Current Status and Future Scope of Clinical Data Standards

Arghya Chattopadhyay, PhD
What is CDISC?

- The Clinical Data Interchange Standards Consortium (CDISC) is a global, open, multidisciplinary, vendor-neutral, non-profit standards developing organization (SDO).
- The CDISC has been developing global standards 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 “to inform patient care and safety through higher quality medical research".
- The CDISC standards supports medical research of any type from protocol through analysis and reporting of results.
Information from healthcare (private, aggregated) to enable research

Healthcare
- Quality healthcare
- Informed decisions
- Personalized medicine
- Patient safety and privacy
- Public health
- Improved therapies
- Efficiencies/reduced costs

Research
- Discovery of new therapies
- Understanding diseases
- Testing/comparing therapies (CER)
- Assessing efficacy
- Monitoring safety
- Understanding responses (genomics, biomarkers)
- Public health/quality evaluations
- Post-marketing surveillance

Research findings to inform healthcare decisions

Inefficient cycle
Vision – Medical Innovation

Subject Data – Enter Once for Multiple Purposes

Data Sources
- EDC
- EHR
- ECG
- X-RAY
- LAB

CDISC Standards Real-time Integration

Regulatory Authority

Public Registries and IRBs

Sponsor

CRO or Partner

Payer

“Rolling” Warehousing, Reporting and Submissions

© 2011
Interoperability

- The interoperability is the ability of exchanging and processing data successfully among systems.
- HL7 has categorized interoperability as technical interoperability, semantic interoperability, and process interoperability.

<table>
<thead>
<tr>
<th>Type</th>
<th>Requirement</th>
<th>Standard</th>
</tr>
</thead>
<tbody>
<tr>
<td>Technical Interoperability</td>
<td>Integrity of Data Transportation</td>
<td>SAS XPORT Transport Format, ODM</td>
</tr>
<tr>
<td>Semantic Interoperability</td>
<td>Technical Interoperability + Integrity of Understanding</td>
<td>Controlled Terminology, BRIDG</td>
</tr>
<tr>
<td>Process Interoperability</td>
<td>Semantic Interoperability + Integrity of Processes</td>
<td>Not Available (dynamic diagrams need to be defined in BRIDG)</td>
</tr>
</tbody>
</table>

- Semantic Interoperability: ‘the ability of two or more computer systems to exchange information and have the meaning of that information automatically interpreted by the receiving system accurately enough to produce useful results, as defined by the end users of both systems.
## CDISC Value by Profession

<table>
<thead>
<tr>
<th>Profession</th>
<th>Why CDISC?</th>
</tr>
</thead>
</table>
| **CEO, Study Sponsor, Program/Project Manager** | a) To initiate your study quickly and economically?  
'b) Have your CRFs easily understood and completed by investigative site personnel?  
'c) Receive high quality data that will readily fit into the format requested by FDA?  
d) To have a protocol with sections that can be re-used (without re-entry) for trial registration, IRBs, generating study reports, publication, eSubmissions  
e) Your data to readily integrate with that of other studies?  
f) To be able to find your data later?  
g) Have your data ready in case of a merger or acquisition?  
h) Be able to use data from past research to improve current/future research? |
## CDISC Value by Profession

<table>
<thead>
<tr>
<th>Profession</th>
<th>Why CDISC?</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Medical Writer</strong></td>
<td>a) To write your protocols and study reports a bit faster?</td>
</tr>
<tr>
<td></td>
<td>b) Re-use information from your protocols without re-entering information, e.g. trial registration, study reports, publications?</td>
</tr>
<tr>
<td></td>
<td>c) To enable others on the team to auto-generate visit schedules and CRFs?</td>
</tr>
<tr>
<td><strong>Data Manager</strong></td>
<td>a) To get your CRFs ready faster and economically?</td>
</tr>
<tr>
<td></td>
<td>b) To create your data validation specifications more quickly and effectively?</td>
</tr>
<tr>
<td></td>
<td>c) To build your databases more efficiently?</td>
</tr>
<tr>
<td></td>
<td>d) To reduce training and improve communication with your CRAs and sites?</td>
</tr>
<tr>
<td></td>
<td>e) To get cleaner data, faster?</td>
</tr>
<tr>
<td></td>
<td>f) To reduce data problems and be able to focus more on the scientific content?</td>
</tr>
<tr>
<td></td>
<td>g) To build more effective partnerships with the whole study team?</td>
</tr>
</tbody>
</table>
# CDISC Value by Profession

<table>
<thead>
<tr>
<th>Profession</th>
<th>Why CDISC?</th>
</tr>
</thead>
</table>
| **Vendor or Information Technologist** | a) To ensure that your system will be able to readily exchange information with another system the sponsor may wish to use?  
b) To be able to provide a system based on industry standards?  
c) To be able to quickly respond to sponsor requests by using standard libraries? |
| **Statistician/Programmer**      | a) To be able to create tables, listings and figures more efficiently?  
b) To be able to integrate data from multiple studies more easily?  
c) To be able to standardize your safety analysis programming? |
| **Others?**                      | What do you want from standards?                                                                                                                                                                          |
FDA’s position statement:

“FDA envisions a **semantically interoperable** and sustainable submission environment that serves both regulated clinical research and health care. To this end, FDA will continue to research and evaluate, with its stakeholders, potential new approaches to current and emerging data standards. FDA does not foresee the replacement of CDISC standards for study data and will not implement new approaches without public input on the cost and utility of those approaches.”
CDISC – Functional Standards
- The content and format standard supporting the interchange of clinical trial protocol information. This is a collaborative effort with Health Level Seven (HL7).

