National Institutes of Health
National Institute of Diabetes and Digestive and Kidney Diseases

Addressing Gaps, Challenges, and Opportunities
Related to Data and Metadata Standards for NIDDK Research Priorities

Virtual Meeting
June 3–4, 2025

EXECUTIVE SUMMARY

The National Institutes of Health (NIH) sponsored a scientific workshop on June 3 and 4, 2025, titled "Addressing Gaps, Challenges, and Opportunities Related to Data and Metadata Standards for NIDDK Research Priorities," which was hosted by the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK). The purpose of the workshop was to identify and address the current gaps, challenges, and opportunities in these areas.

Community-adopted data and metadata standards are crucial for effective data sharing and integration across various domains. Data standards and ontologies for clinical and laboratory observations enable integration of clinical results for improved care and clinical research. However, disease-specific measures in clinical studies are often varied in data collections, data specifications, and data representations, which creates challenges for integrated secondary data analyses. To address these challenges, NIDDK organized a workshop aimed at identifying gaps, challenges, and opportunities related to data and metadata standards for data types relevant to NIDDK research mission areas.

NIDDK brought together experts from industry, academia, regulatory bodies, and other fields who specialize in research related to diabetes; obesity; nutrition; and liver, digestive and kidney diseases. The initial focus of the workshop centered around data and metadata from continuous glucose monitors (CGMs), other wearable devices, dual-energy X-ray absorptiometry (DEXA) scans, diet and nutrition, physical activity, urodynamic testing, and related measures (such as those for fatty liver and kidney disease), but the discussion explored additional topics, as well.

The objectives of the workshop were to—

- Facilitate meaningful discussions with community partners on promoting data interoperability for both primary and secondary use in NIDDK research mission areas.
- Identify gaps and challenges resulting from lack of or inconsistent use of existing data and metadata standards.
- Explore opportunities to develop standards, standardized terminologies, or common data elements (CDEs) where data and metadata standards do not exist.
- Identify next steps to encourage the development or adoption of data and metadata standards for relevant data types.

Following opening remarks from the Deputy Director of NIDDK, leading experts in NIDDK-related scientific areas presented their research across two keynote presentations and six scientific sessions. These sessions encompassed recent advancements, the current landscape, and areas needing improved data and metadata standards. In discussions, participants identified approaches to explore that could advance data and metadata standards related to NIDDK research priorities, which would promote interoperability, foster seamless data sharing, and enable advancements in patient care.

Keynote 1: The Land That Time Forgot: How Bridges to Nowhere, Abandoned Tracks, and Sinkholes Stand Between Us and the Diabetes Care We Deserve (and How to Fix Them)

The speaker discussed a clinical decision support system for diabetes. The Rising T1DE Alliance is a successful initiative that resulted in effective frameworks and support tools (e.g., diabetes data dock structure, scalable model for 14-day time in range). This initiative may provide guidance for data standardization and data sharing in other NIDDK-related research areas.

Gaps, Challenges, and Opportunities

- Designing data interoperability is complex, and businesses are not incentivized to include interoperability in the design process. The challenge of making data interoperable falls on health care and research institutions. The existence of numerous conflicting data standards complicates the creation and implementation of unified standards for data interoperability.

- Numerous ancillary devices are used to support patient care. The device data are stored in proprietary software platforms. Device data have several layers to navigate—including the raw signal, processed signal, derived features from the raw signal, and metadata. Not all patients have continuous connectivity, which hinders data sharing with the receiver.

- A broad range of data types and sources are used during different phases of the patient care continuum (i.e., screening and prevention, diagnosis and prognosis, and treatment and management). Data definitions are subjective and can change based on the source.

- An eight-step, simplified digital health integration framework has been developed to implement integrated diabetes data and technology solutions, which may be a useful tool for other chronic diseases.

- Designing data interoperability and reducing data silos will save health care dollars, drive additional research, and accelerate innovation.

- More specialists are needed to provide sustainable care for patients with metabolic diseases and endocrine conditions. A focus on clinical decision support systems and population health management platforms is needed to address this gap. These tools would allow adaptive interventions, which would improve patient care.

