PhUSE US Connect 2018

Paper SI08

A Successful Meta-Data Repository (MDR) - It's About Managing Relationships

Phil Giangiulio, GSK, Collegeville, PA, USA
Melanie Paules, GSK, Collegeville, PA, USA

ABSTRACT

Building and implementing a clinical Meta-Data Repository or Registry (MDR) is an expensive, exhaustive and time consuming commitment for any pharmaceutical organization. Ultimately, the success of an MDR will be determined by its longevity and return on investment. A well designed, CDISC standards based MDR, although crucial, is only one factor in determining success. Managing metadata, process and people relationships are also vital components. This paper examines a few key expectations of a successful MDR, their connection to effective relationship management, and how those relationships could be utilized within an MDR driven clinical operation environment.

Introduction

Well designed, CDISC based, integrated data standards are the foundation for implementing a successful clinical Meta-Data Repository or Registry (MDR). A successful MDR manages the integrated set of clinical data definitions and the relationships between them, that allows for beginning to end traceability and management and the use of those definitions on a given clinical trial. Although the immediate worth of the MDR is felt by those who managing the organization's clinical trial data standards, the most significant value is determined by how effectively the metadata can be leveraged to increase the quality and decrease the time required to deliver industry standard, compliant data and analyses of clinical trials.

This paper examines how defining and leveraging an MDR's metadata and relationships can enable and realize those desired clinical trial efficiencies. We will discuss a few key expectations of a successful MDR, types of relationships, and some guiding principles for managing those relationships.

MDR Expectations

The clinical trial process of a drug development organization, like any other process, can be characterized by a maturity model. The Capability Maturity Model (CMM) development by research funded by U.S. Department of Defense, describes the term "maturity", as it relates to the degree of formality and optimization of processes, from ad-hoc practices (level 1), to formally defined steps (levels 2 & 3), to managed result metrics (level 4), to active optimization of the processes (level 5).

The most significant return on investment of an MDR is achieved when it is leveraged to move operational processes to levels 4 and 5 of the capability model. In other words, when the MDR enables metrics driven process management and process time and quality optimizations. But, how can an MDR drive this optimization? Let's see.

Wikipedia's Metadata Repository "Reasons for Use" currently stand as follows:

1. **Integration of the metadata across the organization.**
2. **Build relationship between various metadata types**
3. **Build relationship between various disparate systems.**
4. Define business golden copy of definitions.
5. Version control of the changes at structure level.
6. Interaction with Reference data
7. Link view to master data.
8. Automatic synchronization with various authorized metadata source systems.
9. **More control to business decisions.**
10. Validate the structures by overlapping the models
11. Discovering discrepancies, gaps, lineage, metrics at data structure level

Notice that a significant number are dependent, in some fashion, on the definition, codification and use of relationships. Building and leveraging metadata relationships are critical to a highly successful MDR, and to driving the operational efficiency. Drilling down from generic reasons to employ an MDR, here are a few specific abilities that rely on relationships.

Metadata Conformance/Reuse

Effective metadata sharing is enabled by two types of relationships, one is inheritance and the other is versioning. Capturing the relationship between study data specifications and organizational standards, and relationships between organizational standards and industry standards, enables compliance monitoring and management. Managing study specifications with compliance to data standards is essential for driving clinical process capability maturity. Identifying appropriate versions for accurate compliance checking is critical.

Traceability/Impact Analysis

For the purposes of this discussion, we define Beginning to End (B2E) Traceability as the tracking of related clinical data from its origination/entry into the clinical trial process through to the analysis and final reporting of those data. (Eventually the desire is to have the beginning start with protocol creation and the end carry through into the assembly of the clinical study report and beyond, e.g., disclosure.) This tracking is most effectively achieved by defining these clinical data in our industry standard CDISC data standards and then specifying mapping relationships as data moves from acquisition standards (CDASH), to tabulation (SDTM), to analysis (ADaM), to reporting (Analysis Results Metadata – ARM). Once these "data track" relationships are established, one can determine the impact of making a change at any point along the path.

Industry standards evolve, science progresses and regulations can be modified so the metadata used for studies needs to change in order to keep up with new requirements. There is a common misconception that standards are static and that metadata once defined is unchanging. This is far from reality in a clinical trial setting where old and new standards will exist concurrently over an extended period. Relationships and versioning of them is essential to accommodate both on-going and older studies as well as the connection to the standards that was used.

