INFORMATICS PROFESSIONALS. LEADING THE WAY.

May 10, 2021

Patricia Flatley Brennan, RN, PhD
Director
National Library of Medicine (NLM)
8600 Rockville Pike,
Bethesda, MD 20894

Walter J. Koroshetz, MD
Director
National Institute of Neurological Disorders and Stroke (NINDS)
P.O. Box 5801
Bethesda, MD 20824

Re: Request for Information (RFI): Use of Common Data Elements (CDEs) in NIH-funded research

Directors Brennan and Koroshetz:

The American Medical Informatics Association (AMIA) is pleased to provide input on NIH's request for information (RFI) on the use of Common Data Elements (CDEs) in NIH-funded research.

AMIA is the professional home for more than 5,500 informatics professionals, representing frontline clinicians, researchers and public health experts who bring meaning to data, manage information, and generate new knowledge across the health and healthcare enterprise. As the voice of the nation's biomedical and health informatics professionals, AMIA plays a leading role in advancing health and wellness by moving basic research findings from bench to bedside, and evaluating interventions, innovations, and public policy across settings and patient populations.

CDEs are essential to delivering on the promise of the Learning Health System, as they provide a means for consistent use of data elements in various research studies, as well as clinical settings. To this end, we commend NLM on its work with CDEs over the past decade, including its vitally important work on maintaining the NIH CDE Repository and making efforts to improve its usability. However, as evidenced by the number of CDEs present in the repository, the expansion of available data has led to an expansion of the ways CDEs are captured. By way of example, the search term "race" currently generates 65 different results from the CDE Repository. While some of these search results are indeed meaningfully different from one another, this is nonetheless illustrative of the overall lack of harmonization among available CDEs. We urge NLM and NINDS to take the next step in its stewardship of the Repository and focus on how to better harmonize CDEs and make these data standards truly usable and reusable. Scaling and developing CDEs for domains of interest to other NIH institutes (e.g., National Institute of Mental Health) will be critical, as well.

In the table below, we outline additional comments and recommendations gleaned from our members in response to the RFI questions. Should you have any questions or require additional information, please contact Scott Weinberg at scott@amia.org or 240-479-2134. We thank NLM and NINDS for the opportunity to comment and look forward to continued dialogue.

Sincerely,

Patricia C. Dykes

Patricia C. Dykes, PhD, RN, FAAN, FACMI
AMIA President and Board Chair
Program Director Research, Center for Patient Safety, Research, and Practice
Brigham and Women's Hospital
Associate Professor of Medicine, Harvard Medical School

Topic Category	NIH Questions	AMIA Response
Current use of Common Data Elements in research projects. NIH seeks broad input on how researchers are currently using CDEs, regardless of whether research is related to COVID-19.	Explain how you identify, select, deploy, and encourage use of CDEs in your research.	AMIA members report that principal investigators (PIs) will, where possible, identify CDEs for researchers and recommend them. One method is to use the CDE repositories and look for elements for various research purposes. For example, one can use CDEs as building blocks for more complicated knowledge representation, or as metadata in data collection. Other times, funding will require specific CDE subsets to be used. Other researchers encourage data managers to build data collection forms based on a CDE. When this is not possible, they will work with a statistical analysis team to plan, a priori, how to rename, crosswalk, or derive CDE elements from the collected data. They will also encourage internal standardization when formal CDEs do not exist and when a data collection form appears re-usable.
	Describe the benefits to you of using CDEs in your research.	CDEs motivate researchers to think critically about data elements that are needed to answer their research questions. They additionally enable pooling of data across studies, or across sites when data are collected locally prior to aggregation for a multi-center trial.

	Identify existing incentives or resources that have facilitated your use of CDEs.	CDEs availability through repositories, including the NIH CDE Repository,[1] the NLM Value Set Authority Center,[2] and the NCI CDE Browser.[3] Research funding requirements are an additional incentive to use certain CDEs, as are data sharing requirements.
	Describe COVID-19 research activities you have engaged in and how you have used CDEs to collect data.	AMIA counts many members engaged with the NIH National COVID Cohort Collaborative (N3C), where any CDE must be mapped to the Observational Medical Outcomes Partnership (OMOP) CDM.
Challenges to adopting CDEs. NIH seeks broad input on barriers to adopting CDEs in the context of both research activities generally and COVID-19 research specifically.	Describe the major barriers you have encountered in using CDEs.	We note that variables needed for analysis are not data elements. However, the research community often equates the two. A variable presumes a context—meta data that evolve from whom the data were collected and how. The same CDE could mean different things in two different environments, while two CDEs could mean the same thing. At the same time, however, the tight linking of a CDE to specific data type (e.g., date) and context (e.g. PhenX toolkit; or a particular form or survey instrument) leads to confusion and proliferation of CDEs for the same basic

---

[1] https://cde.nlm.nih.gov/home

[2] https://vsac.nlm.nih.gov/context/cs

[3] https://cdebrowser.nci.nih.gov/cdebrowserClient/cdeBrowser.html#/search

		topic. The lack of a consolidated CDE repository only compounds this challenge. Multiple CDE browsers add barriers to searching for and adopting CDEs. For one CDE repository, too many options for one search term adds more time, cost, and ultimately burden to the selection process. The trade-off considerations between all-inclusive detailed descriptions for each element and a concise and specific tabular view for all options may sometimes be the most important factor in deciding the adoption and use of CDEs.
	Provide specific examples of challenges you have encountered in identifying and/or using CDEs in your research.	Additional challenges reported by AMIA members include: CDEs are often not linked with standardized vocabularies, which limits how CDEs are used and realized in research studies. For example, a CDE such as Urinary Creatinine (Id PXedUkcq8B) is not tagged with a LOINC code, so is not “fungible” with other data sources. Cross walking the multiple data requirements can also be challenging. For example, a funder might require a specific subset of NIH CDE, while FDA requires CDISC coding, and medical records or

