

Taming Rave: How to control data collection standards?

Dimitri Kutsenko, Entimo AG, Berlin, Germany

Table of Contents

Introduction	1
How to organize metadata.....	2
How to structure metadata.....	3
How to manage metadata	4
How to check metadata.....	5
How to present metadata	6
Insights and best practices.....	6
Conclusions	7
Contact Information	7

Introduction

The project which is described in this paper was started at the end of February, 2013. The MDR (Metadata Repository) which is implemented in the scope of this project is clearly a new trend in the pharmaceutical industry. Entimo was selected as a vendor to deliver the technology and to implement the MDR at a mid-sized, global pharmaceutical company.

The project has had from the very beginning the following goals: In the first phase, the operational (data collection) metadata will be standardized to enable the organization to effectively and efficiently set up studies in the EDC system. In addition, SDTM and controlled terminology governance shall be implemented in the MDR in the first phase to address the requirement of regulatory submissions in SDTM. In the following project phases, other processes and data models will be included in the MDR. Another strong requirement identified by the project team was to support study level metadata in addition to global and project level standards.

In general, Entimo recommends an iterative project approach where the project is divided into phases. Within each phase, workshops are conducted to develop and review the requirements (the number of workshops depends on the anticipated complexity of the process landscape). Such an iterative approach is possible due to the fact that the MDR and other solutions based on the entimICE platform (entimICE - Entimo Integrated Clinical Environment) provide the flexibility to adopt changes in the configuration (hierarchies, data models, workflows etc) through the administration module without changing the application layer ("configuration versus customization").

PhUSE 2013

Medidata RAVE – the leading industry EDC system - is used by the customer as the EDC system. However, the described approach appears to be generally applicable to other EDC systems (although some system-specific modifications may be necessary).

The author has no intention of completing a project “inventory”. This paper will merely try to illustrate the project as a set of critical “How to” questions raised and answered in the course of the project. The author hopes that these questions transferred into other contexts and projects might be beneficial for readers.

How to organize metadata

The organization of metadata is understood in this paper as the folder hierarchy visible to MDR users.

The primary goal of the initiation workshop is to define the metadata organization. Though the requirements of the standards to be supported in the first configuration release were clear before the project started (as stated above, operational metadata for Rave, SDTM and controlled terminology), additional design aspects were identified in the first workshop.

Let’s start the review from the end of the process.

The final goal was to support the study setup in the MDR. Consequently, the hierarchy was configured to contain studies grouped by projects. A separate area was configured for global standards. Each area is an independent instance of standards (“a copy”). Across different projects and implementations, I often hear the question: Why not inherit metadata properties from the top level down to the study level? This would allow more consistent definition of metadata properties: you only need to change the attribute at one level (e.g. global) and the change would automatically apply in other inherited areas. In a similar way, the CDISC SHARE project claims to use the inheritance concept. Though technically possible in the entimICE-based MDR, such a design decision has wide-reaching consequences. Due to the fact that global standards might evolve over time, the inheritance could mean that study level metadata would be “automatically” updated in studies AFTER the study has been set productive in the EDC system – a clear integrity violation as the study build metadata and the study in EDC should be in synch with each other. A mix of approaches (inherit and copy) could be possible (e.g. stopping inheritance at the project level). This would lead to a more complicated SOP and workflows and would impose additional requirements on the training.

Consequently, Entimo’s best practice is to implement one approach across a system configuration to make user training simpler (the metadata world is complex enough). With the available toolset for impact analysis described below, consistency is enforced within the MDR without metadata inheritance. However, the question “copy or inherit” is merely a philosophical one resulting in differences of the underlying governance process.

Several specificities of the used EDC system influenced the hierarchy configuration. As a matter of fact, RAVE delivers data extracts in a different format than that used for the study setup. Though related, the extract data model contains additional attributes added by RAVE. A separate area for extract metadata was configured in the MDR. The project team created a mechanism to generate extract metadata from the setup metadata – a best practice that significantly improved metadata consistency.

