

PhUSE 2014

Paper PP05

From ODM to SDTM: An End-to-End Approach Applied to Phase I Clinical Trials

Alexandre Mathis, Department of Clinical Pharmacology, Actelion Pharmaceuticals Ltd., Allschwil, Switzerland

Denis Boutin, Department of Clinical Pharmacology, Actelion Pharmaceuticals Ltd., Allschwil, Switzerland

Jochen Lutz, XClinical GmbH, Munich, Germany

Philippe Verplancke, XClinical GmbH, Munich, Germany

Andreas Krause, Department of Clinical Pharmacology, Actelion Pharmaceuticals Ltd., Allschwil, Switzerland

ABSTRACT

Clinical Data Management (CDM) activities performed during phase I clinical trials are generally conducted over a shorter period of time compared to later clinical phases. Typically, participants are healthy subjects, study sizes are smaller, measurements are more numerous and more frequent, and additional measurements and protocol adaptations occur frequently due to the fact that the compound is not as well characterized as in later stages of development. This paper presents the data management model we developed to tackle such challenges.

Our model considers that the responsibilities of the data managers include the design of the electronic Case Report Form (eCRF) and the oversight of the clinical data collection at the clinical site including data cleaning and tabulation of the clinical data according to the Standard Data Tabulation Model (SDTM). Thus, the data manager is responsible for collection and delivery of the raw clinical data for the statistical analysis.

This model is supported by XClinical's integrated software solution covering all these aspects by the use of different modules in a single environment. The solution is based on the Operational Data Model^[1] (ODM). Both, the design of the eCRF and the collection of the clinical data, follow the ODM-XML structure. In a continuous stream of work, the solution bridges this ODM-XML structure to the Standard Data Tabulation Model^[2] (SDTM) by defining paths between ODM data points and SDTM variables. This step overcomes the main difficulty of CDM models, the difference in structure between the way data are captured and the way data are prepared for analysis.

INTRODUCTION

The biggest challenges faced by Clinical Data Management (CDM) during the conduct of phase I clinical trials are linked to timelines and flexibility in implementing clinical protocols. The primary goal in phase I is to assess the characteristics of the compound in man, including safety and bioavailability. The studies typically include healthy subjects for short treatment periods. Duration of enrolment period, study conduct and data cleaning is shorter than in phase II and III. However, data management requirements are identical and activities must be efficient to keep these shorter timelines. We developed our CDM model to be effective, flexible, and fast enough to handle these challenging timelines.

The most common CDM model limits the data management responsibility to CRF design, data collection, database building, and data cleaning. Statistical programmers then have the responsibility of reformatting the clinical data before submission to the statistician or modeling scientist for analysis. This separation between data management and analysis data set generation was justified by the fact that statisticians and modeling scientists

have particular requirements for data formats that need to be known and understood. This task is with the statistical programmers. Now that the SDTM standard is widely used in the industry, such requirements were standardized and are broadly shared. The disadvantage of this setup is that statistical programmers are usually hardly involved in the CRF setup and the clinical database build. Therefore, this usually leads to extensive interactions between statistical programmers and data managers to understand the structure of the clinical data and to plan their mapping. Also, the statistical programmers can face difficulties in mapping improper or incomplete datasets compared to SDTM requirements, leading usually to the use of many supplemental variables, extended code lists or partial datasets. This can be improved by the use of a data acquisition standard such as the Clinical Data Acquisition Standards Harmonization (CDASH) that has been published by the Clinical Data Interchange Standards Consortium (CDISC). However, such a standard is less common than SDTM that is more widely accepted and used. Moreover, it is an additional standard to implement and to maintain in addition to SDTM.

We have therefore developed our CDM model around a data manager role that encompasses all activities from the CRF development to the mapping of the clinical data to SDTM (Figure 1). In fact, the data manager has the oversight on the raw clinical data from collection to submission for statistical analysis. Thus, the data manager can design the CRF having in mind the final SDTM output he/she will have to produce. The input to the clinical protocol takes into account the way data will be submitted to the statistician and the possibilities and restrictions of the SDTM. This gives a more consistent and robust CDM approach compared to collecting the clinical data in CDISC standard.

