS Office of the National Coordinator for Health Information Technology presented on the importance of standardized data for reuse, and Dr. Allan Uribe of the Centers for Disease Control and Prevention outlined the data-driven efforts of the Healthy People 2030 initiative, which provides a framework for prevention and wellness programs across a diverse array of users. Discussions centered on the interoperability challenges associated with the U.S. health care system and the need for grassroots initiatives to support increased interoperability within agencies.

Session II: Current NIH Resources for CDEs

In the second session, Ms. Robin Taylor of NLM presented on the NIH CDE Repository and NIH CDE Governance Committee, which reviews and endorses CDEs for publication in the repository. A panel discussion with members of the Governance Committee included comments on the difficulty of versioning CDEs and potential strategies to increase and track the use of CDEs.

Session III: Overcoming Barriers in CDE Adoption, Mapping, and Use in Community Research

In the third session, Dr. Warren A. Kibbe from Duke University discussed RADx-Underserved Populations and its use of community-developed CDEs to better understand the communities that are being engaged as a part of this effort. Dr. Sirimon O'Charoen of the Crohn's & Colitis Foundation spoke on the importance of CDEs for the IBD Plexus program, a research ecosystem focused on inflammatory bowel disease (IBD) that aims to accelerate progress toward precision medicine by providing researchers integrated clinical patient-reported data, with the goals of accelerating precision medicine and finding a cure for IBD. Dr. Stuart Gansky from the University of California, San Francisco, presented on the use of CDEs within the National Institute on Minority Health and Health Disparities Health Equity Action Network, which is studying comorbid chronic diseases in communities with health disparities and focusing on disease prevention, treatment, and management across a wide variety of health outcomes. In this session, attendees discussed the importance of validating CDEs, engaging with partner communities, and using language that reduces barriers between scientists and the community.

Session IV: Technical Implementation Aspects of Mapping, Transformation and Harmonization

The fourth session opened with a talk by Dr. Shruti Sehgal of the Northwestern Feinberg School of Medicine on efforts to standardize the collection of clinical and research data related to food allergies, which are highly complex. Dr. Anne E. Thessen from the University of Colorado Anschutz Medical Campus outlined her work promoting data integration and interoperability through CDE mapping, which requires data-model alignment and value set alignment. Ms. Kathleen Malum from the National Marrow Donor Program spoke on harmonizing data using CDEs, noting that streamlining data collection and facilitating data sharing generates new knowledge that ultimately yields better clinical outcomes. In the discussion, participants touched on the need for CDE ontologies, the importance of incorporating disease mechanisms into CDE development, and the importance of open access and collaboration.

Session V: Approaches to Improve Interoperability

In the fifth session, Dr. Paul Harris from Vanderbilt University Medical Center shared an overview of REDCap, a software platform for designing research databases to support diverse clinical and translational studies. Ms. Jocelyn Craven and Ms. Sara Meyer from the Medical University of South Carolina discussed lessons learned from incorporating the CDEs into their case report forms through StrokeNet and the Strategies to Innovate Emergency Care Clinical Trials Network (SIREN). Ms. Ginger Riley from Westat discussed considerations related to standardization of the National Cancer Institute Clinical Trials Evaluation Program CDEs, and Dr. Kenneth Gersing from the National Center for Advancing Translational Sciences spoke on common data model harmonization. Workshop attendees discussed the use of data dictionaries, the need to leverage expertise to understand the relevance of existing content, and the importance of metadata.

Session VI: Use Cases for Preparing and Applying CDEs for Intelligent Technologies

The last session opened with a talk from Dr. Deborah Duran of the National Institute of Minority Health and Health Disparities. She discussed the Science Collaborative for Health Disparities and Artificial Intelligence Bias Reduction (ScHARe) data repository, which was built to bring underrepresented people,

including women and people of color, into data science and improve the use of big data in social science and health disparities research. Next, Dr. Katherine Liao of Harvard Medical School provided the perspective of a clinical investigator on CDEs, outlining an artificial intelligence/machine learning project to study treatment response in rheumatoid arthritis, and Dr. Sally Baxter from the University of California, San Diego School of Medicine outlined challenges to making data in the ophthalmology field ready for use with artificial intelligence. The discussion focused on the importance of CDEs for ensuring different projects can use the systems that are most appropriate and the difficulty of integrating many types of data.

Closing and Wrap-Up

In the final keynote, Dr. Victoria Shanmugam of the NIH Office of Autoimmune Disease Research outlined the need for more interoperable longitudinal data to understand the causes and mechanisms of autoimmune diseases in the United States. Dr. Steve Tsang closed the conference by identifying crosscutting themes—including the need to be aware of the community when developing CDEs, the need for better collaboration and communication and platforms through which to do so, and the importance of emerging technologies—and thanking participants for their participation and engagement across many disciplines.