PhUSE 2013

Paper DH10

Time travel for librarians: versioning complex standards library metadata for past, present or future retrieval

Nico De Leeuw, Business & Decision Life Sciences, Brussels, Belgium
Ine Felsenstein, Business & Decision Life Sciences, Brussels, Belgium

ABSTRACT

The approach for maintaining clinical data standards libraries by librarians is usually done using an off-the-shelf medium such as Excel. This is error prone as, generally, the management and quality control of data standards are conducted manually.

With this approach, many problems can potentially arise, for example:

- How best to version data standards records when they have been updated?
- How to retrieve a record and linked records which were applicable at a specific date?
- How to check that the data standards meet the business requirements?
- How do you ensure that data is not duplicated, etc.?

To resolve these problems, an automated system is proposed which is customized to manage all clinical data standards metadata, by means of libraries. Several functionalities are built into the system to link related data standards and to ensure no duplication is required. Company defined checks can be added to the system so as to ensure that ever changing clinical trial data regulations and business requirements can be met at all times.

One of the key elements to versioning the data standards library objects is the use of dates giving users the opportunity to time travel through their libraries.

KEYWORDS

clinical data standards library, version control, technical platform, data governance, performance management

INTRODUCTION

Clinical data standardization is closely interlinked with maintaining data standards metadata. The latter is becoming increasingly complex due to the volume and frequently changing regulatory and data standards requirements.

Due to emerging business challenges, data standardization becomes a necessary step to further improve the process of the clinical trials data submission.

Therefore, a solution to handle the regulatory and business challenges is proposed through the set-up of an automated Clinical Data Standards platform together with a Data Standards Governance framework.

This technical platform allows to version complex data standards metadata. Next to the platform it is equally important to set-up an appropriate Data Standards Governance structure with the corresponding process flow, roles and responsibilities, and performance management objectives.

IMPORTANCE OF CLINICAL DATA STANDARDIZATION

Before we explore the appropriate system and techniques for versioning clinical data standards, it is important to clearly understand why clinical data standardization is crucial in the first place.

EVOLUTION OF THE REGULATORY REQUIREMENTS

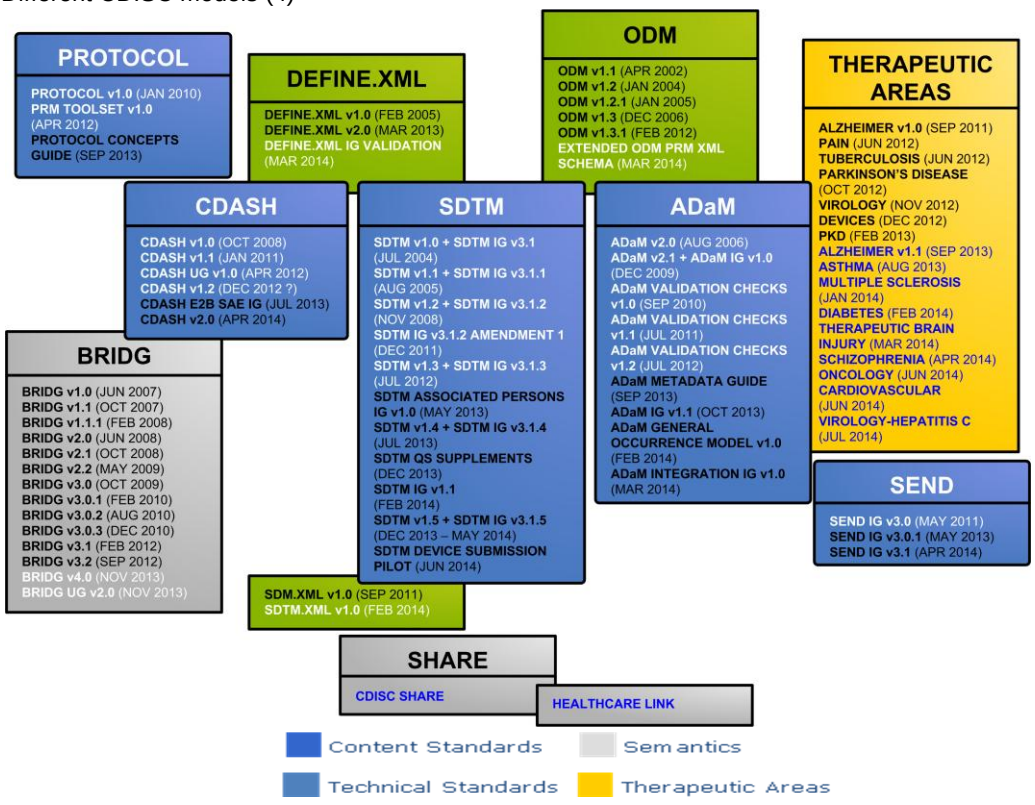
The Food & Drug Administration (FDA) in the USA states that “Standardizing study data makes the data more useful. Data that is standardized is easier to understand, analyze, review, and synthesize in an integrated manner in a single study or multiple studies, thereby enabling more effective regulatory decisions”. (1)

In the early nineties, pharmaceutical companies were gradually recognizing the value of standardizing clinical trial data collection throughout their company. The next logical step was that initiatives were taken to start harmonizing clinical data standards across pharmaceutical companies. One of the most widely known initiatives taken within the industry was the creation of the Clinical Data Interchange Standards Consortium (CDISC) format.