The downside of such a CDM model is that the data manager has to use several software systems, as each individual activity usually requires switching between different tools to produce different outputs in different formats. The most typical example is the programming stage usually performed using SAS®^[3] to map the clinical data as collected in the clinical database in SDTM format. Both environments usually do not share common data format or structure, requiring to set up data standards on one side and script libraries on the other side to facilitate the process and make it effective and consistent across studies.

We have overcome such complications by selecting the XClinical's integrated software solution^[4] which support our CDM process in every aspects. It provides different modules that support all aspects of the data manager role in a single environment. The environment is based on the ODM standard. ODM was developed by the CDISC to standardize the process of electronic acquisition and exchange of clinical trial data. It is based on the Extensible Markup Language (XML) and can represent both, clinical data and metadata. An eCRF designer module, the Study Composer®, supports the configuration of the ODM-XML metadata, which is translated by the MARVIN® module into eCRF data entry screens for the acquisition of the clinical data through a web-based remotely hosted Electronic Data Capture (EDC) interface. The tabulation of the clinical data to the SDTM format is supported by the Tabulator® module that uses the same ODM-XML as input, automating this step of translation from the clinical database to the SDTM.

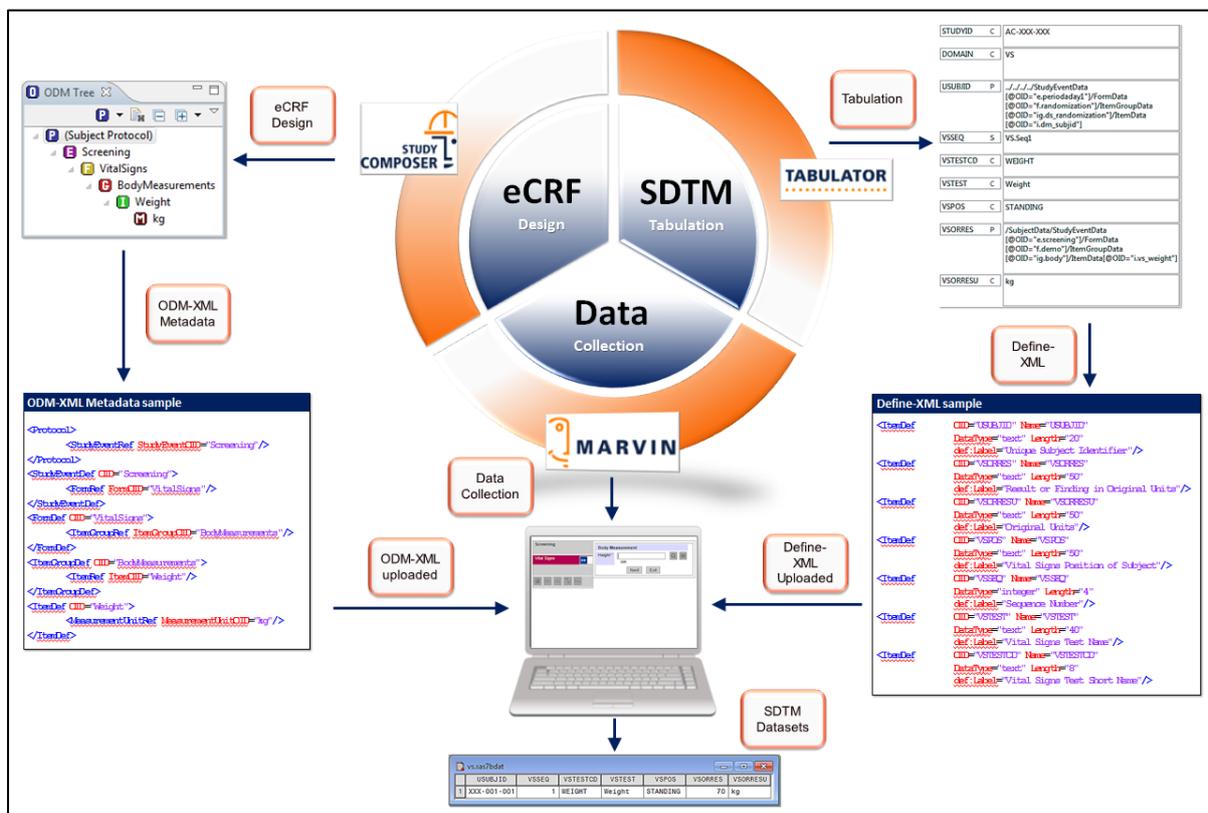


Figure 1. Data manager support by the XClinical's solution: From eCRF to SDTM data sets.

FROM ODM TO SDTM

ODM-XML METADATA CONFIGURATION

As in any CDM model, the first step of the process is to design the CRF for data entry. This is performed using the STUDY COMPOSER® module, and since the solution is based on ODM-XML, the definition of the eCRF follows the structure of ODM (Figure 2). The data manager defines the data points in terms of events, forms, item groups, and items as per clinical protocol requirements. A user interface defines both, the data that will be collected and the visits that will occur. In the background of the application, the ODM-XML code is generated automatically.

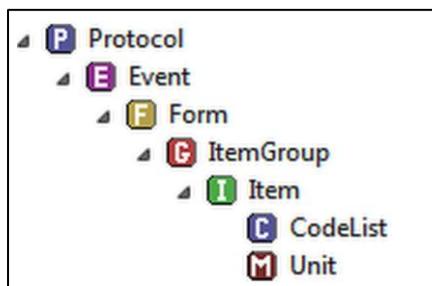


Figure 2. Operational Data Model structure, STUDY COMPOSER® layout.

To prevent any mapping issue due to misconception of the eCRF, the data manager defines the eCRF so that all data required in SDTM are collected and that all data collected can be mapped to SDTM.