Session 1: Success Stories of Data Collection, Integration, and Interoperability, Including Data from Wearable Devices

Speakers outlined the technical, operational, and strategic challenges of data standardization and interoperability. Speakers provided success stories as examples of how data integration and interoperability have been advanced through large collaborative efforts. Presentation topics included the Integration of Continuous Glucose Monitoring Data into the Electronic Health Record (iCoDE), the Diabetes Research Hub (DRH), and the *All of Us* Research Program. These large collaborative efforts to adopt, adapt, and create have worked well for data standardization.

Gaps, Challenges, and Opportunities

- Data acquired from CGMs should be integrated into electronic health records (EHRs) to minimize administrative burden and reduce the introduction of errors. Interoperability is vital, but increased capabilities result in increased complexity. Numerous technical integration barriers are present (e.g., account linkage, data fidelity, data storage, workflows), which has resulted in the slow integration of CGMs into EHRs.

- iCoDE is a comprehensive and practical guide for CGM–EHR integration; it provides a framework for any organizations interested in using it for their clinical workflows. Six working groups, including 130 participants from 60 organizations, were created to develop an adapted interoperability framework for CGM data.

- CGM research is hindered by technical, knowledge, and community barriers.

- The DRH provides a secure and user-friendly platform for CGM research, and the DRH will be expanded to create a comprehensive resource. Broader use of iCoDE and the DRH by researchers will help expedite research breakthroughs.

- When mapping EHR data to the observational medical outcomes model, data extraction is inconsistent, and text data from clinical notes are not available. Opportunities moving forward include leveraging national platforms, exploring EHR acquisition strategies with other NIH institutes and centers, and developing methods to collect participant data outside of health professional online services.

- Wearable devices demonstrate challenges in data ownership. The 21st Century Cures Act and Executive Order 13813 have increased patient access to data. Additionally, the Centers for Medicare and Medicaid Services has a request for information (RIN 0938-AV68) posted until June 16, 2025, regarding beneficiary and stakeholder access to data. However, policy changes are needed to improve device data ownership.

- These challenges with wearable device data ownership and accessibility hinder patient trust and literacy, integration into clinical care, and data standards and interoperability. The inclusion of additional wearables (e.g., smartwatch, fitness device) in research may help bridge the gap between consumer and clinical data ecosystems.

- Patients should be able to access and control how their device data are used. A patient data platform that allows patients to direct their data is currently unavailable.

- Collaborating with other federal agencies (e.g., U.S. Food and Drug Administration [FDA]) could lead to the development of a standardized model for data integration into EHRs and a way to merge data from multiple devices. Reference standards will be needed to merge data from different vendors and across studies.

- Changing device data accessibility and intellectual property policies, increasing standardization advocacy efforts, developing standard CDEs for diabetes technology, identifying shareable standards to incorporate tabular data into time series data models, and creating metadata standards for devices should be prioritized.

Panel Discussion—Gaps, Challenges, and Opportunities

- Artificial intelligence (AI) and machine learning (ML) pattern analyses identify mechanistic relationships in diseases (e.g., metabolic subphenotypes of type 2 diabetes) that allow for specialized treatments. Comorbidities and disease complications should be incorporated into analyses. Extrapolation of data should be translated to actionable therapeutic and dietary recommendations for clinicians to provide to patients. Clinical decision support tools need to be simple for physicians to use.

- Multisource and multimodal patient-generated health data are complex, and their integration into clinical workflows is limited. Device capabilities (e.g., sampling frequency, new analytes) are expanding, but the current standards do not support these expanding capabilities.

- Five distinct areas need additional standardization: patient connection, data pre-processing, discrete structured data and summary statistics, design choices and visualization, and pushing data back to EHR.

- Linkages between workflows and standards are needed, including lateral translation tools and clinical workflows connected to data analytics workflows. Researchers should be able to easily retrieve data from the EHR into a platform that allows analysis.

- Canonical models, which are an abstraction of concepts that each manufacturer maps to the model, are currently used to bring data together, but that process adds an additional layer of complexity. Universal contracting standards would improve accessibility to raw data and reduce variability across devices.

- Collaborations are vital for the development and integration of standards. Building new collaborations with individuals across sectors (e.g., expert panels, working groups, research partners) is necessary to come to consensus about the core health metrics, terms, methods, and approaches to move data standardization forward.

