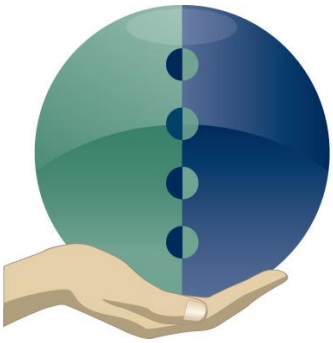


CDISC SHARE

SHARED HEALTH AND CLINICAL RESEARCH ELECTRONIC LIBRARY

CDISC Shared Health and Research Electronic Library

An Overview

Date: 30 October 2009

Version: 1.0

CDISC

Table of Contents

1 Introduction

1.1 Overview

The mission of CDISC is “to develop and support global, platform-independent data standards that enable information system interoperability to improve medical research and related areas of healthcare.”

Over the past decade, CDISC has fulfilled its mission by publishing and supporting a suite of standards that enable the electronic interchange of data throughout the lifecycle of a clinical research study. Specifically, CDISC has developed standards for use across the various points in the research study lifecycle: Protocol Development (Protocol Representation Model Version 1); data collection (Clinical Data Acquisition Standards Harmonization (CDASH)); exchange of operational data (Operational Data Model); exchange of clinical laboratory data (LAB) and data submission to regulatory agencies (Study Data Tabulation Model (SDTM) and Analysis Data Model (ADaM). As adopters have realized the benefits of these standards, it has become apparent that there is a need for a foundational standard to support computable semantic interoperability (CSI) – the predictable exchange of meaning between two or more systems -- across multiple standards including, but not limited to, those developed by CDISC.

In addition to the desire for CSI described above, CDISC’s stakeholders have made it clear that there is a pressing need to fill the gaps in the content of the existing standards, to bring those standards closer together while at the same time developing therapeutic area standards; and, they would like CDISC do all of this at an increased pace. In addition, the ability to use EHR data in medical research is becoming increasingly attractive, which emphasizes the importance and value of having common vocabularies/definitions for research and related healthcare data. Therefore, while this project was originally envisioned as a tool primarily for biopharmaceutical product development, it was soon realized that the stakeholders and scope needed to move beyond this to include other areas, including but not limited to academic research and public health reporting to support clinical decisions in healthcare.

These issues are taken as given. The current emphasis is how CDISC best achieves this and what impact any chosen course of action may have.

To address these challenges it is considered by many that an industry-wide shared semantic library would meet these needs and solve the issues CDISC and its stakeholders face. The value proposition for CDISC SHARE (CDISC Shared Health And Research Electronic Library) is the assumption that the creation of a library of shared semantics will enable CDISC stakeholders – global biopharmaceutical companies, academic institutions and clinical research organizations – to achieve multiple benefits including improved operational efficiency around the collection, processing, exchanging and reporting of data, evaluation of drug safety concerns across traditional organizational boundaries, and, in the end, enhanced scientific capabilities and resulting patient benefits and therapeutic efficacies.

1.2 Vision

The vision for CDISC SHARE is to build a global, accessible electronic library, which through advanced technology, enables precise and standardised data element definitions that can be used in applications and studies to improve biomedical research and its link with healthcare

1.3 Scope

The scope of CDISC SHARE includes the data elements for all protocol driven medical research and the overlapping areas of Healthcare.

2 Business Needs and Benefits

2.1 Business Opportunity

The business/market needs and potential benefits of CDISC SHARE center on the value of sharing information among business partners and secondary use of healthcare information. Semantic interoperability (or the ability of computers/systems to exchange data along with the meaning of that data) is at the core of this information sharing. Information sharing supports a number of needs, including safety reporting or Pharmacovigilance, clinical research studies, biosurveillance, patient and disease registries, regulatory reviews of eSubmissions and other such use cases. To compare and/or aggregate data for comparative effectiveness studies or even to compare information on the same therapy or treatment across studies requires data standards.

An electronic library with a set of unambiguous concepts, such as CDISC SHARE, can provide the following potential benefits.

2.2 Provide a Consistent Approach to Standard Definitions

Mapping legacy data that are collected with different terminologies is often impossible, and if possible, meanings may be lost or misinterpreted thus impacting data quality. The use of consistent, standard definitions will obviously improve the quality of information that is exchanged, integrated, aggregated or compared. The standard definitions also facilitate the aggregation, integration and comparisons of information within and across studies since the terminology, code lists and meanings thereof are consistent,

CDISC SHARE should provide a consistent approach to standard definitions which would then allow for the following:

Unambiguous Definitions –The human and computer communication processes are made significantly easier and more effective if identical words have the same meaning, and differing concepts do not use the same word. If healthcare organizations are to exchange information in a meaningful and useful way, a standard and consistent set of definitions is essential.

Definition Quality – Without a rigorous, consistent process of defining concepts, meanings can be ambiguous and/or incomplete, which negatively impacts the quality of the accuracy and quality of information/data.

Target Data Standards – So that data can be aggregated and/or compared, a “target” standard is needed for legacy data mapping and/or to collect information that should be compared in the same way at the beginning of the research study. Hence, new studies should use these “target” standards so that data can be integrated within studies and/or compared or aggregated across studies. The target standards may be based upon both existing and new definitions.

