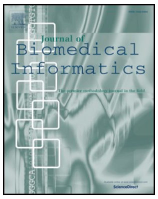
Contents lists available at ScienceDirect

Journal of Biomedical Informatics

journal homepage: www.elsevier.com/locate/yjbin



FAIR data sharing: The roles of common data elements and harmonization

R.D. Kush[a,*], D. Warzel[b], M.A. Kush[c], A. Sherman[d], E.A. Navarro[e], R. Fitzmartin[f], F. Pétavy[g], J. Galvez[h], L.B. Becnel[i], F.L. Zhou[j], N. Harmon[k], B. Jauregui[l], T. Jackson[m], L. Hudson[n]

[a] *Elligo Health Research and Catalysis, USA*
[b] *U.S. National Cancer Institute, USA*
[c] *University of Chicago, USA*
[d] *Massachusetts General Hospital, Harvard Medical School, USA*
[e] *Center for Drug Evaluation and Research, US Food and Drug Administration, USA*
[f] *Center for Biologics Evaluation and Research, US Food and Drug Administration, USA*
[g] *European Medicines Agency, Europe*
[h] *Office of Strategic Programs, CDER, US Food and Drug Administration, Formerly Biomedical Translational Research Informatics, National Institutes of Health Clinical Center, USA*
[i] *Pfizer, USA*
[j] *Sanofi, France*
[k] *Cohen Veterans Bioscience, USA*
[l] *CDISC and Fondation Mérieux USA*
[m] *PPD, Inc., USA*
[n] *Critical Path Institute, USA*

ABSTRACT

The value of robust and responsible data sharing in clinical research and healthcare is recognized by patients, patient advocacy groups, researchers, journal editors, and the healthcare industry globally. Privacy and security concerns acknowledged, the act of exchanging data (interoperability) along with its meaning (semantic interoperability) across studies and between partners has been difficult, if not elusive. For shared data to retain its value, a recommendation has been made to follow the Findable, Accessible, Interoperable, Reusable (FAIR) principles. Without applying appropriate data exchange standards with domain-relevant content standards and accessible rich metadata that uses applicable terminologies, interoperability is burdened by the need for transformation and/or mapping. These obstacles to interoperability limit the findability, accessibility and reusability of data, thus diminishing its value and making it impossible to adhere to FAIR principles.

One effort to standardize data collection has been through common data elements (CDEs). CDEs are data collection units comprising one or more questions together with a set of valid values. Some CDEs contain standardized terminology concepts that define the meaning of the data, and others include links to unique terminology concept identifiers and unique identifiers for each CDE; however, usually CDEs are defined for specific projects or collaborations and lack traceable or machine readable semantics. While the name implies that these are ‘common’, this has not necessarily been a requirement, and many CDEs have not been commonly used. The National Institutes of Health (NIH) CDEs are, in fact, a conglomerate of CDEs developed in silos by various NIH institutes. Therefore, CDEs have not brought the anticipated benefit to the industry through widescale interoperability, nor is there widespread reuse of CDEs. Certain institutes in the NIH recommend, albeit do not enforce, institute-specific preferred CDEs; however, at the NIH level a preponderance of choice and a lack of any overarching harmonization of CDEs or consistency in linking them to controlled terminology or common identifiers create confusion for researchers in their efforts to identify the best CDEs for their protocol. The problem of comparing data among studies is exacerbated when researchers select different CDEs for the same variable or data collection field. This manuscript explores reasons for the disappointingly low adoption of CDEs and the inability of CDEs or other clinical research standards to broadly solve the interoperability and data sharing problems. Recommendations are offered for rectifying this situation to enable responsible data sharing that will help in adherence to FAIR principles and the realization of Learning Health Systems for the sake of all of us as patients.

1. Introduction and background

Many research and healthcare organizations, including patients and patient advocacy groups, and government agencies, have recognized the value of data sharing. The U.S. National Academies of Science, Engineering and Medicine [1]; the CORBEL project [2] and Innovative Medicines Initiative (IMI) [3] in Europe; global patient advocacy groups such as OneMind [4]; and a new peer-reviewed Learning Health Systems Journal [5] are among those who have published on the benefits of data sharing, including support of open science and realization of Learning Health Systems (LHSs). Further, the International Committee of Medical Journal Editors (ICMJE) now requires data sharing of

[*] Corresponding author.
E-mail address: rkush@catalysisresearch.com (R.D. Kush).

https://doi.org/10.1016/j.jbi.2020.103421
Received 11 November 2019; Received in revised form 13 March 2020; Accepted 9 April 2020
Available online 12 May 2020

research results by authors of their publications [6], and the National Institutes of Health (NIH) has been strengthening its data sharing policies [7] with increasing interest in data standards. For shared data to retain its value, a recommendation has been made by FORCE11 [8] to follow the Findable, Accessible, Interoperable, Reusable (FAIR) principles. Adherence to these principles will more likely if data is standardized from the start. Defining how these principles may be implemented is the focus of a new IMI initiative called FAIRplus [9].

Without applying data exchange standards along with appropriate domain-relevant content standards and accessible rich metadata that uses applicable terminologies[1], interoperability is hindered by the need for manual transformation and/or mapping. These obstacles to interoperability limit the findability, accessibility and reusability of data, thus diminishing its value.

This interest in data sharing has resulted in a multitude of repositories, registries and methodologies that, while potentially useful on a site or study basis, severely limit interoperability and reusability; indeed, they risk creating further silos and may even compromise the meaning of data pooled across studies, especially when consistent data standards have not been implemented across the shared studies. Data that are shared in non-standard formats can take considerable time to understand and can lead to interpretation errors. Aggregation of datasets that are in different formats, either across studies or repositories or for inclusion in a data commons, requires mapping, which is not only extremely time-consuming and costly, but can also compromise data quality, integrity, completeness, and traceability. The IMI's European Translational Research Information & Knowledge Management Sharing (eTRIKS) Standards Starter Pack makes a business case for standards, stating that poor data comparability and reproducibility in the life sciences (when data standards are not used) wastes significant resources and impairs scientific research [10]. An article written on this subject by the first Executive Director of IMI and the Founder of CDISC states: *"The precise format of the data to be shared cannot be an afterthought. In an era of increased transparency and integrative analyses of data from multiple origins, data standards are essential to ensure accuracy, reproducibility, and scientific integrity. Their use will help in fostering innovation—and thereby in honoring the sacrifices of research participants everywhere."* [11]

Unfortunately, certain data sharing initiatives [12] have initially declared data standards "out of scope" or have avoided recommending specific standards and implementation methodologies, thus exacerbating the problem of competing standards and a lack of harmonization and consensus building that could encourage adoption of common global standards. The CORBEL Initiative, in its consensus-based project *"Sharing and reuse of individual participant data from clinical trials: principles and recommendations"*, made an effort to address this issue through its recommendations, one of which is: *"To promote interoperability and retain meaning within interpretation and analysis, shared data should, as far as possible, be structured, described and formatted using widely recognized data and metadata standards."* [13] A core recommendation of the EMA Big Data Report refers to data standards to *"promote use of global, harmonized and comprehensive standards to facilitate interoperability of data."* They go further to recommend especially the promotion of open source standards to aid adoption [14]. The U.S. FDA has also published an FDA Data Standards Strategy [15].

The case for standards to improve the quality, meaningful 'shareability' and reproducibility of research was recently made by the Scientific Advisory Committee from the Coalition for Accelerating Standards and Therapies [16]. This international Committee, which included representation from academia, government, regulatory agencies, biopharmaceutical companies, global standards development organizations, the Critical Path Institute, the Innovative Medicines Initiative, and the Pan American Health Organization / World Health Organization, recognized that meaningful data sharing and the return on investment of medical research requires the broad adoption of consensus-based, widely adopted global data standards and terminologies such that the data can be readily exchanged, interpreted, compared and aggregated across studies [17]. This is particularly important and relevant for clinical research since data collected from participating patients are of limited volume and therefore precious.