From its inception in 1997, CDISC has recognized the need for the establishment of standard data models to improve the process of electronic acquisition and exchange of clinical trials information for the benefit of all medical and pharmaceutical stakeholders. This is reflected in the mission statement: "CDISC is a global, open, multidisciplinary, non-profit organization that has established standards to support the acquisition, exchange, submission and archive of clinical research data and metadata. 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. CDISC standards are vendor-neutral, platform-independent and freely available via the CDISC website.” (2)

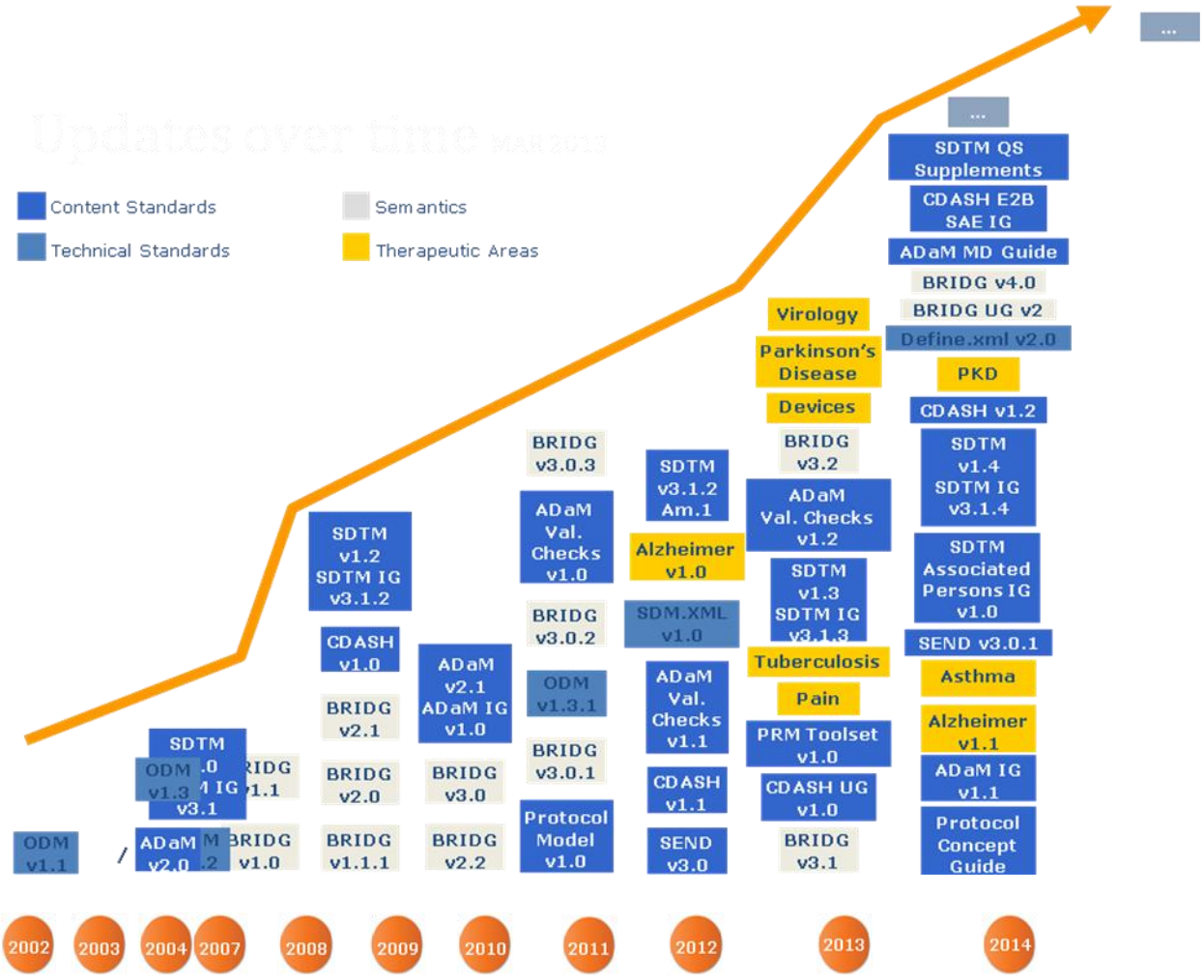
CDISC has developed various models, including content and technical standards, semantics and standards for therapeutic areas (Figure 1). Study Data Tabulation Model (SDTM) is the most widely adopted CDISC standard. This is not surprising since it was the first mature standard released and has also been cited in FDA Final Guidance for eSubmissions and other published guidelines related to FDA data submissions. (3)

Figure 1: Different CDISC models (4)



There is not only a growing variety of CDISC standards but also an accelerated volume as clearly displayed in Figure 2. This makes the adoption of CDISC standards within every pharmaceutical company an increasingly complex activity. Moreover, it is critical for every pharmaceutical company to further mature clinical data standardization in order to prepare efficiently for these current and future challenges.

