

Effective use of a Metadata Repository across data operations: the need for a machine readable form of the protocol

Isabelle de Zegher, PAREXEL Informatics, Wavre, Belgium
Michele Gray, PAREXEL International, Philadelphia, United States
Mark Sullivan, PAREXEL Informatics, Boston, United States
Michael Goedde, PAREXEL International, Boston, United States

ABSTRACT

Many companies have deployed a metadata repository (MDR) to manage their data standards. While an MDR brings value through increased quality and consistency of standards, its effective use for study management remains a challenge: there is limited (or no) integration with downstream systems (EDC, data collection instruments, SDTM mapping tool, etc.) that need to consume these standards for a specific study.

This paper introduces how we managed integration of the PAREXEL[®] Clinical MDR in data operations. We will first explain the options envisaged for integration and then expand into the approach we implemented through the Study Instance Metadata (SIM). The SIM is a machine readable form of part of the protocol created within the MDR, supporting smooth integration of standards in data operations through the following features: availability in multiple formats, linkage of standards, versioning to accommodate protocol amendments and storage in the SIM library for re-use.

INTRODUCTION: OPPORTUNITIES & CHALLENGES FOR AN MDR

Many companies have deployed – or are in the process of deploying – a metadata repository (MDR) to manage their data standards, including data collection (CDASH), data submission (SDTM) and derived data (ADaM) standards.

In most organisations data standards are maintained in silos: data collection /CDASH standards are maintained within Data Management, SDTM is governed by the clinical programmers and ADaM is managed by the statistical programmers. These different groups collaborate to maintain proper mapping between the different standards, but this remains a challenging process across separate groups with disjointed governance processes. In addition, mapping between the standards remains an “art”, based on manual interpretation and experience from the programmers. In a separate paper [1], presented at the PhUSE 2016 conference, we explained how we implemented a different approach, based on End-to-End data standards, in order to increase efficiency and provide a complete data lineage as required by FDA.

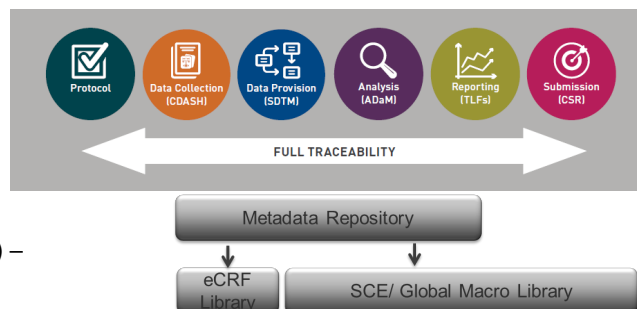
While the MDR brings value in increasing quality and consistency of data standards, it is truly effective when data standards are used in the context of a study. And this is the second challenge faced by organizations deploying an MDR: there is indeed limited (or no) integration with the downstream functions & systems that need to consume data standards for a specific study. This paper is focusing on this second challenge: how to leverage the MDR to support the transformation of (part of) the protocol into a machine readable format that can be utilized by all the downstream systems used in trial execution.

OVERVIEW OF CLINICAL MDR AS PART OF END-TO-END DATA STANDARDS

The PAREXEL[®] End-to-End Data Standards ecosystem is a comprehensive suite of metadata driven capabilities enabling governance and utilization of data standards from the interpretation of the approved protocol, through data collection and analysis, to reporting.