2.3 Improve Access to Standards

CDISC SHARE will be globally accessible and include standard and definitions in a common central library that can be accessed and “consumed” electronically. This will mean that a ‘reference standard’ will be readily accessible. With respect to clinical research, specifically, CDISC standards are currently available in a number of formats. SDTM, CDASH and ADaM are available in pdf, ODM in html, Protocol Representation is in Enterprise Architect UML, and CDISC Controlled Terminology can be accessed and downloaded electronically via the NCI EVS systems and infrastructure. Many CDISC users have asked for the SDTM domains in a machine-readable format. A repository that contains all of the CDISC standards in a consistent, human and machine-readable format would be a much more usable and effective approach for CDISC

users. It will also improve processes and thus reduce costs for CDISC users. The same will be the case for other standards that are incorporated in CDISC SHARE.

2.4 Decrease Costs for Standards Users

Maintenance Costs – Organizations can save maintenance costs if there is a central electronic library of standards with unambiguous definitions that can be leveraged consistently across the entire organization.

Process Improvement – Eliminating the need for mapping legacy data and improving quality always improves processes and decreases costs

Downloadable Metadata - A number of organizations would like to be able to download the standard metadata and use these as the foundation for their own repositories.

2.5 Facilitate Data Re-use (Secondary Use)

Data Aggregation and Mining - Running a clinical research study is quite costly. If data can be aggregated across studies or if legacy data can be mined to answer new questions, the cost savings would be tremendous. Some examples of reviewing aggregated or legacy data are:

1. Evaluate safety issues.
2. Review completed studies to understand effects of placebo or to obtain information on a placebo population.
3. Compare treatments for a similar indication.
4. Assess pre-clinical studies of drugs in different classes to predict probability of failure of studies in human subjects. At a certain probability of failure, the human study would not be run.
5. Assess legacy studies to calculate sample size based on variance of endpoints.
6. Assess the opportunity for new indications.

Sharing Data from EHRs for Research Purposes – Additional efficiencies can be realized by using data directly from EHRs vs. re-entering research information into a separate system. Such data can be used to support clinical research studies, safety reporting, biosurveillance, clinical trial registration, study or patient registries or other research needs.

2.6 Alignment of Clinical Research and Healthcare Standards

The use of EHR data to support research could shorten the time needed for research information to inform healthcare. For this reason it is essential that the data that supports both healthcare and research be aligned. The need to address this alignment is consistent with the CDISC Mission Statement. This is also consistent with collaborations that CDISC has deemed important and increasingly added since 2001 (e.g. with HL7, ISO and the Joint Initiative Council).

2.7 Improve Standards Lifecycle Management

CDISC SHARE should improve the way that standards are developed and maintained more efficient. Such benefits may be manifested in a number of ways or processes as follows:

Initial Development timeframes (in particular new areas, e.g. efficacy) – Development of new therapeutic area standards and aligning with controlled terminology is time-consuming and can take 1-2 years (or longer depending on complexity per domain. The stakeholders need a more timely delivery of new domains and are increasingly requesting new therapeutic standards (e.g., tuberculosis, cardiology).

Approval times – The approval cycle for a new standard is quite lengthy. Depending on how much new content is in a draft release of a standard, reviewers have thirty to sixty days to review and provide comments. Then the standards development team must address each comment, develop consensus about the resolution, and provide a rationale for either changing the standard or not. The approval cycle and subsequent update of the draft standard can take a year or more.

Maintenance of a central repository and also company-specific concepts – In addition to industry standards, each company has their own company-specific content. A large company can have a team of several standards maintenance staff, which are very experienced and expensive.

Governance – The process of maintaining standards often more of an art that draws on the experience and intuition of the staff than a clear repeatable process that can be consistently used for standards-related tasks.

2.8 Enable Computable Semantic Interoperability

There are four pillars of computer semantic interoperability (CSI), which are required but not sufficient to obtain CSI. The four pillars are: 1) a common information model spanning all domains of interest, 2) a computationally robust data type specification, 3) a robust infrastructure for specifying and binding to controlled terminology and, 4) a formal, top-down development process. ***CDISC SHARE is the computationally tractable implementation of the first 3 pillars.***

3 Stakeholders and Stakeholder Analysis

Over the last few months considerable time and effort has been utilized to talk to CDISC members and other SHARE stakeholders. This analysis included a) ~ 50 teleconference with well over 70 individuals including participants from Denmark, Canada, Australia, Europe, US, about 10% from the CDISC Advisory Board; b) a survey with the CDISC Advisory Board; c) a Consensus Building session during the HL7 Working Group Meetings; a webinar on 6 October and informal discussions. This resulted in a tremendous amount of data, a representative sampling of which we have included below. This stakeholder analysis also provided information in the areas of risks, benefits, business models, potential partners and concurrent related work.