For standard data types such as safety data (adverse events, vital signs, etc.) the SDTM format is now fairly stable and the CRF structure does not hold many pitfalls. Most of the efforts are focused on specific data types not explicitly defined in SDTM implementation guides. In such a case, the SDTM structure is thought through and drafted before CRF design so that the CRF is specified based on SDTM requirements, e.g., in the case of clinical assessments such as pulmonary function testing, it has been decided that the final tabulation would be done in a supplemental domain, however, following the structure of the vital signs domain and therefore the eCRF structure was based on the vital signs eCRF entry forms. This reverse process of defining how to collect the data based on the specifications of the final output avoids friction at the later stage of SDTM mapping.

An additional advantage of the STUDY COMPOSER® is that it allows isolating any individual element and its subsequent elements from the hierarchical structure of the ODM-XML metadata of an eCRF. Therefore, it can be stored as independent element of a library to be reused from an eCRF to the other. This step speeds up the design of the eCRFs and maintenance of their consistency across studies.

The module also supports the core task of data management, data cleaning. The data manager defines different "actions" triggered by conditions on the data points to define data discrepancies and entry form dependencies and dynamics such as hidden fields available only if further information are needed (e.g., the end date of a resolved adverse event is only shown if the start date of the adverse event is entered). Moreover, the STUDY COMPOSER® can extract from the ODM-XML metadata the relevant information to automatically generate documents such as the blank CRF printout that represents a paper version of the eCRF or the Data Validation Plan. The paper version is in turn the list of automatic data validation checks embedded into the eCRF to ensure the clinical data consistency.

CLINICAL DATA COLLECTION

The ODM-XML file generated automatically by the STUDY COMPOSER® is directly uploaded into the web-based EDC interface MARVIN®^[4]. MARVIN® translates the ODM-XML automatically into entry screens without the need of additional setup. Via the web interface of the EDC system, the clinical site personnel enters into the entry screens the clinical data that are stored into the clinical database. MARVIN® includes all functions usually present in data management systems such as query management, data upload, and data coding to obtain the complete set of clinical data.