Two additional remarks on the hierarchy defined in the project are worth mentioning. CDISC delivers standards in “generic” form where some specific attributes are defined by sponsors (e.g. variable length which are required in the down-stream processes in order to create SAS datasets). The project team decided to create a special area called “references” where standards are stored as they come from CDISC and other external organizations (e.g. NCI). These are released, stable standards and proposed standards under public review. The global area serving as the basis for projects and studies contains the metadata already adapted to organizational purposes.

PhUSE 2013

Another best practice developed in the project is to use metadata templates as subsets of the global area to make study setup easier and less laborious. Though criteria for such templates might vary from organization to organization, possible grouping criteria might be study phase and therapeutic area, for example.

How to structure metadata

In this paper the structure of metadata is defined as the data model used to store metadata in the MDR.

The SDTM and controlled terminology (CTerm) was an easier case with regard to the data model in this project and will be discussed first. The SDTM and terminology attributes were configured in the project following the published CDISC specification. In both cases, the project team made the decision to enhance the standard definition with several additional attributes with the purpose of storing organization-specific information.

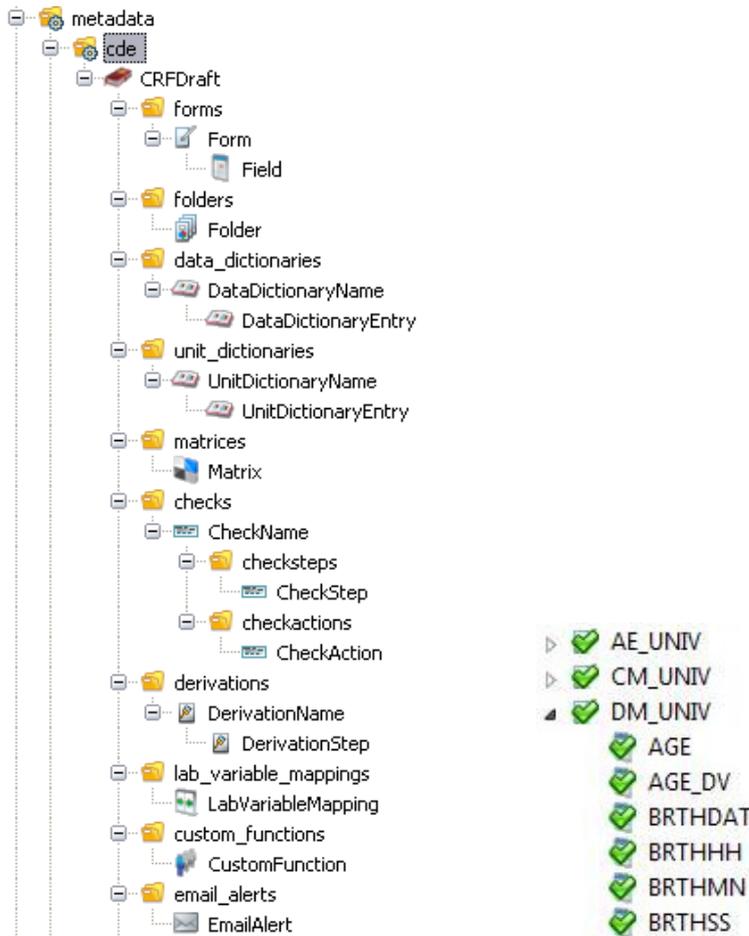
The operational data model turned out to be far more complicated than expected and required intensive discussions.

ODM is a popular CDISC standard that is gaining more and more attention among vendors and the pharmaceutical community. ODM has some clear advantages such as easier interoperability and interfacing between systems in the IT landscape due to its machine-readable, XML-based nature. On the other hand, ODM imposes additional requirements on the metadata structure and content managed in the MDR. In addition, if a sponsor works with a number of CROs and service providers creating content for the repository outside of the MDR, the introduction of ODM-compatible software is needed by all partners across the network – a challenging organizational undertaking.

Due to the fact that Medidata RAVE was the targeted EDC system in this project, ODM had “competition” in the form of ALS (Architect Loader Specification) – an MS Excel-based format supported by RAVE which allows the exporting and importing of the complete study design information required to build studies in RAVE. Also machine-readable, the advantage of ALS is easier communication across the sponsor network without additional tooling needs.