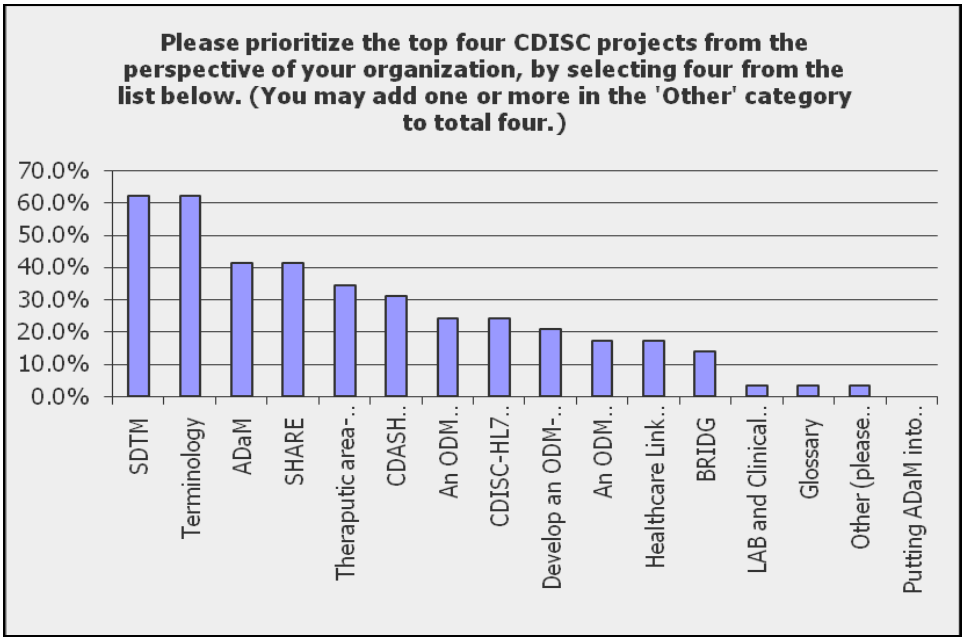
3.1 Online Survey with CDISC Advisory Board

Our advisory board played a key part in the stakeholder analysis and some of the results of an online survey that was conducted with them are included below (again full details are available on request).

Please prioritize the top four CDISC projects from the perspective of your organization, by selecting four from the list below. (You may add one or more in the 'Other' category to total four.)		
Answer Options	Response Percent	Response Count
SDTM	62.1%	18
Terminology	62.1%	18
ADaM	41.4%	12
CDISC SHARE	41.4%	12

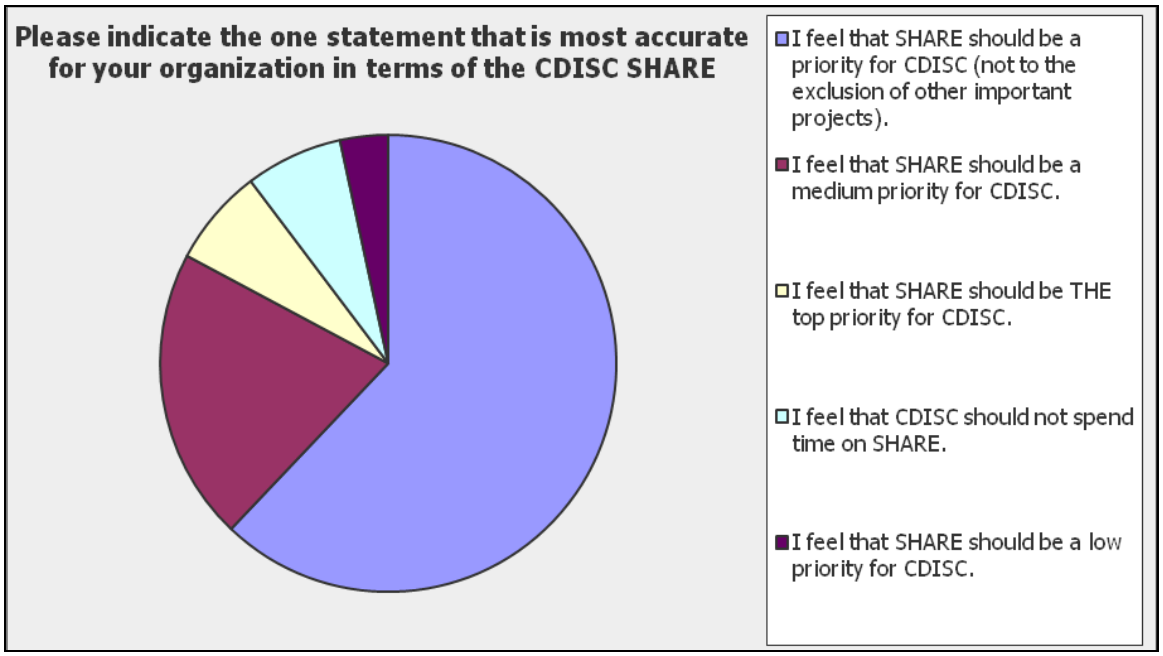
Theraputic area-specific standards (with or without CDISC SHARE)	34.5%	10
CDASH Implementation Guide	31.0%	9
An ODM Implementation for CDASH	24.1%	7
CDISC-HL7 Messages (trial design, study data, i.e. SDTM content)	24.1%	7
Develop an ODM-based eSubmission Mechanism (SDTM and ADaM)	20.7%	6
An ODM Implementation for the Protocol Representation Model	17.2%	5
Healthcare Link (EHR-Clinical Research standards and processes)	17.2%	5
BRIDG	13.8%	4
LAB and Clinical Genomics	3.4%	1
Glossary	3.4%	1
Other (please specify)	3.4%	1
Putting ADaM into BRIDG (Statistics Domain Analysis Model)	0.0%	0
answered question		29
skipped question		0

No.	Other (please specify)
1	We would like to see the final version of the metadata implementation guide



Please indicate the one statement that is most accurate for your organization in terms of the CDISC SHARE (CDISC SHARE Health and Research Electronic Library) project (which has also been called the CDISC SHARE semantic repository or metadata repository). Select one response

Answer Options	Response Percent	Response Count
I feel that CDISC SHARE should be a priority for CDISC (not to the exclusion of other important projects).	62.1%	18
I feel that CDISC SHARE should be a medium priority for CDISC.	20.7%	6
I feel that CDISC SHARE should be THE top priority for CDISC.	6.9%	2
I feel that CDISC should not spend time on CDISC SHARE.	6.9%	2
I feel that CDISC SHARE should be a low priority for CDISC.	3.4%	1
answered question		29
skipped question		0