It includes 3 main components

- The **PAREXEL[®] Clinical Metadata Repository (MDR)** – based on Sycamore MDR, completed with PAREXEL data standards content and a set of SOPs supporting metadata driven management of standards, end-to-end



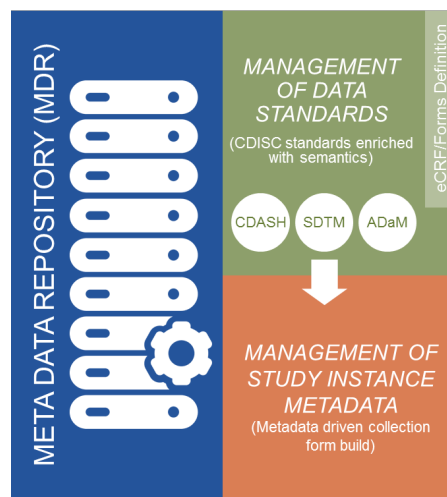
PhUSE 2016

from protocol to submission. The content of the MDR is tightly synchronized with multiple EDC specific eCRF libraries.

- The **PAREXEL® Statistical Computing Environment (SCE)** - based on Sycamore SCE and the PAREXEL® Global Macros Library, and developed in accordance with the PAREXEL® data standards, stored into the MDR
- The **PAREXEL® Data Standards governance framework** – including a dedicated unit of Data Standards Analysts as well as optimized clinical operations SOP/work instructions to ensure proper use of the MDR and SCE.

The Clinical MDR has 2 main sets of functionalities

- Management of data standards through semantic clinical concepts. This includes the classical import/export functions from other MDR/data standards libraries. In addition to the ISO11179 management of variables, the conceptual layer enables grouping of variables into “concepts” or Domain hubs, allowing them to be managed together in a semantically meaningful manner. A hub is equivalent to the specification of a CDISC domain and is standards agnostic. Each domain hub is mapped to the CDISC standards (CDASH, SDTM, ADaM) through predefined patterns. There are about 20 patterns that can be re-used to support mapping of all domains. A separate paper, presented at the PhUSE 2016 conference, explains in more details how we implemented this conceptual layer.
- Creation and management of a machine-readable form of part of the protocol, consisting of: study design, study descriptive metadata, time & events definition, forms to be collected for the study (including pre-defined mapping to SDTM and ADaM), and complete visit schedule. This is the Study Instance Metadata (SIM) further described in this paper.

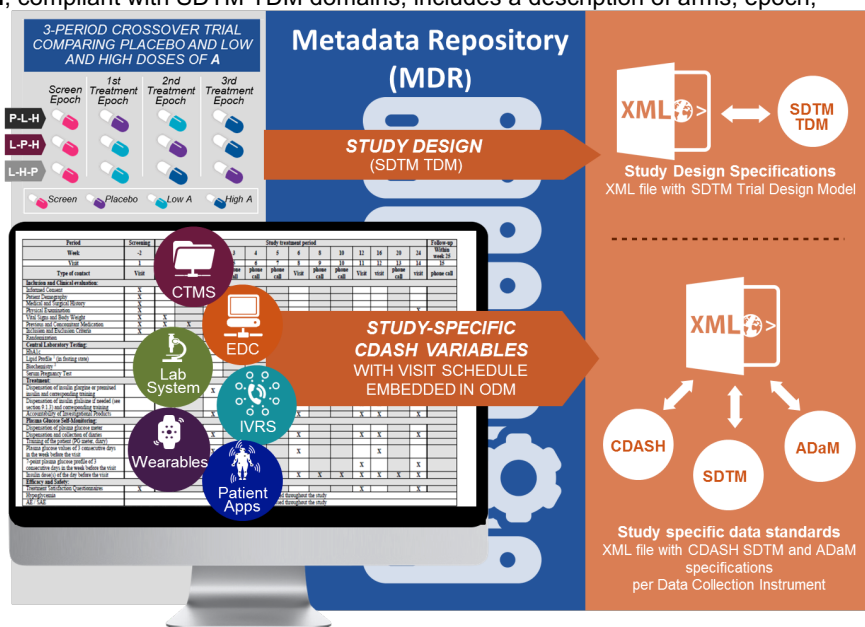


WHAT IS THE STUDY INSTANCE METADATA (SIM)

OVERVIEW

As mentioned above, the SIM is a machine readable form of part of the protocol with 2 main components.