Figure 2: Evolution of the CDISC models (4)



The FDA has not only adopted the use of CDISC standards. Now there is also a similar evolution within the European Medicines Agency (EMA).

PHARMACEUTICAL RESPONSE ON THE CHALLENGING REGULATORY AND BUSINESS REQUIREMENTS

Preparing clinical data in a CDISC format is a basic requirement for FDA submission. To respond to the challenging regulatory requirements, pharmaceutical companies are handling this request in a few different ways. Some of them are converting all clean clinical data from a company into a CDISC format at the end of the clinical trials process. This is obviously not a scalable and long-term solution. Therefore pharmaceutical companies are trying to move integration of CDISC standards from a back-end conversion into front-end solution. The final step of this challenge is to fully integrate the CDISC implementation into the company processes, technical infrastructure and resourcing. Ultimately, there is a need for efficient CDISC clinical data collection, comparison & reporting processes.

Moreover, pharmaceutical companies are experiencing increased competition. Therefore, it is critical to achieve overall operational excellence with optimized processes, a robust technical infrastructure and the adequate resources.

From a business perspective it is increasingly important to strive for reduced turnaround times for clinical trial set-up, reporting and submission, improved clinical data quality and consistency, and a reduced cost for conducting clinical trials.

BUSINESS CHALLENGES WITH DATA STANDARDIZATION

The following section clearly describes the underlying business challenges throughout the different data standardization processes:

SETTING UP CLINICAL TRIALS WITH NEW CLINICAL DATA STANDARDS

Possible challenges are:

- Which CDISC data standards need to be used for set-up? Which version at Clinical Data Acquisition Standards Harmonization (CDASH), SDTM, and Analysis Data Model (ADaM) level? All of the same date or not?
- How to handle data standards for protocol amendments while new standards are active?
- It is critical to know when a clinical data standard was used for a given clinical trial. How to track all of this?
- What about future clinical trials? Maybe verify the existing data standards Library: compare with other studies to identify the differences. Change or leave it? How to process all of this? Consistency is important.
- What about study specifications for the eDC application (if front-end integration is applied)?
- Do we keep the data standards generic or eDC specific?
- How far do we standardize? What about Edit checks, Completion Guidelines etc.?
- How to validate clinical data outputs against the applicable CDISC standards.

PREPARING CLINICAL TRIALS STANDARDS FOR A SUBMISSION DOSSIER

Possible challenges are:

- Verify clinical trials selected to go into the submission dossier: compare with other clinical trials to identify the differences.
- Change or leave the clinical data standards version? How to process all of these?
- How to validate clinical data outputs against the applicable CDISC standards.

UPDATING THE COMPANY DATA STANDARDS BASED ON A NEW CDISC RELEASE

Possible challenges are:

- Identify what is different for the new CDISC releases; these are becoming increasingly complex.
- Update the existing company standards? How to do this most efficiently?
- How to validate clinical data outputs against the applicable CDISC standards.