3.2 Key Information from Stakeholders

There were some key themes that came out of the various interviews and surveys that were conducted and these have driven CDISC SHARE forward:

1. **All stakeholders interviewed felt that CDISC SHARE is basically a good idea and that it is needed.**
2. **Many stakeholders commented on the fact that this is big, it will not be easy, it will take time and it will be costly. A few cautioned that the scope needs to be managed.**
3. **CDISC SHARE needs to address clinical research broadly (biopharmaceutical companies as well as academic institutions) and the related areas of healthcare to ensure alignment.**
4. **CDISC SHARE is not as much a technology issue as a political challenge. CDISC is not a technology company.**
5. **Adoption will require careful communication about what we are doing, the benefits of CDISC SHARE and why this will bring value to the stakeholders**
6. **Many current related efforts were mentioned; most of these could either turn into collaborators or they could be competitors, depending on how CDISC proceeds and engages them (or not). (See Section on Current Related Work.)**
7. **Several are anxious to partner in some way (e.g. AHIMA, HL7, OpenEHR, BioIT Alliance, NCI, NICHD, certain pharmaceutical companies and technology providers); others cautiously offered support at least in terms of collaboration from the perspective of the representatives with whom we spoke (e.g. AMIA, IHTSDO, ONC, HITSP, CDC, certain biopharmaceutical companies, FDA); while others are not yet certain if/how to support this (e.g. SCO, NLM, certain academic institutions/CTSAs).**

The information from 3.1 and 3.2 was made available for the SHARE Scope and Vision document and were presented to the CDISC Board on 18 September 2009. Further stakeholder initiatives took place subsequently.

3.3 SHARE Consensus Reaching Discussion

A session at the HL7 Working Group Meetings in Atlanta, GA in September 2009 was designed to build consensus around the value of SHARE for the vision of entering data once for multiple purposes, with a specific goal of aligning standards for use of EHR data for research, quality, public health and safety. Representatives were included from AHIMA, CDISC, HL7, Centers for Disease Control, National Quality Forum and others. The presentation to stimulate discussion is available.

3.4 Global Webinar

A webinar on SHARE and next steps was presented to an audience of well over 200 participants globally on 6 October 2009. There was a Q&A session. The slides from this webinar and the Q&A session are available.

3.5 Stakeholder Analysis Conclusion

The stakeholders with whom we spoke were encouraging and generally very positive about CDISC SHARE and about CDISC being an appropriate leader. There were concerns expressed

about the scope and a long-term business model along with engaging the right partners. There was broad agreement that CDISC SHARE needs to accommodate clinical research as well as the significant amount of overlapping clinical care/ healthcare terminology and concepts. There were also cautions about how this should be communicated to ensure buy-in and adoption, particularly for the academic researchers. There are numerous efforts that are related to CDISC SHARE. For the most part, there is an opportunity to turn the majority into collaborators (see the section on Current Related Work); however, if not careful, some of these could well end up becoming competitors.

4 Related Current Work

Based upon the CDISC Collaborations and the Stakeholder Analysis, there are several organizations or initiatives that are doing work that would be considered related to the CDISC SHARE project.

The Business Models section of this document provides certain possible opportunities for the potential collaborators, which would leverage the CDISC proven "Strength Through Collaboration".

4.1 Standards Development Organizations (SDOs)

1. Health Level Seven (HL7) – www.hl7.org: has had an Associate Charter Agreement (MOU) with CDISC since 2001. Potential related efforts are the Clinical Interoperability Council (CIC), the Clinical Information Interchange Committee (CIIC), Clinical Document Architecture (CDA) Templates, Regulated Clinical Research Information Management (RCRIM) Workgroup, EHR Workgroup, Terminology.
2. International Health Terminology SDO (IHTSDO) - www.ihtsdo.org: is part of the Joint Initiative Council (along with ISO, CDISC, CEN, HL7); have an MOU with CDISC since mid-2009. Offers SNOMED free for appropriate clinical research, regardless of the country. U.S. has a country wide license.
3. Other SDOs and their collaborations, including ISO, JIC, JWG, CEN, BioIT Alliance and Pistoia Alliance (discovery standards), the SDO Charter Organization (SCO)
4. Universal Data Element Framework (UDEF) – www.udef.com (www.OpenGroup.org)

   The Data Indexing Standard to Reduce the Costs of Applications Integration and to Improve Data Discovery; UDEF provides semantic links, through assigning an intelligent, derived ID as an attribute of the data element, essentially labeling the element as a specific data element concept. When this UDEF ID exists in both source and target formats, it can then be used as an easy analysis point via a match report, and then as the primary pivot point for transformations between source and target.

   The Open Group assumed from AFEI the right to grant public use licensing of the UDEF. Ron Schuldt, Sr. Enterprise Data Architect, Lockheed Martin, originated the UDEF concept based on ISO/IEC 11179 Metadata standards approximately 15 years ago.
5. HITSP - Note that the Health Information Standards Panel (HITSP) is not an SDO; they identify standards to be used to support capabilities and use cases for EHRs. They work with SDOs.

4.2 Non-SDO Organizations or Initiatives

1. OpenEHR - http://www.openehr.org/home.html:

   Technically, *open*EHR is about creating **specifications, open source software and tools** for such a platform. In the clinical space, it is about creating high-quality, re-usable

clinical models of content and process - known as **archetypes** - along with formal interfaces to terminology.