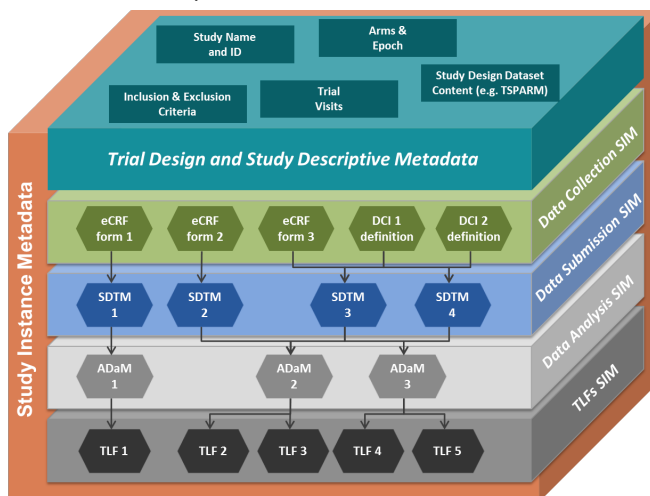
- The **study design specification**, compliant with SDTM TDM domains, includes a description of arms, epoch, time & events, inclusion/exclusions criteria and more than 50 parameters defined in the TSPARM domain. Most of this information – required by the FDA – is needed in systems used within set-up, execution and reporting of a clinical trial. There is therefore a major benefit to define it once through the MDR and re-use it consistently across all these systems.
- The **study specific data standards** including the domains extracted from the MDR conceptual layer, adapted for the specific study and linked to the time & events defined in the study design specification



KEY FEATURES OF THE SIM

The SIM managed within the Clinical MDR has some features of particular interest

- Maintenance of data standards is driven by the need of study – and therefore the SIM – through a strict governance process. For any new study, the Clinical Data Lead selects the necessary domains and forms. In case of modification (addition, deletion, changes) of the standards, the form needs to be approved by a Data Standards Analyst before it can be used in a study. Following the Data Standards governance model this happens within a very short time frame to ensure there is no impact on the overall time line for the study set-up.
- Each approved SIM is stored within a SIM library with attributes – defined in TSPARM – that allow easy retrieval. SIMs stored in the library can be used as a basis to design a new SIM for a similar study. Before starting a new study, Clinical Data Leads can search the attributes that characterize their study, and select a former SIM close to the one they are developing; this allows them to develop the SIM for a new study more efficiently and in a more consistent manner.
- The SIM is versioned. Each time there is a protocol amendment, the MDR generates a new version of the SIM and the “diff” between the 2 versions. This enables full traceability of the changes while helping the clinical programmer to identify precisely the changes that need to be implemented.
- The SIM can be extracted from the MDR in various formats and extensions to ensure consumption by destination systems. In addition, there is an attribute overlay mechanism which allows enriching the SIM with additional attributes needed for a specific data collection instrument. This increases the possibility of automatic set-ups of EDC and other data collection instruments.
- Because CDASH is linked to SDTM and to ADaM in the Data Standards Management layer, when the Clinical Data Lead selects a form for a study, the Clinical MDR automatically generates not only the CDASH data collection specification, but also the SDTM and ADaM specification for the study. The CDASH specification can be used by the data collection instruments, while the SDTM and ADaM specification can be directly used by the Global Macros Library managed within the SCE for analysis & reporting



