

HIT Standards Committee Vocabulary Task Force 1 September Meeting – Written Testimony submitted by Bron Kisler, CDISC

Opening Remarks: Putting end users and patients first

It is clear standard vocabularies and value sets are key ingredients for semantic interoperability and can lead to greater efficiencies in data collection, exchange, aggregation and analysis. However, to ensure adoption they must be easy to access, download and culminate into useful products for end users. Clinical and scientific knowledge must be translated effectively into standards and electronic representations. Clinicians and clinical researchers would benefit greatly from common forms, comprised of standard and easy-to-use data collection elements. Pushing data standards (including terminology and value sets) upstream to data collection is the best way to ensure semantic interoperability down the full data chain to regulatory reporting and review and improve the symbiosis of data across clinical domains. This will create efficiencies not yet realized. Consistency in electronic data collection as well as data collection in resource poor settings will facilitate the ability to aggregate critical information downstream for analysis.

The US FDA is taking significant steps to: (1) improve the efficiency and quality of regulatory reviews for potential new therapies; and (2) improve the efficiency and feasibility of developing new and innovative therapies. FDA CDER and CBER divisions are seeking ways to substantially reduce the time it takes to aggregate, access and analyze data in regulatory submissions, and The Office of Critical Path Programs is seeking to bridge the gap between discovery of new compounds and the testing of those compounds in humans. Increasingly, the FDA and international health organizations are supporting and encouraging new initiatives to address problems in specific disease categories. Data Standards are recognized as a key enabler in the FDA achieving a key public health goal – “provide timely patient access to safe and effective new drugs”.

FDA Commissioner Margaret Hamburg recently hosted the launch of a new initiative for Tuberculosis (TB) called Critical Path to TB Regimens (CPTTR), which represents a partnership between the Bill & Melinda Gates Foundation, Global TB Alliance, Critical Path Institute and many other organizations. CPTTR seeks to create an innovative drug development framework that will cut research, development and approval time for 4 TB drugs and combination regimens from 24-years to 6-years. Once again data standards are a critical component. With TB killing more than 1.8 million people each year (primarily in developing countries), the development of new, simpler and more effective drug regimens is not just an option, but a major public health imperative.

By putting patients ahead of profits and politics, CPTTR has been able to quickly bring together 100’s international regulators, global health organizations, pharmaceutical industry organizations and clinicians representing academia, professional societies and foundations from around the world. Focus on patients has proved the key denominator for getting so many organizations (some with competing interests) moving quickly in the same direction with a common and determined purpose.

As the US moves forward with HIT standards, it’s important to keep the welfare of patients front and center, particularly as federal health agencies from developing and emerging countries look to the US for global leadership and direction.

1. What are the requirements for a centralized infrastructure to implement “one-stop shopping” for obtaining value sets, subsets, and vocabularies for meaningful use?

CDISC has been developing and deploying standard controlled terminology and value sets through a partnership with NCI since 2005. NCI EVS provides a suite of terminology services to CDISC and other key stakeholder organizations (e.g., US FDA, NIH institutes) ensuring rapid development and deployment of new terminology sets. CDISC controlled terminology (housed by the NCI EVS organization) provides a consistent semantic foundation for CDISC data standards across the full clinical research data chain from protocol representation and data collection through analysis, regulatory reporting and review. CDISC’s standard for data collection – CDASH – was developed as part of the FDA Critical Path Initiative, and CDISC’s standard for regulatory submission and review – SDTM – is referenced in FDA Regulatory Guidance and the recently released CDER Data Standards Plan. These standards have been implemented around world with downloads tracked nearly 15,000 times in over 60 countries. To keep pace with implementation demand, CDISC has established coordinating committees and user groups across North America, Europe and Asia.

As part of the SHARE project, CDISC has compiled detailed user and system requirements across a broad clinical stakeholder community, including content requirements, governance requirements, requirements testing with Mayo using LexEVS and detailed stakeholder analysis. CDISC SHARE seeks to provide a single reference standard not only for CDISC and regulatory clinical research but also for other clinical use cases including public health and quality reporting. The idea is for standard data elements to support multiple purposes and clinical use cases, where value sets may be reused and repurposed. CDISC SHARE and the new NCI Semantic Infrastructure consider terminology in the context of a layered framework that enables standard data models, clinical data elements with value sets and controlled terminology to be coupled and aligned electronically. A centralized infrastructure to implement “one stop shopping” should:

- Leverage existing work, federal infrastructure and key emerging projects
- Support global clinical research
- Support development and deployment of a common reference standard that can be used across different clinical use cases such as safety reporting, public health and quality measures
- Include links to and between selected controlled vocabularies, data elements and value sets
- Have an intuitive interface with low barrier of entry for access, download and use of content through common web-browsers
- Provide content that is open and free of proprietary licensing encumbrances
- Incorporate a collaborative semantic framework and services that enable key stakeholder organizations to share standards, including terminology and value sets, for cross-harmonization purposes (e.g. CDISC, US FDA, NCI, NQF etc.)

