PhUSE US Connect 2019

Paper DS03

A paradigm shift to represent and manage Clinical Data Standards for creating a data driven specification

Veena Nataraj, Shire, Takeda, Lexington, USA

ABSTRACT

The clinical research industry uses Clinical Data Standards published by CDISC to represent the clinical data collected in clinical trials. These standards are interpreted by companies when implementing them within their organizations. More companies are assessing how to represent and managed these standards within a Metadata Repository (MDR). This paper looks at how Shire has utilized a MDR to create a data driven specification which would streamline processes when compared to a more traditional approach. The MDR within Shire is also being used to manage governance processes and workflow to enhance traceability, consistency, conformance and to generate an impact analysis of beginning to end (B2E) standards development.

INTRODUCTION

The data collected during clinical trials form the basis for determining whether the experimental treatment or device is safe and effective for patient use. This study data is standardized to be conformant to Clinical Data Interchange Standards Consortium (CDISC) Standards before submission to the regulatory agencies so that it is traceable and consistent. CDISC has played a major role in developing standards so that data can be analyzed and summarized (e.g., pattern analysis and data mining) using automated tools. These data standards can be used at different stages in the clinical research life cycle. Plus, the efficient and effective management of the standards can streamline the submission and approval process.

Let us look at why data standards are important to regulatory agencies and how are they equipped to support the data review process. The Food and Drug Administration (FDA) has amended the Federal Food, Drug, and Cosmetic Act (FD&C Act), Section 745(a) and has signed the Food and Drug Administration Reauthorization Act (FDARA) into a law. This new law includes the reauthorization of the Prescription Drug User Fee Act (PDUFA) (2018-2022) which governs the requirement for data standards. The FDA has communicated to the industry about their Data Standards Strategy (2018-2022), to reinforce the ongoing commitment to the development, implementation, and maintenance of a comprehensive data standards program, and the study data requirements, which are on the “FDA Study Data Resources” page. This page also includes the Data Standards Catalog (DSC), and the Study Data Technical Conformance Guide (TCG). The DSC lists the FDA acceptable standards, formats, and terminologies which are used to represent electronically submitted study data. The Pharmaceuticals and Medical Devices Agency (PMDA) has published their own requirements which are similar to the FDA’s DSC and TCG.

This paper discusses how Shire has implemented the metadata repository and a methodology which provides a paradigm shift in how the CDISC and Shire-specific standards are reviewed and managed.

CRITERIA FOR PARADIGM SHIFT

A paradigm shift is a concept identified by the American physicist and philosopher Thomas Kuhn. He identified it as a fundamental change in the basic concepts and experimental practices of a scientific discipline. While recognizing that the term has been widely used in contexts that do not align to Kuhn’s intended meaning, it is a useful term in describing the fundamental change we see coming in how organizations are able to support and implement the CDISC standards.

Historically, the clinical data standards metadata model and implementation metadata have been provided by CDISC to member organizations in the form of the spreadsheets. Sponsor organizations which contract clinical trial activities to Contract Research Organization (CRO), have tried to ensure adherence to CDISC data standards either by providing guidelines to the CRO on how to interpret and implement the standards across their clinical trials or alternatively they provide the interpretation metadata as a guidance. This latter approach inevitably involves discussions, manual review, and process challenges, relating to how a CRO consumes and incorporates the metadata into their day-to-day processes. These processes are often specific to a CRO and do not necessarily support the ability to manipulate metadata electronically. Regardless of a sponsor’s internal tools and utilities often

the metadata exchange will still rely on the use of spreadsheets. While providing many advantages, spreadsheets lack formal control, and many times sponsor struggle to set up a consistent process to oversee the quality of the deliverable, with the ever-present risk of someone, either internally or at the CRO, modifying the content or structure by adding or dropping columns and rows, thereby disrupting the management and traceability of the metadata.

Ideally, tools such as an MDR would significantly enable process improvement, but the tools alone cannot solve all the challenges. It requires a paradigm shift in thinking that allows sponsor organizations to represent their interpretations in broadly scoped and robust tools, while engaging the CROs to leverage more automated approaches for incorporating the metadata into their processes. While yet not necessarily a full reality, the aim is to transform these complex interactions from the current state of specifications and requirements that is implemented as resource-heavy programming tasks, to a more streamlined mapping activity that is largely based on configuration rather than original programming. This change impacts the resources needed to support the process, both in terms of volume and expertise. To make this a reality, the sponsor companies need to invest in tools to represent the metadata, and in engaging with CROs to implement bridging tools and processes capable of automating and integrating with the sponsor's systems. Having the sponsor represent the metadata in a tool such as MDR, is the first step in this paradigm shift.