2. United States Health Information Knowledgebase (USHIK) - http://ushik.ahrq.gov/registry/index.html?Referer=Index:

   Catalog of US standards, e.g. CHI, HITSP; hosted by AHRQ – Agency for Healthcare Research and Quality; not a reference standard

3. Public Health Information Network (**PHIN**) Vocabulary Access and Distribution System (**VADS**) (PHIN/VADS) - http://phinvads.cdc.gov/vads/WebHelp/Welcome_to_PHIN_VADS.htm :

   " A vocabulary repository and server which allows **CDC**'s public health partners to browse, search, and download vocabulary concepts required for PHIN messaging and applications"; hosted by Center for Disease Control (CDC); not a reference standard.

4. Informatics for Integrating Biology and the Bedside (i2b2) - https://www.i2b2.org/

   An NIH-funded National Center for Biomedical Computing based at Partners HealthCare System. The i2b2 Center is developing a scalable informatics framework that will bridge clinical research data and the vast data banks arising from basic science research in order to better understand the genetic bases of complex diseases. This knowledge will facilitate the design of targeted therapies for individual patients with diseases having genetic origins. The i2b2 is funded as a cooperative agreement with the National Institutes of Health.

5. Research Electronic Data Capture (RedCAP) – http://www.project-redcap.org/

   The **REDCap Consortium** is comprised of 57 active institutional partners from CTSA, GCRC, RCMI and other institutions, and it supports two secure, web-based applications (REDCap and REDCap Survey) designed exclusively to support data capture for research studies.

6. Clinical and Translational Science Awardees (CTSAs) - http://www.ncrr.nih.gov/clinical_research_resources/clinical_and_translational_science_awards/

   A national consortium of medical research institutions, funded through **Clinical and Translational Science Awards** (CTSA), is working together to improve the way biomedical research is conducted nationwide. Consortium members share a common vision to reduce the time it takes for laboratory discoveries to become treatments for patients, to engage communities in clinical research efforts and to train clinical and translational researchers. CTSA is funded through National Center for Research Resources (NCRR).

7. PhenX – https://www.phenx.org/

   PhenX is a three year project led by RTI International and funded by the National Human Genome Research Institute (NHGRI) to contribute to the integration of genetics and epidemiologic research

   PhenX has prioritized 20 research domains related to complex diseases and environmental exposures

   Consensus building is being used to develop a recommended minimal set of high priority measures for use in Genome-wide Association Studies (GWAS) and other large-scale genomic research efforts

High priority measures will maximize benefits of future research by enabling cross-study comparisons and analysis

8. National Cancer Institute (NCI) Enterprise Vocabulary Services (EVS) and NCI cancer Data Standards Repository (caDSR)/Mayo–LexGrid – http://ncicb.nci.nih.gov/NCICB/infrastructure/cacore_overview/vocabulary

   CBIIT bases its data semantics on controlled terminology supplied by the NCI Enterprise Vocabulary Services (EVS) Project. The NCI EVS represents a set of services and resources that address NCI's needs for controlled vocabulary as well as that of other key stakeholders. The EVS Project is a service of the Center for Biomedical Informatics and Information Technology (CBIIT).

   These services encompass terminology development and coding, terminology licensing, software development and licensing and operations support activities. From its inception, EVS has sought to address the broad spectrum of terminology needs at NCI. EVS provides the base upon which the data semantics of caCORE and caBIG initiatives depends, and houses the CDISC, FDA, HITSP and other controlled terminologies.

5 Major Features and Capabilities

CDISC SHARE is a warehouse of scientific concepts used in biomedical research and healthcare that includes all information about those concepts including: concept meaning, concept definition, variables associated with those concepts, code lists, data types, and relationships between concepts. The high level business capabilities and the business quality requirements are described below. A more detailed look at the business requirements can be found in the related Appendices zip file.

5.1 High Level Business Capabilities

Atomic Definition and Grouping: The solution will provide a definition for each concept and variable. The solution will also provide the ability to group concepts and variables.

Content Curation: The solution will need to provide a mechanism to curate the content and/or upload content from external parties. Prior to the curation process there will be content that will need to be uploaded as individual and/or collections of concepts at one time. There will also need to a mechanism to add, delete or modify concepts already in the library.

Content Accessibility (includes Storage and Retrieval): The solution shall provide 24/7/365 accessibility to the current version of the shared semantics in a view mode. All concepts and its current mappings and annotations should be available real-time. Any content that is in development will be accessible at the point of validation. The solution shall provide a mechanism for searching subsets of the information based on user-defined criteria. In addition to real-time access to the content, the solution will need to provide the capability to export the entire repository for input to an organization-specific repository.

Duplicate Resolution: The solution should prevent duplicate concepts being entered into the repository. In the past, there have been duplicate concepts within organization-specific dictionaries that have led to intra-organization harmonization efforts.

Authoritative Sources: The solution shall as much as possible provide content by reference (not content by value), so that the authoritative source of the concept is controlled. The versioning of content should be maintained in one location and linked in all applicable contexts.

Concept Definitions: The solution will include concept names, concept definitions, and relationships between concepts, and relationships between concepts and variables. The solution

will also address how to work with existing standards, variable and concept synonyms, and alternate definitions.

Variables: The solution will include variable names, variable definitions, valid value sets, data types, variable lengths and relationships between concepts and variables, variable grouping, and variable optionality.