NOTE: For more detail on CDISC SHARE, please refer to CDISC’s HIT Standards Committee Vocabulary Task Force testimony 23 February 2010.

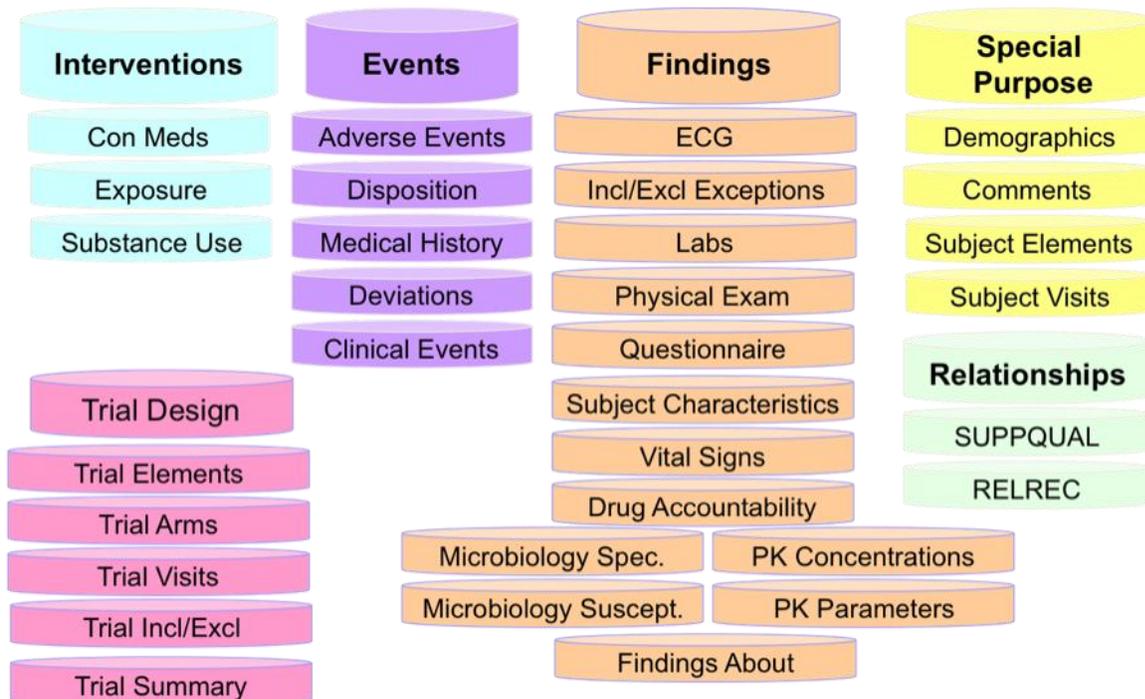
2. Which requirements or functionalities are urgent, i.e., absolutely required to support “meaningful use”? Which would be most useful immediately? What would be a staged approach over time to get to the desired end state?

As part of the SHARE project CDISC has documented and tested many requirements applicable to “meaning use”. If helpful, CDISC will be glad to share this documentation with the HIT Standards Committee and Vocabulary Task Force. Requirements addressed first should include:

- Public domain open content with license; platform and vendor neutral
- Sustainable business model and infrastructure insulated from funding fluctuations and not subject to changing political wind
- Straightforward and intuitive process for user community access and download
- Dedicated staff and support to address new and evolving requirements for large user communities with the understanding that some communities extend beyond the US
- Ability for users to extend standard value sets to meet local clinical trial needs
- Services and resources to ensure successful value set deployment as well as ongoing management and alignment between value sets, super sets and subsets
- Ability for users to provide feedback on terminology and value sets with rapid turnaround response for requests
- Linking together of multiple vocabularies into a central repository space, bringing together best-of-breed to meet different clinical use cases
- System uptime and security of community-approved content that is tagged and stored centrally
- Version control and tracking for both standard vocabulary and value sets

3. Where are you using value sets and subsets? For what domains? How many value set and subsets?

CDISC has developed standard data models, controlled terminology and associated value sets (aka codelists) for use across the clinical trial continuum from data collection (CDASH standard) thru regulatory submissions (SDTM standard) in human clinical trials. These standards are currently in production and being used around the world. To date, the CDASH and SDTM standards have primarily focused on “safety data” domains or the information collected across all clinical trials. The SDTM standard contains over 150 unique and shared standard data elements across the 32 domains depicted in the diagram below for SDTM ver3.1.2.