SHIRE DATA STANDARDS AND METADATA REPOSITORY (MDR) IMPLEMENTATION

Many organizations strive to create infrastructure and processes to ensure that the CDISC Standards are consistently interpreted and maintained within their organizations to ensure that the data can be traced within the clinical trials. Shire has an established a Standards Governance model with a Clinical Standards Governance Board and a Governance Team for each standard. Within the Shire Governance Model, each standards governance team interprets each of the standards and publishes the new Shire-specific metadata. This ensures the beginning to end standards can be used to gain business efficiency across the entire clinical development life cycle through automation, which supports consistency and reduction in cycle times.

Maintaining this metadata can be challenging. A common starting approach is to keep such metadata in spreadsheets. A more robust solution would then be to develop or purchase tools to manage this content. At first, starting with CDASH, we represented metadata in a metadata tool. As we expanded to support other CDISC standards, and while still leveraging the tool for CDASH, we used spreadsheets and custom programs to support other processes metadata. To better support this approach, we are currently implementing an MDR as part of our CORE Data Warehouse solution for representing the defined metadata.

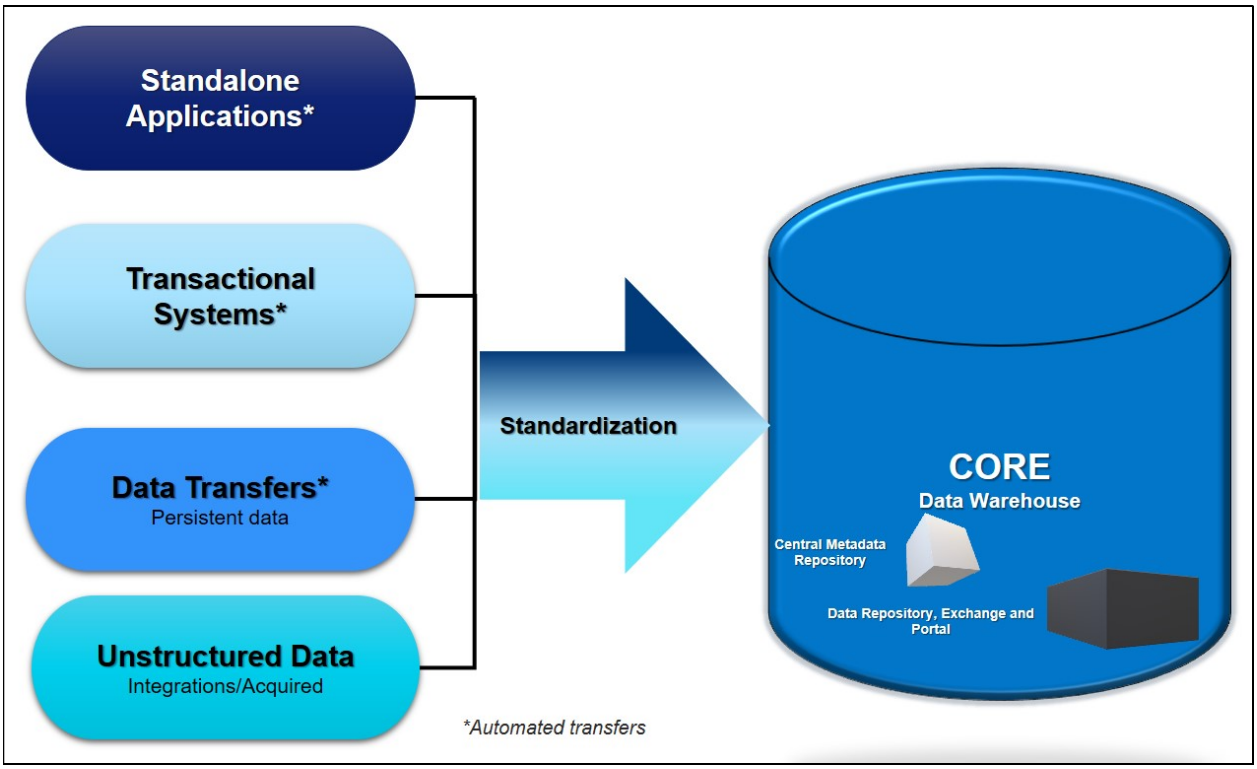

Fig-1: CORE Data Warehouse with Central Metadata Repository

The CORE Data Warehouse allows us to standardize and integrate information from various data sources such as standalone applications, transactional systems like clinical trial management systems, external data transfers and any other unstructured data sources. The CORE Data Warehouse has components such as Central Repository (MDR) to define metadata, Data Repository to store information from data sources, Exchange to extract, transform and load data, and a portal to collectively bring different components together. The Central Repository is also being used to store data standards.

The process starts by taking the published CDISC defined metadata models and their associated Implementation Guides and interpreting them for Shire specific needs. The CDISC metadata content is made available to CDISC members through Shared Health and Research Electronic (SHARE) Library, a metadata repository which exports files in excel and other formats, along with differences across versions. We focused on the CDISC published standards of CDASH, SDTM, ADaM, and Controlled Terminology.

Utilizing our governance model, we established company-specific interpretations to address any gaps or ambiguities in the published standards.

IMPLEMENTATION CONSIDERATIONS

REPRESENTING METADATA AND ORGANIZING CONCEPTS INTO FOLDERS:

CDISC published standards should always be considered as the primary source. The use of a primary source is an important concept within the industry because everyone including the regulatory agencies will consume the metadata information from CDISC as a source. It is always useful to keep this version as “Read Only” so that an end user cannot change it, and everyone can reference this information.

