PharmaSUG 2019 - Paper DS-336

Metadata Repository V1.0 – A Case Study in Standards Governance

Aparna Venkataraman, Celgene Corporation

ABSTRACT

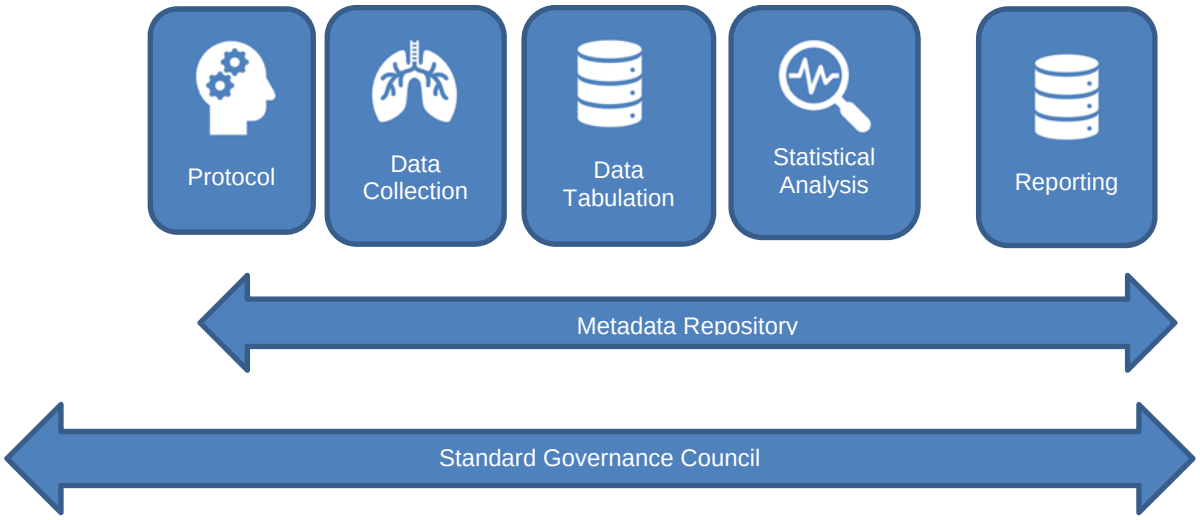
Over the last couple of decades Metadata Repository (MDR) tools have been playing a growing role in the Pharmaceutical/Bio-Technology space. It has been long since Regulatory agencies as well as the sponsor companies have, for various reasons, widely recognized the need for integrating data standards starting at Protocol Development and Data Collection all the way through to Submission. However, we still face tremendous challenges in maintaining and managing Clinical Data Standards that fall far beyond the capabilities of any single MDR tool that is currently available in the market.

We will explore in this paper, the tools that need to be in the Standards Manager's toolbelt to successfully dispense End-to-End Clinical Data Standards across the company In the multi-dimensional world of Clinical Data Standards, a nimble Governance Model will play an undeniable role in getting that critical drug in the hands of the patient at the right time.

INTRODUCTION

In 2019, 3 years after the FDA requirement for study data standards for NDAs, BLAs, ANDAs took effect, the term End-to-End data standardization and governance is more than just an industry jargon. Though the core concepts of metadata repository and data standardization is sought after in the bio-tech/ pharmaceutical industry, a comprehensive approach to maintain standard objects and build necessary governance structure to support the development of the standards is still an area of many riddled with many challenges.

Figure 1. End-to-End Data Standardization



As turnaround time for drug development shrinks, a clear vision to manage cross-functional standards can play a major role in getting critical drugs to the market. Sponsors should look at Clinical Study Data Standards at the beginning of protocol design up to CSR and beyond. Investment in standards management tools and processes will start paying out when companies can establish a core standard models which can provide standard structures which are also flexible enough to ensure quick turnaround for the clinical trials.

Since end-to-end metadata standardization and governance could vary from sponsor to sponsor, for the sake of this paper, we will consider data standards and governance that span from protocol and data collection into tabulation, analysis and reporting as noted in the figure below.

METADTA REPOSITORY

Metadata repository (MDR) is considered key to effectively managing standards across the organization. Based on the size of the company, this might mean a dedicated metadata repository with well-integrated set of systems (CTMS, CDR, Reporting tools etc) or a hodge podge of Excel sheets, MS Access or other database for metadata storage, along with homegrown macros, and well-established process steps for metadata management and governance.

Here are some of the key attributes for a good metadata repository. It is important to determine the structure of the database to determine the metadata that will be governed. For end-to-end standards, determining early in the process the kind of EDC systems to use can serve as an anchor for the collection metadata.