Referenced Standards: The solution will contain the content from all CDISC standards including the Protocol Representation Model (PRM), Clinical Data Acquisition Standards Harmonization (CDASH), Study Data Tabulation Model (SDTM), Standard for the Exchange of Non-clinical Data (SEND), Analysis Data Model (ADaM), and Controlled Terminology.

Internationalization/Globalization: The solution will contain information pertinent to biomedical research and healthcare globally. The solution shall provide the appropriate references and/or attributes to enable use of concepts internationally.

Traceability: The solution shall provide a mechanism for tracing the origin of the concept back to its owner, and/or contributing organization. The information shall be provided in order to provide traceability and transparency for the users of the solution.

Governance: The solution will include a plan for governance at three levels:

1. Organizational Governance the governance as it relates to the provision of the entire CDISC SHARE service to a range of organizations (content providers and content consumers).
2. Content Governance the ongoing stewardship, support, and maintenance of the CDISC SHARE content. The governance should provide a mechanism for managing changes (additions, deletions, modifications, or merges) to concepts in the repository as well as concept resolution.
3. Technical Governance the ongoing input into the development of the tools used within the provision of the CDISC SHARE service such that the tools meet the needs of the service.

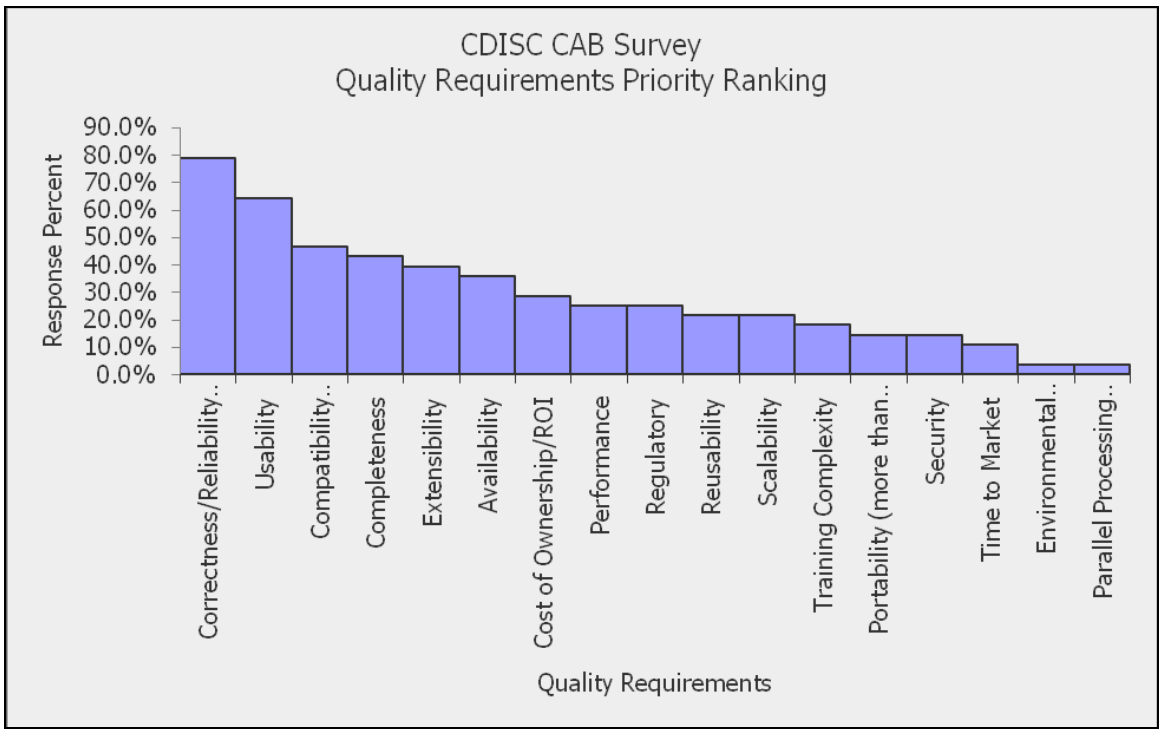
Change Management: The solution will include a well-documented process that ensures that standardized methods and procedures are used for efficient and prompt handling of CDISC SHARE change requests in order to minimize the number and impact of any incidents on the repository.

5.2 Quality Requirements

Quality Requirements are those requirements that focus most on the system architecture of a solution. The 90-member CDISC Advisory Board was asked to prioritize the top 5 quality requirements. The results of this survey are shown below. For a definition of each of these requirements see the appendices.

All of the following will be Quality Requirements for the technology to support CDISC SHARE. Please help us prioritize their importance by selecting the top five of the following Quality Requirements for the technology to support CDISC SHARE.		
Answer Options	Response Percent	Response Count
Correctness/Reliability for use in mapping data	78.6%	22
Usability	64.3%	18
Compatibility (interaction with other sponsor systems)	46.4%	13
Completeness	42.9%	12
Extensibility	39.3%	11

Availability	35.7%	10
Cost of Ownership/ROI	28.6%	8
Performance	25.0%	7
Regulatory	25.0%	7
Reusability	21.4%	6
Scalability	21.4%	6
Training Complexity	17.9%	5
Portability (more than one operating environment)	14.3%	4
Security	14.3%	4
Time to Market	10.7%	3
Environmental (conditions in which the system must function)	3.6%	1
Parallel Processing (fulfill requirements simultaneously using duplicated rather than CDISC SHAREd resources)	3.6%	1
	answered question	28
	skipped question	1



5.3 Detailed Business Requirements

The CDISC SHARE Business Requirements activities included broad input from many stakeholders across the industry, and in particular global biopharmaceutical companies and CROs. Requirements development was distributed across two primary teams: (1) Business Requirements and (2) Content Governance Requirements. Smaller sub-teams were formed as needed to focus on specific CDISC SHARE storyboards and specify the requirements for particular industry roles or "actors." Due to the large amount of material developed the information is presented in a separate document.

