CDISC as the global standards

CDISC standards and innovations*

Andrea Vadakin   Rebecca D. Kush

CDISC (Clinical Data Interchange Standards Consortium)

Abstract

The Clinical Data Interchange Standards Consortium (CDISC) is an open, multidisciplinary neutral non-profit standards developing organization (SDO) that has been working through productive, consensus-based collaborative teams, since its formation in 1997, to develop global standards and innovations to streamline medical research and ensure a link with healthcare. The CDISC mission is to develop and support global, platform-independent data standards that enable information system interoperability to improve medical research and related areas of healthcare. The CDISC vision is informing patient care and safety through higher quality medical research. The CDISC suite of standards supports medical research of any type from protocol through analysis and reporting of results. They have been shown to decrease resources needed by 60% overall and 70-90% in the start-up stages when they are implemented at the beginning of the research process. They are harmonized through a model that is now not only a CDISC standard but also an HL7 (Health Level 7) standard on the path to becoming an ISO (International Organization for Standardization) /CEN (European Committee for Standardization) standard, thus giving the CDISC standards (harmonized together through BRIDG (Biomedical Research Integrated Domain Group)) an international status and accreditation. This publication provides a summary of each of the primary CDISC standards in addition to two CDISC innovations that are designed to improve the value of the standards in terms of enabling higher quality medical research done faster and with fewer resources.

Key words
medical research, data standard, innovation, data interchange, interoperability

---

1.　The CDISC Standards

The CDISC standards are summarized, beginning with the data collection standard (CDASH), then the transport standard (ODM), and followed by LAB for laboratory data, SDTM, SEND, and ADaM for tabulated CRF data and analysis datasets. The Protocol Representation standard supports the study plan and the submission of elements of the plan to enable evaluation of the actual data in accordance with the plan. The BRIDG model for harmonizing the CDISC standards and the Controlled Terminology and Glossary, which support the semantic interoperability across the suite are then summarized. Two innovations — Healthcare Link and SHARE are then described at the end of this publication. The end-to-end support for medical research of any type should become apparent as these standards are envisioned to work together, building quality in from the start of a study.

1.1　Clinical Data Acquisition Standards Harmonization (CDASH)

CDASH supports clinical and medical research data collection to improve data interoperability and quality. This is accomplished by focusing on the development of consensus-based "content standards" for Case Report Forms (CRF). CDASH has been developed in response to issues dealing with the overwhelming diversity of data formats and conventions across CRFs and computer operating systems. This diversity led to a great deal of inefficiencies and opportunities for error, an issue recognized by the U.S. Food and Drug Administration (FDA), among others.

In 2006, CDISC was approached by the Association of Clinical Research Organizations (ACRO) and FDA to lead in the development of 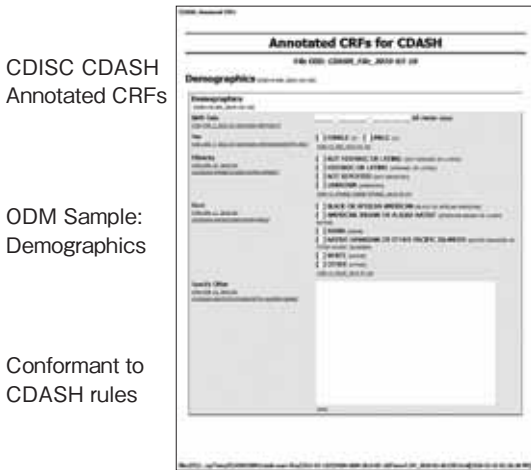standardized case report form content, as the global clinical research standards development organization. A multidisciplinary collaborative group of 17 organizations across the industry to was formed to oversee the development of CDASH. The primary aim was to identify the minimum set of data collection fields needed from a clinical, scientific and regulatory data collection perspective to enable efficient data collection at the investigative sites. It is acknowledged that additional data fields will be needed to address study-specific requirements, and this approach allows sponsors the flexibility to determine which fields should be added to CDASH recommended fields.