Session 2: Data and Metadata Standards Integration and Interoperability Gaps, Challenges, and Opportunities for NIDDK-specific Disease Areas

In the second session, speakers outlined integration and interoperability efforts for data and metadata standards specific to NIDDK-related disease research areas, including diabetes, obesity, nutrition, and urologic disease. Presentations addressed the measurement of body composition by DEXA in single-site and multisite trials, how to make -omics data FAIR (findable, accessible, interoperable, and reusable), lower urinary tract dysfunction, and preclinical studies on obesity, metabolism, and energy balance.

Digestive Diseases and Nutrition Research Areas—Gaps, Challenges, and Opportunities

- Instrument performance throughout software updates and maintenance visits should be monitored to ensure accuracy and determine if data normalization is required. Recommended approaches for instrumentation use and data analysis should be documented in methodological papers.

- Data outputs can vary across manufacturers and instrument models. Multi-institutional trials should use the same manufacturer and instrument model to acquire data, if possible. Additional factors (e.g., operator variability, analysis mode, scan mode) can result in data differences among institutions, so cross-calibration methods must be implemented. Direct comparison studies are needed to ensure accurate cross-calibration.

- Data generated from NIDDK-funded projects should provide metadata that follow defined ontologies, controlled vocabularies, and standards. Metadata descriptions and ontologies are critical to find data that users are interested in for secondary analysis.

- Data interoperability across distinct data repositories (e.g., gene, drug, glycome) and data harmonization across types are both needed to allow researchers to complete integrative analyses. Application programming interfaces (APIs), metadata structure, and new workflow tools will help overcome this challenge.

- Use cases would help the community learn how to use repository data.

- A minimal clinical metadata model would aid cohort selection for secondary analyses.

- Large language models (LLM) cannot capture context with currently available data. Further context is needed for AI-LLM analytics.

- User-friendly interfaces should be biologically meaningful and intuitive.

Kidney, Urologic, and Hematologic Diseases Research Areas—Gaps, Challenges, and Opportunities

- Several factors hinder translational and clinical urology research efforts: (1) Research funding for surgical departments is limited, (2) residencies offer fewer research years

than previously, and (3) patients experience a high level of physical invasiveness and emotional burden.

- Data collection standards for urodynamic studies are not strictly followed (e.g., regarding catheter placement), and a lack of training exacerbates this issue. The field lacks evidence that links diagnosis standards to the correct treatment approaches. A predictive model would help standardize treatments and improve monitoring of patient outcomes.

- Additional clinical tests should be included to obtain data in neglected areas and to accurately diagnose patients into urological dysfunction subtypes (e.g., mixed incontinence).

- Data sharing standards are limited. Urodynamics uses subjective assessments, and when clinical interpretations vary it becomes difficult to understand the impact on patients.

- Structured data entry into EHRs could be useful, but such entries are an abstraction of the raw data. A centralized repository for raw data should be created to help generate hypotheses, develop algorithms, and archive electronic data.

- A common minimal data set needs to be established that includes easy data entry and data sharing. A committee or initiative could help facilitate this effort.

- Clinicians and the health care community should be encouraged to embrace data standards and sharing.

Diabetes, Endocrinology, and Metabolic Diseases Research Areas—Gaps, Challenges, and Opportunities

- The International Indirect Calorimetry Consensus Committee (IICCC) developed a community-driven, multinational agreement of common data standards (i.e., a guide to preclinical indirect calorimetry experiments).

- Standards for data formats, metadata, analysis methods, and graphical representations are being established. The CalR Project is a starting point for this effort, but it cannot accommodate all experimental designs. User feedback helps shape updates to CalR.

- A consensus paper in a high-visibility journal will help ensure researchers use the standards established by the IICCC when submitting research manuscripts. Researchers should be cautious when comparing new studies to previous studies that did not follow data standards.

- A large-scale data repository for aggregation and creation of indirect calorimetry data is a missing resource.

- Preclinical use of human diagnostic tests and dosing calculations are not ideal. These test results and calculations must be interpreted or modified correctly (e.g., for a glucose tolerance test).

- Consistency is needed when normalizing or presenting data and when reporting research factors (e.g., mouse traits, procedural parameters). Linear mixed models and other variance-based approaches can help with batch corrections.