USING THE SIM IN THE CLINICAL DEVELOPMENT PROCESS

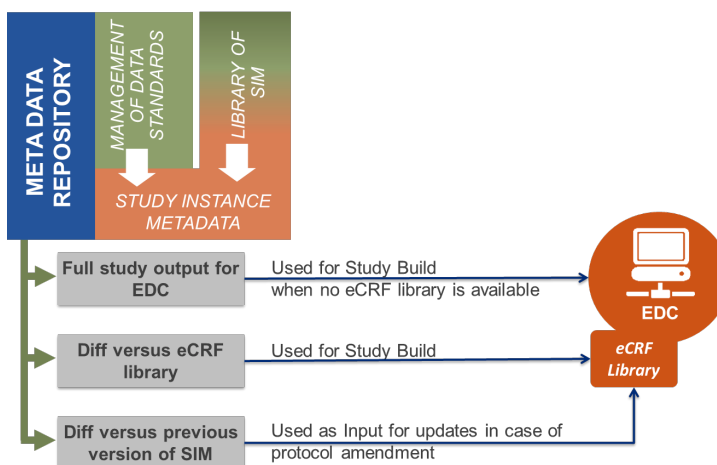
MAIN SCENARIOS AT PAREXEL

The most important use case of the Clinical MDR and SIM relates to faster and more consistent study set-up across multiple data collection instruments. For the initial set-up, a SIM is generated in XML format. At minimum, the SIM can be considered as formal specifications of the study design and data collection standards. At best the SIM can be used for ½ automatic set-up of a data collection instrument.

Below, we explore the case of EDC set-up.

The same scenario applies for any form based data collection instrument – including wearables and mobile apps.

- First, the attributes overlay mechanism included in the MDR enables enrichment of the SIM for a specific EDC system.
- Second, we have identified several possibilities to support EDC set-up with the SIM
 - If there is no available eCRF library, the full SIM can be downloaded within the EDC to speed up the work of the clinical programmer
 - If there is an available validated eCRF library – synchronized with the content of the MDR – the SIM will only contain the difference between the

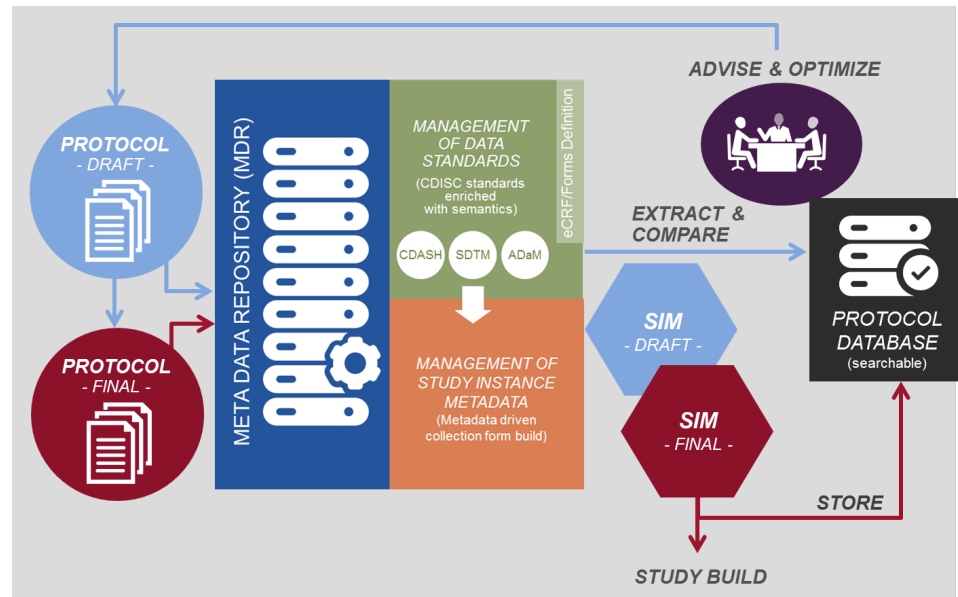


PhUSE 2016

approved eCRF library and what is needed for the study. The Clinical Programmer has therefore clear specifications on how to use the available eCRF library

- In the case of a protocol amendment, the SIM will only contain the difference against the previous version of the SIM. Again, the Clinical Programmer has therefore clear specification on how to program the protocol amendments.