CDASH is based upon the Study Data Tabulation Model (SDTM), as SDTM has been established as the standard for submitting tabulated data content from CRFs to the FDA. While these standards are closely related, they have differing purposes, specifically submission (SDTM) versus data collection (CDASH). Basic data collection fields identified by CDASH project teams are mapped into SDTM and are compliant with the SDTM Implementation Guide (SDTM IG).

The scope of CDASH does not include a standard physical design for CRFs, as CDASH was originally designed for the development of standards for the content of 16 CRF safety data domains. Two additional domains were later added. A sub-project team, CDASH-ODM, has been initiated to develop machine-readable electronic CRFs (Fig. 1).

The CDASH standard is designed to improve workflow and quality from the very start of a study. One of its greatest benefits is to physician investigators at a site who gain by seeing a standardized CRF fields every time, whether on paper or electronic, thereby reducing the number of multiple data entry requirements of the past. Having a standard will improve the quality of data collected and improve communication and efficiency throughout a

Fig. 1 Sample of CDASH demographics CRF in ODM transport format

CDISC CDASH
Annotated CRFs

ODM Sample:
Demographics

Conformant to
CDASH rules

study, through the collection of quality data from the study's initiation. In addition, this benefits all involved in a biopharmaceutical development program, since data are collected in essentially the same format in which it will be submitted. Hence, FDA reviewers will receive electronic data in a standard format (SDTM), creating a more streamlined process from the outset to the completion of a study.

1.2 The Operational Data Model (ODM)

ODM is designed to facilitate the interchange and archive of metadata and data for clinical research. The success of ODM is most apparent when data are being collected from multiple sources. ODM is compliant with the FDA's 21 CRF Part 11, meaning that it meets the requirements needed to replace traditional paper records with electronic records.

ODM uses extensible markup language (XML). XML has wide acceptance as a data interchange framework in other industries as well as for numerous vendors of clinical research software products,

hence CDISC has developed ODM to represent clinical research data and metadata in XML. XML uses entities called tags to identify the actual data content. Another XML feature, "attributes," allows for extra information to be provided if necessary. For instance, in order to be able to interchange data with another system or company, there is a need for each party to know how to build the document and how to interpret it. This is the where XML comes in. The ODM schema allows a machine to check that a received XML document meets the ODM standard.

ODM brings several advantages to the clinical study process. It has a single format, providing all components needed to describe clinical research data with attribution requirements mandated by regulatory agencies. It reduces the number of unique file formats a clinical application needs to support. ODM also has a number of unique features that can improve the integrity and reliability of clinical data interchange. It has a flexible design that can be adapted to any clinical study application where data may be needed.

In addition to these advantages is the ability of ODM to assist in managing and archiving an audit trail. ODM keeps all clinical and associated administrative data, old and new, erasing nothing. This is vital to the regulatory process as specific data points throughout a study may be subject to FDA review (Fig. 2).

In addition to having features that work with any current clinical data management system (CDM), CDISC has initiated the ODM Certification Program to ensure interoperability among available ODM toolsets to promote user confidence.

Define.xml is another instance where XML has been used to assist reviewers in preventing the need to match up dataset documentation and is able to handle repeat values, links and code lists, facilitating the exchange of dataset information. The

Fig. 2　A portion of the ODM structure showing the audit trail support

define.xml specification defines the metadata structures that are to be used to describe the submission datasets and variables in a manner that meets or exceeds the minimum FDA requirements outlined in the FDA's Data Definition Document.

1.3　The Clinical Laboratory Data Model (LAB)

One of the largest components of clinical research data is that which is collected from laboratories. Through the encouragement of its members, CDISC brought together a group of volunteers to develop a standard model for the acquisition and interchange of laboratory data — **LAB**. Prior existing standards for laboratory data were not widely accepted by the industry in that they are typically considered difficult to use, inefficient, have inadequate field definitions, non-matching population rules with clinical research data, and are unnecessarily complex.