2. What is a CDE?

By consensus, the NIH defines CDEs as "discrete, clearly defined and reusable data collection units" [18]. This is fairly close to the ISO/IEC 11179-3 Metadata registry model and basic attributes (ISO/IEC 11179) standard which defines a data element as "a unit of data for which the definition, identification, representation and value domain are specified by means of a set of attributes." [19] The "common" in CDE was originally intended to convey the fact that it has been agreed to be used by more than one group; unfortunately, this no longer applies in practice.

In an effort to improve data sharing and 'FAIRness', repositories in the US have been developed to manage common data elements (CDEs). The NIH CDE Repository includes CDEs that are largely based on ISO/IEC 11179, are constituted by a number of attributes, and may include a question (e.g., "What is the patient's sex?), concept linkage, data type (e.g., integer, text or enumerated lists), unit of measure, and, for enumerated lists, a set of permissible values [20]. Case report form (CRF) questions or data fields are examples of implemented CDEs. The value of CDEs is described in detail by Sheehan et al. [21], who also recognized that the CDEs should be "linked to accepted data standards and terminologies". CDEs based on data standards can minimize data transformations, and CDEs linked to standard and controlled terminology can form the basis for machine readable semantics to aid in aligning similar CDEs across studies as well as exploring deeper meaning of the data, but NIH requires neither. In short, clinical and translational research CDEs have the promise of facilitating interoperability and reuse. However, the current CDEs are not necessarily standards, nor are they always 'common'; they sometimes represent how data were collected for just a single study or database. CDEs can deliver more value when they conform to accepted data standards, are bound to terminologies and are used consistently across studies; this practice would avoid much of the tedious transformation and/or mapping at the end of the study for broader sharing.

The NIH National Cancer Institute (NCI) has maintained the Cancer Data Standards Repository (caDSR) [22] since 2000 to manage CDEs for standardizing data collected in NCI funded clinical trials and from a number of oncology research teams, individual oncology investigators, cancer specific standards such as the American Joint Committee on Cancer (AJCC-cancerstaging.org) and the North American Association of Central Cancer Registries (naccr.org), as well as other NIH Institutes such as the National Institute of Neurological Disorders and Stroke (NINDS). Thus, the caDSR now contains over 67,000 CDEs that are far wider ranging than oncology-specific data.

Recognizing that harmonization across the NIH Institutes or Centers (IC) would be beneficial, in 2012 the NIH National Library of Medicine (NLM) proposed a broader scope for the aforementioned NIH CDE Repository [20]. The NLM CDE repository includes CDEs from the NCI caDSR, NINDS, PROMIS (Patient Reported Outcomes Measurement Information System), PhenX (an online catalog of standard measurement protocols for use in biomedical research), FITBIR (Federal Interagency Traumatic Brain Injury Research) Informatics System, and other sources including other NIH ICs. However, the CDEs from these various sources are currently not harmonized within the NIH CDE Repository; there are numerous redundancies, and most of the NIH ICs have not contributed CDEs to this Repository. For example, the Agency for

[1] **For the purpose of this paper, we intend "terminologies" to include medical terminologies and ontologies, the latter which we define as controlled terminology together with definitions, relational expressions, and other formal specifications.**

Healthcare Research and Quality has contributed 91 CDEs; National Eye Institute, 235 CDEs; National Institute of Nursing Research, 141 CDEs; National Institute on Drug Abuse, 121 CDEs; and, the National Institute of Neurological Disease and Stroke (NINDS) has contributed by far the most with 18,021 CDEs. In addition, many of the CDEs in this repository are 'incomplete' in that they do not include important metadata (such as units of measure) or linkage to terminologies for precise semantics. This becomes a major issue when research data using these CDEs is collected and aggregated into databases for tabulation and analysis; without the critical metadata, the resulting databases will be incomplete and difficult to interpret.

3. A plethora of CDEs

Despite the promise and promulgation of CDEs over the past two decades, most are essentially a local resource and are not suitable for wholesale adoption and global reuse. As a consequence, although thousands of CDEs from multiple sources already exist, and some large collaborations seek to harmonize with existing CDEs in order to aggregate data with existing repositories, individual investigators routinely create their own CDEs. There are a number of reasons that existing CDEs have not been broadly and uniformly adopted. These reasons include, but are not limited, to those identified in Table 1.

For such reasons a large number of diverse organizations and investigators create their own CDEs for specific studies, without an overarching consistent means or effort to harmonize, which means their data cannot be readily aggregated or reused. ISO/IEC 11179 standardizes CDE repositories, including the structure for recording the semantics to facilitate harmonization, provisions for versioning content, and tagging it with details about where it was used; unfortunately, few repositories have implemented ISO/IED 11170 with discipline, nor is there collaboration across repositories to harmonize at higher levels. The lack of appropriate management of CDE repositories to help identify standards versus study specific CDEs can also lead to CDE misuse.

In addition to reusing nationally or internationally standardized CDEs, adequate modeling of data is required to correctly interpret the data once collected and aggregated for secondary analysis. Study specific CDEs may not have been adequately constructed, completed, and tested for these purposes. The lack of appropriate data modeling can lead to the lack of important data and metadata in the database and thus limit its reuse without costly harmonization, curation and transformation (Fig. 1).

The examples below illustrate how data model designs can limit CDE reuse and/or how CDEs can be insufficiently robust:

- There are two elements in the NIH CDE Repository of "External Forms": "Birth control method at exit Reported –at exit" and "Birth control method at intake Reported –at intake". While these two elements are designed to be used on a form to collect data at different timepoints, and share a code list, the difference in timepoints is embedded in the question text, and thus important metadata (timing) for reuse does not appear in the database, as with the example in Fig. 1. Instead, there could be one element with the same variable name, referencing the same codelist, collected along with another variable to represent the timepoint: "Birth Control Method", and "Reporting Period" with choices "at exit", "at intake" to represent the timing. (The Clinical Data Interchange Standards Consortium variable name would be RPORRES_BCMETHOD). Two different names for two questions (vs. one question that does not include a timing variable) introduces unnecessary confusion and compromises the resulting database. Timing information (such as dates, time points, or visit information) should be collected in separate fields to facilitate a complete and more robust database for use in subsequent analysis.
- Another example from the NIH CDE repository is the element of "Body Height". Within the Properties section, there are PhenX Variables, with two variable names: PX150203_Height_Feet and PX150203_Height_Inches, where the unit value is included only within the CDE name, but not included as the unit of measure attribute for the CDE. The availability of these two CDEs allows for a study to be set up differently across sites (where the sites may use different units); but, including the units in the CDE name(s) and not as another attribute in the data model to capture units can lead to disastrous results. In contrast, another approach would be that of the Clinical Data Interchange Standards Consortium (CDISC), which has one variable for Height (VSORRES_HEIGHT) and one variable for Height unit (VSORRESU_HEIGHT), allowing more flexibility for collecting these variables in a consistent manner across studies and countries. The codelist used for these two variables could be Unified Code for Units of Measure (UCUM) [23].