Panel Discussion: Crosscutting Gaps, Challenges, and Opportunities for Integrating Data and Metadata Standards for NIDDK-specific Disease Areas

Gaps, Challenges, and Opportunities

- Core common metadata standards and CDEs are necessary for both clinical and preclinical studies. A trial-and-error approach likely will be needed to identify CDEs that are relevant to the community.

- Researchers should implement accepted ontologies and controlled vocabularies to ensure that the CDEs are as standardized as possible.

- Researchers can play a key role in developing resources for integrating data and metadata standards, and NIH can help coordinate these efforts and set best practices.

- NIDDK could establish small working groups on focused topics (e.g., phenotypic data, -omics data, clinical data). These groups could work toward larger NIDDK recommendations for extramural researchers. Additionally, a wiki resource would allow community members to post use cases.

- Several established NIH working groups are setting preclinical data standards in other domains. NIDDK can leverage these past efforts and foster discussions with users on lessons learned.

- Integrative studies can provide insight into coordination across body systems and diseases. By characterizing common data standards across body systems and disease areas, researchers can work toward achieving a goal of whole-body health.

- Repositories for CDEs have been established, but they are largely focused on clinical data. Approaches for incorporating preclinical data into repositories will be crucial, and the data type (e.g., phenotypic, imaging, cellular assays) must be considered.

- Efforts will be needed to apply standards within both specialist and generalist repositories. Data access and sustainability also represent long-term concerns. Quality control metrics also must be considered, particularly across various repositories and domains.

Keynote 2: Better Health Enabled by Data

The keynote speaker discussed data interoperability efforts facilitated by the U.S. Department of Health and Human Services (HHS) Assistant Secretary for Technology Policy (ASTP) through data standards, API standards, and network standards. Coordination efforts with standards development organizations (e.g., National Council for Prescription Drug Programs), federal

agency partners, and industry ensure that established data standards meet agency programmatic goals and are widely adopted into practice. ASTP has focused on building a foundation for the digital health care infrastructure.

Gaps, Challenges, and Opportunities

- The United States Core Data for Interoperability (USCDI) is a collection of core data elements, managed by ASTP, that is updated annually to meet community needs. Requested data elements undergo a public, transparent review process to determine inclusion in the next release. USCDI is the minimum data set for key EHR functions, interoperability networks, and patient access to data.

- Federal agencies can use USCDI as a foundation and expand it with additional data elements for specific needs and priorities (e.g., cancer-related activities).

- Fast Healthcare Interoperability Resources (FHIR) API standardization and semantic interoperability reduce a researcher's time spent on data cleanup and manipulation, providing additional time for innovative research.

- ASTP coordinates with federal agencies (e.g., FDA, Centers for Disease Control and Prevention, U.S. Department of Veterans Affairs [VA]) to leverage FHIR APIs and implement additional APIs for program-specific needs.

- Health information networks vary in multiple ways, and support from the federal government is required to enhance national network interoperability.

- The Trusted Exchange Framework and Common Agreement (TEFCA) provides a unified policy framework with common technology infrastructure to support and simplify data exchange. It is currently being used for patient treatment, and additional work is needed to expand TEFCA's use for other purposes (e.g., research).

- Collaborations between NIDDK and ASTP would advance new standards and capabilities in certified health information technology, which supports investigators conducting NIDDK-relevant research.

- Patient access to data records has substantially increased, and continued patient engagement will promote positive health care outcomes.

- Incentive programs promote data interoperability.

Session 3: Resources for Electronic Health Records and Clinical Research Data Integration and Interoperability

In the third session, speakers discussed leveraging traditional and technical approaches to advance the adoption, integration, and interoperability of clinical research data and EHR data standards. Presentation topics included Logical Observation Identifiers Names and Codes (LOINC) data standards and resources for research purposes, the Center for Linkage and

Acquisition of Data (CLAD), and the challenges of compiling data and treating patients with multiple chronic conditions (MCC).

Gaps, Challenges, and Opportunities

- Development of standardized LOINC terminology is achieved through collaborative efforts among several groups—the LOINC team at the Regenstrief Institute, LOINC committees, and the international user community. Within the LOINC database, several names are connected to one LOINC concept, which provides flexibility to users.