Clinical Process Efficiency

Although study metadata conformance and B2E traceability clearly enable clinical process efficiencies, the most significant productivity gains are realized when validated, clinical system programming can be reliably employed for a given study. Since most of the processes involve receiving clinical data in one data standard and transforming it to another data standard, unambiguous specifications for those standard to standard mapping transformations are essential. It is not enough to simply identify the track data takes along its B2E standards path, but how those data are to be transformed must also be incorporated in the mapping relationships. MDRs that accurately instantiate and communicate these standards mapping relationships will be the best enablers of clinical system productivity.

Metadata Relationship Types & Constructs

Now that we've identified inheritance and mapping transformations as key relationships for a successful MDR, let us examine components of those relationships, along with versioning, and how they might be implemented.

Versioning

The ability to identify a specific version of any MDR object (including codes, code lists, variables, data sets and mapping relationships), at any given in timepoint, is a basic requirement. It is needed for other capabilities including Inheritance, B2E Traceability and Mapping. The MDR, must have versioning of objects at the system level to keep track of changes or corrections. However, for the management of MDR objects for use in study specifications and dealing with the evolving industry standards, object versioning based on effective begin and end dates provides great flexibility for implementing this functionality.

Inheritance

Inheritance can be considered a vertical type of relationship. It captures child/parent associations including study to organization {company}, and organization to industry {CDISC}. These relationships are indispensable for compliance monitoring and management and impact analysis.

B2E Traceability

As beginning to end implies, this traceability can be considered a horizontal type of relationship. They are links or mappings between source and target elements and element sets of data standard models.

Example source to target links include:

- Source data acquisition item groups and items mapped to Target review model data sets and variables.
- Source SDTM data sets and variables mapped to Target ADaM data sets and variables.
- ADaM data sets and variables mapped to Target ARM display and results.

Source to target links need to accommodate multiple sources required for a single target and single sources used for multiple targets. The description and sharing of these, sometimes complex, relationships are key to an MDRs success. For discussion purposes, we will identify them as Standards Mapping.

Standards Mapping

Let's look at one standards mapping for the Adverse Event data, on the B2E data track. More specifically, consider the transformation from an AE Data Review Model to the SDTM AE data domain. The mapping needs to identify the source for each target variable, its attributes, and any transformations required. One transformation example would be the conversion of date and times stored in collection system formats into CDISC compliant ISO 8601 date-time format. Additionally, some variable mappings require pulling in non-AE information for completion. Examples include MedDRA terminology for preferred terms and subject study start reference date for study day derivations. The mappings to include non-AE information require instructions which specify how to pull together sets to enable other mappings

This information is fundamental in accurately describing mapping relationships. Instantiating it as well formed metadata is to key enabling metadata driven data transformations. One strategy for organizing mapping information is to categorize or group them as either element (e.g. variable) or element set (e.g. data set) transformations. Set transformations could describe the required data points from source data sets and instructions on how to access and select them. Element transformations could then describe how source elements are to be used to create the desired target elements. Element transformations could also include or reference other relevant metadata like controlled terminology and value level clinical content. An MDR containing these types of mapping relationships and this organization/structure is well positioned to inform clinical system transformations. Currently, no industry-defined standard exists for these mappings.

The following example is a segment of mapping between an Adverse Event (AE) Data Review Model (DRM) data set, as a source, being mapped to a target of an AE SDTM data set. It includes variable mappings (direct, transformation and derivations) as element transforms. Some of the element derivations require elements (variables) from another element set (data set). The pre-requisite inclusion of these variables is shown as set transforms.