Across SDTM ver3.1.2 and CDASH ver1.0, there are 90 shared terminology codelists and supersets of terms that comprise nearly 4500 coded terms and definitions. Codelists contain standard lists of coded terms approved by the CDISC global user community.

The CDASH, SDTM and Terminology standards are currently being augmented and extended for “efficacy data” and disease-specific content. Standards are currently in production for Tuberculosis and Acute Coronary Syndrome. These were jointly developed and approved by both CDISC and HL7. CDISC, FDA and several NIH institutes are currently applying lessons learned from these initial disease-specific projects, and working with clinicians to develop new standards for Oncology, Cardiovascular Disease, Kidney Disease, Neurological Disorders and Imaging data. The intent is to enhance and optimize the development paradigm such that clinical and scientific knowledge can be effectively translated into useful electronic representations. Each new disease area will extend the CDASH and SDTM with new clinically-specific data elements and value sets.

Additionally, CDISC has mirrored and extended the SDTM standard to address animal studies. This standard is known as SEND or Standards for the Exchange of Non-Clinical Data. The SEND standard includes 50 unique codelists with 12 others shared with SDTM. SEND is comprised of nearly 3800 terms; 2225 of which are unique to the SEND standard.

Finally, the BRIDG model is an overarching data model that links clinical research with the HL7 RIM. It was developed collaboratively between CDISC, NCI, FDA and HL7, and passed the standards-balloting process for CDISC, HL7 and ISO earlier this year. The BRIDG model includes more than 1600 data elements represented in the NCI Semantic Infrastructure. BRIDG data elements are currently being aligned with standard controlled terminology and values sets developed by CDISC, FDA and others and published in the NCI EVS terminology space.

4. In your experience with creating, disseminating, updating and/or using value sets, subsets, and entire vocabularies, what works and what does not work?

To ensure uptake and success it’s critical to keep the needs of implementers and end users in mind. A process must be in place to address evolving end user needs as well as a growing and diverse group of global stakeholders – international health organizations; global drug developers and other industry organizations; academic and clinical communities including foundations and professional societies. Dissemination and use of terminology value sets and subsets must be intuitive to the end user. Extensive training should not be required, or significant uptake is unlikely.

When considering a central repository for standards, strict content control must be maintained at all times to ensure any changes to approved standards (terminology, value sets etc.) is handled by the owning organization and its community through a formal change control and maintenance process.

Complex controlled vocabularies are of little use without dedicated experts and support, providing a suite of terminology services that ensure the vocabulary meets the needs of its user community. Terminology services must be wrapped around controlled terminology to support ongoing adaptation and development, publication, distribution and change management. Examples of key terminology services include:

- Subject matter expertise extending into specific disease specialties
- Definition writing and analysis
- Terminology tagging and sub-setting

- Terminology coding that provides clean cross-harmonization between key stakeholders
- The ability to share standards across stakeholder organization and clinical communities without costly mapping is critically important
- Handling of terminology requests and maintenance
- Links to other controlled terminologies as needed (e.g. FDA, MedDRA, LOINC, ISO etc.)

Achieving and maintaining value set alignment and harmonization across different key stakeholder organizations has proved to be hard work. However, CDISC has been able to achieve this through shared use and services of the NCI EVS Organization and Terminology infrastructure. This has enabled alignment between CDISC, FDA, NCI and others. Value set harmonization and standardization that cuts across clinical research, clinical care, public health, and quality metrics will prove even more challenging and will likely only be achieved using a similar central command center approach such as that of NCI Semantic Infrastructure and CDISC SHARE.

Terminology Governance must be in place to support terminology evolution, allowing for terminology and value sets to be adapted and extended to meet user needs.

5. What human resources does it take to implement and manage value sets, subsets, and entire vocabularies? Informaticists? Clinicians? IT people? How are you organized?