- A comprehensive data set for an individual's life course could improve health and research, but many hurdles exist (e.g., different formats and standards, regulatory and legal requirements, security and technical compatibility). Patient data are heterogeneous despite using common data models and exchange formats (e.g., FHIR).

- Actively contributed data are only from the point of enrollment onward in a patient-based research study. Passively acquired data (i.e., historical records) are necessary for innovative analytics, but EHR data may have improper or missing units. Algorithmic repair fixes the data so they can be included in studies.

- Multisite collaborations improve algorithmic repair and data quality, but multisite analytics are not feasible without commonly defined visits. A microvisit to macrovisit map allows analyzing visits across different sites in a consistent format.

- EHR data collection is expensive and fragmented. eHealth Exchange, through TEFCA, is being used to collect EHR data from health information exchanges and health information networks.

- Hex 3 allows the linkage of place-based environmental and health data through spatially layered geocoding. Hex 3 aggregation allows more comprehensive coverage than ZIP codes, avoiding the creation of dead zones and maintaining privacy.

- Numerous repositories of CDEs exist, comprising 1,800 databases, 1,500 standards, 900 ontologies, and more than 23,000 CDEs. This volume hinders data navigation and harmonization. Improved standardization tools and updated mapping and harmonizing of CDEs are necessary for data interoperability.

- Patients with MCC have complex health care needs. Disease-specific approaches to treatment are misaligned with the person-centered care planning. The MCC MyCarePlanner, which is being developed with support from the Agency for Healthcare Research and Quality, enables patients and their caregivers—who may have varying levels of health literacy—to understand their health record data. Improvements are needed to obtain and record medication history, standardize the care team list of practitioners, update display systems, and accurately track hospital discharges to edit patient goals.

- Behavioral health care settings require substantial improvement—including collecting and exchanging data, monitoring longitudinal care, and accessing discharge summaries to adapt patient care.

- Identifying specific data standardization use cases will reduce complexity and help achieve outcomes. Analyzing the current landscape will help determine where new standards are preferable to the implementation of current standards.

- A team-based approach is needed for integration and interoperability. Government-funded opportunities to develop open-source standards, along with public–private partnerships, are needed to create sustainable standards.

Session 4: Moving the Field Forward, Hybrid Approach (Traditional and Technology)—Focus on Technology

In the fourth session, speakers discussed leveraging hybrid approaches, emphasizing technological solutions to implement data and metadata standards for interoperability. Presentation topics included the use of an AI-enabled framework to support interoperability and the role of CDEs in AI.

Gaps, Challenges, and Opportunities

- CDEs facilitate data sharing, consistency in data collection, and integration across different research studies. Creating a centralized, comprehensive reference resource of CDEs is needed because data from different sources vary, the appropriate resources can be hard to find, obtaining all relevant information from a single source is challenging, and efforts across platforms can be redundant.

- Vector embedding facilitates the systematic aggregation of CDEs. New data sources and repositories can be integrated into this system as needed. A pilot project that built a vector database for inflammation-related information demonstrated the utility of vector embedding. Such a resource would be beneficial in other clinical topics, as well.

- The cardinality of each CDE can be diverse. It can be complex to associate a CDE with a single vector when using high-dimensional data (e.g., -omics data) in a CDE vector-based AI model approach. Potential ways to address this challenge include defining a standard representation for CDEs (e.g., a numerical scale), guiding and encoding metadata, and normalizing dimensionality.

- Use of AI requires large volumes of high-quality data. Incorrect labels, inconsistent terminology, and missing data validation lead to substantial time spent manually cleaning data prior to training the AI model. Inconsistent data formats, heterogeneous data models, and lack of interoperability reduce data reuse in AI pipelines.

- LLMs are susceptible to hallucinations if insufficient or poor-quality data are used to provide outputs. CDEs enhance AI training and reduce LLM hallucinations, which results

in superior AI model performance and accelerates discoveries to advance research and health care.

- Context is important for conveying the meaning of words. CDEs linked to standard concept codes give domain-specific context and provide unambiguous, shared meaning.

- CDE use should be expanded across NIH institutes and cores and integrated into different research domains. Integration across different domains would improve the identification of patterns and relationships between different diseases.

- Regular updates and refinements of CDEs should be completed to keep pace with evolving research needs and to provide continuous improvement.