CUSTOMIZABLE

Expect that in addition to setting up standard structure to the Metadata Repository, that there will be non-standard information coming from some of the common external data such as Lab, Biomarkers, ECG. Based on the size of organization, many vendors will work on a standard structure or template. But a agile metadata repository needs to handle changes in structure to accommodate non-standard data without a lot of overhead. This will include the dataset format as well as in some cases mapping of the vendor test codes to controlled terminology. External metadata will have a generic SAS dataset format.

OBJECT INHERITANCE

In object-oriented programming, the property of inheritance enables new objects to take on the properties of existing objects. In setting up metadata repository to handle standard models for metadata management consideration given to inheritance is crucial. The core principle of any standards management should be to review and define standard objects, and reuse with limited intervention to set standards.

A class that is used as the basis for inheritance is called a *superclass* or *base class.* A class that inherits from a superclass is called a *subclass* or *derived class*. The terms *parent class* and *child class* are also acceptable terms to use respectively. A child inherits visible properties and methods from its parent while adding additional properties and methods of its own.

In this example below, defining the UNIT codelist once and setting up the necessary entries at the outset before even standard CRFs are designed allows the codelist to be used across collections, submission, reporting standards e.t.c., For future changes, when new units are added to the codelist, updates should be made once, impact analysis should be done to assess and update related objects.

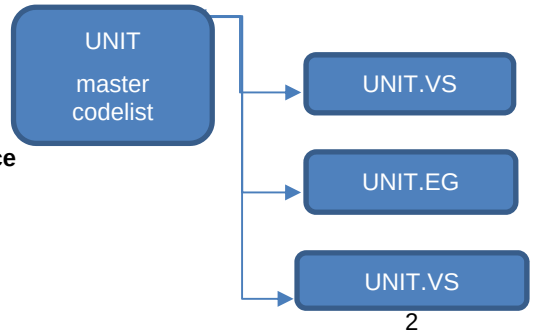


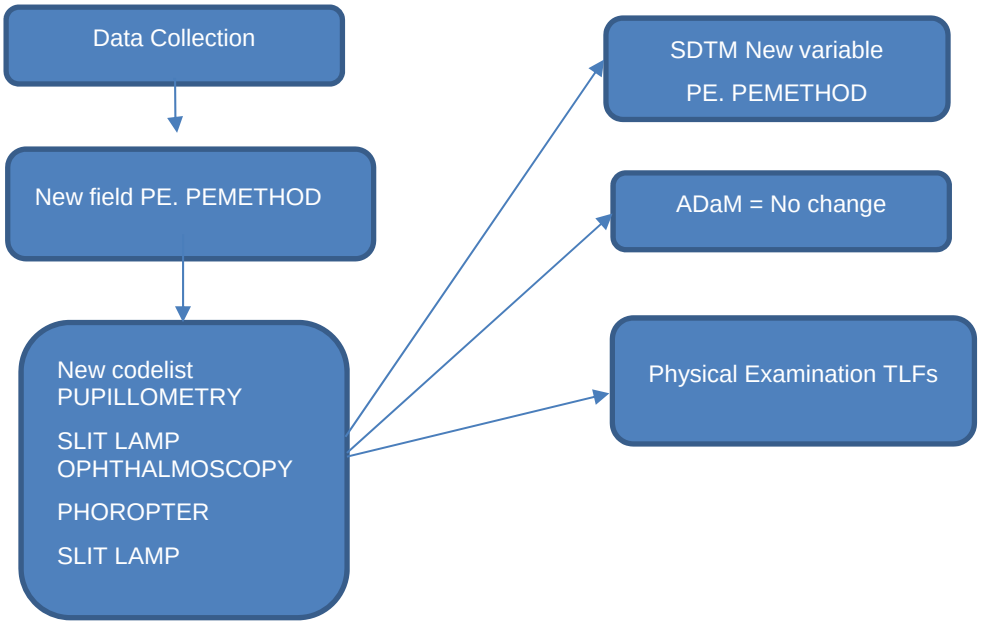
Figure 2.Inheritance

IMPACT ANALYSIS

One of the key objectives of metadata management should be ensuring impact analysis across the various objects and standards within the repository. For example, a question such as *"what are all the downstream or upstream metadata and process that will be impacted.* Analysis should be performed on all the different metadata repositories. Impact analysis should be built in to determine both downstream as well as upstream impact on all data objects maintained within the repository such as columns, tables, external files, information maps, reports, stored processes, Traceability of metadata across different models will be possible with a well-structured MDR.