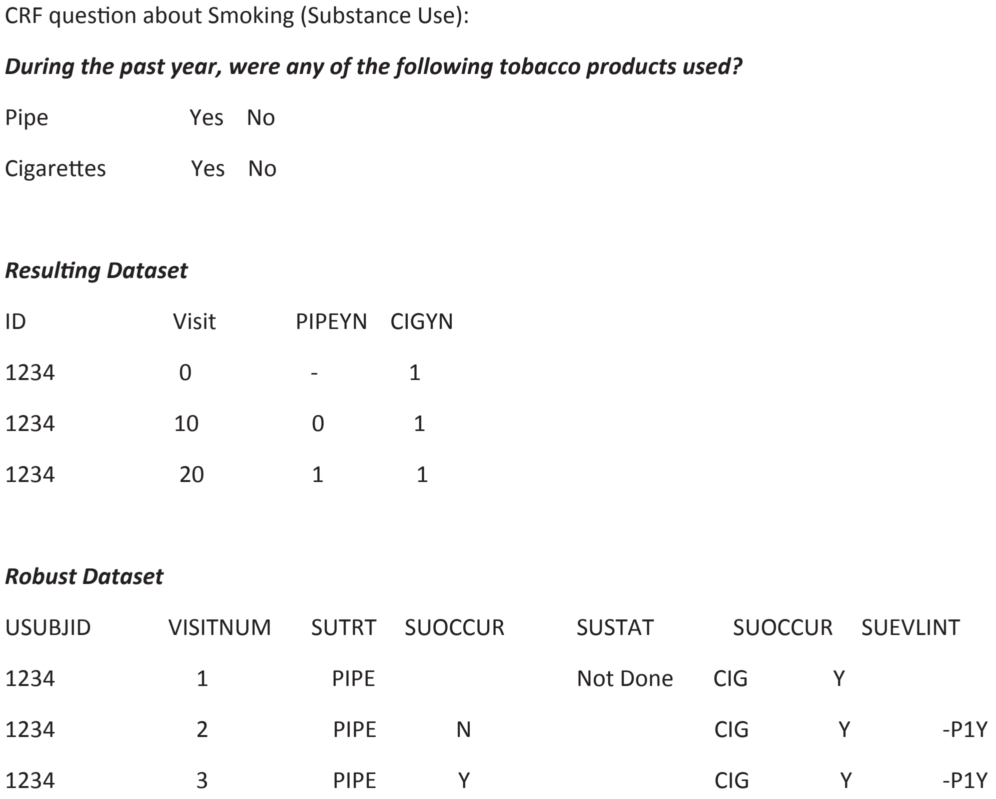
Table 1
Reasons for Lack of Broad Adoption of Existing CDEs.

Lack of awareness of existing CDEs
Perception of investigators that they need unique or better CDEs
Need for highly-specific concepts and valid values utilized in a given study
Inability to easily and rapidly find and deploy existing CDE
Availability of multiple CDEs that may be applicable without adequate context/ information for selecting the best one
Easier to create a new CDE than harmonizing existing CDEs
CDEs for the same or similar biomedical CDEs by different organizations using different codelists

In an arena outside of clinical research, the Mars Space Orbiter provides a great example of an actual disaster that occurred due to lacking or inadequate metadata. When two groups of scientists misinterpreted metadata – one used Imperial units while the other interpreted them as metric units –the Orbiter crashed, and the project wasted millions of dollars [24].

The NIH CDE Repository groups CDEs into collections by institute, center, or project, including a "TEST" section and the aforementioned "External Forms" section. The latter is related to demographics, vital signs and other administrative information as well as patient-reported outcome (PRO) elements. It also includes elements such as social security number and patient name, which are considered confidential information and are not to be published in repositories used for clinical research, thus raising concerns about the instructions related to the use of CDEs that are posted in this Repository. Again, there are no 'preferred' or recommended CDEs at the NIH level, nor is there guidance about the use of CDEs in this Repository. If a study is conducted that requires the inclusion of personal health information (PHI), the regulations for PHI (including HIPAA in the U.S. and GDPR in the EU) should be enforced and followed [25,26].

Contribution to the NIH CDE Repository is on a voluntary basis, IC by IC. Upon inspection at the time of this paper, there were 8 NIH Institute/Centers with registered CDEs (most of them from NINDS) and 19 with none. A preponderance of choice in the absence of recommended or preferred NIH CDEs that were modeled to optimize data sharing and reuse, and a lack of harmonization of CDEs across NIH Institutes are often confusing for researchers in their efforts to identify the best CDEs for their protocol among the identical or near-identical available and frequently promoted CDEs. Despite helpful information, such as a CDE 'ownership', CDE utilization history, and features to compare CDEs, the selection of a CDE in the NIH CDE Repository by



CRF question about Smoking (Substance Use):

During the past year, were any of the following tobacco products used?

Pipe	Yes	No
Cigarettes	Yes	No

Resulting Dataset

ID	Visit	PIPEYN	CIGYN
1234	0	-	1
1234	10	0	1
1234	20	1	1

Robust Dataset

USUBJID	VISITNUM	SUTRT	SUOCCUR	SUSTAT		SUOCCUR	SUEVLINT
1234	1	PIPE		Not Done	CIG	Y	
1234	2	PIPE	N		CIG	Y	-P1Y
1234	3	PIPE	Y		CIG	Y	-P1Y

Fig. 1. CRF question about Smoking (Substance Use): This example indicates the dangers of using a single CDE without knowledge of the study design for data collection. The Resulting Dataset above may not provide enough detail to interpret the data because the time period for the smoking question is embedded in the question prompt (“during the past year”). The standalone CDE does not guide the potential user about how to capture this detail. The Robust Dataset makes use of the CDISC standard which guides the user to collect information on the specific Substance used (PIPE or CIG) in addition to the period of time (SUEVLINT, 1 Year), which is embedded in the question and thus lost in the initial dataset. The CDISC codes also make clear to a human whether the response is yes or no (SUOCCUR), which may be lost with responses of 0 and 1 if the key is not available.

investigators or their data managers is still a significant challenge. One unfortunate consequence is that investigators may find the process too burdensome and cease attempts to implement existing NIH CDEs. With no impetus or incentive to utilize existing CDEs, investigators might customize them. Such a one-off change or adjustment to a CDE question actually results in the generation of a new CDE because the data values represent different semantics and thus cannot easily be compared; this results in the generation of new data elements that are ‘common’ only to that investigator’s project. This lack of governance also does not encourage or enable investigators to take advantage of the experiences and best practices of others in CDE modeling for data sharing.

Although CDEs are a step in the direction of ‘FAIRness or FAIR Data Sharing’, the lack of standardization and governance to harmonize and elevate specific/preferred CDEs to standards, limited engagement in the research community, inconsistent enforcement efforts, difficulty in implementation and use, and issues with harmonization or curation remain barriers and deterrents to broader adoption. Sheehan et al. states that “currently there are no formal international specifications governing their construction or use.” The ISO/IEC 11179 standard does address CDE construction so that semantics can be more easily compared and provides guidance for naming, defining and governing CDEs. For example, a registrar is supposed to ensure that there is no overlap or redundancy in their repository. However, it is clear that in the existing CDE repositories, including the NIH CDE Repository, conformance to this standard can be improved to help address many of the issues identified in Table 1. For examples regarding the issue of “Availability of multiple CDEs that may be applicable without adequate context/information for selecting best one” see Table 2**,** For examples related to the issue “CDEs for the same or similar biomedical CDEs by different organizations using different codelists” see Table 3. Indeed, the Sheehan el al. manuscript concludes that the responsible development, use and promotion of CDEs should be supported by bindings to mature data standards and controlled terminologies (codelists), engagement with and feedback from the research community on CDEs, and encouragement and enforcement from regulatory authorities [21].

Table 2
Incomplete CDEs.

Concept = Patient Age at Diagnosis CDE	Prompt	Response Type
CDE A	Is the patient greater than or equal to 8 and less than or equal to 30 at the time of diagnosis	Yes/No
CDE B	Patient Age at Diagnosis (years)	Years
CDE C	Patient age at Diagnosis (Months)	Months
CDE D	Is the patient between the ages of 18 and 65?	Yes/No

These are four CDEs that all refer to Patient Age at Diagnosis, yet the meaning of each is not complete without the context of the study for which it is being used. Hence, the data in the database may also be meaningless. To ensure that a dataset contains meaningful information, the appropriate metadata must be included with the data and kept in the database.

4. Beyond CDEs

The research and healthcare communities have been working on standards, including semantics, for several decades, although they have not yet built consensus around a common shared set. Since early 2000, the NCI caDSR and Enterprise Vocabulary Services (EVS) group [27] have been collaborating with and supporting the development of controlled terminology for the Clinical Data Interchange Standards Consortium (CDISC) [28], a global standards development organization (SDO).