- **Trial Design Model (TDM)**: The content standard that defines the structure for representing the planned sequence of events and the treatment plan of a trial. This is a subset of the SDTM and Protocol Representation.
<table>
<thead>
<tr>
<th>Protocol Section</th>
<th>CRF Development</th>
<th>Data Collection</th>
<th>Data Analysis</th>
<th>Report or eSubmission</th>
</tr>
</thead>
<tbody>
<tr>
<td>Info for Trial Registration</td>
<td></td>
<td></td>
<td></td>
<td>Basic Info/Trial</td>
</tr>
<tr>
<td>Eligibility Criteria</td>
<td></td>
<td></td>
<td></td>
<td>Summary (Registration)</td>
</tr>
<tr>
<td>Study Design: Arms, Epochs</td>
<td></td>
<td></td>
<td>SDTM Data</td>
<td></td>
</tr>
<tr>
<td>Study Design: Planned Events</td>
<td>SDASH CRFs</td>
<td></td>
<td>Data Tabulation</td>
<td></td>
</tr>
<tr>
<td>Statistical Analysis Plan</td>
<td></td>
<td>Data Analysis</td>
<td></td>
<td>ADaM Datasets</td>
</tr>
<tr>
<td>Appendices, etc.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Information Re-Use
Improved Quality and Efficiency

PR Version 1.0

CDASH
CRFs

SDTM

ADaM
Datasets
• FDA CRITICAL PATH INITIATIVE: STREAMLINING CLINICAL TRIALS
  ▪ Creating Innovative and Efficient Clinical Trials and Improved Clinical Endpoints
  ▪ 45. Consensus on Standards for Case Report Forms. Clinical trial data collection, analysis, and submission can be inefficient and unnecessarily expensive. A wide array of different forms and formats are used to collect clinical trial information, and most data are submitted to the FDA on paper. Differences in case report forms across sponsors and trials creates opportunities for confusion and error. Standardization of the look and feel of case report forms could reduce these inefficiencies and also help accelerate progress toward electronic data capture and submission.

CDISC LAB Model

• Primary AIMS
  • Interchange of test results & reference ranges
  • Incremental and cumulative data interchange
  • Full range of transaction types
  • Interchange data from 1+ studies in single file
  • Support the bulk transfer of laboratory data
SEND

- SEND is an implementation of SDTM for animal data
- SEND defines domains and variables for submitting all data generated from animal toxicity studies
  - Includes: single- and repeat-dose toxicity, carcinogenicity, reproductive toxicity, and rodent micronucleus
  - Does not include data generated from in vitro studies or as part of basic pharmacology or efficacy studies conducted in animals

- CRADA (April 2002) between PharmQuest and CDER to develop and evaluate software tools for receiving, storing, viewing and analyzing nonclinical (i.e., animal toxicity) data based on SEND model
SEND v2.3 Findings Domains

- Animal Characteristics
- Water Consumption
- Clinical Signs
- Clinical Pathology
- Organ Weights
- Fetal Data
- Group Observations
- Drug/Metabolite Levels
- Tumor Analysis
- Vital Signs

- Food Consumption
- Body Weights
- Animal Disposition
- Macroscopic Findings
- Microscopic Findings
- Fertility
- Group Characteristics
- Study Summary
- Rodent Micronucleus
Key Principles for Analysis Dataset Creation

Analysis datasets should:

- facilitate clear and unambiguous communication
- be useable by currently available tools
- be linked to machine-readable metadata
- be analysis-ready
- include subject-level analysis dataset named ADSL
- use the convention: ADxxxxxxx for naming
- have optimum number of datasets so minor programming needed
- maintain SDTM variable attributes for same variables
- use SDTM naming fragments where feasible
CDISC Operational Data Model

• Transport Standard (XML)
  ▪ Developed to carry case report form data
  ▪ Carries complete audit trail information (21CFR11)
  ▪ Supports electronic signatures
  ▪ Archives electronic data without need to archive original system at sites
  ▪ Can automate generation of eCRFs
  ▪ Enables remote monitoring or auditing
  ▪ Facilitates exchange of data between different technologies that are ODM (supports features common to all CDM and EDC systems)
eXtensible Markup Language