Based on the governance model, any new or changed metadata defined within Shire is represented in the tool under the global organizational standards folder. This helps the end users to quickly differentiate the global organizational standards metadata from the CDISC published metadata and also makes it easier to be used globally across all the Shire clinical trials. Any defined Therapeutic specific standards are represented under the Study Standards folder.

TEMPLATE AND FORMATS:

Any standard that needs to be represented in the MDR should be in a specific format. A format file is considered a template. It is recommended that the templates should be reviewed and evaluated before loading so that any unstructured representation can be organized in a structured format. For example, a company can receive the excel format of the CDISC Controlled Terminology (CT) and then if the organization needs to add a new term to this terminology, for example “Completion/Reason for Non-Completion” with C-code C66727, it can be represented in as shaded in gray in Table-1:

Code	Codelist Code	Codelist Extensible (Yes/No)	Codelist Name	CDISC Submission Value	CDISC Synonym(s)	CDISC Definition	NCI Preferred Term
C66727	C41331		Completion/Reason for Non-Completion	ADVERSE EVENT		Any untoward medical occurrence in a patient or clinical investigation subject administered a pharmaceutical product and which does not necessarily have a causal relationship with this treatment. An adverse event (AE) can therefore be any unintended sign	Adverse Event

Code	Codelist Code	Codelist Extensible (Yes/No)	Codelist Name	CDISC Submission Value	CDISC Synonym(s)	CDISC Definition	NCI Preferred Term
C66727	C25250		Completion/Reason for Non-Completion	COMPLETED		To possess every necessary or normal part or component or step; having come or been brought to a conclusion. (NCI)	Complete
C66727	C28554		Completion/Reason for Non-Completion	DEATH	Died	The absence of life or state of being dead. (NCI)	Dead
C66727	C38155		Completion/Reason for Non-Completion	DISEASE RELAPSE	Recurrent Disease	The return of a disease after a period of remission.	Recurrent Disease
C66727			Completion/Reason for Non-Completion	DRUG INTERRUPTION		Interruption of study medication dosing not related to an Adverse Event.	

Table-1: Representation of codelist

Typically, in the absence of a MDR, the codelist is represented and maintained as spreadsheets in a central location like a SharePoint site. When users access, download the excel spreadsheet, review and ensure the content has been used in a study, home-grown tools or macros can ensure the compliance and consistent representation of this extended codelist. These home-grown tools need to work across all studies and programs and will have to be maintained separately, openly available and updated regularly, to make them smart and ensure there are no bugs.