CDISC has created a robust suite of data standards for clinical and translational research. These standards have been vetted globally following an ISO-recognized process; they are unique/non-redundant and harmonized among one another to ensure integrity and adequate metadata, from protocol development and data collection through aggregation into tables and analysis for statistical results. The highly curated CDISC data standards yield standard data representations at each stage in the clinical research lifecycle. The CDISC study data tabulation model (SDTM) and analysis dataset model (ADaM) standards are now required by the FDA and Japan’s Pharmaceuticals and Medical Devices

Table 3
Twenty-three CDEs from different organizations for similar biomedical concepts using non-standard codelists create confusion about which CDEs to use, establishing barriers to data sharing and aggregation. CDISC and HL7 are also not aligned with respect to this data element.

Name	Permissible Values	Steward
1. Gender [HL7v3.0]	Female; Male; Undifferentiated	NLM
2. Gender Code	0; 1; 2; 9	NCI
3. Sibling gender type	Female; Male; Unknown; Unspecified; Not reported	NINDS
4. Gender type	Female; Male; Unknown; Unspecified; Not reported	NINDS
5. Parent gender type	Female; Male; Unknown; Unspecified; Not reported	NINDS
6. Family history gender type	Female; Male	NINDS
7. Person Gender Other Specify		NCI
8. Person Biological Entity Or Sex Gender Code OMOP CDM Gender Concept Identifier	M; O; F; U; A	NCI
9. Person Biological Entity Or Sex Gender Code OMOP CDM Provider Gender Concept Identifier		NCI
10. Person Sample/Specimen Gender Text Type	Female; Female-to-male transsexual; Intersexed; Male; Male-to-female transsexual; Not Specified; Other, specify; Pooled; Unknown	NCI
11. Child Behavior Checklist (CBLC) - Respondent gender category	Female; Male	NINDS
12. Quality of Life in Swallowing Disorders (SWAL-QOL) - gender type	1; 2	NINDS
13. Quality of Life in Swallowing Disorders (SWAL-QOL) - gender indicator	1; 2	NINDS
14. Person Biological Entity Or Sex Gender Code PCORnet CDM Sex Code	M; F; A; NI; UN; OT	NCI
15. Family History Research Diagnostic Criteria (FH-RDC) - Gender type code	1; 2	NINDS
16. Person Biological Entity Or Sex Gender Code Sentinel CDM Sex Code	A; M; F; U	NCI
17. Person Biological Entity Or Sex Gender Code ACT I2B2 CDM Sex Code	A; F; M; NI; O	NCI
18. Person Biological Entity Or Sex Gender Code PCORnet CDM Provider Sex Code	A; F; M; NI; OT; UN	NCI
19. Family History Research Diagnostic Criteria (FH-RDC) - Shared child gender type	Male; Female	NINDS
20. Sex	1; 2; 3	NICHD
21. Sex [AHRQ]	a; b; c	NLM
22. Sex of relative	1; 2	LOINC
23. Sex	1; 2	LOINC

Agency (PMDA), endorsed by the National Medical Products Administration (NMPA) and acknowledged by the European Medicines Agency (EMA).[2] An OPTIMAL framework has been proposed to address OPerational, TechnIcal, and MethodologicAL challenges in both designing, running, and assessing a study to enhance the quality of evidence generated and the consistency of regulatory decision making [29]. Standardizing and validating data retrospectively is expensive, time consuming, and potentially introduces errors and biases, hence it is important to consider in advance the scope, depth, and quality of data that will be required to generate reliable evidence suitable, especially for regulatory use cases.

More recently, the CDISC Glossary Group, which was initiated in 2002, has worked with NCI EVS to create and link the CDISC Glossary [30] terms with controlled terminology concepts. Controlled terminology by itself does not sufficiently standardize data. Data managers and study designers need to understand the variables and their meaning, along with the controlled terminology specified for data collection. To help meet this goal, CDISC worked with NCI to create over 1500 CDEs based on the EVS controlled terminology; these are hosted by NCI to simplify study design and data collection prospectively. Following ISO/IEC 11179, this controlled terminology annotates the semantics of the CDISC CDEs making them comparable and mappable to other CDEs that use the same terminology.

To facilitate digital exchange, each unique term in the Glossary has a single definition and concept code (C-code) [31]; and synonyms, or similar terms, are noted. As in the ISO/IEC 11179–6 Registration standard, CDISC acts as a Steward and manages the Glossary seeking to eliminate semantic confusion in human information exchange and disambiguate the meaning of terms, acronyms, abbreviations, and initials used in the various foundational standards developed by CDISC for clinical research. The CDISC Glossary also serves as an educational resource for the clinical research community by defining relevant terms related to global clinical research. The Glossary term definitions and controlled terminology are curated and supported through the NIH NCI EVS, and used to create CDEs, providing robust metadata and semantics to ensure that the meaning of each data element is shared along with the data; this helps to minimize mapping and maximize integrity, quality, and meaning when data are shared.

In the healthcare arena, Health Level Seven (HL7) [32] is now offering a standard called, Fast Healthcare Interoperable Resources (FHIR), which is designed to improve interoperability among electronic health records [33]. There are also ISO standards related to EHRs that include research as a key principle [34]. NIH has begun to investigate how to use FHIR for research, issuing a Request for Information: “Use of the Health Level Seven International (HL7®) Fast Healthcare Interoperability Resources (FHIR®) for Capturing and Sharing Clinical Data for Research Purposes Notice [35]. With recent encouragement from the U.S. Congress through the 21st Century Cures Act and FDA to better leverage Real-World Data (RWD) [36,37], there has been progress in developing FHIR ‘Research Study’ and ‘Research Subject’ resources such that FHIR could eventually be able to support clinical research needs. Unfortunately, gaps and inconsistencies remain between research and healthcare standards, even in the most basic areas such as demographics, and there are still concerns relating to the quality of healthcare data for use in regulated research. In addition, while providing optimal flexibility, FHIR supports deviations from the primary resources it defines through community defined “Profiles”. These profiles, which are based on the base FHIR resources, can modify the base resources, including adding data elements, defining profile specific codelists for data elements, even redefining datatypes, creating a situation in which even FHIR adopters must perform mapping and transformation to make data interoperable between implementations using

[2] EMA does not require that raw data be included in regulatory submissions for new product approval, yet European data is challenged by a multitude of different standards, terminologies, structure and mechanisms of access.

Table 4
The CDISC/CDASH Demographics Domain Variables as represented by NCI CDEs. Many of these CDEs are reused across CDASH Domains facilitating interoperability. The codelists could be used by other CDEs for the same or similar fields to help standardize how data is collected.