In this project, ALS was chosen as the basis for the configuration of the operational data model in the repository. The screenshot below illustrates the hierarchy which was configured as the project outcome (structural template definition on the left side, a form example on the right side):

PhUSE 2013



Though the “nested” structure of ALS broken down to single data elements made the configuration more laborious, ultimately it allows all study artifacts to be maintained in a clear-cut, understandable and easily searchable structure and makes metadata management (e.g. the creation of new forms) effortless. Also on the operational side, the “standard” specification was extended by several additional, non-ALS attributes for organization-specific information.

How to manage metadata

Metadata management includes definition of the lifecycle per object and the workflow of promoting objects between defined areas (e.g. dev, prod) and levels (e.g. global, project, study) which follow defined access rights.

To make user administration easy, entimICE-based solutions including the described MDR manage access rights based on roles. In addition and on specific occasions, authorized users may overwrite default, role-based access rights settings. With the idea of keeping the number of roles as small as possible, the project team defined in the governance workshop a list of roles which would be configured in the MDR. In the course of the workshop, questions discussed included:

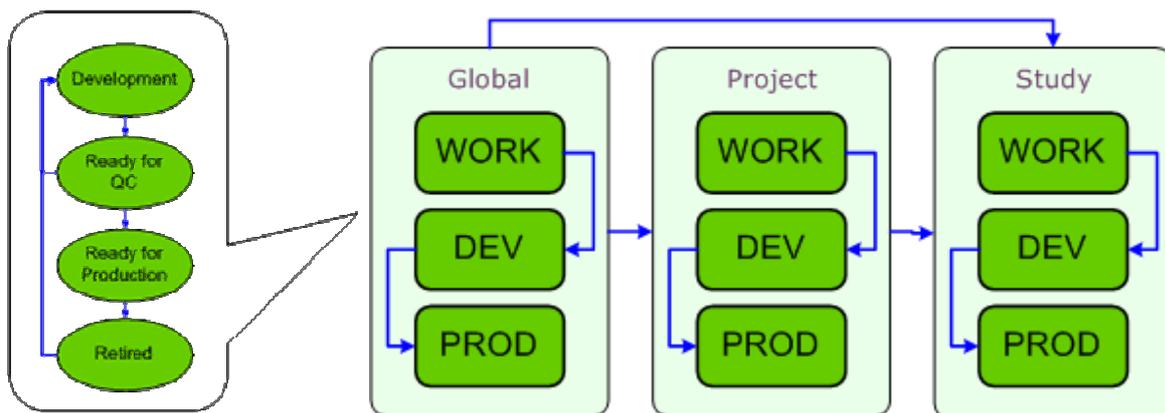
- Who is allowed to access what in which area? (Related to this, additional questions were answered: Are users allowed to copy standards only from a higher level or from anywhere? Are users allowed to copy only from the prod areas or from any accessible folder?)

PhUSE 2013

- Who is allowed to modify what at what level? (The project team quickly came to the common understanding that some fields from the global level are not allowed to be modified in projects and studies.)
- Who carries out quality control in the metadata lifecycle, at which level?

Due to the configurability of the system, the lifecycle definition itself offered design options: As the most prominent example, ISO11170 suggests a definition of the lifecycle for metadata standards. As the definition is ambiguous, Entimo's best practice is to separate lifecycle and workflow. In the project, Entimo's best practice was implemented as follows:

- Per level and data model, the folder structure separates work, dev and prod areas.
- The work area is not quality-controlled. It serves for experimental work and supports versioning of all metadata. The dev area contains a lifecycle from development to the production-ready stage. When ready for production, artifacts are allowed to be copied to write-protected prod areas.
- At the study level, an additional area is available to store metadata from outside of the MDR and is used for comparisons and integration.
- In both the dev and prod areas, metadata can be retired and re-opened within the lifecycle.



The governance of controlled terminology shall reflect regular updates from CDISC. For controlled terminology, the project team worked out a workflow different from the other data models. However, the description of the governance process for controlled terminology is deliberately left out of the scope of this paper.