The CDISC LAB model was therefore developed to allow interchange of test results and reference ranges and to be flexible enough to keep pace with industry changes. CDISC LAB is able to support both cumulative data and incremental data to meet stakeholder needs, and extensions have been added
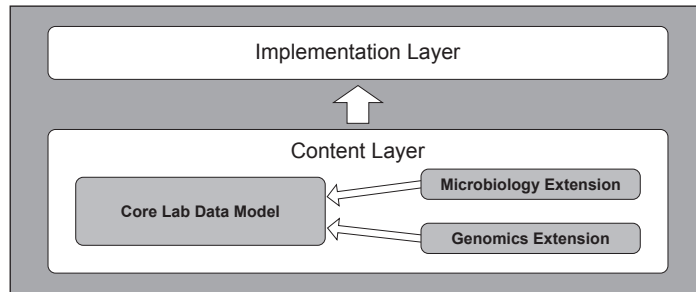
to the model to handle even more complex tests.

LAB is a content standard that can be transported in a number of different ways through different transport standards already in use for the exchange of LAB data. ASCII, HL7 and SAS Transport can all be used, as well as ODM. An HL7 V3 message was also developed, based upon the CDISC LAB content standard; however, it was deprecated by HL7 RCRIM (Regulated Clinical Research Information Management) in 2011 due to lack of use.

The LAB model approach consists of two layers: the implementation layer and the content layer. The content layer is fixed, in that it will have standard, set fields. The implementation layer can be changed dependent on which interchange/transport method is being used (Fig. 3).

1.4　Study Data Tabulation Model (SDTM)

SDTM provides a general framework for describing the organization of information collected during human and animal studies. Other CDISC standards have an impact on SDTM and it is important to consider the way data flows through to SDTM: from CDASH and LAB into SDTM. CDISC Controlled Terminology is also vital to SDTM, as

Fig. 3　CDISC LAB model layers

consistency in terminology creates seamless data flow across the standards.
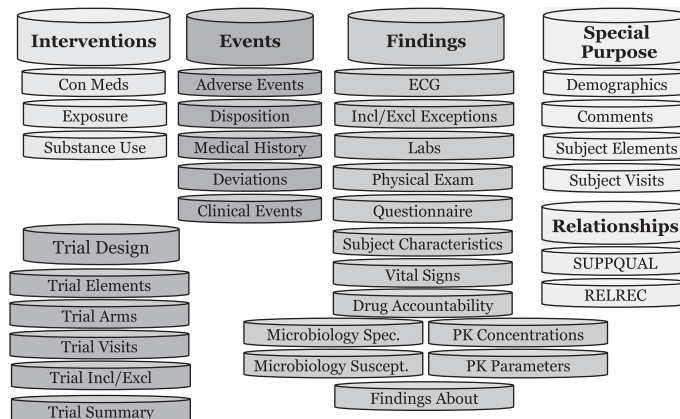
SDTM is built around observations gathered about subjects participating in a clinical study (i.e. through CRFs/CDASH). Each observation is described by a series of variables, corresponding to a row in a dataset or table. The metadata for each variable includes a variable name, label, formats or decodes and origin. These variables are classified into four major roles: identifier variables, topic variables, timing variables, and qualifier variables. A fifth type of variable role, "rule," is currently only used with the Trial Design Model.

Observations in SDTM are normally organized into a series of domains consisting of groups of related data, such as adverse events or concomitant medications. There are two categories of domains: CDISC standard domains and custom domains, the latter of which can be created for areas not yet modeled and allow more flexibility to cover all the areas needed for submission. Most observations collected during a human clinical study fall into one of three general classes: interventions, events, or findings. Precise definitions of these three terms, as well as additional Special Purpose domains can be found in the SDTM Implementation Guide (Fig. 4).

In addition to supporting FDA reviews, SDTM is also very useful for supporting data aggregation across studies (e.g. in data warehouses) and for exchanging/sharing study data across partners, for example, in cases of collaborative studies, acquisi-

Fig. 4　SDTM v 3.1.2 domain types and domains

tions and mergers or when contracting out studies. Data can be mapped to SDTM at the end of a study or when the need for aggregation occurs; however, this is inefficient and there is a loss of quality and integrity in the data. It is much more efficient to collect the data in a CDASH format with Controlled Terminology in the beginning, thus enabling efficient end to end data flow and higher quality data.