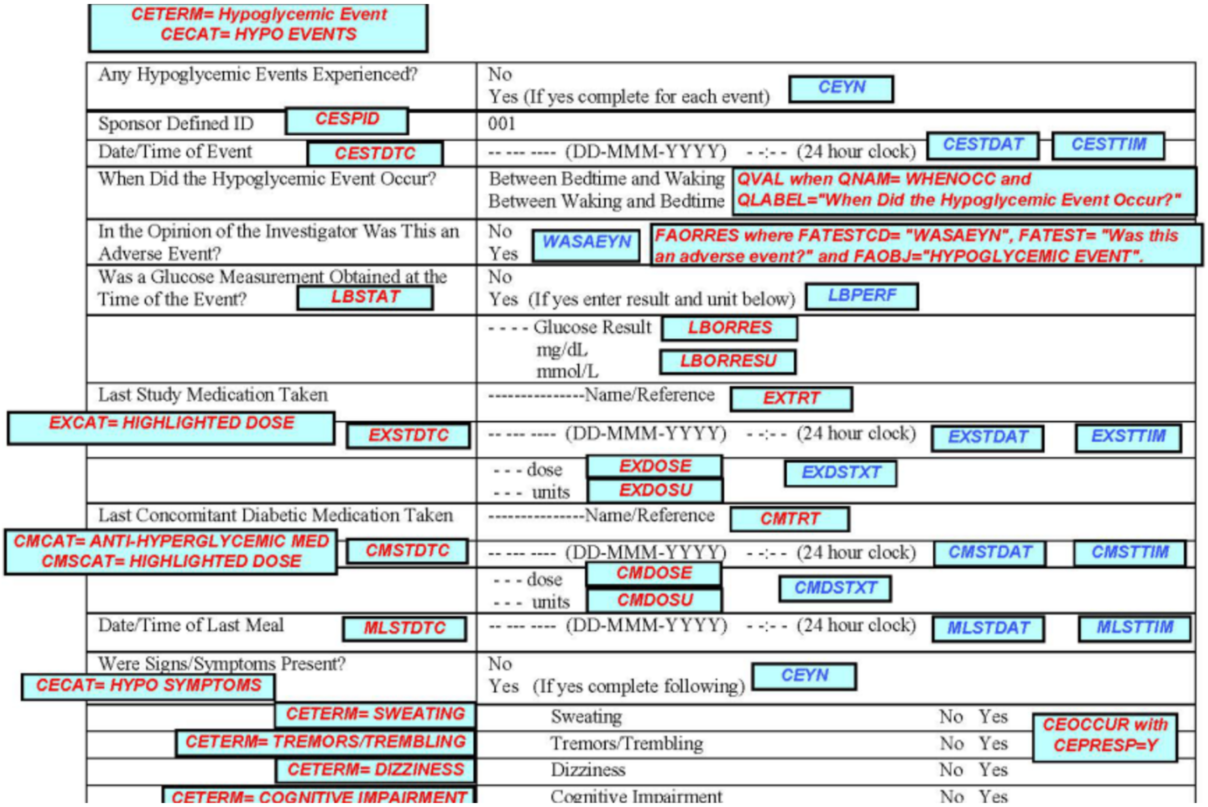
CDASH Variable Name	Long Name	Preferred Question Text	Identifier and Version	Valid Values	Value NCI Concepts
AGE	Age	What is the subject's age?	6412753v1	NUMBER	
AGEU	Age Units	What is the age unit used?		WEEKS	C29844
				MONTHS	C29846
				YEARS	C29848
				DAYS	C25301
				HOURS	C25529
BRTHDAT	Date of Birth	What is the subject's date of birth?	6341138v1	DATE	
BRTHDD	Day of Birth	What is the subject's day of birth?		CHARACTER	
BRTHMO	Month of Birth	What is the subject's month of birth?	6412736v1	CHARACTER	
BRTHTIM	Time of Birth	What is the subject's time of birth?	6409556v1	CHARACTER	
BRTHYY	Year of Birth	What is the subject's year of birth?	6412768v1	CHARACTER	
CETHNIC	Collected Ethnicity	What is the ethnicity of the subject?		ASHKENAZI JEW	C17950
				CENTRAL AMERICAN	C67118
				CUBAN	C107608
				…	…
CRACE	Collected Race	Which of the following racial designations best describes you? (More than one choice is acceptable.)	6412503v1	BHUTANESE	C43673
				BARBADIAN	C43823
				BANGLADESHI	C43671
				BAHAMIAN	C67271
				…	
DMDAT	Date of Demographics Collection	What is the date of collection?		DATE	
ETHNIC	Ethnicity	Do you consider yourself Hispanic/Latino or not Hispanic/Latino?		NOT HISPANIC OR LATINO	C41222
				UNKNOWN	C17998
				NOT REPORTED	C43234
				HISPANIC OR LATINO	C17459
RACE	Race	Which of the following five racial designations best describes you? (More than one choice is acceptable.)	6343384v1	BLACK OR AFRICAN AMERICAN	C16352
				NATIVE HAWAIIAN OR OTHER PACIFIC ISLANDER	C41219
				ASIAN	C41260
				…	
RACEOTH	Race Other	What was the other race?		CHARACTER	
SEX	Sex	What is the sex of the subject?	6343385v1	UNDIFFERENTIATED	C45908
				F	C16576
				M	C20197
				U	C17998
SITEID	Study Site Identifier	What is the site identifier?	6380048v1	CHARACTER	
SUBJID	Subject Identifier for the Study	What is the subject identifier?	6380049v1	CHARACTER	
STUDYID	Study Identifier	What is the study identifier?	6380045v1	CHARACTER	

different FHIR profiles. As of this writing there were over 100 such FHIR profiles and extensions [38]. Use cases are needed to demonstrate the potential of FHIR resources as a solution to data sharing and bi-directional data integration between healthcare and research systems.

One harmonization goal is to ensure that the FHIR resources are aligned with the Biomedical Research Integrated Domain Group (BRIDG) model [39,40], which was collaboratively developed over the past 15 years by FDA, NCI, CDISC and HL7 and is now a standard vetted and approved through three SDOs (CDISC, HL7 and ISO). The scope of the BRIDG model is protocol-driven research, and it has been made increasingly robust over the years by incorporating genomics modeling and additional domains that are important to clinical and translational research. With the BRIDG model as a central shared model for research and its link to healthcare, and FHIR as a central set of resources for healthcare data, there is an opportunity to improve semantic interoperability when exchanging healthcare and research data, especially when the BRIDG model is used in conjunction with harmonized terminologies at the points of overlap.

These standards developed by SDOs and others, such as the Integrating the Healthcare Enterprise Retrieve Form for Data Capture (RFD) [41], which was developed to support a use case for pulling data from an electronic health record for use in secondary systems, must currently rely on individual mapping exercises due to the disparate implementations of EHRs, similar to the problems reflected by the numerous FHIR Profiles. However, efforts are being made to leverage FHIR and identify which data in EHRs can be mapped to a research standard such that the resulting data collected is in a standard format that can be readily pooled and analyzed (e.g. mapping EHR data into CDISC CDASH format, a standard used for data acquisition) [42]. Table 4 shows a partial list of CDEs in the CDISC CDASH Demographics domain. Studies can then reference the CDASH variables in case report forms to ensure interoperability. Fig. 2 shows a case report form for a diabetes study with annotations for CDISC CDASH in blue and CDISC SDTM in red, indicating how collecting by-patient data in CDASH format enables the production of SDTM tables when the study is over and the patient-level data are aggregated for review and analysis purposes.

Not only are biopharmaceutical companies submitting data to regulators in CDISC format, but some academic organizations are also implementing the CDISC standards. For example, the Academic Research Organization (ARO) Council, which was initiated in Japan, has now extended into other Asian countries, Europe and the U.S.



CETERM= Hypoglycemic Event
CECAT= HYPO EVENTS

Any Hypoglycemic Events Experienced?	No Yes (If yes complete for each event) CEYN
Sponsor Defined ID CESPID	001
Date/Time of Event CESTDTC	-- --- ---- (DD-MMM-YYYY) - -:- - (24 hour clock) CESTDAT CESTTIM
When Did the Hypoglycemic Event Occur?	Between Bedtime and Waking Between Waking and Bedtime QVAL when QNAM= WHENOCC and QLABEL="When Did the Hypoglycemic Event Occur?"
In the Opinion of the Investigator Was This an Adverse Event?	No Yes WASAEYN FAORRES where FATESTCD= "WASAEYN", FATEST= "Was this an adverse event?" and FAOBJ="HYPOGLYCEMIC EVENT".
Was a Glucose Measurement Obtained at the Time of the Event? LBSTAT	No Yes (If yes enter result and unit below) LBPERF
	- - - - Glucose Result LBORRES mg/dL mmol/L LBORRESU
Last Study Medication Taken	---------------Name/Reference EXTRT
EXCAT= HIGHLIGHTED DOSE EXSTDTC	-- --- ---- (DD-MMM-YYYY) - -:- - (24 hour clock) EXSTDAT EXSTTIM
	- - - dose EXDOSE - - - units EXDOSU EXDSTXT
Last Concomitant Diabetic Medication Taken	---------------Name/Reference CMTRT

CMCAT= ANTI-HYPERGLYCEMIC MED
CMSCAT= HIGHLIGHTED DOSE CMSTDTC
	-- --- ---- (DD-MMM-YYYY) - -:- - (24 hour clock) CMSTDAT CMSTTIM