The fundamental change is in bringing the standard, as represented in an MS-Excel® file, into a repository where the template governs the structure of the file. When any structural changes are encountered the MDR automatically recognizes the issue and provides a quick feedback. This prevents any structural change creep that easily occurred when using MS-Excel ®. Another use of such a framework would be supporting the creation of organization specific standards, where any existing relationships between the terms, synonyms or meanings can be maintained and any discrepancies can be assessed and corrected. There is no need for any special programming to do this task. This aligns to the concept of having a single source of truth that can be accessed by multiple users in an organization.

COMPARISONS AND SHIRE BUSINESS COMPLIANCE:

As industry progresses, standards evolve and CDISC releases versions of metadata spreadsheets for different foundational standard models, implementations and Controlled terminology, it becomes very important to understand the version differences. To ensure such changes are easily identifiable, CDISC releases files which highlight the differences from the current to the previous version. When there is a need to identify differences between current and any prior version, it is not always easy to check for differences without custom programming in SAS®, Visual Basic® and would require resource with this programming knowledge. As ongoing and older clinical trials use different versions of standards, it is important to understand how they relate to each other.

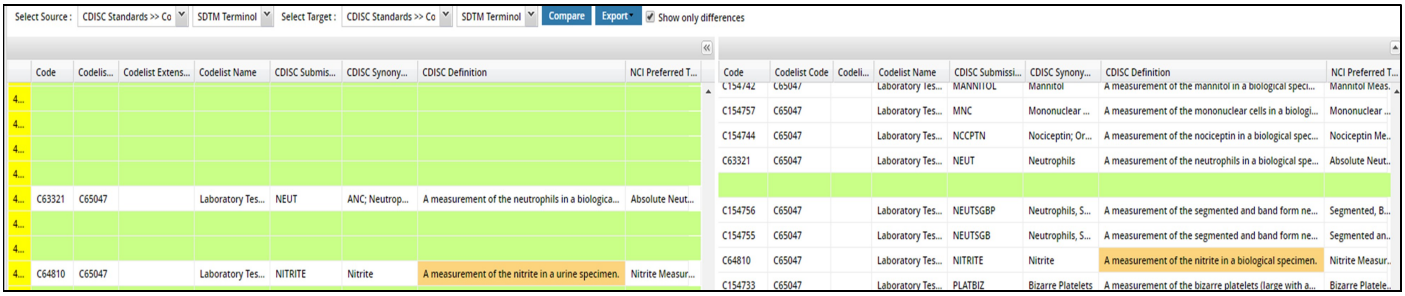

Fig-2: Comparison of the controlled terninology with “new terms” in green and “changes” in yellow

When standards are loaded in the MDR, a comparison feature is available to all the end users, including those with read-only privileges. This becomes a self-serving feature that allows anyone to understand the differences they are interested in, without requiring technical expertise.

When metadata is published either from CDISC or represented within an organization, it needs to be curated and reviewed to ensure Shire business rules are followed. For example, ensuring 8-character for SAS variable name and 40-character for SAS variable label. This ensures the metadata defined in the specifications can meet the needs of down-stream processes, enables early detection of issues thereby streamlining the submission needs. The business checks ensure traceability between standards and that source and targets are defined to ensure horizontal type of relationship. This sort of checks can be made available to all users as standard tools and reports.

INTERPRETATION REPRESENTATION:

The Shire SDTM Governance Team is responsible for interpreting the CDISC SDTM concepts and representing the standards metadata as specifications. Shire's SDTM Toolkit is the name of the SDTM Standards Library created within Shire. This SDTM Toolkit is part of the beginning-to-end standards implementation. It maps the collected information and is also the source for the ADaM datasets. As part of the internal Standards Development cycle, a vendor-neutral SDTM Toolkit is updated and published each year. The SDTM Toolkit consists of metadata specifications, business rules and best practices for study standards implementations. It is provided to multiple strategic CRO Partners, who implement these standards within their environment.

To support Shire's vendor-neutral SDTM standards, three Microsoft® Excel® utilities were developed, which along with other materials make up the Shire SDTM Toolkit:

- The Shire Study Start tool
- The Shire Trial Design tool
- The Shire SDTM Interpretation Metadata

Representing the standards in a spreadsheet format is not optimal since relationships cannot be maintained, hence it does not have the ability to check for integrity. Suppose one uses a foundational standard such as SDTM to represent the specification in spreadsheet and needs to check if the CDISC controlled terminology is used, this can be done using custom programming in SAS®, Visual Basic ® or other programming utility. This requires a technical programming resource.

