PharmaSUG 2019 - Paper DS-268

# Best Practices in Data Standards Governance

Melissa R. Martinez, SAS Institute

## ABSTRACT

Most organizations working in the pharmaceutical and biotechnology industries have adopted CDISC submission data standards by now, but the challenge of effective governance and compliance within an organization remains high. CDISC data standards are notoriously open for interpretation and the assumptions and understanding of the data standards can vary widely among users. Adopting CDISC standards is more than just making CDISC-style submission datasets and using tools to help with the programming and compliance checks. Each organization needs to define its own interpretation of CDISC standards to ensure consistency among its studies, put in place workflows and processes to facilitate the governance process, determine what it means to be compliant with the CDISC data standards, and find tools to help with the governance process and compliance determination.

## INTRODUCTION

Regulatory agencies around the world have made it clear that new drug applications should be using CDISC data submission standards, making it required for submission soon if not already. Most pharmaceutical organizations have already adopted the standards, but many continue to struggle to have a consistent understanding of the standards and compliance with them. The task of managing the organization's use of the data standards is often either left to individual study managers or is overwhelming to try to manage centrally. Without central guidance and oversight, the use of CDISC data standards can result in wildly different interpretations from one study to the next (while still technically being compliant with the standard), which is counter to the entire spirit of standardizing data in the first place. The best solution is to have a data standards governance review board and a dedicated staff of data standards governors responsible for the organization's data standards.

## DATA STANDARDS GOVERNANCE BOARD

The responsibilities of the data standards governance board include determining the organization's interpretation of CDISC data standards, maintaining the standards through updates from CDISC and internal requests for modification, and enforcing compliance to the standards. Members of the board should include subject matter experts in CDISC standards, representatives from functional groups responsible for implementing CDISC standards in their daily work such as statistical programmers, data managers, and statisticians, and management ultimately responsible for choosing the strategy for implementing CDISC data standards and determining the organization's internal standards.

In general, 1-3 individuals from each of these groups is a reasonable size for the board. Consider how mature the data standards governance process already is in your organization, how large your end user base is in your organization, which may include contract research organizations, and how complex your needs are. For example, if your organization is developing a single drug that has not yet submitted its first New Drug Application, your governance board will probably be composed quite differently than a multi-national organization with dozens of drugs under multiple therapeutic areas. If you do have an organization more akin to the latter, consider including working groups as a part of your governance board. These working groups can tackle problems and have one member present the results back to the governance board.

## INTERPRETING CDISC DATA STANDARDS

When the governance board is established, its first order of business must be defining what data standards mean to the organization. At a recent CDISC Advisory Council meeting, interpretation of data standards was identified as the most significant challenge to adopting CDISC data standards. Several people could review the Implementation Guides and each come up with very different ideas of how to

implement the standards, all of which would be technically compliant with the standards. The governance board must first agree on the organization's interpretation and planned implementation of each data standard. It will take a significant investment of time, but it is crucial to long term success.

Determining how to interpret a data standard includes determining what data belongs in each submission data set and deciding how to map source data into the standardized data sets. The source data needs to be evaluated to understand how consistently data is being collected across the organization. Consider whether it makes sense to group data standard interpretations by therapeutic area; source data may be consistent within a therapeutic area but not across the entire organization. Relatively consistent source data is important for being able to provide realistic guidelines for data mapping. You may find that some domains may be universally defined, while others need different interpretations for each therapeutic area. A hierarchy of related data standards might make sense in this case.

## EXTENDING DATA STANDARDS BEYOND CDISC

CDISC data standards are in many ways a starting point, a basic set of requirements for submissions. Many organizations have additional requirements for their users that are more stringent or more detailed than the basic CDISC data standards. There are also ways to extend data standards to facilitate other data processes. Some examples include:

- Variables listed as Expected or Permissible in the CDISC data standard considered Required within the organization
- Additional controlled terminology code list references beyond the CDISC data standard
- Implementing an SDTM-Plus model that allows for extra variables in SDTM-like domains while a study is active, which will be converted to true SDTM for submission
- De-identification instructions for preparing data for public sharing/transparency initiatives

The data standards governance board should identify opportunities to standardize this information and maintain it with the basic CDISC data standard metadata. In my paper "Designing Flexible Data Standard Models" (PharmaSUG 2019) I discuss including additional metadata like the examples above in data standards. The information is complementary to the metadata provided in data standards and is typically used by the same audience responsible for creating the standardized datasets. It saves time and effort for both the data standards governance board and end users to maintain the standardized metadata together.

## ENFORCING COMPLIANCE WITH DATA STANDARDS

Data standards governance does not mean much without a way to verify that the data standards are being used appropriately across the organization. The definition of appropriate use will vary from one organization to another.

The data standards governance board should define what it means to be compliant with a data standard. Compliance can include a comparison of a study's metadata to the organization's data standard metadata. Additionally, third party tools like Pinnacle 21 may be used to run compliance checks against the basic CDISC data standard. Reports of the results of these adherence checks may be a required step in study deliverable and submission processes so the data standards governance board can evaluate the degree to which studies are adhering to the organization's data standards.