1.5　Standard for the Exchange of Non-Clinical Data (SEND)

The **SEND** model is based on SDTM, and is used exclusively to organize, structure, and format non-clinical data from animal toxicology studies submitted to the FDA. SEND facilitates transfer of non-clinical data from the sponsor to the FDA. Just like SDTM, observations are arranged into observation classes: interventions, findings, and special purpose domains. Due to the differing nature of the subject and studies, SEND necessarily has differences in domain contents. There are four categories that describe observation variables: identifier, topic, timing, and qualifier — identical to SDTM (Table 1).

1.6　The Analysis Data Model (ADaM)

ADaM supports the statistical analysis and subsequent statistical review of clinical research data. Analysis datasets and their associated metadata

are one of the four types of data that can be submitted to the FDA, along with study tabulation datasets, subject profiles, and listings. SDTM datasets are intended to represent the collected clinical data, and are not sufficient to support statistical analysis and testing of the scientific hypotheses that are the purpose of conducting clinical trials.
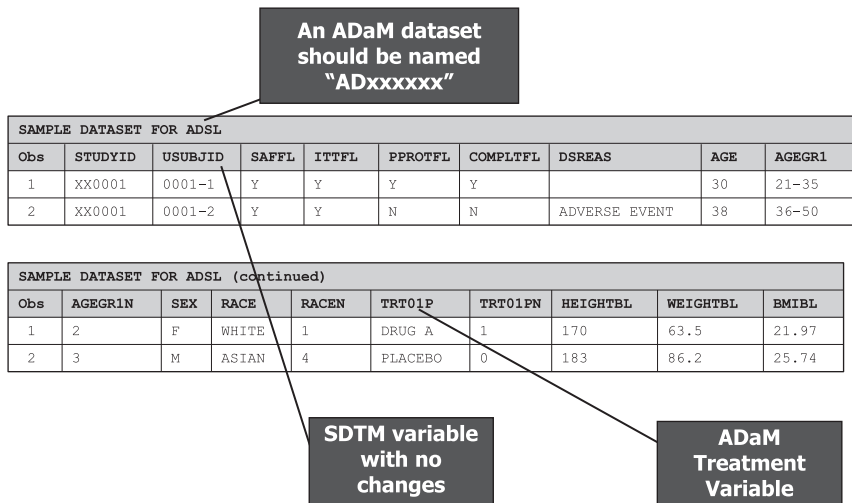
ADaM specifies data structures and associated metadata for analysis datasets. In ADaM, variables, observations and even entire datasets may be derived to enable the scientific evaluation of the objectives of the clinical study. ADaM allows reviewers to have a clear understanding of how the analysis datasets were derived and how they are to be used to produce the statistical results provided in the submission.

Through standardization of the structures of analysis datasets, software tools can be created that allow access, manipulation and viewing of analysis datasets. This further increases efficiency to assist all parties involved in the creation and use of analysis datasets. There are 4 key principles of a well-constructed dataset: 1) facilitate clear and unambiguous communication and provide a level of traceability, 2) be analysis ready, 3) be linked to machine-readable metadata in order to facilitate software development, and 4) work with tools currently available (Fig. 5).

Table 1　SEND v 2.3 findings domains

• Animal Characteristics	• Food Consumption
• Water Consumption	• Body Weights
• Clinical Signs	• Animal Disposition
• Clinical Pathology	• Macroscopic Findings
• Organ Weights	• Microscopic Findings
• Fetal Data	• Fertility
• Group Observations	• Group Characteristics
• Drug/Metabolite Levels	• Study Summary
• Tumor Analysis	• Rodent Micronucleus
• Vital Signs	

Fig. 5　Sample of the CDISC Analysis Data Model (ADaM)

An ADaM dataset should be named "ADxxxxxx"

SAMPLE DATASET FOR ADSL

Obs	STUDYID	USUBJID	SAFFL	ITTFL	PPROTFL	COMPLTFL	DSREAS	AGE	AGEGR1
1	XX0001	0001-1	Y	Y	Y	Y		30	21-35
2	XX0001	0001-2	Y	Y	N	N	ADVERSE EVENT	38	36-50

SAMPLE DATASET FOR ADSL (continued)