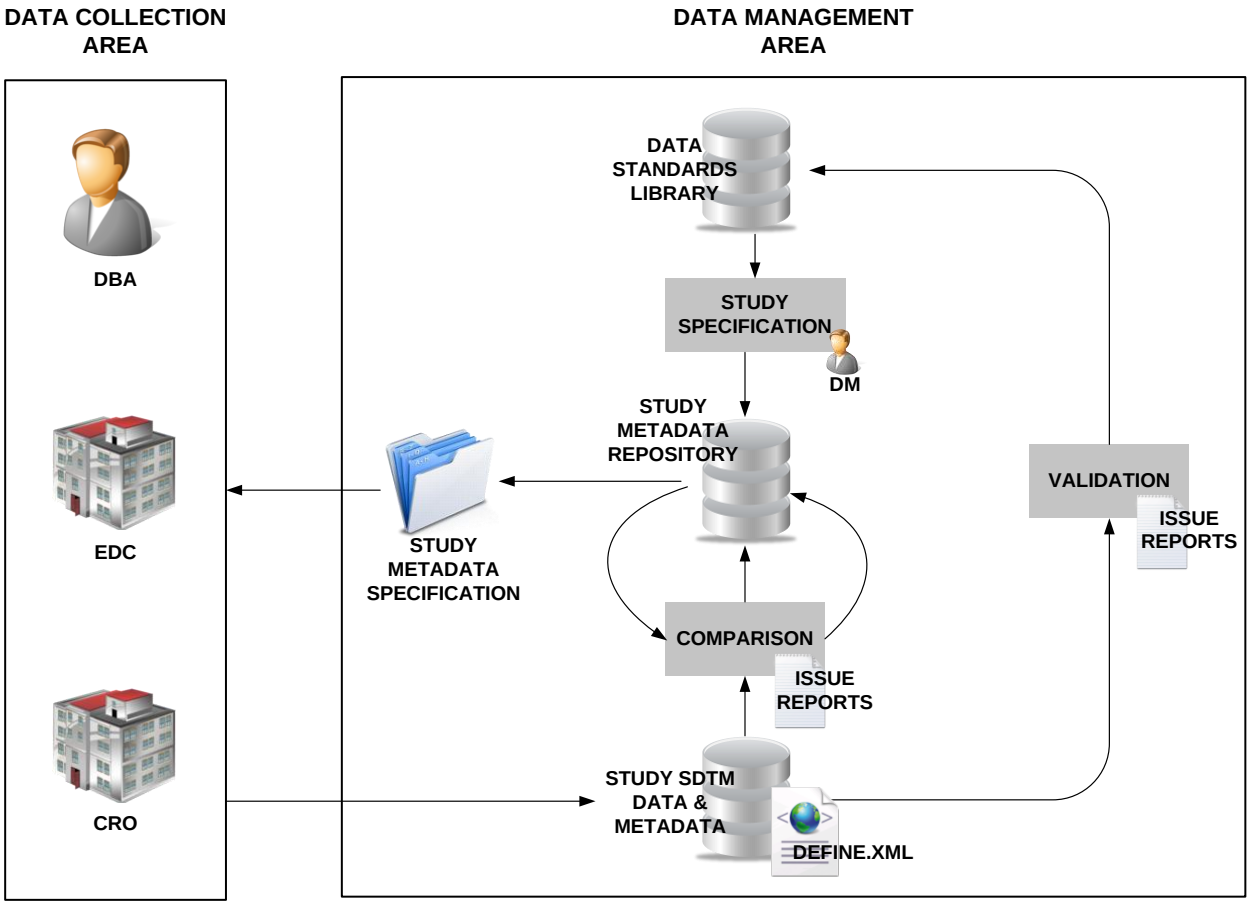
PROPOSED DATA STANDARDS FRAMEWORK SOLUTION

DATA STANDARDS PLATFORM

To handle the previously described data standardization business challenges, there is an obvious need to automate data standardization activities through a technical platform (Figure 3).

Automation of the CDISC data standards Library facilitates all respective data standards creation, maintenance and governance processes. More specifically, the system automation drives eDC study specification, study metadata generation and validation, and allows sponsors to dictate what data standards are to be used within the library. It also enables a user to build study specifications for a study by simply selecting the metadata from the library and copying it into a study metadata repository, resulting in re-usable clinical data standards across many studies within the company.

Figure 3: Data model – high level flow (5)



DATA STANDARDS CREATION AND LINKING

The CDISC data standards Library is built to ensure accuracy and traceability of metadata and results. Therefore it is critical to link clinical data standards objects across the different levels of metadata (Figure 4).

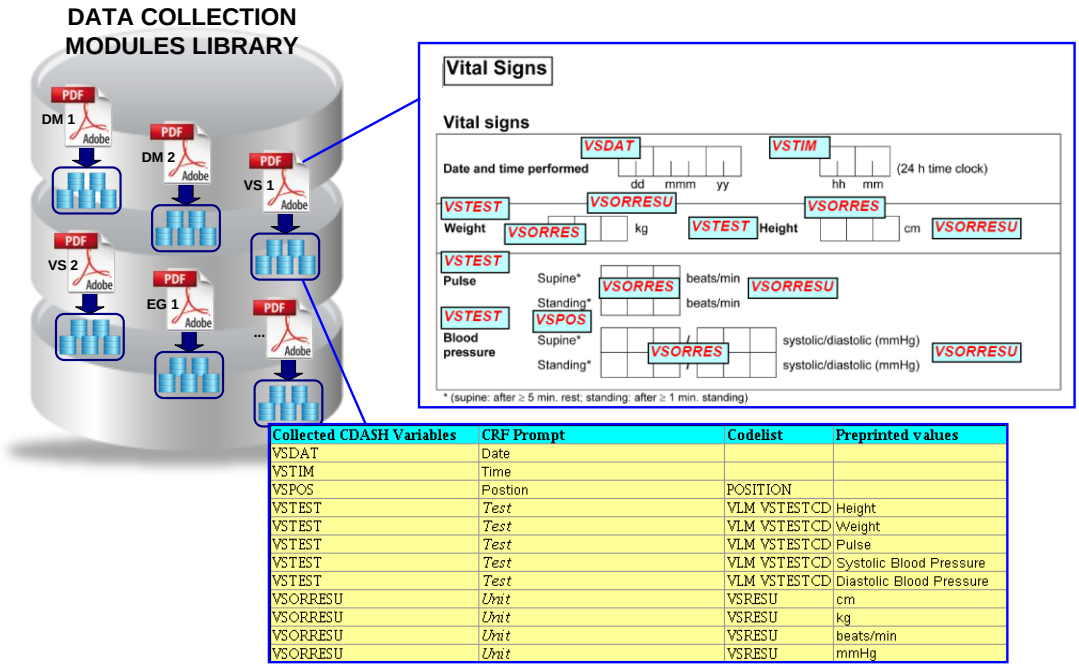
The Case Report Form (CRF) modules library contains all standard CRF questions having both a CDASH and SDTM annotation. The CRF modules in the library will focus on standardized content to facilitate the use of data standards during data collection.