By using the MDR, a template that defines Shire's interpretation can be represented. The tool provides a way to check for integrity between different items. The need to represent this information in vertical format would be necessary. This frees up the programming resourcing to work on tasks that better utilizes their expertise. Currently we are exploring ways to make the metadata smart and define relationships so that the exported file can be consumed by the end users as a data-driven specification.

STANDARD COMPLIANCE CHECK:

The regulatory agencies (FDA and PMDA) have released the validation checks and business check definitions which provides the industry with an overview of the submission expectations to the agencies. Similarly, CDISC submission standards, have defined compliance checks that provides us with an overview of the metadata, the industry has to comply with. There are tools developed by software vendors such as Pinnacle21 that contain some of the defined checks. When defining the standards' metadata, it is not straightforward to interface with them unless these are implemented as a library. Alternatively, custom programs can be developed to check for compliance and ensure the metadata is curated.

By using the MDR, any of the defined metadata checks can be custom programmed in the tool to understand how they can be checked for compliance irrespective of the source from which they came. There may be data related checks that can be programmed when checking for study data compliance. These business checks can be created once and can be applied multiple times on multiple clinical trials.

STREAMLINE PROCESSES AND GENERATE DATA DRIVEN SPECIFICATIONS

Sponsor organizations rely heavily on the deliverables from CROs to support their submissions. Based on the recent ICH E6 R2 requirements, it is important for sponsor companies to oversee these CRO on the execution and quality of such deliverables. When companies work with the CRO Partners processes should be established and the CRO business practices should be thoroughly understood. Sponsors should from the start share their expectations on how to use data standards.