	- - - dose CMDOSE - - - units CMDOSU CMDSTXT
Date/Time of Last Meal MLSTDTC	-- --- ---- (DD-MMM-YYYY) - -:- - (24 hour clock) MLSTDAT MLSTTIM
Were Signs/Symptoms Present? CECAT= HYPO SYMPTOMS	No Yes (If yes complete following) CEYN
CETERM= SWEATING	Sweating No Yes CEOCCUR with CEPRESP=Y
CETERM= TREMORS/TREMBLING	Tremors/Trembling No Yes
CETERM= DIZZINESS	Dizziness No Yes
CETERM= COGNITIVE IMPAIRMENT	Cognitive Impairment No Yes

Fig. 2. Example of CRF for Diabetes Study Annotated using CDISC Standards. Each element on the individual patient CRF that can be annotated with CDASH (CDISC data acquisition standard) is shown in red above, and the SDTM (the CDISC study data tabulation model) annotations are shown in blue. The use of these standards will result in a database (created from aggregated case report form data) that will be meaningful and robust (containing all important metadata) and support the generation downstream of tables and statistical analyses for reporting.

through the Global ARO Network [43,44]. Key goals for this organization are harmonization and standardization, and they have adopted and been educated on CDISC standards. Unfortunately, most NIH ICs have not adopted similar standards-centric approaches in their CDE development efforts. And, while there have been efforts to access data from EHRs for certain clinical research networks, including Sentinel [45], i2b2 [46,47], OHDSI [48] and PCORNet [49], these research networks that often involve academic institutions have ended up with four different data models to serve each of their purposes, increasing the burden for academics to participate in multiple networks.

A recent project funded through the PCOR Trust Fund and funding from the 21st Century Cures Act has focused on harmonizing these data models and mapping them to the BRIDG Model [50,51]. The specific goals for this Common Data Model Harmonization (CDMH) project were to explore ways to make it easier for institutions using disparate CDMs to provide real world data (RWD) that FDA could use to augment the knowledge gained from traditional randomized clinical trials to enhance decision-making. Useful deliverables from the CDMH project (Phase 1) are the BRIDG mappings and the harmonized terminology across all of these models, including partial mappings to FHIR resources, which are now accessible through a publicly available website [52]. See Table 5 for a list of derived FHIR / BRIDG mappings created by NCI using the FHIR CDMH Implementation Guide. These products can facilitate the harmonization that is critical between healthcare and research, particularly with respect to RWD. In May 2019, FDA issued draft guidance on submitting RWD and Real World Evidence (RWE) to FDA for drugs and biologics [53].

The IMI's eTRIKS initiative recommends CDISC standards for both non-regulated and regulated research [10]. Moreover, to complement the CDISC foundational standards, CDISC standards for specific therapeutic areas and specific purposes have been developed over the past 15 years; these therapeutic-area specific standards augmentations have typically been developed in collaboration with other organizations, including the Critical Path Institute [54], Cohen Veterans Bioscience [55], World Wide Antimalarial Network (WWARN) [56], and Danone yogurt [57]. The CDISC standards are now electronically available through the CDISC Library (formerly known as the CDISC Shared

Table 5
This list of derived FHIR / BRIDG mappings was created by NCI using the FHIR CDMH IG (http://hl7.org/fhir/us/cdmh/2019May/profiles.html) and the CDM data elements that were mapped to BRIDG and registered in the NCI caDSR. It illustrates that the use of a high-level conceptual model may help facilitate harmonization between research and healthcare.

FHIR Resource	FHIR Element	BRIDG Identifier	BRIDG Class	BRIDG Attr
AdverseEvent	date	cadsr:3759985v1	AdverseEvent	occurrenceDateRange
Observation	bodySite	cadsr:3174965v1	PerformedObservation	bodyPositionCode
Specimen	collection.collectedDateTime	cadsr:3760208v1	PerformedSpecimenCollection	dateRange
Patient	gender	cadsr:3175307v1	Person	administrativeGenderCode
Patient	extension: us-core-race	cadsr:2868138v1	Person	raceCode
Procedure	performedDateTime	cadsr:3760199	PerformedProcedure	dateRange

Health and Research Electronic Library or SHARE) [58]. Many of them are cited in the FDA's Data Standards Catalog [59].

There are numerous examples of how academia, regulators and industry (research and health care) have successfully collaborated with the 'common good' in mind. These include, but are not limited to, the Innovative Medicines Initiative, the CORBEL project, the BRIDG model, and certain TransCelerate initiatives, including the common protocol template (CPT) [60], which is now known as Clinical Content & Reuse (CC&R) and is forming the basis of a guideline of the International Council for Harmonization (ICH) [61].

5. Recommendations to enhance data sharing and responsible use of standards

To overcome the barriers to interoperability and responsible data sharing in research and healthcare, which continue to increase costs and hinder availability of information throughout the healthcare system for all stakeholders, appropriate use and adoption of robust data standards and appropriate terminologies is critical. A concerted collaborative global effort is essential.

These recommendations are designed to expand the use of healthcare and research data, assist researchers in designing studies that follow current best practices, facilitate data aggregation and reuse, and enhance FAIR data sharing to advance the goals of a learning healthcare system (LHS) [62].

5.1. Establish a global infrastructure that encourages the acceptance, adoption and re-use of harmonized and preferred CDEs and global data standards

Substantial efforts on the part of researchers, investigator teams, NIH, SDOs and other groups are currently necessary to create, manage, and maintain CDEs and data standards. New approaches are needed to make CDEs more robust, harmonized, accessible, and understandable by the average investigator, to realize the full potential of these CDEs in facilitating FAIR data sharing. The technology required to achieve this goal is available, and the continued integration of existing global standards into the CDE management process can help to encourage the acceptance, adoption and re-use of harmonized, preferred CDEs and global data standards.

Basu et al. [63] describe a national cancer harmonization infrastructure comprised of specific components:

a) common data dictionaries hosting standard information models and CDEs, including CDISC, FHIR and other standards that link to harmonized terminology;
b) BRIDG as a higher level organizing conceptual model to link disparate models and their CDEs and mappings between them;
c) tools that semi-automate matching new variables to these standards to support data harmonization and transformation; and
d) access to Subject Matter Experts (SMEs) to assist with alignment and harmonization activities.

The "Common Data Model Harmonization (CDMH) and Open Standards for Evidence Generation" [51] (CDMH) project in which four common data models, OHDSI/OMOP, i2b2ACT , Sentinel and PCORNet, were manually harmonized and annotated with BRIDG terminology demonstrates the feasibility and power of the approach. Evidence that common terminology annotations of reusable ISO/IEC 11179 structured CDEs to support data harmonization is available through a visualization tool developed by NCI and FDA [52]. The visualization exposes the how CDE annotations provide semantic linkages between data represented using different models and codelists. In addition, new projects are demonstrating that FHIR resources can be developed for research, such as Phenopackets on FHIR [64].

5.2. Remove political and social barriers to data sharing

Re-aligning incentives and funding opportunities, organizations stepping forward as ISO Registrars, Submitting and Steward Organizations as described in ISO/IEC 11179–6 Registration [65]and working together to help align disparate standards within a common framework will address key challenges at this point to help remove political and social barriers by encouraging research centers globally to work together to build and deploy infrastructure to support harmonization. The HMA/EMA Joint Big Data Taskforce summary report provides certain incentives for sharing, resources for transforming and anonymizing the data, clear and robust data governance and other recommendations [66]. This means transitioning existing siloed, IC-specific efforts towards collaborative approaches, working synergistically with global standards organizations and research organizations, including patient advocacy groups and other parties with varied expertise that can facilitate progress. One of the goals of the NIH Strategic Plan for Data Science is directed at that challenge: "With community input, develop, promote—and refine as needed—data standards, including standardized data vocabularies and ontologies, applicable to a broad range of fields." [67].