Obs	AGEGR1N	SEX	RACE	RACEN	TRT01P	TRT01PN	HEIGHTBL	WEIGHTBL	BMIBL
1	2	F	WHITE	1	DRUG A	1	170	63.5	21.97
2	3	M	ASIAN	4	PLACEBO	0	183	86.2	25.74

SDTM variable with no changes

ADaM Treatment Variable

1.7　Protocol Representation Model (PRM)

The protocol is the plan or blueprint for the clinical research study. It is the document that describes the objective(s), design, methodology, statistical considerations and organization of the trial or study. Ensuring that the protocol is well designed is vital to the overall success of a research study.

CDISC initiated the process of developing the **Protocol Representation** standard in order to support a machine-readable protocol with a set of standard elements that can be reused without the need for re-entry. The PRM supports trial registration, CRF development, setting up a clinical trial management system (CTMS) for managing data and tracking studies, and reporting results. Using the protocol to define data collection tools, being able to track and readily exchange information, from CDASH right through to SDTM and ADaM, can complete the full life cycle of a research program — a true end to end process.

The **Trial Design Model (TDM) or Study Design Model (SDM)** is at the core of the Protocol Representation standard and also in SDTM. To assist reviewers in interpreting data received, the FDA requested that structured information about trial design be submitted along with clinical study datasets. As the TDM began to emerge, it was quickly realized that trial design has an impact on both submission data and protocol representation, and the first version of SDTM was released in 2004 including the TDM (or SDM).

A Trial Design is a plan for what will be done to subjects and what data will be collected about them in order to address the study's directives as laid out in the protocol. Key elements of a trial design are epochs, arms, study cells, elements, branches, treatments, visits, and inclusion/exclusion. Trial designs can be stored as XML and be submitted as a Trial Design Matrix for viewing by a regulatory authority. Further information on TDM is available in the SDTM Implementation Guide (SDTM IG) and the Protocol Representation standard.

TDM can be used for the purposes of planning baselines to support study/project management, define data collection requirements for a study, and is important to the statistical analysis plan. TDM

allows reviewers to 1) clearly and quickly grasp the design of a clinical study, 2) compare the designs of different studies, 3) search a data warehouse for clinical studies with certain features, and 4) compare planned and actual treatments and visits for subjects in a clinical study.

One of the key aims of the Protocol Representation Group was to ensure that common protocol content can be communicated and exchanged without compromising the innovation behind each protocol. The standard has also been made generic enough to accommodate all types of protocols for clinical research, and has utilized existing standards-based information to guarantee that the Protocol Representation standard achieves semantic interoperability amongst all systems. There is now a library detailing around 360 common elements that can be used repeatedly for new protocols or other documents. The Protocol Representation Model version 1.0 has been represented in the BRIDG model to express the relationships among the key PR ele-
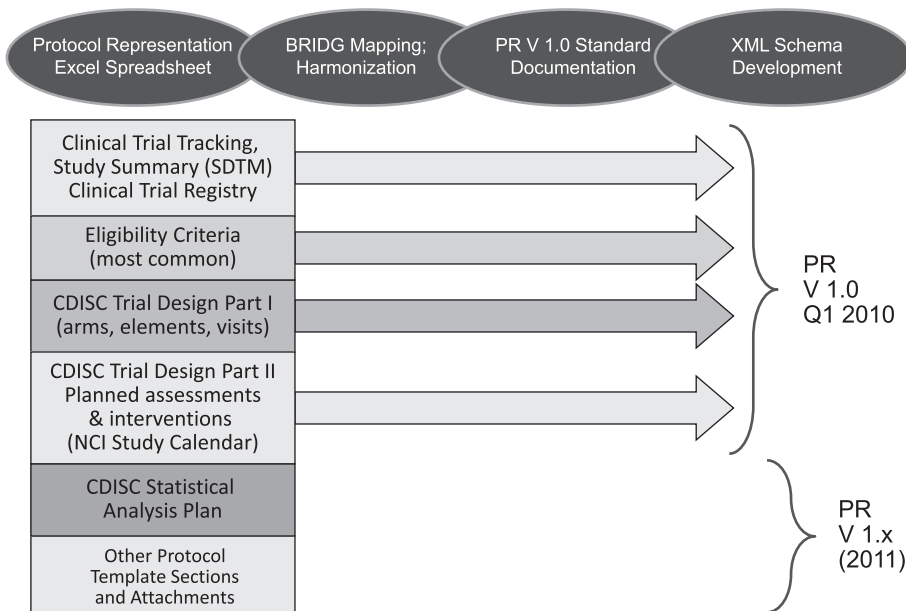
ments (Fig. 6).