Each unique CRF question (or object) is linked to a cluster of relevant metadata. The metadata cluster library will contain all standard domain structures, variables, code lists, value level metadata and computational algorithms.

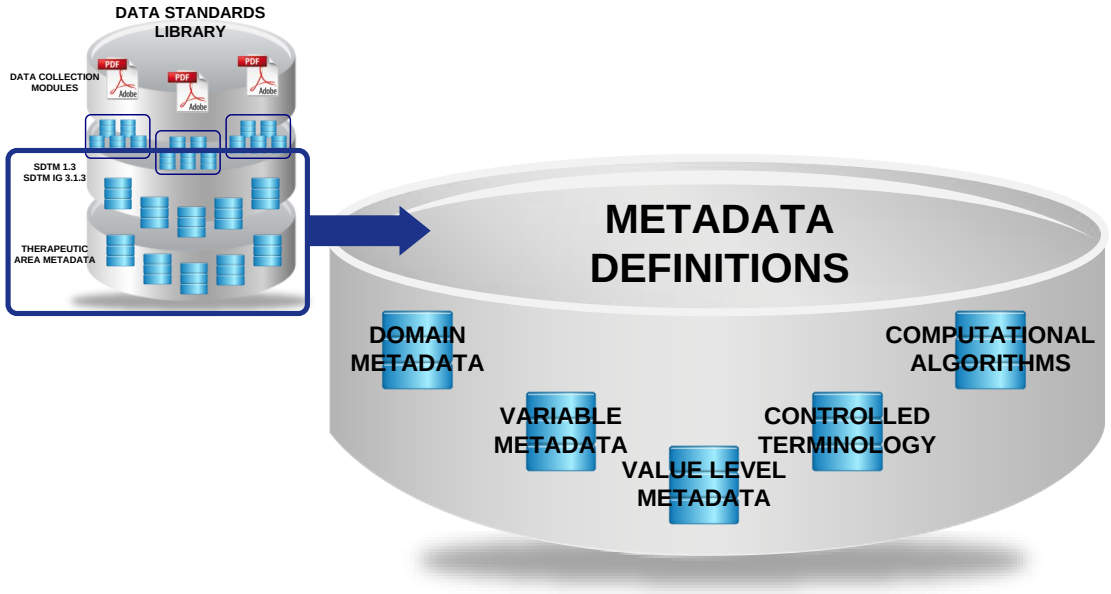
Related information like edit checks and completion guidelines are also classified on an object level and are as such integrated in the cluster of metadata.

The data standards Library will include all standardized CRF and metadata objects and as such will facilitate: the study specification process, the comparison of metadata across studies for consistency, the interaction between different departments and between CRO(s).

Figure 4: Data Standards Library with CRF and metadata clusters (5)



Collected CDASH Variables	CRF Prompt	Codelist	Preprinted values
VSDAT	Date		
VSTIM	Time		
VSPOS	Postion	POSITION	
VSTEST	*Test*	VLM VSTESTCD	Height
VSTEST	*Test*	VLM VSTESTCD	Weight
VSTEST	*Test*	VLM VSTESTCD	Pulse
VSTEST	*Test*	VLM VSTESTCD	Systolic Blood Pressure
VSTEST	*Test*	VLM VSTESTCD	Diastolic Blood Pressure
VSORRESU	*Unit*	VSRESU	cm
VSORRESU	*Unit*	VSRESU	kg
VSORRESU	*Unit*	VSRESU	beats/min
VSORRESU	*Unit*	VSRESU	mmHg



Updating/maintaining the Data Standards Library involves the possible modification to all of the following sections:

Table 1: Possible modifications

CRF library	SDTM metadata library	And the defined links between those sections
• CRF module/question (delete, insert, update) • CDASH metadata (delete, insert, update)	• Domain metadata (delete, insert, update) • Variable metadata (delete, insert, update)	• CRF module versus CDASH/SDTM annotation (delete, insert, update) • CRF module versus SDTM