Consider this example of collecting different methods of ophthalmologic testing. A need to collect this information as identified in the protocol should then trigger an impact analysis that checks across data collection, tabulation, analysis and reporting models and identify the objects that will be impacted by the identified change. The knowing the objects that will be impacted by a new object or change to existing object will allow for proactive update to metadata and processes. This approach will save time by avoiding rework or coming up with solutions once a part of the chain is completed.

Figure 3. Impact Analysis



CHANGE REQUEST MANAGEMENT

Change request process should be setup within or in conjunction with, metadata repository to help users/study team members request change to the existing standard objects implemented through established governance processes. Study teams should be able to request anything from new entries to a codelist or changes to CRF labels to addition of entire forms/ corresponding SDTM domains.

Ideally when configuring change request systems, it is good to keep in mind that during real world study build, change requests can be generated by different functions and will have to go through direct, iterative or looped approval process. Identify the various functions that can/need to initiate changes to standard metadata.

VERSIONING STANDARDS

FDA accepts electronic submissions that provide study data using the standards, formats, and terminologies described in the FDA Data Standards Catalog. The latest version 5.2 released on Dec 2018, supports 3 different versions of SDTM standards, one version of ADAM and 2 different versions of Define.xml for the use of Clinical Study Datasets.

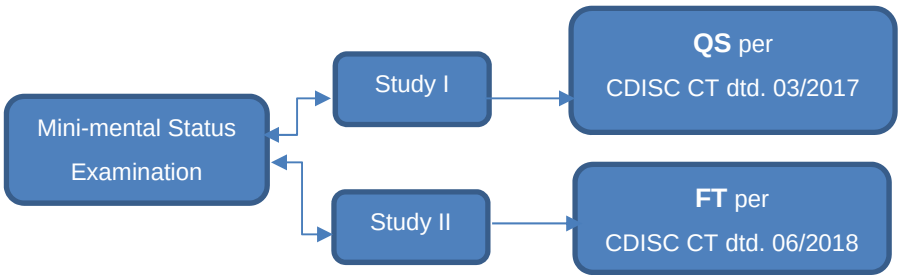
It should be noted here that though a capable MDR should be able to handle the maintenance of different version of standards across the organization, it is by only setting a clear standards governance roadmap many of the questions around versioning can be resolved.

Here is a partial list of the different category of standards that might have to be maintained within an MDR.

- CDISC Foundational Standards
- Quarterly releases of CDISC-CT
- Therapeutic Area specific Standards
- Sponsor defined standards
    - Though FDA/PMDA prescribe using CDISC foundational standards for submission, there can be a persistent gap between regulatory guidance and scientific data collected for various therapeutic areas. These are addressed by creating sponsor defined domains, supplemental fields etc.
- Study level standards that differ from company level standards.

At any given point in time, study teams will have to work with different versions of industry standards while utilizing internal standards to cover gaps in CDISC standards. Consider the case where Mini-Mental Status Examination is performed in two studies a year apart from each other. Note that CDISC recommendation to collect the necessary information as Questionnaire has changed to a collecting the same information now as a Functional Test. Following this, new standards are implemented at the organizational level. There is now a problem where studies under the same project are collecting and submitting the same information as different domains.

Figure 4. Versioning



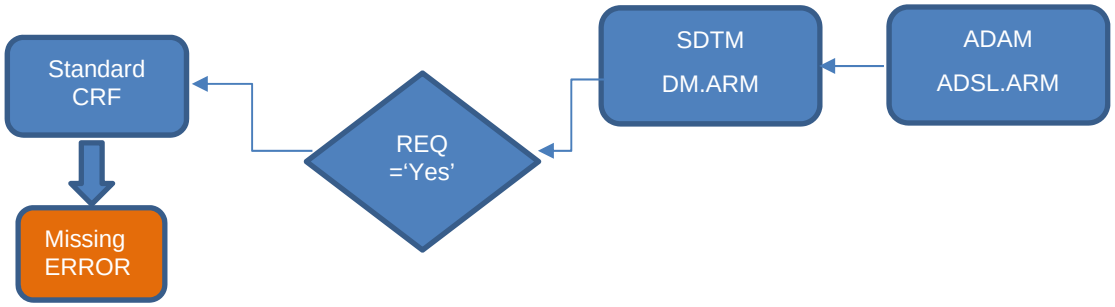
Early impact assessment both at the study level and organization level standards should indicate that there is a in the metadata prior to study build. This information should then be factored in maintaining different versions of the same standard CRF, one as Questionnaire for legacy studies and one as Functional Test for future studies. If you can further imagine, how a change in CRF can then affect a change in SDTM Mapping, programming and TLF generation, you will begin to appreciate the need for clear versioning of various standards within an organization.