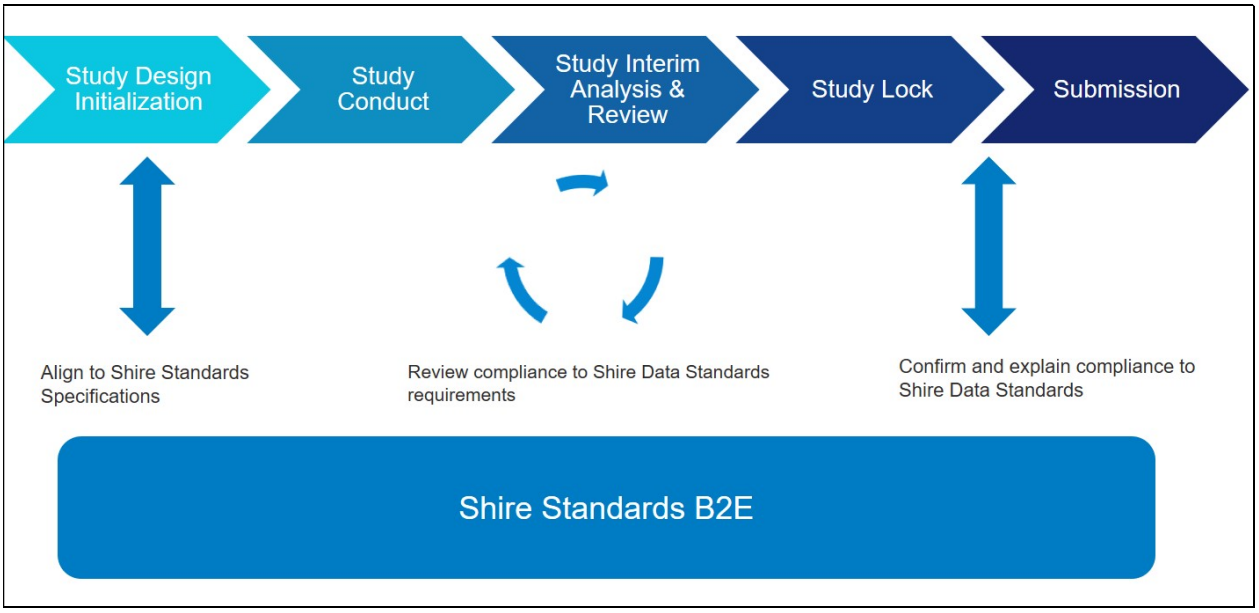

Fig-3: Shire Standards use during Clinical Development Process

The above figure shows different stages where Shire Standards is used in the clinical development process. During the study design initialization, alignment to Shire Standards and its expectations can be shared via interpretation specifications which becomes very crucial when receiving submission ready deliverables. During study conduct, Study Interim Analysis and Review, the CROs typically send the mapping specifications to a sponsor who reviews and agrees to for each study. This may also include the review deliverables such as SDTM annotations to ensure compliance to the Shire data standards. At the Study submission, the study deliverables are compliant to Shire Standards that includes CDISC and regulatory needs. During study analysis, it is important to review how the study is compliant to Shire Standards and confirm at the end of the study during submission. This ensures the quality of the submission package and ensures resolving and discussing data issues earlier. Sometimes a study or program will not be able to follow Shire Standards and the study team can ask for waiver to the Governance Team as “waiver requests”.

There seems to be different implementations and processes used within the CRO’s which complicates the automation of review of this information and cause many challenges. Often CRO processes are manual or semi-automatic for mappings to pre-defined specifications which are sent as Microsoft® Excel® sheets. Each study team reviews study mappings to ensure the accuracy. Since it is using spreadsheets, there is a chance of unstructured data in them which causes a challenge for Sponsors to create a process to automate this review. Sponsors often create transformation utilities or custom tools or programming to extract, transform and loading of the mapping specifications received.

By using the MDR, sponsor expectations to how to receive mapped information can be sent as a requirement to the CROs to receive structured data. This provides an opportunity for sponsor companies to streamline their processes to review of mappings, which enables the programming team to semi-automate the review of submission-ready deliverables thereby saving time and resources. This ensures the overall oversight of external vendors and compliance to Shire standards.

STUDY COMPLIANCE CHECK

The recent ICH E6 R2 requirements and expectations for sponsor companies to oversee the CRO on the execution and quality is very critical to ensure the safety and efficacy of a study. At Shire, we are serious about ensuring the quality through compliance check. This involves not only defining the organizational interpretations but also to identify what it means when Shire Standards are not followed. Like FDA and PMDA, Shire has identified within its metadata the key variables and its non-compliance classifications to be either error or violation. Such metadata was used in custom programming to evaluate pattern analysis of the study mappings to check how much of study is compliant to the standards defined.

When a study uses the global standard represented in the MDR, it can inherit the standards as is with little code change. Suppose a study needs to divert from the global or therapeutic standard the metadata will have to be updated which may cause a divergence to the code and automation.

By using the MDR, Shire has semi-automated this process by using the transformation utilities within the tool the compliance checking to the received mapping specifications. This will provide sponsor companies oversight and ensure better quality for analysis of the submission deliverables which supports the reduction in time and resources and in-turn aid in reducing the costs over time.

STANDARDS DEVELOPMENT AND GOVERNANCE WORKFLOW PROCESS

The Shire Clinical Standards Governance model oversees the standards development process. The Standards governance teams are responsible for defining and reviewing the shire standards to ensure the metadata is compliant and conformant across different standards. This information is sent to the Clinical Standards Governance Board which reviews and approve the Shire interpretations, best practices and business rules. Once the metadata is approved, it is released to the study teams for use. The review of metadata is semi-automated with the use of custom programs and had to be manually reviewed throughout the life cycle.