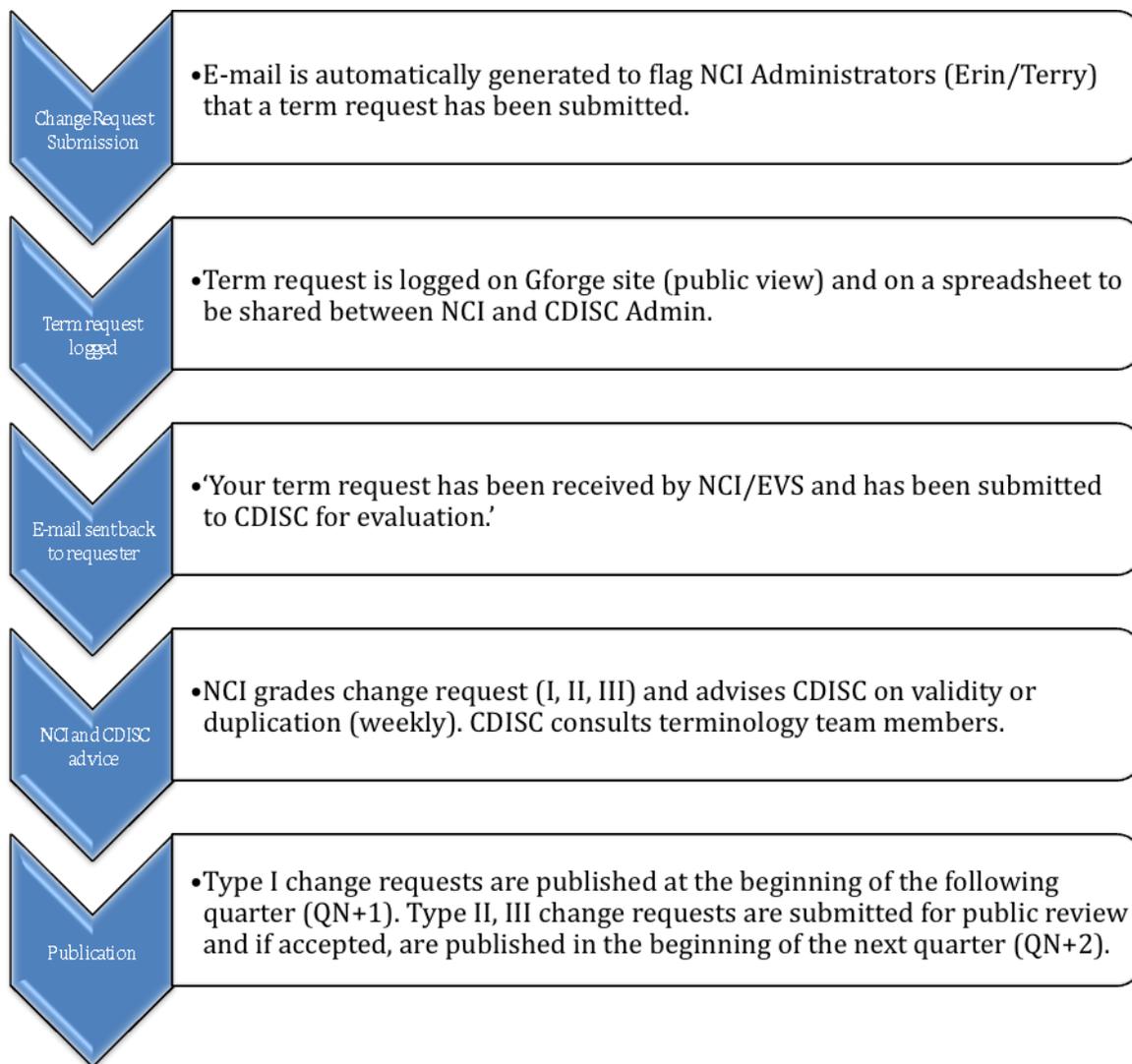
CDISC identifies, adapts and approves standard terminology and value sets via a global consensus process, and partners with NCI EVS for controlled terminology services, distribution, publication, maintenance as well as terminology and value set management. Standards development teams are multi-disciplinary and include standards and terminology experts, clinicians, clinical data managers, regulatory representatives as well as international cross-representation.

6. What national resources and services could be leveraged to reduce the level of effort required for local implementations? What is the irreducible minimum of local work at an implementation site, or within an organization or system?

- NCI's lessons learned in Oncology as well as existing terminology infrastructure, services and staff experience.
- Existing CDISC standards and expertise in developing, adjudicating and delivering new standard content.
- FDA standards and expertise in evaluating efficacy and safety of therapies.