		other source documents might use different coding schemes. As mentioned, the sheer number of CDEs for a given concept is a challenge, in that it impacts how to select between them and prioritize. In another example, typing “diabetes” into the NLM CDE repository, returns 209 results, while the NCI CDE browser returns 87 results. Although both repositories include relatively concise descriptions for each result, it is a very time-consuming process to go through the lists.
	Describe why you have or have not chosen to use CDEs in your research.	The difficulty in navigating the number of CDEs has led to some of our members choosing to forego using them in their research entirely. There is an additional pragmatic component, as well, in that CDEs can provide inconsistent coverage of some needed elements. Further, to get studies up and running, data managers may build data collection forms to fit with study workflows, and not necessarily to match with CDEs. Data managers and statisticians also may not be aware of or understand what CDEs are and why they are beneficial. Their main benefit – data sharing – may be perceived as less critical or a downstream activity when trying to get a study launched.

		As for choosing to use CDEs in research, they can facilitate research and promote a better understanding of novel phenomena and conditions, like COVID-19.[4]
CDE guidance, tools, infrastructure, and incentives. NIH seeks broad input on what guidance, tools, and incentives would facilitate wider uptake of CDEs.	Suggest specific guidance or information that would enhance your understanding and use of CDEs.	We suggest adding in metadata about how data were captured. Additionally, NIH should supply guidance on analysis, such as maximum missingness acceptable, data type (some ordinal data can be analyzed as interval), and any preferred methods for adopting CDEs. For novel conditions like SARS-CoV2, there needs to be better coordination and collaboration with multiple stakeholders (clinical, informatics, policy, payer, regulatory, provider) to evaluate the system of care and the data journey that results from care delivery to inform CDEs. The use-cases for CDEs are multiple and multiple stakeholders should be engaged to help determine them.
	Describe how the [NIH CDE Repository](#) could be enhanced to facilitate discovery and use of relevant CDEs.	The Repository should provide an integrated hierarchy (tree), with high-level categories like biomedicine and Social Determinants of Health and/or various clinical domains organized, for example, by various NIH institutes. It should also provide links to published articles whose research took advantage of a CDE. The

---

[4] Edlow BL, Boly M, Chou SH, Fischer D, Kondziella D, Li LM, Mac Donald CL, McNett M, Newcombe VFJ, Stevens RD, Menon DK; GCS-NeuroCOVID. Common Data Elements for COVID-19 Neuroimaging: A GCS-NeuroCOVID Proposal. Neurocrit Care. 2021 Feb 11:1–6. doi: 10.1007/s12028-021-01192-6. Epub ahead of print. PMID: 33575956; PMCID: PMC7878171.

		Repository should additionally implement a metadata model applying the FAIR principles that guarantee the traceability and reproducibility of the experiments described in the articles based on the data used. The model will also allow measuring the FAIR maturity level of the datasets according to the maturity model of the Research Data Alliance.[5] As for the Repository's search function, we suggest an advanced search function with more filters to aid in narrowing down search results. Search results should also be able to be viewed in different ways. In addition to the current view, there should be an option to view results based on usage frequency. Finally, the results should include a context note. This could be medical specialty, or a note indicating whether the CDE is a diagnostic test or diagnostic result.
	Describe resources that would enable use of CDEs, such as toolkits for including CDEs in studies or tools that support researchers' mapping existing data to CDEs.	Standards-based APIs are one such tool, while links to other repositories (LOINC, SNOMED, NCI Terms) would be even better. NIH should enable a library of CDEs within REDCap, so that CDEs come "out of the box" and investigators do not have to create (or recreate) existing variables. A small number of CDEs

---

[5] https://www.rd-alliance.org/group/fair-data-maturity-model-wg/outcomes/fair-data-maturity-model-specification-and-guidelines

		should be standardized for the same concept (e.g., race), making it easier (standard structure) to define variables from a CDE.
	Describe incentives that would encourage and increase use of CDEs.	Educational resources should be made be made available, in order to increase awareness of the importance of CDEs in improving the aggregation value and portability of researchers' data sets. We believe a concerted awareness campaign to educate various stakeholders involved in research will result in an uptake of CDE use. Such a campaign could include both infographic materials to promote CDE awareness broadly and in-depth tutorials to train researchers on effective use of CDEs. NIH should also make it easier to satisfy data sharing requirements if CDEs are used or recommend CDEs to be a part of data sharing plans for NIH research proposals. NIH should consider financial incentives for their use, as well. NIH can further launch a similar effort to the National Information Exchange Model (NIEM),[6] to enable the use of certain CDEs across other federal agencies and non-federal stakeholders. This can additionally allow CDE concepts more

---

[6] https://www.niem.gov/

		exposure to physicians, nurses, and public health professionals. Many stakeholder organizations make large research investments, such as in disease registries. If the request for proposal (or during the project implementation) can recommend using CDEs in the new data collection efforts, CDEs can be recognized and understood by a much broader audience. Finally, NIH should promote publications of studies that conduct methodologically sound research studies using CDEs. This will illustrate to research and decision-making stakeholders the value of CDE standards and resources, as they potentially provide trust in results from standards-driven data sources.
	Provide specific examples of questions you have about how to identify, select, deploy and use CDEs.	The lack of CDE hierarchy makes it difficult to get a broad view of what is available. Similarly, not having standard coding systems makes the CDEs "invisible."