How to check metadata

Entimo came very early to the understanding that controlling the quality of metadata was a key factor in improving the quality of data in down-stream processes. The concept of meta-metadata was developed, which can be understood as the design model and rules for metadata. Addressing the requirement to support different metadata models, the meta-metadata engine was extracted from the application layer and is accessible to administrators with special privileges. This gives the repository an enormous flexibility to enhance data models with changing requirements "on the fly". The "meta-meta" engine supports a set of integrity rules for metadata such as checks of mandatory values, key checks and checks of attribute types. In the MDR such rules are used for all configured data models, such as extract or SDTM.

In the scope of this project, numerous additional integrity checks were configured for the operational data model such as cross-table checks (e.g. referenced forms, linked dictionaries, controls and dependencies) or RAVE-specific content checks (e.g. FieldOID vs.

PhUSE 2013

VariableOID). The ALS validation module of RAVE was basically re-implemented in the MDR to allow fast validation without the need to export metadata from the MDR and to import it into RAVE in order to confirm metadata validity. Running the ALS checks is a mandatory part of the metadata QC process.

In order to ensure metadata consistency in the MDR, impact analysis was identified in the project as another key step in the QC process. The available comparison tool helps answer such questions as: Are all critical attributes from Global the same as at lower levels in the MDR? Is the delivered study build from the MDR equal to the EDC exports from UAT/PROD in RAVE? The comparison is also a mandatory part of the QC process. The flexibility of the comparison tool which supports the usage of different comparison keys and compares variables depending on the comparison level was very beneficial to the process. As a project outcome, several sets of rules were identified and configured as parameters per area (global, project, study levels) and comparison type (e.g. inbound vs. outbound for studies, project to global standards).

How to present metadata

A few words need to be said about metadata presentation. Though the metadata delivery process is very straightforward and allows the export of different vertical levels of granularity (single data elements, domains, complete data models), the experience of the project team showed that several perspectives on the operational metadata exist: metadata-centric and CRF-centric. The former relates to the question: What is needed to describe a data model in a consistent way? The latter focuses on the question: How will a form look once set up in the EDC? The EDC-centric view requires only a subset of metadata, but includes additional requirements of metadata layout which follow EDC-specific rules.

As the system supports both views on metadata, the CRF preview is created from ALS metadata in this project and contains a number of CRF validation rules developed in the course of the project and included in the QC process for operational metadata.

Insights and best practices

In the course of the described project, numerous best practices were identified by the project team. Only a few of those selected will be presented in this paper.

Best practice: Plan sufficient maturation time for the business process and configuration based on the agile approach.

To address the anticipated high complexity of the MDR world and of the processes involved, "agile" methodology was used in this project to configure the MDR. The environment was configured in a step-wise process, which was followed by intensive review cycles. Though it took several iterations of configuration review and business process review (as it became clear several times in the course of the project that the initially designed process needed modifications), it was possible to finally create the configuration following the business process – excellent news for adherents of the metadata-driven paradigm! The regular configuration review meetings (conducted by the core team daily in the "high season") were very helpful, with each meeting focusing on a selected, small part of the whole process. Alternatively, such reviews can be carried out as face-to-face workshops. It is important to reflect review time or review workshops in project planning!

Best practice: Have a motivated, stable team which represents all processes needed

This project confirmed the common wisdom from extensive literature on organizational behavior – the most valuable project asset and the key success factor is a motivated project team. The core team was stable from the very beginning (a few people joined the core team in the course of the project) and was composed of individuals (SMEs) representing all

PhUSE 2013

involved business processes and who were thus strongly motivated to make the project a success!

Conclusions

Today, now that the main MDR configuration work is completed, all environments are installed and the UAT is accomplished - the production environment is ready to go. The goals of the first phase have been fulfilled. Looking back, the clear goals with a manageable scope belonged among other things to the success factors which allowed the first project phase to be successfully finished within a quite restricted time period. A new project phase begins, and a productive pilot will start at the end of the month with selected studies. But that is another story...

Contact Information

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

Dimitri Kutsenko
Entimo AG
Stralauer Platz 33-34
Berlin / 10243
Germany
Work Phone: +49 30 520 024 100
Fax: +49 30 520 024 101
Web: www.entimo.com

Brand and product names are trademarks of their respective companies.