• STDM related metadata (delete, insert, update) • Usage/Completion Guidelines (delete, insert, update) • Edit checks (delete, insert, update)	• Value Level metadata (delete, insert, update) • Codelist metadata (delete, insert, update) • Computational algorithms (delete, insert, update)	related metadata (delete, insert, update) • CRF module versus Edit check/Completion guideline (delete, insert, update) • SDTM metadata versus CDASH metadata versus CRF module (delete, insert, update)

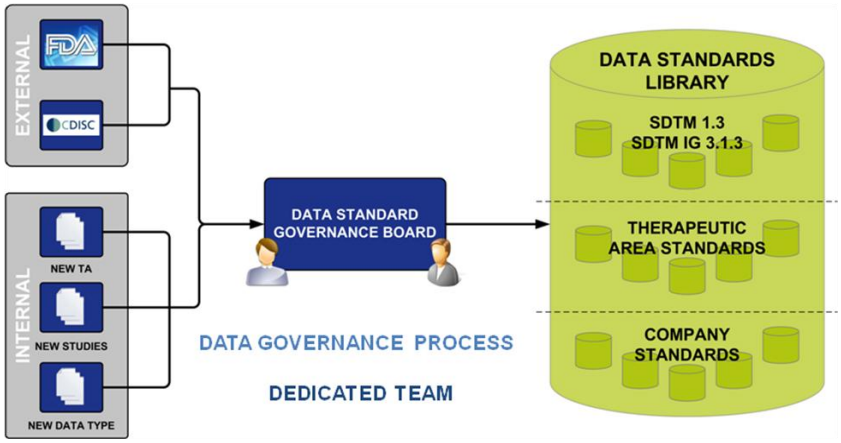
DATA STANDARDS GOVERNANCE

It is recognized that we need data standards processes, resources, roles and responsibilities next to a global and robust technical platform (Figure 5).

Data governance is paramount to the data standards themselves, as without data governance, the data standards cannot be maintained. The main reason to employ data governance is to identify the various roles and responsibilities of the persons who will act as librarians in governing the overall maintenance and proper handling of the data standards within the company.

It is also critical that the data standards and Governance program is well planned, implemented and controlled based on strategic planning & performance management principles. Inherent to this approach is that these objectives should be defined in a well balanced manner across all relevant company perspectives i.e.; innovation, internal processes, stakeholder value and financial outcomes.

Figure 5: Data Standards Governance (6)



ROLE OF KEY PERFORMANCE INDICATORS FOR DATA STANDARDS MANAGEMENT

The data standards Library team monitors, measures and reports on the operational performance of the Data Standards Governance model.

First of all, the objectives of the Data Standards Governance process need to be defined. Based on the objectives, Key Performance Indicators, Targets and Initiatives for improvement are defined, for example:

Table 2: Example of a Data Standards Governance objective and the corresponding parameters

Data Standards Governance objectives	Key Performance Indicators	Targets	Initiatives
• To develop, implement, and maintain an efficient Data Standards Governance process	• Data Standards update cycle times	• Level 1: 3 week turnaround rule for normal requests • Level 2: 3-4 days turnaround rule for exceptional/urgent study requests • Level 3: 6 week turnaround rule for maintenance requests	• Implement an Escalation Scheme if the cycle times are not respected • Identify additional User Training

To manage data standards from a high level perspective it is critical to install a performance management system.

When building an efficient system, it is important to consider:

- Which KPIs are important?
- How frequent do KPIs need to be generated? By whom?
- What to do? How to use escalation schemes

MANAGEMENT OF VERSIONING THROUGH FROM AND TO DATES

To create and maintain data standards version control, it is highly critical to make data standard objects unique without duplication or gaps, and use *from and to dates.*

A valid *From and To date* functionality will be implemented at record level to version the clinical data standards. This versioning will be applied on the 5 levels of metadata. This enables the end-user to select the current data standards, but also provides the possibility to go back to a standard applicable at a defined moment in the past. No records will be deleted when an updated metadata version becomes available. The end-user will still be able to use the data standards even upon expiration. Warnings will be introduced into the system to indicate expired data standards were selected during the study specification process.