Standards Mapping Example

Target Model	Target Set	Target Element	Source Model	Source Set	Source Element	Transform Level	Transform Name	Transform Category	Transform Item	Set Operation	Set Selection	Set Condition	Set Element Selection
SDTM	SDTM.AE		DRM	Review.DM		Set	SDTM_AE		DM_JOIN	LEFT JOIN	Review.AE AS A <OPERATION> Review.DM AS B	A.USUBJID EQ B.USUBJID	A.*, B.REFSTRDT, B.REFENDDT
SDTM	SDTM.AE		DRM	Reference.MEDDRA		Set	SDTM_AE		MEDDRA_JOIN	LEFT JOIN	Review.AE AS A <OPERATION> Reference.MEDDRA AS B	A.AELLTCD EQ B.AELLTCD AND B.AEPATHCD EQ '1' AND B.AENC EQ 'C' AND B.AEVER EQ <Version Parameter>	A.*, B.AELLT, B.AEPTCD, B.AEPT, B.AESOCCD, B.AESOC, B.HLGTCD, B.HLGT, B.HLTCD, B.HLT
SDTM	SDTM.AE		DRM	Review.SE		Set	SDTM_AE		SE_JOIN	…	…	…	…
SDTM	SDTM.AE	AETERM	DRM	Review.AE	AETERM	Element	DIRECT_MAP						
SDTM	SDTM.AE	AESTDTC	DRM	Review.AE	AESTDTC_DT	Element	ISO_MAKE_DTC	Transformation					
SDTM	SDTM.AE	AESTDY	…	SDTM_AE.DM_JOIN	AESTDTC_DT	Element	CALC_STDY	Derivation					
SDTM	SDTM.AE	AESTDY	…	SDTM_AE.DM_JOIN	RFSTDTC	Element	CALC_STDY	Derivation					
SDTM	SDTM.AE	AELLT	…	SDTM_AE.MEDDRA_JOIN	AELLT	Element	DIRECT_MAP						
SDTM	SDTM.AE	AEPT	…	SDTM_AE.MEDDRA_JOIN	AEPT	Element	DIRECT_MAP						

Clinical Content/Value Level Metadata

Much of the data collected for clinical purposes are best understood in clusters or associations. A classic example are subject vital signs. A blood pressure measurement, the units of the measure, and the position of subject during the measurement provide a complete picture/understanding of the reading. For findings data, CDISC includes these clusters or groupings in their Biomedical Concepts. Storing these clinical associations as discrete objects allows for their consistent and repeated referencing beginning with their Data Acquisition specifications through to their Analysis and Reporting specifications. This is critically important in ensuring clinical data reporting reflects the data as it was collected.

Due to the normalized structure of SDTM findings class data domains, values and associated metadata of variables in these clinical content clusters vary by the findings test being conducted. These variable attributes appear as Value Level Metadata (VLM) in the define.xml descriptions. An MDR that stores and manages Clinical Content/VLM, and references it in standards definitions and mappings, is well positioned for success.

The example below is a segment of a clinical content for Systolic Blood Pressure measurements in the Vital Sign group. Notice that the SYSBP_POS item includes the entire POSITION_VS code list while the SYSBP_TEST and SYSBP_TESTCD items contain specific codes of their respective code lists. Often, the collection form will collect a

single data point, and using a relationship to the clinical content cluster of clinical content, a mapping relationship can provide instructions for creating a "complete" SDTM record in a findings dataset. Referencing this clinical content definition in the Data Acquisition, Data Review Model, SDTM and ADaM specifications, ensures continuity from data collection, through reporting.

Clinical Content / Value Level Metadata Example

Content Category	Content Set	Content Item	Code List	Code	Assigned Value	Element Transform	VLM
VSTESTCD	BMI	…	…	…	…	…	…
VSTESTCD	BSA	…	…	…	…	…	…
VSTESTCD	DIABP	…	…	…	…	…	…
VSTESTCD	SYSBP	SYSBP_TESTCD	VSTESTCD	SYSBP		ASSIGN_CODE	Y
VSTESTCD	SYSBP	SYSBP_TEST	VSTEST	Systolic Blood Pressure		ASSIGN_CODE	Y
VSTESTCD	SYSBP	SYSBP_ORRES				DIRECT_MAP	Y
VSTESTCD	SYSBP	SYSBP_ORRESU	VSRESU	mmHg		ASSIGN_CODE	Y
VSTESTCD	SYSBP	SYSBP_STRESN				CONV_TO_STD	Y
VSTESTCD	SYSBP	SYSBP_STRESU	VSRESU	Pa		ASSIGN_CODE	Y
VSTESTCD	SYSBP	SYSBP_POS	POSITION_VS			LOOKUP_DECODE	Y
VSTESTCD	RESP	…	…	…	…	…	…
VSTESTCD	TEMP	…	…	…	…	…	…

Relationship Management / Sharing

Now that we have identified Inheritance, B2E Traceability, Standards Mapping as important relationships and discussed some of their key components like Versioning and Clinical Content, let us consider a few aspects of managing them.