Metrics may be useful for some organizations to track adoption of both data standards and new processes for implementing them within studies. Tools that evaluate compliance may be used several times during a project, with the results from each run gathered and reported to the data standards governance board. The results across many studies over time can show trends toward or away from data standard compliance and help evaluate whether studies are reaching acceptable compliance thresholds earlier in their lifecycle.

Another question to consider is what it means to have consistency across studies or the organization. As already mentioned, it is possible to have many studies that are all technically compliant with a data

standard, but whose implementations are significantly different. Consistency within an organization is essential for reusable programming and macros, making it easy for programmers to work on a variety of studies, and for combining data for integrated summaries. The data standards governance board may wish to obtain reports comparing metadata across similar studies on a regular basis. If significant inconsistencies are discovered, the board should determine whether they need to reevaluate the data standards to ensure they are providing appropriate details or interpretation or to provide additional training for end users on proper use of the data standards.

## MAINTAINING DATA STANDARDS

Another important aspect of data standards governance is establishing processes for maintaining the data standards. Some considerations for long-term maintenance include:

- When and how to adopt new versions of data standards and metadata being released by CDISC, NCI, etc.
- When and how to adopt new data standard types being released by CDISC (such as new therapeutic area standards)
- How to manage requests and suggestions from users to change or update data standards
- How to roll out new data standards and changes to existing data standards to users
- Evaluating effectiveness of the organization's data standards by reviewing compliance reports and consistency reports

Managing conflict is another important aspect of the data standards governance process. With so many interpretations of the Implementation Guides possible, individuals can be quite passionate about their personal interpretation. The data standards governance board should establish a process for managing disagreements over interpretation that allows for all positions to be heard but has an ultimate deciding authority.

## TOOLS TO FACILITATE GOVERNANCE ACTIVITIES

The data standards governance board should evaluate tools that can help with both the governance and the implementation of data standards. Some important tools include a metadata repository or other application for defining data standard metadata, ticketing system for the data standards governance workflows, a tool to check studies' adherence to data standards, versioning software for tracking and maintaining changes to data standards, and a comparison tool for comparing versions of data standards or study metadata

## HOW CAN SAS LIFE SCIENCE ANALYTICS FRAMEWORK HELP?

SAS Life Science Analytics Framework 5.1 was released in October 2018. This new major release includes a new clinical metadata management component, allowing customers to manage their data standards, controlled terminology, and study metadata in the same application where they do their statistical programming work.

Data standards, including controlled terminology dictionaries, may be defined and maintained globally, and users need a specific privilege to be able to create, modify, and delete data standards. Data standards are created and modified in a shared workspace area available only to those with the Govern Data Standards privilege. Once data standards are placed into a Production state, they are visible and usable by end users. Data standards may be placed into a Retired state, meaning they may not be used for additional studies, although existing study associations will remain in place.

Data standards are versioned in SAS Life Science Analytics Framework 5.1. Additionally, the SAS Life Science Analytics Framework Repository is a versioned repository where data standards governors may store and version files related to data standards.

Comparison reports are available to allow comparison between data standard versions and between two different data standards. Coming in future releases are adherence reports to compare study metadata to the study's associated data standard(s) and impact analysis reports to itemize the impact of changes to a data standard on its associated studies.

Studies obtain data standard metadata from the global data standard and then further refine and populate the metadata to match the study's needs. Tables and variables that are not needed for the study are removed. Metadata that cannot be standardized, like the reference Case Report Form page number for the define.xml file are populated. Users must have a specific privilege in the study to modify study metadata.

Data standards and study metadata may be extracted to SAS datasets. Metadata stored in data standards can be read from these SAS datasets to drive a variety of business processes from data mapping to data de-identification.

SAS Life Science Analytics Framework includes a Process Flow module. Users can define business workflows using a third-party modeling tool to create a Business Process Model and Notation format XML file that can be deployed in SAS Life Science Analytics Framework. Process Flows could be defined to help with the data standards governance board's activities like modifying data standards or taking user requests for updates to data standards.

## CONCLUSION

A data standards governance board is essential for an organization to be successful with consistent implementation of CDISC data standards. The data standards governance board should start by defining their interpretation of data standards and what it means to be compliant with data standards. An ongoing maintenance process should be created and tools should be adopted to make the governance process easier. SAS Life Science Analytics Framework 5.1 was designed to fully support clinical metadata management and has a variety of tools to facilitate the data standards governance process right within the application used for statistical programming activities.

## RECOMMENDED READING

Martinez, Melissa R. 2019. "Designing Flexible Data Standards Models." *Proceedings of the PharmaSUG 2018 Conference*, Philadelphia, PA.

## ACKNOWLEDGMENTS

Thank you, Bill Gibson for your collaboration and feedback. Thank you, Sharon Trevoy and Donna Dutton for your consistent support and helpful feedback. Thank you, Angela Lightfoot for your continued support of my professional development and career goals.

## CONTACT INFORMATION

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

Melissa R. Martinez
Principal Health and Life Sciences Industry Consultant
SAS Institute
Melissa.Martinez@sas.com