There is now an available XML (ODM) representation of the Study Design model. The next step is to develop a tool to help protocol authors (who are not typically familiar with UML (Unified Modeling Language) models) experience the value of this standard.

1.8　The Biomedical Research Integrated Domain Group (BRIDG) Model

The **BRIDG** Model is the combined work of four major stakeholders: CDISC, HL7 RCRIM, NCI (National Cancer Institute) and FDA. BRIDG effectively enables information system interoperability by harmonizing the CDISC standards to work together, as well as to ensure that developers can establish applications that will work with the CDISC standards.

The scope of the BRIDG model is ambitious and vital to the success of future tool development that will work alongside the standards. It is a formal

Fig. 6　Diagram of the sections of the PR Model V1.0 and development steps

UML Model, the semantic foundation for software applications and message development. The BRIDG Model should ensure that systems are able to communicate with each other and transfer data seamlessly. This model can remove the semantic ambiguities present in the complex world of medical research by defining the domain semantics, serving as the foundation for all standards, software or service specifications in CDISC, HL7, NCI and FDA. In order to achieve full semantic interoperability, it is essential that healthcare, research and regulatory authorities have a shared semantic view. That shared semantic view is BRIDG.
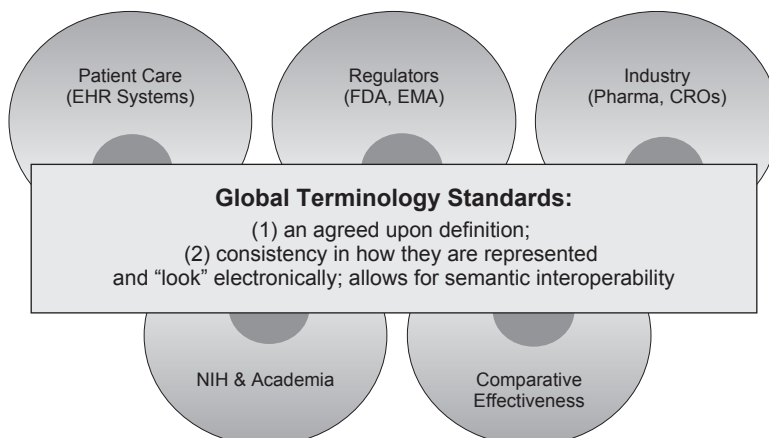
1.9　Standard Controlled Terminology

Physicians working in the area of medical research are challenged by the various terminologies/vocabularies that are required. Due to the use of several concurrent codings for healthcare and medical research, there is a lack of interoperability, precipitating a confusing and time-consuming harmonization process. The **CDISC Standard Controlled Terminology** initiative had an initial primary objective to support the terminology needs

of the CDISC models across the clinical research continuum (from Protocol and CDASH through LAB to SDTM and ADaM). The terminology initiative is a collaboration of volunteers from stakeholders including the FDA, global clinical research sponsors, contract research organizations, academia and others. The terminology team also works in partnership with the US NCI Enterprise Vocabulary Services (EVS), as well as actively coordinating with ICH, EMA (European Medicines Agency), ISO and HL7 RCRIM (Fig. 7).

Many companies have worked with proprietary standards for many years and have developed their own terminology within their organizations, creating different sets of codes and terms. In order to achieve full semantic interoperability across the standards, the terminology needs to be agreed upon by stakeholders on a global level. To achieve this, the Standard Controlled Terminology team developed terminology that supports all CDISC data models, utilized existing open-source terminology, consulted key industry stakeholders, and collaborated with standards organizations and vocabulary developers in order to ensure that terminology has international usage. The terminology is currently

Fig. 7　Controlled Terminology broadly supporting semantic interoperability in clinical research

maintained through the NCI EVS and is continually updated as CDISC continues to develop new standards, processes and procedures.