- CDEs should be based on defined ontologies or controlled variables as long as the terminology is accurate. Mapping among terminologies using the Unified Medical Language System is helpful when new users are incorporating different terminology.

- Numerous data sets exist that have yet to be annotated with CDEs. Repositories and NIH should work to harmonize previous data with existing CDEs. A hybrid approach using AI and manual review would help with retrospective annotations.

Session 5: Moving the Field Forward, Traditional and Technology—Focus on a Multidisciplinary Approach

In the fifth session, speakers discussed multidisciplinary approaches to advance data and metadata standards. Presentation topics included a biomedical engineering approach to developing informatics tools for research and clinical trials, AI use in biomedical imaging, and the NIH Biomedical Research Informatics Computing System (BRICS).

Gaps, Challenges, and Opportunities

- Collecting unstructured data from independent studies and harmonizing the data afterward is highly inefficient and limits the return on research investment. Consistent, structured data collection using CDEs supports FAIR principles and reduces data manipulation.

- Assigning Concept Unique Identifiers and permissible values to CDEs allows easier aggregation with data from other studies.

- Stakeholder input about user interfaces leads to usability improvements in systems. A multidisciplinary approach with subject-matter experts from diverse fields is needed to implement workflows and update systems for efficacy and efficiency.

- Data collection practices are highly variable, especially in multi-institutional trials; the variability requires substantial efforts to harmonize data for AI use. Standardized infrastructure is needed to overcome this concern.

- AI has transformed biomedical image analysis, but the lack of available clinical images and text in certain areas (e.g., radiology) remains a challenge. Well-labeled biomedical image data are also expensive to produce. Fostering access to publicly available clinical data would benefit further AI development.

- Many AI systems are siloed but need to be shared. Publications may share some code, but this information has limited utility without the AI model.

- A uniform NIH-wide solution is absent. Imaging data archives differ in image formats, user interfaces, user agreements, availability and format of metadata, and storage costs. A centralized image archive that makes clinical data publicly available and fosters reuse of data would result in improved patient health.

- Manual review of clinical text files is not feasible. Standards for clinical text anonymization are desired.

- NIH BRICS offers researchers a secure, robust platform and tool suite for use across different disease research areas. It is a unified, secure, and collaborative platform that requires a single set of hardware, does not have licensing fees, stores data on government-hosted infrastructure, and meets compliance requirements.

- Leveraging partnerships (e.g., Microsoft, Google) will promote robust data standards.

- As new data standards are implemented, a large effort in computation will be needed to restructure data from the previous format.

- Technical implementation, governance, and culture must be modified to advance data and metadata standards.

Session 6: Building Partnerships and Facilitating the Establishment and Adoption of Data and Metadata Standards

In the sixth session, panelists discussed the role of government agencies in building and facilitating the establishment and adoption of data and metadata standards. Government agencies must balance their dual responsibilities in regulating and facilitating the development of standards with external organizations. Partnerships among government agencies are crucial to promoting new standards that regulate imaging devices and electronic health data exchanges and, in turn, improving patient outcomes.

Gaps, Challenges, and Opportunities

- Clinics manage multiple imaging systems and medical devices, and proprietary formats exist for various instruments, which hinders research efforts.

- The National Eye Institute (NEI) has included the integration of data standards in upcoming funding strategies. Notices of funding opportunities encourage applicants to

address standards and conform when using imaging methods. NEI also has launched an open-access GitHub repository for sharing coding tools.

- Standards are not adopted consistently across companies, and differences in infrastructure make it difficult to find a solution that will work for all companies. Adopting new standards may require additional investment.

- The problem must be clearly defined, and the appropriate policy or program to leverage must be identified. To implement certain policies or programs, evidence must be generated that can be used as a foundation to inform regulatory decisions.

- Collaborations across different public and private stakeholder groups (e.g., professional societies, industry, the VA) in health care is needed. Mechanisms to promote discussion, receive stakeholder input, and define a unified goal—including workshops, joint position papers, and requests for information—are needed.

- Implementing standards creates predictability in product development pathways, which helps global partners and manufacturers.

- Unique device identifiers (UDI) provide important details (e.g., manufacturer, product version, serial number). UDIs can be leveraged to promote innovation, enhance traceability, and expand access to various patient populations.