STANDARD COMPLIANCE

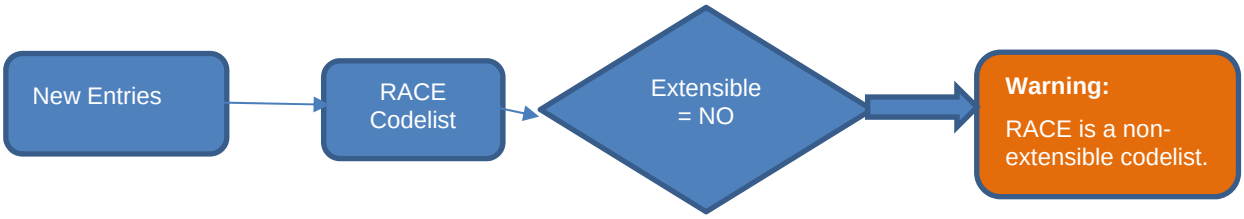
No MDR system is complete without validation checks to ensure that the output artifacts meet the standard expectations. In addition to expected validation checks to ensure metadata governance compliance, it is a good idea to build in early compliance checks for downstream data models. This is to say that standards compliance for SDTM can be built at the Data Collection or even protocol development. The consistency, integrity and compliance to the various standards models is paramount for effectively managing.

- Consider this example, Description of Planned Arm (ARM) is a required variable to be included in the Demographics domain in SDTM. Building a check at the standard CRF and enforcing this at the study level would ensure information for this required variable is available before this issue is identified in the data at SDTM or ADAM datasets.

Figure 5. Standard Compliance



- In another case, consider that a new protocol that requires the collection of subcategories for RACE. As of 03/2019 RACE is a non-extensible codelist within CDISC. Checking for compliance early in the process will help identify this deviation from standards and provide guidance for the study teams.



A WORD ON GOVERNANCE

In addition to building a customized metadata repository that fits the organization's needs, it is important to put together a standards strategy and governance team to help implement and enforce such a strategy. No amount of carefully customized technology, validation checks or system integrations can stand in to make critical decisions on deviations, exceptions and other challenges faced in the implementation of set standards. Sponsor companies need a comprehensive approach to Clinical Data Governance to make informed business decisions, including Metadata Governance. Metadata Governance involves looking at Metadata roles and responsibilities, standards, lifecycles, and statistics, in addition to how operational activities and related Data Management projects to integrate Metadata.

Though many sponsors acknowledge the value of having Standardized metadata end-to-end standards and corresponding implementation is very much a struggle. Metadata standards in place, a crucial piece of Metadata Governance. Formal roles, such as an Executive Sponsor or champion assist stakeholders in understanding the importance of standards and Metadata Management. Finding ways to track and view Metadata quality through completeness, accuracy, currency/timeline, consistency, accountability, integrity, privacy, and usability can show strengths and improvements needed in Metadata management.

Effectively governed Metadata provides a view into the flow of data, the ability to perform an impact analysis, and finally an audit trail for compliance, ensuring trust in a firm's data. Good Metadata Management becomes central to holistic Data Governance. At a fundamental level, data govenance should be handled at strategic, tactical and operational levels.

Here are some of the core principles upon which to build the governance model:

- Multi-tiered, cross-functional approach
- Enforce the data standards adoption on every project
- Reduce deviations, rework and confusion on implementation of study standards.
- Define metadata as an asset
- Guide the management of enterprise data across therapeutic areas
- Provide common processes and policies
- Provide escalation point for unresolved issues

CONCLUSION

The management of Clinical Data standards is going to play an ever-increasing role in running efficient clinical trials. We have tried to outline the core attributes of a metadata repository that can handle changing industry standards and company needs.

- Customizable
- Object Inheritance
- Change request Management
- Impact analysis
- Versioning Standards
- Standards compliance

In addition to any MDR System, a strong governance body with the ability to enforce standards across studies and projects, provide guidance for deviations and bridge the gap between industry standards and clinical needs is crucial to successfully organize and maintain clinical data standards within an organization.

REFERENCES

CDISC Foundational Standards: https://www.cdisc.org/standards/foundational

FDA Resources for Data Standards: https://www.fda.gov/industry/fda-resources-data-standards

FDA Technical Conformance Guide: https://www.fda.gov/drugs/electronic-regulatory-submission-and-review/electronic-common-technical-document-ectd

ACKNOWLEDGMENTS

The authors would like to thank Georgina Wood, Eileen Gardenhire for reviewing the paper and providing valuable feedback.

CONTACT INFORMATION

Your comments and questions are valued and encouraged.

Contact the author at:

Aparna Venkataraman
Celgene Corporation
avenkataraman@celgene.com
https://www.linkedin.com/in/aparna-venkataraman-6534a28/