1.10　CDISC Glossary

The **CDISC Glossary** was actually the first initiative with demonstrated success within a newly established CDISC organization. This very useful CDISC standard dictionary is now updated annually and published on the CDISC website and in Applied Clinical Trials. It includes dictionary-style definitions of terms frequently used in clinical research, along with a set of Abbreviations and Acronyms used in this arena.

The BRIDG Model, Glossary and Controlled Terminology all support the end-to-end suite of CDISC standards to ensure harmonization and semantic interoperability across these models (Fig. 8).

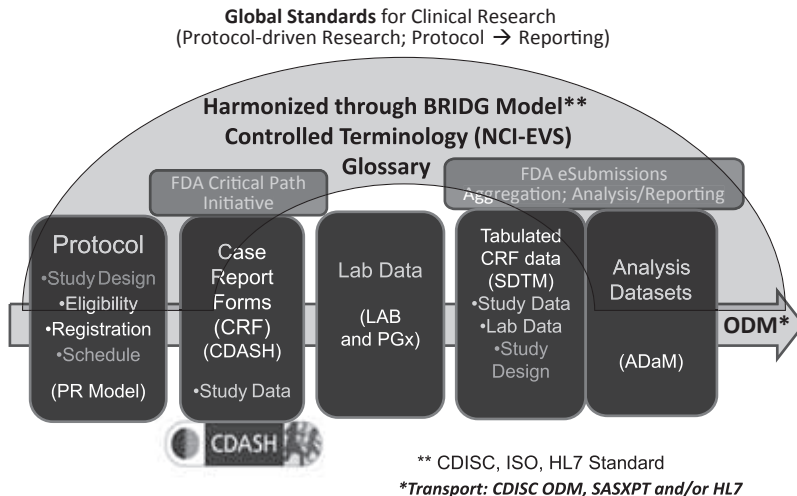## 2.　CDISC Innovations

2.1　Healthcare Link Initiative

Electronic Health Records (EHR) have many benefits, not least of which is giving physicians the ability to cut down time reentering information written on paper forms and the ability to ensure standardized fields for entry. An additional, future use of the EHR may be the potential for EHR data to be utilized in the realm of medical research, yet much of the valuable data in EHRs is not accessed for regulated research at this given time. Reaching the point where research is more closely linked to the workflow of medical professionals is the key driver behind the CDISC **Healthcare Link Initiative**.

CDISC is working toward the time when medical research workflow is seamlessly linked with the healthcare workflow. This development will significantly improve recruitment of physicians who will conduct research and also patients who will participate in research studies, in turn bringing about better patient care. Some exciting developments in the Healthcare Link Initiative to form a basis for linking research and healthcare are:

1) **The Electronic Source Data Interchange (eSDI)** document — a seminal work on eSource aimed at bringing about a paperless process

Fig. 8　The CDISC suite of standards: Protocol through reporting

Global Standards for Clinical Research
(Protocol-driven Research; Protocol → Reporting)

Harmonized through BRIDG Model**
Controlled Terminology (NCI-EVS)
Glossary

FDA Critical Path Initiative

FDA eSubmissions
Aggregation; Analysis/Reporting

Protocol
•Study Design
•Eligibility
•Registration
•Schedule
(PR Model)

Case Report Forms (CRF)
(CDASH)
•Study Data

Lab Data
(LAB and PGx)

Tabulated CRF data (SDTM)
•Study Data
•Lab Data
•Study Design

Analysis Datasets
(ADaM)

ODM*

CDASH

** CDISC, ISO, HL7 Standard
*Transport: CDISC ODM, SASXPT and/or HL7

for research.

2) **The Biomedical Research Integrated Domain Group (BRIDG)** model — a collaborative project using unified modeling language, creating a reference map that harmonizes clinical research standards and healthcare standards.