7. What is your maintenance process? How do you manage updates?

Creation and adaptation of CDISC controlled terminology and value sets begins with input and feedback from the CDISC global user community. Terminology and value set requests may be submitted by any organization or individual via the CDISC website and web-link to the NCI term request webpage. Formal changes to and additions of terms and values sets are tightly controlled. The diagram on the next page depicts the high-level change request process established between CDISC and NCI. All new and revised sets of terms are distributed and approved via CDISC's standards development process, ensuring industry consensus before publication of new version releases. Maintenance of CDISC terminology is conducted on a quarterly basis. Per the request of global drug developers 2 updates are aligned with the MedDRA release cycles in March and November.



8. What metadata do you maintain and how do you maintain versioning?

CDISC metadata is currently maintained as part of the standards specification and versioning of such. CDISC maintains metadata standard specifications for data tabulations (Study Data Tabulation Model), analysis data (Analysis Data Model), and case report form specifications (CDASH). Specifications are currently maintained as PDF documents. When new versions are implemented, version numbers are incremented. CDISC Metadata will be represented electronically when CDISC SHARE goes live in 2011. This is a key area where CDISC continues to gather and refine requirements for SHARE development. CDISC welcomes input from other organizations.

9. Is there a difference between versioning for clinical documentation vs. versioning for reported measures, i.e., when do you go live with a change in the EHR vs. when do you use the new version of measures?

This seems implementation specific and not applicable to CDISC. However, in the future CDISC will begin to develop and publish standard data collection forms. It would seem prudent to control and version common forms that contain collections of data elements for a specific purpose (e.g. Oncology). CDISC SHARE is being developed to support this capability as an end user requirement.

10. How do you manage versioning in clinical decision support vs. changes in value sets?

CDISC has an established terminology request mechanism that may be accessed via the website. This is described in section 7 above.

11. How does an application know which value set is for which purpose? How is the specific context for a value set maintained at the message data element level of specificity? How is the English language intent of the value set context documented and maintained?

12. What are lessons learned about web links vs. storage of the vocabulary or other artifact in a physical repository?

CDISC has chosen an approach using a centralized terminology repository, resources and services provided by NCI. CDISC terms are tagged in NCI Thesaurus where CDISC terminology subsets can be easily published in multiple formats. CDISC terms are linked to other external controlled terminologies as needed. This central command center approach reduces the resource burden required for terminology maintenance and helps provide necessary quality control over community-approved terms and value sets. Also, since other key stakeholders (e.g. US FDA) use NCI terminology services, this provides a mechanism to share standard terminology and value sets and keep them aligned.

13. How do you manage distribution of update to multiple sites?

CDISC terminology may be accessed via the CDISC website and a web-link directly to published CDISC terminology subsets and value set hosted by NCI EVS. Terminology and value sets are managed centrally, ensuring continuity of access, download and content around the world. This is critically important for global organizations implementing CDISC standards across many geographic regions simultaneously.

Variables in the standards specification contain web-links directly to applicable terminology value sets. CDISC provides notification of new quarterly terminology releases via the home page of the CDISC website and CDISC-HL7 team distribution lists.

14. Where is local customization appropriate and how much customization is acceptable?

CDISC is a standards development organization, and hence does not disseminate customized versions of standards. CDISC provides standard value sets capturing the most commonly used terms for a particular field. If a particular region or organization has a specific need they wish addressed, they can submit a terminology request. This ensures a standard remains precisely that...a Standard. On the other hand, if a local or trial-specific variation is needed that does not fit within the standards paradigm, many CDISC value sets are extensible allowing organizations to extend and add terms as needed. Customization should be minimized and only done as an extension to a standard (or perhaps a translation into another language); otherwise, it is no longer a standard.

15. How do you manage distribution of updates with local variations and optionality? Unique subsets? Local mappings?

16. What has to be local in an EHR implementation vs. what can be external in a vocabulary repository?

If a core dataset can be agreed to and standardized, this would help EHR vendors immensely. Additionally, if an EHR can support the CDASH dataset (essentially a subset of the CCD), then the

majority of the information for research studies can be readily provided by any EHR vendor through an integration profile. This can be the same for the quality datasets or AE reports.

17. What functions are required that users have not yet appreciated?

The need for a central terminology and semantic infrastructure, terminology services and dedicated resources as well as the costs associated with this business model

Standards initiatives and products linking clinical research with healthcare such as: the BRIDG model; IHE Integration profile RFD (HITSP # 158) that aligns core CDASH data collection elements with EHR systems; and disease-specific standards aligned across CDISC and HL7. The CDISC community primarily focuses on the research aspects, not fully appreciating the value of connecting to healthcare.