6 Pilot

6.1 Purpose

The purpose of the pilot is to address two risks identified at the start of the inception phase:

1. Can definitions taken from multiple sources be merged into a single version agreed to by all parties and can this be done within a timeframe that makes business sense
2. Can high-quality definitions be created and can ontologies help in ensuring such and avoid duplicate definitions being created.

A secondary aim of the pilot was to provide any relevant lessons to subsequent development work.

6.2 Technology Selection

To address the above risks it was necessary to obtain a piece of technology that supported the aims of the pilot. The selection of the technology commenced with a meeting in San Francisco on the 27th May At that time it was presumed that Tolven would supply the technology for the pilot. A proposal was received from Tolven on the 4th June. On the 5th June Mayo indicated that they also wished to submit a proposal. This was received on the 17th June. At the same time NCI indicated that they would provide financial support for the project. Key technology issues were identified during this period and are summarised below.

Costs from Mayo and Tolven were similar, the Mayo proposal was more considered and supportive of CDISC's aims but feedback from third parties regarding Tolven were not positive. Further conversations with Mayo resulted in their agreement to support BRIDG 2.x semantics.

Thus CDISC had the choice of two comparable technical solutions at equivalent costs. The Mayo option had the advantage of greater cooperation with NCI support.

Question	***Tolven***	***Mayo***
1. Regarding its underlying ontological representations:		
a) -- does the system support RIM semantics?	Yes	No -> Yes
b) -- does the system support BRIDG 2.x semantics?	Yes	No -> Yes
c) -- does the system support ISO 21090 DTs (or HL7R2 ADTs)?	Yes	Yes
d) – does the system support SNOMED?	Yes	Yes
e) – does the system support MedDRA?	Yes	Yes
f) – does the system support CDASH?	Yes	Yes

2. RE system functionality: a) – does the system support cross-terminology searching?	Yes	Yes
3. What is your estimation of the time involved in having a system installed and operational?	2 Weeks	2 Weeks
4. What is your estimation of the time/complexity of training -- <<based on previous experience within inexperienced domain experts entering terminology>> -- involved in terms of the goals of the pilot?	8 Hours	8 Hours
5. Please list any other relevant positive or negative issues that have been inadvertently left out of this brief list that you believe would affect the cost, efficiency, effectiveness, or overall quality of the use of the system in the context of achieving the two goals of the CDISC pilot project.		

6.3 Method

Using the semantic wiki tool (LexGRID) provided by Mayo, 50 oncology data elements from 5 volunteer organizations (Mayo Clinic, GSK, MD Anderson, Eli Lilly and Genzyme) along with valid value lists have been identified and loaded into the system. The team has been following a process to align equivalent data elements resulting in a single consensus version. As the team undertake the alignment work the process is refined and the wiki amended to better support the process. Metrics to evaluate the process and use of the wiki, along with benefits and risks, will be collected and reported. Specifically these metrics will include time needed to prepare, load and add concept references as well as the time needed for the harmonization process.

The wiki has been loaded with various terminologies/dictionaries such as the NCI Thesaurus, CDISC Controlled Terminology (CT), the SNOMED CT, and ICD 9 and 10 along with the BRIDG structure. This will permit an assessment of how these support the process of aligning the definitions from the various contributing organizations and permit an assessment of how these support the improvement in the quality of the definitions created and prevent duplicate definitions being created.

The wiki can be found at http://informatics.mayo.edu/cshare/index.php/Main_Page

6.4 Results to Date

The work on the pilot is still ongoing at the time of writing. Significant progress has been made since the start of the project in April. As noted above the various terminologies and source data elements have been loaded. Good progress has also been made on refining the process of harmonization of the elements into a consensus version.

It has quickly become apparent that this process is not as easy as it first appears and that the merge and search functions that are offered to users of the tool are a key component in delivering both ease-of-use and quality in the development process.

There are also issues resulting from the sheer scale of information presented to users and making it easier for users to see what they are working on; to provide tools that subdivide the problem space to make it manageable. There is also a need to provide an ability to collect data elements into meaningful collections, the concept idea.

These issues are now being investigated and improvements to the process implemented within the tool. This will then allow the various metrics to be collected and definitive conclusions to be determined.

At the time of writing it is possible to provide an informal assessment of the pilot outcomes:

1. Can definitions taken from multiple sources be merged into a single version agreed to by all parties and can this be done within a timeframe that makes business sense – Yes, 70% confidence.
2. Can high-quality definitions be created and can ontologies help in ensuring such and avoid duplicate definitions being created. – Yes 50% confidence.

In terms of informing any subsequent development valuable lessons are being learnt. Initial thoughts have been captured in notes included within the appendices. It should be noted that these are preliminary and have yet to be formally reviewed.

7 Next Steps

The Next Steps for SHARE include completion and continuation of the pilot ; development of the tool ; development of collaborations, governance and new content.

7.1 Pilot Next Steps

The first next step will be completion of the Pilot, specifically development of the data elements; investigation of collections ("clumps") of information; collection of metrics for the stages of the process; collection of user experiences after which the process should be refined as much as possible and then testing of different scenarios such as support for EHR integration and CDISC standards such as SDTM and CDASH.