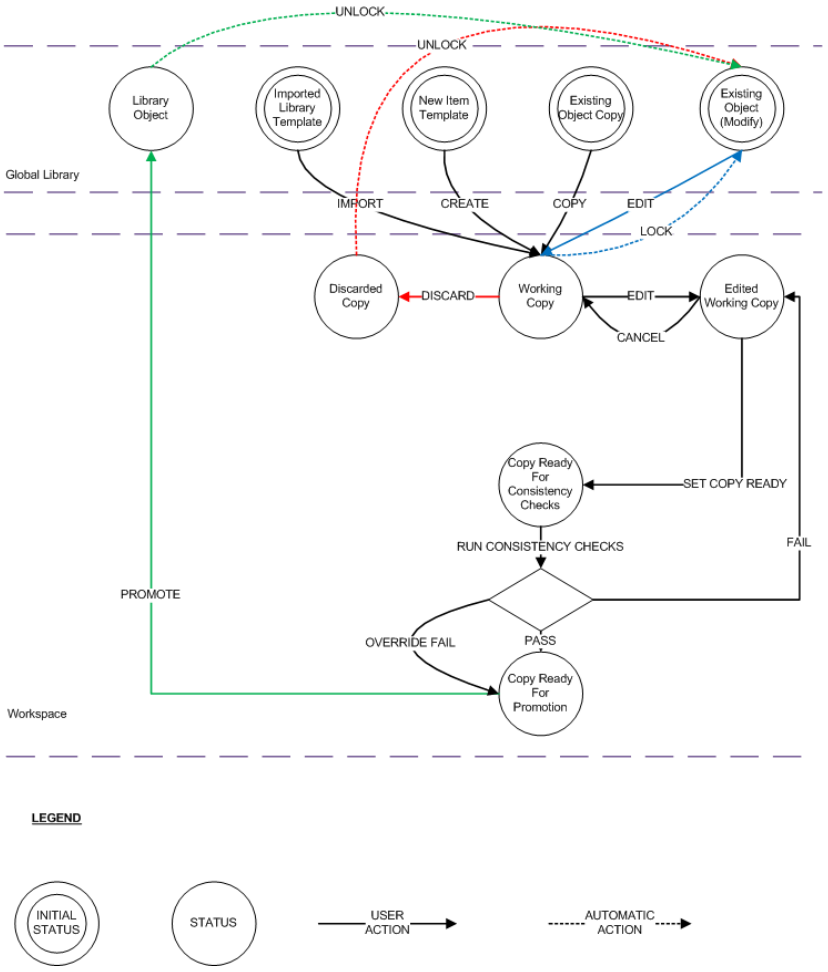
After logging into the data standards Library system, the user selects a specific module and then selects the library version where to perform an action, e.g. view, edit, create, copy, run consistency checks, make invalid or import (Figure 6).

After selecting the appropriate data standards Library version, the user selects the appropriate data standards Library Object Type where an action is required.

The Library Objects Types that the user can select are:

- DCM (Data Collection Module)
- Domain
- Variables (CDASH and SDTM Metadata)
- CA (Computational Algorithm)
- CT (Controlled Terminology): Code List, Code Value, VLM (Value Level Metadata) Name, VLM Value
- Data Validation Rules

Figure 6: Life cycle of a Library Object



The way version control is implemented in the system is through date fields where for each record there are additional fields added which are labeled ‘Active from’ and ‘Active to’.

The ‘Active from’ field contains the date that the record can be used. The date could potentially be a date in the future, if the data record is already created in the system prior to the date when it is released for use. In this field a date in the past is not permitted.

The ‘Active to’ field contains the date until the record can be used. This date can remain blank, if at the date of record creation the record is expected to be used indefinitely.

The system can also automatically complete the ‘Active to’ date of a record, if the record is edited and a new version of the record becomes available in the library. The system then sets the ‘Active From’ date of the edited record to minus 1 day.

Using date and not date/times in the Active From and Active To fields ensures that the system can be used in a global context as each record created in the library can only have an ‘Active From’ date set to a date in the future.

In the example in Table 3, the librarian initially created the record SDTM 1, which was available in the library from 02Jan2013. Due to changes in the CDISC implementation guide the SDTM metadata for the record needs to be updated and will be active from 15Apr2013. The librarian then has to edit the SDTM 1 record and create a new record with the updated data and ‘Active from’ date. Once the librarian confirms that the record is ready to be added to the library, the system adds the edited record SDTM 2 to the library and sets the ‘Active To’ date of the original edited record to 1 day before the ‘Active From’ date of the new record.

Table 3: Librarian initially created the records SDTM 1 and 2

Record	Data Field	Active From	Active To
SDTM 1	Metadata for SDTM 1 record	2013-Jan-02	2013-Apr-14
SDTM 2	Updated metadata for SDTM 1, eg. metadata for SDTM record 2	2013-Apr-15	

Editing of a library object is done in a user specific workspace. In the workspace the librarian can update the data of the record as many times as needed until the data is in line with the update needed. When a librarian selects a record for editing, and the system copies the record to the user's workspace, the system automatically locks the original record in the library to ensure that an object is only updated by one user at a time. As soon as the record in the workspace is placed in the library by the librarian, the system unlocks the record that was locked. The locked record is also unlocked if the librarian deletes the record from the workspace. A locked record is no longer available for editing.

In the example in Table 4, the librarian has locked the record SDTM 3 for editing. This implies that the record can still be used in a trial; however the record is no longer available for editing by anybody else. In the system there is a locked symbol present so the librarians know that the record is already being edited by another user. The user that has locked the record is identifiable by hovering over the lock symbol and the user and date that the record was locked becomes visible. This enable the different librarians to work together to ensure that when a record needs to be updated all changes can be made in one edit of the record.