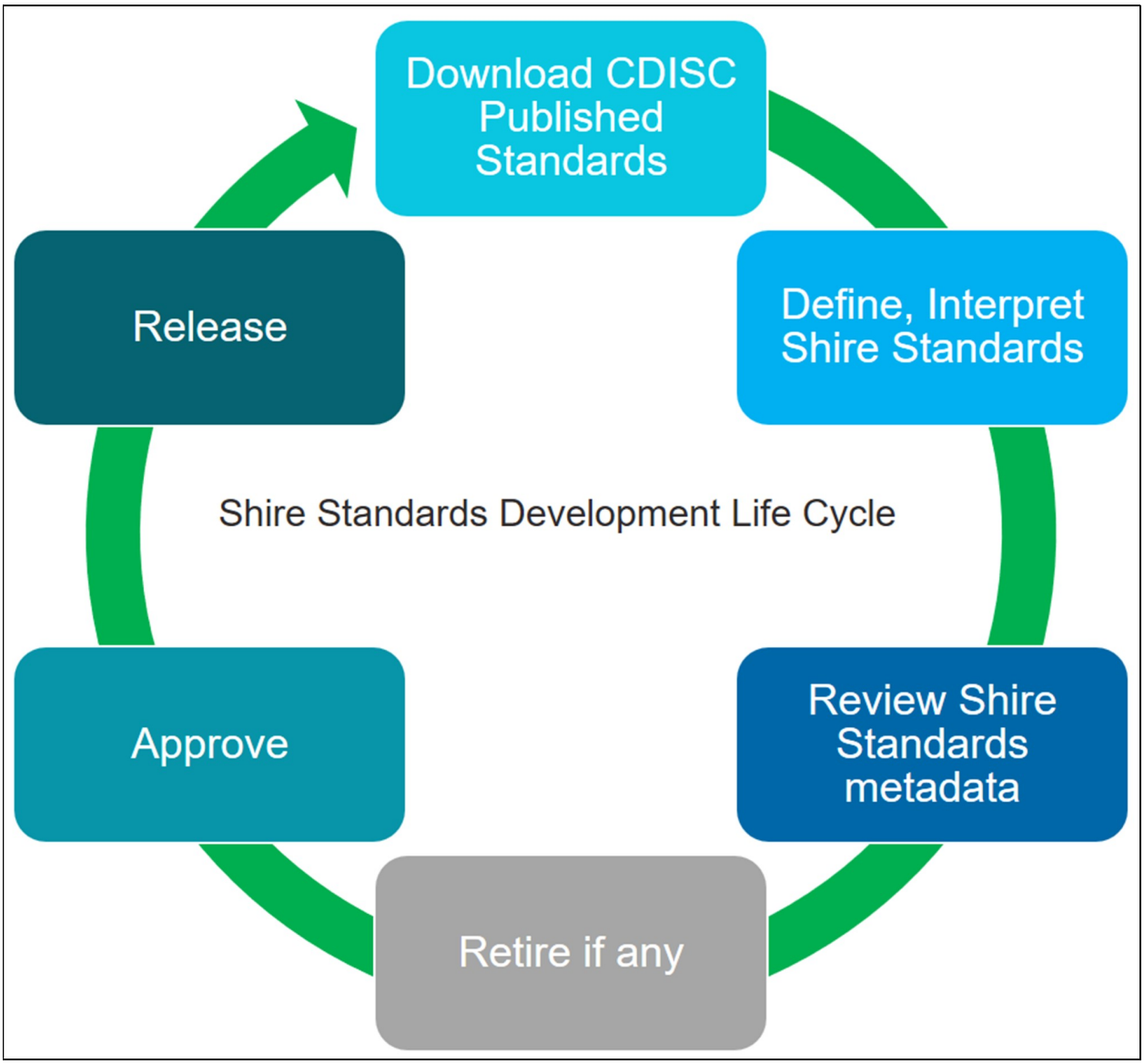

Fig-4: Shire Standards Development Life Cycle

The Shire Standards development starts with the **downloading** and reviewing of the CDISC published standards then the governance teams **define and interpret** any new standards needed for Shire, **review** the metadata and curate it which may lead to **retiring** any old metadata. Then it is sent for **approval** to the Clinical Standards Governance Board. After approval, the standards are published, making them available to the entire organization.

With the use of the MDR the standards development life cycle is automated. Groups of trained users can access the metadata, review and compare the changes as well as send comments. The MDR workflow feature ensures that the metadata changes are communicated via auto-generated notifications and changes are transparent in the audit trail. The Clinical Standards Governance Board members can also review and approve the metadata within the tool, which triggers the standards lead to release the metadata as a toolkit.

PROCESS, TRAINING AND COMMUNICATION

As new tools are implemented in Shire's environment, it is important to create new processes, trainings and communications, to ensure that the end users are properly equipped. The MDR has new features which includes workflow, email notifications, comparison features which the end users should be trained on.

NEXT STEPS:

The MDR implementation uses agile methodology and it is phased. As the end users have more experience with the tool, the next phase will include integrating with external systems, where common definitions representing both clinical and operational metadata are defined. It will also be useful to understand if CRO systems can be integrated using APIs so that the information from the MDR can be transferred to the CROs seamlessly. Defining the mappings earlier enables the CRO configuration to be applied across multiple clinical trials therefore leading to less code change.