7.2 Development of the SHARE Tool

Another important step is that the National Cancer Institute has agreed to fund the development of the toolset. NCI's metadata repository development will be based on requirements from a wide variety of groups to include research and broader healthcare standards; SDOs - HL7, CDISC and other; regulatory entities; pharmaceutical; providers and vendors. The new ISO 11179 standard MDR will be based on a federated, distributed architecture; meaning it will be decentralized, allowing for multiple peer repositories. It will also be a platform independent model, to be openly shareable and will allow for modular development of many and varied customized applications and services for different users, but with a common foundation and generic API. It will be open, platform and vendor neutral, distributable, and shareable. CDISC will participate in the development team for the SHARE tool.

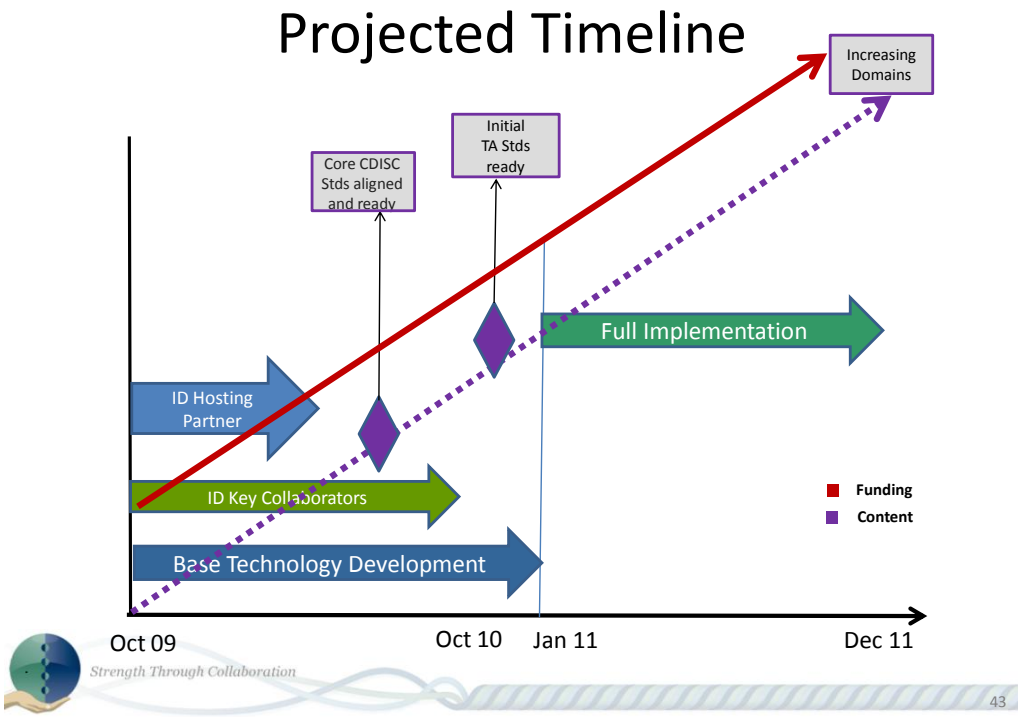
7.3 SHARE as a Reference Standard Repository

CDISC brings to the value of SHARE its expertise in productive collaboration and its process as a standards development organization with a Liaison A status to ISO, formal relationship with HL7 and member of the US ISO TAG and UK National Standards Body. Hence, CDISC can provide the means to developing a global reference standard library – SHARE.

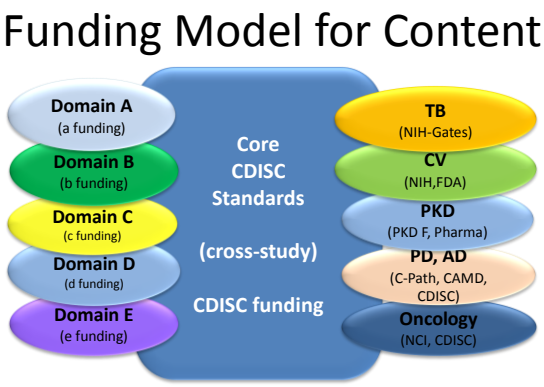
In addition the existing CDISC standards will be migrated into SHARE and will be aligned through a domain analysis model, initially the Biomedical Research Integrated Domain Group (BRIDG) Model.

7.4 Timeline and Funding

While the tool is being developed and CDISC standards are being migrated into SHARE, the development of content standards (particularly therapeutic area/efficacy standards) will proceed. These activities will require funding. There are different areas where funding and support are needed. These include the formation of a governance body for SHARE, a governance process for development of new standards that will go into share (ideally without duplication of other standards), technology and hosting and a collaborative environment and process for funding and resources to contribute to SHARE. They are depicted in the following graph, along with tentative timelines for steps to make SHARE a reality.



In order to fund the content development, CDISC plans to use the model that they have been following to date, which includes a variety of opportunities to obtain the needed funding and resources. This is depicted in the following figure.



Funding Options (a, b, c, d, e) = foundations, NIH centers or other government funding (e.g. ONC, AHRQ, IMI), biopharma, PhRMA, BIO, contributed resources, or other additional creative solutions.

7.5 To Get Involved

Those wishing to get involved can join CDISC, join a team, become a key stakeholder partner, review content standards as they come out, lead a team, contribute funding or many other ideas to support the continued development of terminology and content standards and SHARE.

CDISC, Inc.
15907 Two Rivers Cove, Austin, Texas 78717
http://www.cdisc.org