The other use case we are envisioning with our scientific staff is Protocol feasibility. Whenever there is a draft protocol available, it is transformed through the Clinical MDR into a SIM. This SIM can be then compared with a database of protocols (fundamentally a SIM library) by a team of scientific staff responsible for protocol optimization. Additional functionalities, such as feasibility based on inclusion/exclusion criteria, could also be implemented. As a result, the scientific team can confirm the protocol design which is then updated in the Clinical MDR and a new version of the SIM is generated for study set-up.



DEPLOYMENT STATUS

We started the development of the Clinical MDR in collaboration with Sycamore in February 2015. Following the AGILE approach, we received 2 usable increments that we were able to test before we received the first full release of the product in August 2016. Throughout development, the usable increments were updated with corrections and new features which allowed a lot of interactions between the development team and the end-users.

Working through the AGILE approach with interim deliveries proved extremely useful to support process updates and change management. Our users could check how the system would behave and write their SOP/Work Instructions as the software was developed, as well as identify quickly if features from the systems that did not correspond to what was required and request needed updates into the software. In addition, through performing regular testing of the tools we built a team of power users – including the team of Data Standards Analyst – that can now support wider deployment.

We started the pilot at the time of writing this document. We are now introducing PAREXEL data standards content into the data standard layer, entering domain hubs and patterns to build the standards. In terms of studies we will follow a stepped approach: we will first do a prototype set-up with a mock study, then in a second step we work on a real life study in parallel with the current process and tools, and finally in a third phase we will work only with the MDR. It is only then that we expect to achieve the benefits identified in the original business case.

As part of the deployment we will include finalization of the process updates and related work instruction, followed by training of the staff on the new process and the CLINICAL MDR.

BENEFITS OF USING A SIM

While we have only limited experience in the usage of the SIM, we foresee that standards driven data operations activities – through the generation and update of the SIM – will bring the following benefits

- Comparison of the new study design with prior studies, enabling study design optimization and a decrease in costly protocol amendments
- Faster design of a study by utilizing previously validated study designs stored within the SIM library
- Faster study set-up, be it for EDC or other data collection instruments (such as imaging, wearables, mobile apps) by utilizing the machine readable XML specification of the study, either completely or as a “diff” versus the forms already defined and validated in the EDC eCRF library. The downside of this is that the eCRF library must be synchronized with the MDR content.
- Availability of the SDTM and ADaM specification related to the CDASH/data collection specification, ensuring full data lineage as required by the FDA while allowing smoother integration for analysis & reporting.

PhUSE 2016

- Possibility to store the SIM with the underlying study data in a data lake, aimed at secondary data analysis. The availability of the SIM – which can be considered as a rich metadata layer – will support just in time transformation of the data lake into a specific data mart.
- Consistent standards across studies enabling secondary data analysis through advanced analytics, meta-analysis and data mining.

CONCLUSION

The Clinical MDR was developed with a wide team of experts from PAREXEL, ranging from QA, Data Standards, clinical programmers, statistical programmer and IT staff. One of our key concerns was to ensure proper integration of the MDR within operations; indeed we realized that many MDR tools currently in place are used only by an handful of data standards experts and are therefore of limited value in operations. We believe that the SIM – as a machine readable form or (part of the) protocol- generated and managed within the Clinical MDR is a core component of true integration of data standards within data operations and is paving the way for a new generation of MDR solutions.

REFERENCES

- [1]. Pattern based MDR: toward high quality data standards. PhUSE 2016.
Alan Cantrell, Julius Kusserow, Julie James, Michael Goedde, Isabelle de Zegher.

CONTACT INFORMATION

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

Isabelle de Zegher
PAREXEL International
Jagerslaan 23
3080 Tervuren - Belgium
+32 478 48 28 54
Email: isabelle.dezegher@parexel.com
Web: www.parexel.com

Brand and product names are trademarks of their respective companies.