Table 4: Librarian has locked the record SDTM 3 for editing

Record	Locked	Data Field	Active From	Active To
SDTM 3	√	Metadata for SDTM 3 record	2013-Jul-02	

Per object type, specific fields are defined to ensure that records with the same data are not duplicated. Overlaps in dates between records, with the same data, is not possible in the system eliminating the duplication of records in a library.

In the system specific fields are available to link different library objects. The code list record can be created and needs to be linked to code value(s). By imposing linking between the different library objects data does not need to be duplicated in a library. For example the code value 'No' is available once in the system however many code list records can link to this one code value. If the record is updated then the update of the record only needs to be done for one record.

Through the use of dates for versioning of library objects the users of the system can always retrieve which records where available at which point in time, e.g. on which date a specific record was in use.

SYSTEM CHECKS

There are three different kinds of checks defined in the system; i.e. the integrity, the consistency and the validation checks.

INTEGRITY CHECKS

As many objects are linked in the system, there are built-in database integrity checks which should be met prior to being able to add an object in the library.

For example, when a code list has multiple code values attached to it, the life span of the code value should be equal or greater than the life span of the code list.

In the example in Table 5, all code values attached to the code list record have an active from date that is earlier or equal to the active from date of the code list. Code value records with an active from date which is after the active from date of the code list would fail integrity checks and will never be available in the system.

Table 5: All code values attached to the code list record have an active from date that is earlier or equal to the active from date of the code list

Record code list	Active From code list	Active To code list	Linked code value record	Active From	Active To
CL	2013-Jul-13		CV 1	2013-Jul-02	
			CV 2	2013-Jul-13	
			CV 3	2012-Dec-22	

Another example of a system integrity check relates to the variable order for the SDTM metadata, where for each record in the SDTM metadata table a variable order is expected. The system will not allow a SDTM record to be moved from the workspace to the library unless a variable order is available.

Integrity checks are also defined to check that objects meet the defined rule for versioning library objects.

These types of checks cannot be bypassed by any user as this impacts the database integrity.

CONSISTENCY CHECKS

In the system there can also be consistency checks added. Consistency checks are checks that can be defined to meet specific business needs.

The consistency checks that are available in the system relate to certain fields expected to have data added in the system. For example there are Change Request fields added to all records where a librarian has the possibility to add a reason for editing a record available in the library. If the business needs specifies that the Change Request field needs to be completed by the librarians then a consistency check can be used to trigger the user to complete this field.

Consistency Checks can be bypassed by the user, however the system requires the user to provide a reason for overriding the consistency checks defined.

VALIDATION CHECKS

Finally, in the system there can also be validation checks added. Validation checks are checks that can verify the metadata and data available against the official CDISC libraries.

CONCLUSION

Due to the ever increasing pharmaceutical competition, data standardization becomes a necessary step to further improve the process of the clinical trials data submission.

Clinical data standardization is closely interlinked with maintaining data standards metadata. The latter is becoming increasingly complex due to the volume and frequently changing regulatory and data standards requirements.

To handle the data standardization business challenges, there is an obvious need to automate data standardization activities through a robust technical platform supported by a data standards governance framework and the corresponding performance management objectives.

To create and maintain data standards version control, it is highly critical to make data standard objects unique without duplication or gaps, and use *from and to dates.*

REFERENCES

(1) FDA, 2012. *"Guidance for Industry, Providing Regulatory Submissions in Electronic Format, Standardized Study Data"*, U.S. Department of Health and Human Services, February 2012, http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM292334.pdf
(2) CDISC, 2013. *"Missions & Principles"*. http://www.cdisc.org/mission-and-principles
(3) Friggle William, Li Feng, Labout Shannon and Kush Rebecca, 2011. "CDISC: Adoption Trends, Benefits and Addressing Barriers". CDISC Journal, October 2011, p1-17
(4) Van Reusel Peter, 2013. *"The Evolution of SDTM – What's new"*. Powerpoint presentation at Business & Decision Life Sciences, Brussels, Belgium
(5) Van Reusel Peter, 2013. *"Getting the most out of CDASH and SDTM – Automating the specification and validation process"*. Powerpoint presentation at Business & Decision Life Sciences, Brussels, Belgium
(6) Van Reusel Peter, 2013. *"Data Standards Governance"*. Powerpoint presentation at Business & Decision Life Sciences, Brussels, Belgium

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Nico De Leeuw
Business & Decision Life Sciences
Sint-Lambertusstraat 141 Rue Saint-Lambert
1200 Brussels
Tel +32 2 774 11 00
Fax +32 2 774 11 99
Mobile +32 478 99 61 54
nico.deleeuw@businessdecision.com