3) **Integrating the Healthcare Enterprise (IHE)-CDISC, Retrieve Form for Data Capture Integration Profile (RFD)** — a real world implementation that brings EHR closer to the world of medical research through "button-ready" technology, just-in-time workflow, and bite-sized semantics to easily and efficiently pre-populate CRFs up to 80%.

4) An **Interoperability Specification (IS #158)** documenting the use of three standards for the use case of providing a core research dataset from an EHR. This IS, developed through the HITSP (Healthcare Information Technology Standards Panel) process that was managed through ANSI (American National Standards Institute) in the U.S., cites the implementation of the Continuity of Care Document (CCD), RFD and CDASH (producing CDASH from CCD through a Clinical Research Document

(CRD) that maps CCD to CDASH.

This approach, which makes use of RFD as a workflow enabler to produce CDASH and other standards-based data such as safety information) is now being used in other countries around the world (Fig. 9).
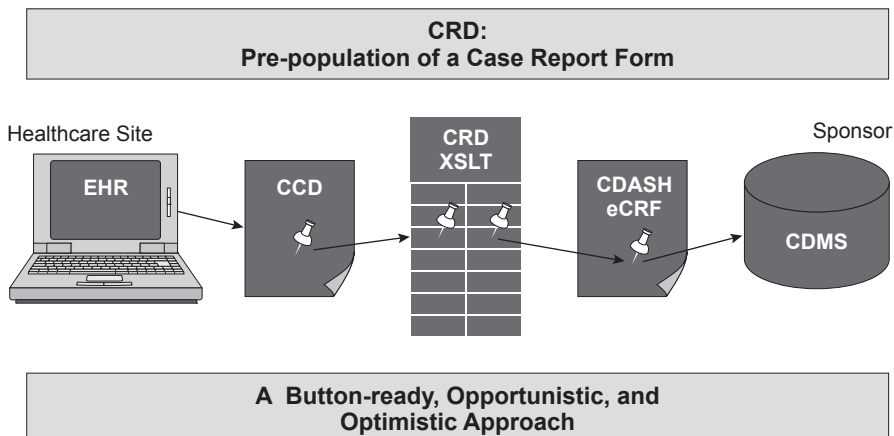
2.2 Shared Health And Clinical Research Electronic Library (SHARE)

SHARE is a global, accessible electronic library, which through advanced technology enables precise and standardized data element definitions that can be used in applications and studies to improve biomedical research and its link with healthcare. Built on BRIDG, SHARE is intended to be a healthcare-biomedical research enriched data dictionary complemented by detailed clinical content, ISO data types and terminologies.

SHARE will provide computable semantic interoperability enabling the reuse of data, and is structured on four pillars:

1. Common information model (BRIDG)
2. Strong data typing (ISO 21090)
3. Common terminologies/value sets (CDISC, HL7, SNOMED (Systematized Nomenclature

Fig. 9 The approach through which EHRs that are 'RFD-ready' provide CDASH (core research dataset)

of Medicine), etc.)

4. Processes supporting exchange of information

A Pilot conducted with Mayo Clinic provided CDISC information upon which SHARE can be developed (Fig. 10).

SHARE will provide the opportunity to improve upon data quality through the establishment of better definitions, enhancing data consistency to allow data integration and aggregation, as well as comparison across research organizations, regulators and partners. This library focuses on reducing the costs associated with the maintenance of diverse dictionaries within separate research organizations, and fosters improved interactions with partners through the mutual use of standardized definitions. This electronic and collaborative infrastructure will speed up the process in developing standardized definitions, improve governance of those definitions, and provide an outlet in which to

access the definitions on a 24/7 basis. Perhaps most importantly, SHARE makes it possible to further enable semantic interoperability and provides a link between research and healthcare.

Conclusion

CDISC Standards and Innovations have been developed to streamline global clinical research processes. The next steps will include development of additional domains to support the individual therapeutic areas. In other words, the efficacy standards will be added to the safety standards over the coming years. CDISC is made possible through the global support of volunteers and organizations that participate in CDISC activities, take CDISC educational courses and join CDISC to demonstrate their interest in improving the processes and quality of medical research for the benefit of patients everywhere.

Fig. 10　Diagram of SHARE pilot