5.3. Build better "Bridges" between research and healthcare

There are ongoing efforts to more closely bridge research with healthcare semantics through initiatives such as: a) the BRIDG model; b) the System for Accelerating Research (SOAR) [68] and the Learning Health Community [69]; c) efforts to share computable biomedical knowledge (MCBK) [70] and the Learning Health Systems Journal [5]; d) achieving semantic interoperability of all structured healthcare information through initiatives such as the Yosemite Project [71], which could be a component of a global infrastructure such as that described by Basu et al. [63]; e) developing of FHIR Resources for Research and leveraging existing therapeutic area standards; f) the EMA's report and recommendations relating to Big Data; g) the FDA's RWE program for developers interested in using RWD to develop RWE to support agency regulatory decisions. As part of the RWE program, the FDA will work on identifying relevant standards and methodologies for collection and analysis of RWD. The aforementioned CDMH project is in its Phase 2 and, with its goal to leverage FHIR, provides another example of activities to build bridges between research and healthcare. In addition, the NIH/NCI and NIH/NCATS have awarded significant grants to develop standardized approaches to address barriers to data sharing, to enable data harmonization and to translate the products of this work to accelerating translation and providing better health. These efforts will no doubt also be beneficial to changing the culture, educating researchers and encouraging and facilitating the adoption or harmonization with standards (including robust, harmonized, and tested CDEs) where the metadata can be used 'off the shelf' across many different organizations. Such positive and collaborative efforts should help remove political barriers, in addition to building bridges between healthcare and research.

6. Conclusion

To resolve overarching issues related to data sharing, CDEs can play a valuable role, especially if they are harmonized, 'preferred' CDEs that are bound to appropriate controlled terminology and common definitions and implemented synergistically with global data standards through a global infrastructure comprised of technology, processes and expert support. However, to achieve this, political barriers must be removed, incentives aligned, and broad collaboration across multiple types of organizations must build bridges between research and healthcare. Current CDE governance can seize the opportunity to broaden CDE impact beyond niche implementation by engaging the broader research and healthcare communities. The collaborators must include SDOs, government agencies, EHR and software vendors, organizations that create and maintain terminologies, and organizations representing both research and healthcare. Achieving recommendations identified in this paper will be essential to efficiently and responsibly share meaningful data that can ultimately evolve into Learning Health Systems through which research more rapidly informs care decisions for the benefit of all of us as patients.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This collaboration was funded in part by 1U01FD005855 awarded to the Critical Path Institute to develop therapeutic area data standards through the Coalition for Accelerating Standards and Therapies. The authors would like to express their sincere gratitude to others who have reviewed and contributed valuable additions and comments to improve this manuscript, including Helle Gawrylewski, Art Gertel and Steve Raymond of the CDISC Glossary Group; Dr. Yuki Ando (PMDA), Prof. Masanori Fukushima (FBRI/TRI and Global ARO Network), Alison Cave (Innovate UK), Nick Halsey (EMA). The efforts of those who collaborated with NCI and Elligo on the PCORTF Common Data Model Harmonization project, especially Mitra Rocca (FDA) and Ken Gersing (NCATS) and including Lisa Lang (NLM), and Albert Taylor (ONC) constitute valuable and significant contributions to the entire field. The expertise of Nathalie Seigneuret (Innovative Medicines Initiative) during her tenure on the CFAST SAC has been greatly appreciated.

Disclaimer

The views expressed in this article are the personal views of the authors and may not be understood or quoted as being made on behalf of or reflecting the position of the US Food and Drug Administration, the position of the National Cancer Institute, the position of Fondation Mérieux USA, the position of the European Medicines Agency or one of its committees or working parties, or the position of the Innovative Medicines Initiative (IMI) nor the European Union, EFPIA, or any Associated Partners.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jbi.2020.103421.

References

[1] National Academies of Sciences, Medicine and Engineering https://nationalacademies.org/.
[2] CORBEL - Coordinated Research Infrastructures Building Enduring Life-science Services. elixir. https://www.elixir-europe.org/about/eu-projects/corbel.
[3] IMI - Innovative Medicines Initiative – https://www.imi.europa.eu.
[4] OneMind https://onemind.org/.
[5] Learning Health Systems Journal (LHS Journal)- https://onlinelibrary.wiley.com/journal/23796146.
[6] "Data Sharing Statements for Clinical Trials: A Requirement of the International Committee of Medical Journal Editors". s.l. : Annals of Internal Medicine, 2016. doi:10.7326/M17-1028.
[7] NIH Data Sharing Policy https://grants.nih.gov/grants/policy/data_sharing/.
[8] FAIR Data Principles https://www.force11.org/group/fairgroup/fairprinciples.
[9] FAIRplus Project https://fairplus-project.eu/.
[10] Innovative Medicines Initiative, eTRIKS Standards Starter Pack https://www.etriks.org/standards-starter-pack/.
[11] "Fostering Responsible Data Shring through Standards". Kush, Rebecca D and Goldman, Michel. 5 June 2014, New England Journal of Medicine, pp. pages 2163-2164.
[12] Sharing Clinical Trial Data: Maximizing Benefits, Minimizing Risks, Committee on Strategies for Responsible Sharing of Clinical Trial Data, Institute of Medicine, National Academies Press, 2015.
[13] " Sharing and reuse of individual participant data from clinical trials: principles and recommendations", Ohmann, C., Banzi, R., Canham, S., Battaglia, S., Matei, M., Ariyo, D., Becnel, L., Bierer, B., Bowers, S., Clivio, L., Dias, M., Druml, C., Faure, H., Fenner, M., Galvez, J., Gheris, D., Gluud, C., Groves, T., Houston, P., Karam, G., Kalra, D., Knowles, R., Kreleza-Jeric, K., Kubiak, D., Kushinke, W., Kush, R., Lukkarinen, A., Marques, P.S., Newbigging, A., O'Callaghan, J., Ravaud, P., Schulunder, I., Shanahan, D., Sitter, H., Spalding, D., Tudur-Smith, C., van Reusel, P., van Veen, E., Visser, G.R., Wilson, J., Demotes-Mainard, J., British Medical Journal Open, 2017:7:e018647, doi: 10.1126/bmjopen-2017-018647.
[14] EMA Joint Task Force on Big Data [Online] Summary Report https://www.ema.europa.eu/en/documents/minutes/hma/ema-joint-task-force-big-data-summary-report_en.pdf.
[15] FDA Data Standards Strategy FY2018-FY2022 (CDER and CBER) https://www.fda.gov/media/110928/download.
[16] Coalition for Accelerating Standards and Therapies (CFAST) https://c-path.org/programs/cfast/.
[17] "Global Standards to Expedite Learning From Medical Research Data". Hudson, Lynn D., Kush, Rebecca D., Navarro Almario, Eileen, Seigneuret, Nathalie, Jackson, Tammy, Jauregui, Barbara, Jordan, David, Fitzmartin, Ronald, Zhou, F. Liz, Malone, James K., Galvez, Jose, Becnel, Lauren B., Clin. Transl. Sci. (2018) 11, 342–344; doi:10.1111/cts.12556. 17b)The Turning Point for Clinical Research: Global Data Standardization", Jauregui, Barbara, Hudson, Lynn D., Becnel, Lauren B, Navarro Almario, Eileen, Fitzmartin, Ronald, Petavy, Frank, Seigneuret, Nathalie, Malone, James, Zhou, Liz F, Galvez, Jose, Jackson, Tammy, Harmon, Nicole, Kush, Rebecca D., Applied Clinical Trials, 22 January 2019 [Online] www.appliedclinicaltrialsonline.com.
[18] NIH definition of CDE National Institutes of Health. What is a CDE? http://www.nlm.nih.gov/cde/glossary.html#cdedefinition (2015).
[19] ISO/IEC 11179-3 Metadata registry model and basic attributes http://metadata-standards.org/11179/.
[20] NIH CDE Repository https://cde.nlm.nih.gov/.
[21] "Improving the value of clinical research through the use of Common Data Elements". Sheehan, J. Hirschfeld, S., Foster, E. Ghitza, U., Goetz, K., Karpinski, J., Lang, L., Moser, R.P., Odenkirchen, J., Reeves, D., Rubinstein, Y., Werner, E., Huerta, M., Clinical Trials 1-6 (2016), DOI: 10.1177/17407745/6653238.
[22] NIH NCI caDSR Data Standards Repository (caDSR) https://wiki.nci.nih.gov/display/caDSR/caDSR + Content.
[23] UCUM Units of Measure https://unitsofmeasure.org/.
[24] "Mystery of Orbiter Crash Solved" K. Sawyer, Washington Post, 1 October 1999 http://www.washingtonpost.com/wp-srv/national/longterm/space/stories/orbiter100199.htm.
[25] Health Insurance Portability and Accountability (HIPAA) [Online] HIPAA for Professionals https://www.hhs.gov/hipaa/for-professionals/index.html.
[26] EU General Data Protection Regulation (GDPR) https://www.eugdpr.org.
[27] NCI EVS NIH/NCI Enterprise Vocabulary Services https://evs.nci.nih.gov/.
[28] CDISC Clinical Data Interchange Standards Consortium. http://www.cdisc.org.
[29] Framework to address Operational, Technical and MethodologicAL Challenges (OPTIMAL) https://www.ncbi.nlm.nih.gov/pubmed/30970161.
[30] CDISC Glossary https://www.cdisc.org/standards/glossary and Gertel, A., Gawrylewski, H., Raymond, S., Muhlbradt, E., Applied Clinical Trials, V. 26, Issue 21 (Dec 2017) http://www.appliedclinicaltrialsonline.com/cdisc-glossary-clinical-research-terminology.
[31] CDISC Terminology https://www.cancer.gov/research/resources/terminology/cdisc.
[32] Health Level Seven (HL7) https://hl7.org.
[33] FHIR Fast Health Interoperability Resources (FHIR) [Online] https://www.hl7.org/fhir/overview.html.
[34] ISO/HL7 10781 EHR Standards [Online] https://www.isoorg/standard/57757.html.
[35] NIH Notice RFP for FHIR [Online] https://grants.nih.gov/grants/guide/notice-files/NOT-19-150.html.
[36] 21st Century Cures Act [Online] https://www.fda.gov/regulatory-information/selected-amendments-fdc-act/21st-century-cures-act.
[37] FDA's Real-World Evidence Program [Online] https://www.fda.gov/downloads/ScienceResearch/SpecialTopics/RealWorldEvidence/UCM627769.pdf December, 2018.
[38] FHIR Profile List [Online] https://www.hl7.org/fhir/profilelist.html.
[39] BRIDG Biomedical Research Integrated Domain Group (BRIDG) Model . [Online] https://bridgmodel.nci.nih.gov/about-bridg.
[40] Becnel, LB, Hastak, S, Ver Hoef, W, Milius, RP, Slack, M, Wold, D, Glickman, ML, Brodsky, B, Jaffe, C, Kush, R, Helton, E, "BRIDG: a domain information model for translational and clinical protocol-driven research", J Am Med Inform Assoc (JAMIA), 26 Feb (2017).
[41] IHE, Retrieve Form for Data Capture IHE Committee, ITI Technical. IHE IT infrastructure technical framework supplement retrieve form for data capture. s.l, IHE International Inc., 2010.
[42] "The Use of FHIR in Clinical Research" [Online] https://www.phusewiki.org/wiki/index.php?title = Investigating_the_use_of_FHIR_in_Clinical_Research.
[43] Academic Research Organization Council [Online] https://www.google.co.jp/search?sa = G&q = ARO + Council + site:tri-kobe.org&tbm = isch&source = univ&hl = ja&ved = 2ahUKEwiQs_rUzdHlAhUMS60KHcCOCYoQsAR6BAgJEAE&biw = 1280&bih = 607&dpr = 1.5.
[44] "The Global academic research organization network: Data sharing to cure diseases and enable learning health systems" Fukushima, M., Austin, C., Sato, N., Maruyama,