- Funding should be provided for proposals that use instruments (e.g., medical devices) that conform to standards. Grant funding to projects that use non-conforming instruments should be minimized because the resulting data would remain siloed.

- Data standardization within EHRs remains a work in progress. There is an important distinction between data elements that are disease specific and data elements that establish safety and efficacy.

- Use of at-home medical devices requires data standardization and interoperability. Patients should be involved in collaborative data standardization efforts because they are leveraging their own data to alter their lifestyle and health care choices.

- The current research landscape rewards investigators for working independently. This culture hinders collaboration because of the competition to publish first and obtain senior authorship. To shift this culture, federal agencies should develop a metric to reward investigators for sharing high-quality data (e.g., a data-sharing index challenge).

- If applicable, a broad range of sample populations should be used in studies (e.g., older adults, Medicare beneficiaries).

Summary and Conclusion

Current efforts by NIH—including the implementation of data management and data sharing policies for NIH-funded investigators—promote a culture of data reuse. Further development of

data and metadata standards will be a community-based effort, which can be coordinated, accelerated, and incentivized by NIDDK (e.g., formation of working groups). Factors influencing progress include (1) agreeing on metadata types, (2) using the existing ontologies and vocabulary for data convertibility, (3) sharing infrastructure to avoid data duplication, (4) creating guidelines for data quality, (5) using AI and API to convert different data into a standard format, (6) including patients as research partners, and (7) distinguishing between semantic and syntax standards. Broad implementation of standards will enhance scientific discoveries and provide opportunities at disease domain intersections and across MCC to develop personalized care and enhance treatment.

Organizing Committee

- Daniel Gossett, Ph.D., NIDDK, NIH
- Veerasamy Ravichandran, Ph.D., NIDDK, NIH
- Rebecca Rodriguez, Ph.D., NIDDK, NIH
- Ashley Xia, M.D., Ph.D., NIDDK, NIH

Speakers and Panelists

- Guillermo Arreaza-Rubin, M.D., NIDDK, NIH
- Julio Ayala, Ph.D., Vanderbilt University
- Maria Bagonis, Ph.D., Center for Information Technology (CIT), NIH
- Alex Banks, Ph.D., Harvard Medical School
- Roy Beck, M.D., Ph.D., Jaeb Center for Health Research
- Rich Bergenstal, M.D., HealthPartners Institute
- Michael Chiang, M.D., NEI, NIH
- Mark Clements, M.D., Ph.D., CPI, FAAP, University of Missouri–Kansas City
- David Dorr, Ph.D., Oregon Health & Science University
- Juan Carlos Espinoza, M.D., Northwestern University
- Jeffrey Grethe, Ph.D., University of California, San Diego
- Melissa Haendel, Ph.D., FACMI, The University of North Carolina at Chapel Hill
- Eza Hafeza, M.D., Regenstrief Institute
- Syed Haqqani, NEI, NIH
- James Hokanson, Ph.D., Medical College of Wisconsin
- David Klonoff, M.D., FACP, FRCP, University of California, San Francisco
- Wendy M. Kohrt, Ph.D., University of Colorado Anschutz Medical Campus
- Shari Ling, M.D., Centers for Medicare and Medicaid Services
- Matthew McAuliffe, Ph.D., CIT, NIH
- Dominic Nathan, Ph.D., Uniformed Services University
- Jenna Norton, Ph.D., NIDDK, NIH
- Veerasamy Ravichandran, Ph.D., NIDDK, NIH
- Amber Reed, NEI, NIH
- Chandan Sastry, Ph.D., *All of Us* Research Program, NIH
- Krishnakant Shanbhag, M.S., Office of Standards, ASTP, Office of National Coordinator for Health IT, HHS
- Shankar Subramaniam, Ph.D., University of California, San Diego
- Ronald Summers, M.D., Ph.D., Clinical Center, NIH
- Michelle Tarver, M.D., Ph.D., Center for Devices and Radiological Health, FDA
- Xujing Wang, Ph.D., NIDDK, NIH
- Denise Warzel, M.Sc., National Cancer Institute, NIH
- Kelly Watson, Tidepool
- Claire Yang, M.D., University of Washington