Standards Governance

As mentioned at the onset, well designed, CDISC based, data standards are the foundation of successful Clinical MDR. The minute those standards are used in the organization's production environment is the minute that their governance and maintenance begins. The standards governing structures and processes, incorporating conflict resolution, need to be defined and understood by all affected parties. A successful MDR should include workflow functionality that is configured to support the organizations standards governance and management processes. When properly implemented, version, inheritance, and B2E data track relationships can be leveraged to provide valuable impact analysis assessments crucial for business decisions such as when to implement a new CDISC standard update.

Global to Study Metadata Sharing

Study building functionality, that utilizes an organization's standards and establishes inheritance relationships for the creation of initial study specifications, is essential to achieving any meaningful clinical process capability maturity. Once study specific modifications are finalized, inheritance relationships to the organizational standards can be leveraged for compliance checking and metrics. A good standards governance practice for necessary study specific metadata objects are to identify them as "provisional" and include them in periodic standards update reviews. This division between provisional and standard can be challenging when study specific context in represented in an industry standard datasets.

MDR to Clinical System Sharing

The final MDR link, in completing the efficiency chain, is the effective communication or sharing of standards definitions, including clinical content/VLM, and mappings to clinical system transformation processes. Well-formed standard mapping relationships can be utilized/consumed by clinical systems as specifications for data transformation programs. A clinical system library of functions and programs, developed to use MDR standards mappings as specifications and inputs, can eliminate hardcoding clinical content for collection forms, and greatly reduce the need for custom programming or, when needed, simplify programming modifications.

Effective, reliable MDR to Clinical System communication is essential for metadata sharing.
Standards mappings in an unambiguous, easily consumed format is essential. For clinical trials metadata, ODM xml, using appropriate extensions, would be a good choice. Providing an API utilizing REST services to access MDR standards mappings, would enable easy, standardized access for clinical systems.

Risk Mitigation

Like all large, complex projects, effective risk mitigation could differentiate success from failure. Here are a few risks that merit consideration when building an MDR.

Unclear roles and processes for Standards Governance

Insufficient procedures regarding the maintenance and up-versioning of standards and the lack of a conflict resolution processes could lead to lower quality, outdated standards.

Gaps in Standards and Mapping and Non-harmonized (B2E) Standards lead to needless customization that reduce both quality and process efficiency

Weak/Ineffectual compliance monitoring

Effective compliance monitoring and management are arguably the most important mechanisms for ensuring that well-constructed, highly effective standards remain that way.

Unclear delineation of standards vs clinical system, operational metadata

Forcing operational metadata, specific to a particular clinical system, and its management into the MDR could have the effect of incorporating MDR development workflows into clinical processes there by decreasing efficiency.

No mechanism for effective metadata sharing

The highest quality standards housed in a well-managed MDR, must be available to and leveraged by clinical systems to drive process efficiency. Providing clinical systems users, the capability for on-demand extracts of MDR study specifications, and the tools necessary to consume those specifications in their day to day work, is essential for multiplying the MDRs return on investment.

Conclusion

Building a well formed MDR is only the first step in cultivating its success. Effective management of the standards held within, and compliance to them, adds to its value. An MDR that defines a complete description of the mapping of data between those standards, and can effectively communicate those mapping transformations to clinical systems, can drive significant process efficiency that lower clinical study timings and cost. Building and leveraging the type of relationships discussed above is the key to ensuring your MDR's success.

References

Characteristics of Capability Maturity Model (From Wikimedia Commons, the free media repository)
https://commons.wikimedia.org/wiki/File:Characteristics_of_Capability_Maturity_Model.svg

Metadata repository (From Wikipedia, the free encyclopedia)
https://en.wikipedia.org/wiki/Metadata_repository

Contact Information

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

Philip Giangiulio

GlaxoSmithKline
1250 South Collegeville Road
PO Box 5089
Collegeville, PA 19426-0989

Email: philip.j.giangiulio@gsk.com

Melanie Paules

GlaxoSmithKline
1250 South Collegeville Road
PO Box 5089
Collegeville, PA 19426-0989

Email: Melanie.A.Paules@gsk.com