- XML - method for putting structured data in a text file
- Looks similar to HTML
  - Tags "<" ">
  - Attributes name="Value"
- Very flexible standard for data/metadata exchange
- Text based & readable by humans and machines
  - Vendor neutral
  - Computer system neutral
CDISC Terminology

• Formalized CDISC Terminology Initiative in 2005

• Primary Objective: to define and support the terminology needs of the CDISC models across the clinical trial continuum (CDASH → SDTM), Focus on “standard” terminology codelist development and publication

• Terminology Initiative comprised of 45 team members (FDA, NCI, Global Sponsors & CROs, Academia) distributed across 4 project teams

• Key partnership with NCI Enterprise Vocabulary Services (NCI EVS) with dedicated CDISC / FDA resources
CDISC-FDA-NCI EVS Terminology Harmonization

http://www.cancer.gov/cancertopics/terminologyresources/

Structured product Labeling (SPL)
Unique Ingredient Identifier (UNII)
Individual Case Safety Report (ICSR)
Center for Device and Radiological Health (CDRH)

SDTM, ADaM, QS
CDASH, SEND

NCI Thesaurus
BRIDG Scope

Protocol-driven research and its associated regulatory artifacts:

i.e. the data, organization, resources, rules, and processes involved in the formal assessment of the utility, impact, or other pharmacological, physiological, or psychological effects of a drug, procedure, process, subject characteristic, or device on a human, animal, or other subject or substance plus all associated regulatory artifacts required for or derived from this effort, including data specifically associated with post-marketing adverse event reporting.
Domain-Friendly, Subdomain-Specific Business Models

Layer 1
- Adverse Event
- Study Conduct
- Common
- Protocol Representation
- Regulatory

Separate EA/XML file for each subdomain

Layer 2
BRIDG Domain Analysis Model (DAM)

Single Enterprise Comprehensive View
Contains subdomain level tags on classes and attributes to support subdomain-specific names for Layer 1 models

- Adverse Event
- Study Conduct
- Common
- Protocol Representation
- Regulatory

Single EA file with comprehensive and subdomain Views

Layer 3
BRIDG OWL-DL File (semantics)

RIM-Based BRIDG Model
Class and attribute level mappings to the Enterprise Comprehensive View will be captured in the Visio tool. The Visio tool generates mapping reports for documentation.

Equivalent to an HL7 DMIM (HL7 Visio)
CDISC Standards and Data Flow

**Healthcare Standards**

Health Level 7 (HL7) Reference Information Model (RIM)

**Clinical Research Standards (CDISC)**

- Patient Info
- Protocol Representation
- Study Design Analysis Plan
- Study Protocol
- Electronic Health Record
- HL7 or ODM (XML)
- CDASH - eCRF Study Data (defined by SDTM)
- (e)Source Document

**Legend**

- ODM XML (transport)
- SDTM & ADaM (content)
- Protocol Information (content)
- Source Data (other than SDTM/eCRF data)
A global, accessible electronic library, which through advanced technology, enables precise and standardised data element definitions (including value sets) that can be used in applications and studies to improve biomedical research and its link with healthcare.

*Key purposes:* Develop efficacy standards faster and make the CDISC standards more accessible.
CDISC Healthcare Link

Original Process

**EHR**
- access patient
- access EHR data
- print out EHR data
- enter EHR data
- submit eCRF
- print out CRF

**Source**
- initialize form

**EDC**
- access patient

Improved Process

**EHR**
- access patient
- retrieve form
- enter new data

**Source**
- archive source
- receive EHR data
- upload EHR data

**EDC**
- enter new data
- submit eCRF
- archive eCRF
- eCRF complete
Patient Value: Quality of Healthcare, Safety
*Research informs healthcare more effectively*
*Build quality into process at beginning*

**Research Results, eSubmission Standard Formats**

**Public Registries, IRB, DSMBs**

**Regulatory Authority**

**Scientific Publication**

**Research Site**
(Healthcare Location, Investigator, Site Personnel)

**De-identified Data**

**Research Data**

**Study Sponsor**
(e.g. ARO, CRO, Vendor, Principal Investigator, potentially AHRQ...)

**Reviewers**
(e.g. Research Partner, Sponsor, Registry, Regulator, IRB, DSMB)

**Site Research Archive**

CDISC Standards are NOT just for FDA eSummissions!
Challenges Faced Within Current Environment

- Limitations of the dataset standard for submission (SAS Transport V5)
  - Variable name length limited (8)
  - Variable label length limited (40)
  - Variable size limited
  - Variable size pre-allocated based on length

- Dataset files size – increasing dramatically

![Pie chart showing 70% empty and 30% data]
Challenges Faced Within Current Environment

- Flat, two-dimensional data structure for hierarchical, multi-relational data
- Clinical data is hierarchical and multi-relational – “round”
- Important meaning is lost when exchanging 2 dimensional flat files, making some interpretations and analyses difficult or impossible, i.e. decreased semantic interoperability
- Just like flat maps are useful for relatively short distances, they are not useful in navigating the globe