T., Navarro, E., Rocca, M., Demotes, J., Sato, N., Haendel, M., Volchenboum, S.L., Cowperthwaite, M., Silverstein, J.C. Webb, C., Sim, I., Chase, M., Speakman, J., Augustine, E., Ford, D. E., Learning health Systems Journal, Vol. 3, Issue 1, First published: 03 December 2018, https://doi.org/10.1002/lrh2.10073.

[45] Sentinel FDA's Sentinel Initiative. U.S. Food and Drug Administration . [Online] https://www.fda.gov/safety/fdas-sentinel-initiative/fdas-sentinel-initiative-news-and-events.

[46] Serving the enterprise and beyond with informatics for integrating biology and the bedside (i2b2),. Shawn N Murphy, Griffin Weber, Michael Mendis, Vivian Gainer, Henry C Chueh, Susanne Churchill, Isaac Kohane. s.l. : Journal of the American Medical Informatics , 201079.

[47] I2b2ACT Common Data Model [Online] https://ctsicn.org/i2b2-shrine-act.

[48] ODHSI/OMOP Observational Health Data Sciences and Informatics (OHDSI)-OMOP Model. [Online] https://ohdsi.org/.

[49] Patient Centered Outcomes Research Institute. [Online] https://www.pcori.org/.

[50] Common Data Model Harmonization, FHIR Implementation Guide [Online] http://build.fhir.org/ig/HL7/cdmh/cdmh-overview.html.

[51] Common Data Model, Harmonization (CDMH) and Open Standards for Evidence, Generation (2017–2018).

[52] BRIDG-CDM-CDISC mappings visualization tool [Online] https://vis-review-si.nci.nih.gov/.

[53] FDA Draft Guidance on Submitting RWD and RWE to FDA for Drugs and Biologics. [Online] https://www.fda.gov/media/12475/download.

[54] Critical Path institute. Critical Path Institute. [Online] https://c-path.org/about/.

[55] Cohen Veterans Bioscience [Online] https://www.cohenveteransbioscience.org/2018/12/12/cdisc-cvb-announce-first-data-standard-for-ptsd/.

[56] Worldwide Antimalarial Resistance Network (WWARN) [Online] - https://www.wwarn.org/.

[57] Danone Yogurt – https://www.google.com/search?q = danone + yogurt&gws_rd = ssl.

[58] CDISC Library (formerly CDISC SHARE) [Online] https://www.cdisc.org/cdisc-library.

[59] FDA Data Standards Catalog [Online] https://www.fda.gov/forindustry/datastandards/studydatastandards/default.htm.

[60] Common Protocol Template. TransCelerate Biopharma Inc. . [Online] http://www.transceleratebiopharmainc.com/assets/common-protocol-template/.

[61] The International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use (ICH)[Online] https://ich.org/page/multidisciplinary-guidelines.

[62] IOM (Institute of Medicine): The Learning Healthcare System: Workshop Summary. Washington, DC: The National Academies Press. NAM-AHRQ-Learning-Health-Systems-Meeting-Summary. 2007. pdf.

[63] A. Basu, D. Warzel, A. Eftekhari, J.S. Kirby, J. Freymann, J. Knable, A. Sharma, P. Jacobs, Call for Data Standardization: Lessons Learned and Recommendations in an Imaging Study, JCO Clin Cancer Inform. 3 (2019 Nov) 1–11, https://doi.org/10.1200/CCI.19.00056.

[64] Phenopackets on FHIR [Online] https://phenopackets-schema.readthedocs.io/en/latest/introduction.html.

[65] Information technology — Metadata registries (MDR) — Part 6: Registration, https://standards.iso.org/ittf/PubliclyAvailableStandards/c060342_ISO_IEC_11179-6_2015.zip.

[66] EMA Joint Task Force on Big Data – Summary Report https://www.ema.europa.eu/en/documents/minutes/hma/ema-joint-task-force-big-data-summary-report_en.pdf).

[67] NIH Strategic Plan for Data Science [Online] https://datascience.nih.gov/sites/default/files/NIH_Strategic_Plan_for_Data_Science_Final_508.pdf.

[68] System of Accelerated Research (SOAR) [Online] https://dcri.org/our-work/analytics-and-data-science/data-sharing/.

[69] Learning Health Community [Online] http://www.learninghealth.org/.

[70] Mobilizing Computable Biomedical Knowledge (MCBK) [Online] https://medicine.umich.edu/dept/lhs/service-outreach/mobilizing-computable-biomedical-knowledge.

[71] Yosemite Project [Online] https://yosemiteproject.org.