MDR BENEFITS AND CHALLENGES:

The benefits of MDR include

- The ability to easily manage standards across the clinical development life cycle. Defining the mappings, business rules to support the upstream and downstream processes.
- When the information contained in the MDR is maintained to be current and accurate, it is the one stop shop and is "the ultimate source". The standard templates and formats representing the metadata can be used to drive automation of business processes, such as eCRF annotation, dataset generation and compliance checking.
- Labor intensive processes using spreadsheets are simplified and it forces teams to think about metadata in a new way which can lead to easier transformation and analysis.
- The resourcing needs for maintaining the MDR is changed from technical (knowing programming language such as SAS®) to understanding the standards representation and relationship modelling to represent standards.

The challenges of MDR include

- When integrating with disparate sources, understanding the common metadata and how the information will be represented in the other systems.
- Since the templates dictate how the information is organized within the MDR, any changes such as new column to the template, needs careful consideration since this was an easy modification when using spreadsheets.
- Clear roles and responsibilities need to be identified and new processes based on the team size should be planned.

CONCLUSION

This paper describes Shire's representation of Clinical Data Standards and the implementation of the metadata in a MDR. This is the first step towards the journey for a paradigm shift to review and manage standards metadata. The paradigm shift within clinical standards is when tools such as a MDR will prompt how we manage and work with metadata so that the processes within and across organizations can be automated and streamlined. It also changes the resourcing needs and knowledge necessary, e.g. instead of hiring programming resources, now data analyst resources with mapping, data analytics and overall data standards knowledge would be essential.

Systems and applications can be developed where data can be analyzed and summarized (e.g., pattern analysis and data mining) using automated tools, to ensure the quality and safety of the standardized clinical data. Sponsor companies can invest in such tools where they can oversee the CROs and can satisfy ICH E6 R2 requirements. This paper discussed a number of considerations when implementing a MDR, such as how to organize the information, templates to represent relationship which help in the comparison and compliance checking of business rules.

Features to run compliance and custom standards checks support the need for the data to be compliant with CDISC and regulatory submission needs. The metadata representations and its use within clinical trials to capture data, supports traceability and thus better data quality. Data traceability, automated streamlined processes and regulatory requirements, are critical reasons for which representations within a MDR and processes to create data driven

specifications become a foundation towards building a future paradigm shift. This allows organizations to create one source of truth for metadata and data requirements.

REFERENCES

1. Kuhn, Thomas (1970). The Structure of Scientific Revolutions (2nd, enlarged ed.). University of Chicago Press. ISBN 978-0-226-45804-5.
2. Veena Nataraj, Erica Davis (2018): Avoid chasing one's Tail - Challenges and Solutions for managing changes in SDTM Standards Development - PharmaSUG 2018 – https://www.lexjansen.com/pharmasug/2018/DS/PharmaSUG-2018-DS11.pdf
3. Veena Nataraj, Erica Davis (2018): Improving Metadata Compliance and Assessing Quality Metrics with a Standards Library – PharmaSUG 2018 – https://www.lexjansen.com/pharmasug/2018/SS/PharmaSUG-2018-SS12.pdf
4. CDISC SDTM Model 1.4 - https://www.cdisc.org/system/files/members/standard/foundational/sdtm/study_data_tabulation_model_v1_4.pdf
5. CDISC Study Data Tabulation Model Implementation Guide (SDTMIG) Version 3.2
6. FDA Data Standards Resources - https://www.fda.gov/forindustry/datastandards/studydatastandards/default.htm
7. PMDA Study Data Resources –
    a. Technical Conformance Guide
    b. FAQs on electronic data submissions
8. ICH E6 R2 - https://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Efficacy/E6/E6_R2__Step_4_2016_1109.pdf

ACKNOWLEDGMENTS

I would like to acknowledge my colleagues, Governance Teams and especially our leader, Diane Piper who has championed Clinical Standards and Clinical Standards Governance Model and its implementation.

CONTACT INFORMATION

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

Veena Nataraj
Shire, Takeda
300 Shire Way
Lexington, MA 02420
Work Phone: 781-482-0906
Email: veena.nataraj@takeda.com