SDTM MAPPING

The mapping of the clinical data in ODM-XML structure to the SDTM format holds three major challenges. The first is inherent to any conversion of clinical data to the SDTM format and resides in the fact that the CRF is not meant to match SDTM requirements but. It is designed to match the clinical protocol requirements so that all clinical data needed for evaluation of the clinical trials objectives are collected. This can result in clinical data that do not find a match in the SDTM implementation guides and need extensive programming work to bend these data to an SDTM-like format acceptable to health authorities. This difficulty cannot yet be overcome by technical means, so in our model, the data manager has the responsibility when reviewing the clinical protocol and, later

on, when designing the CRF, to anticipate these difficulties by ensuring that all collected data can be mapped to existing SDTM variables or that supplemental domains and variables can be created for that purpose.

The second major challenge is due to the cut in the processing of the clinical data done before the mapping to SDTM. Usually, mapping of the clinical data to the SDTM is the responsibility of statistical programmers working in close collaboration with the statisticians and modeling scientists that will analyze the clinical data. This cut is usually justified by the need of interaction between the statistical programmers and the analysts to produce suitable data format for the analysis. In this setup, statistical programmers are usually barely involved in the development of the CRF, therefore, before starting the mapping to SDTM, the programmers have to interact with the data manager to understand the design of CRF and the structure of the clinical database, but also to clarify the data inconsistencies that may remain after data cleaning due to study specifics. Since the release of SDTM and after several versions of the standard, this cut seems less and less needed as the output format of the analysis data sets is more and more robust and consistent over the time. Therefore we decided to give the responsibility of the mapping to the data manager who designed the CRF and therefore knows it best. Moreover, because the data cleaning is part of the CDM responsibilities, the data manager is aware of study specific cases that can impact SDTM mapping.

Finally, the third major challenge is linked to the use of ODM-XML as its structure differs at time substantially from SDTM formats. ODM defines data points using a hierarchical structure matching the CRF structure, grouping data points by subjects, events, forms, item groups and items (Figure 2) whereas SDTM describes clinical data in standard domains, variables and records that are independent of the CRF structure. These differences between the way data are captured and the way data are delivered for analysis require a substantial programming effort to generate SDTM datasets and therefore can cause major delays. If possible, the most frequent solution brought to this problem is usually to start the mapping of the SDTM in parallel to the data acquisition. The programming is done on partial data and, as the study progresses, the SDTM output is checked several times to include more data and verify that all actual clinical cases are mapped in a proper way, e.g., to verify that subjects' study completion status is captured and mapped in a proper way. The mapping step usually requires the use of a third party software such as SAS®. This adds to the difficulty of aligning different formats, the difficulty of switching between different platforms. The most important feature of the XClinical's solution is that it includes in its environment a mapping module called the TABULATOR®. This module simplifies this step by defining the tabulations as paths between ODM-XML data points and SDTM variables, defining the SDTM variables as a calculation between several ODM-XML data points. The data manager defines the tabulations using the ODM-XML defined in the STUDY COMPOSER®. The Define-XML^[5] containing the specifications of the SDTM domain is generated automatically in the background. In addition to the standard SDTM specifications, the Define-XML documents the mapping details by listing all paths between the SDTM variables and the ODM-XML data points. This ensures a complete traceability of the SDTM mapping as reviewers can check back the mapping process from the SDTM variable to the ODM-XML data points. The Define-XML is then uploaded into MARVIN® to automatically fill SDTM data set tables based on the collected ODM-XML clinical data points.

As for the CRF design, the tabulation can be divided into separate tabulation files for each individual domain, facilitating the building of a mapping library in which every element can be reused from one study to another. Such a library improves the consistency of the mapping and saves programming time.

CONCLUSION

We have designed our CDM model around a data manager role that is based on the three main activities related to the handling of raw clinical data: CRF design, data collection and cleaning, and data programming for statistical analysis. The data manager has the complete oversight of the raw clinical data from collection to delivery of the analysis data set. This offers a unique opportunity to align consistently the data management processes in an end-to-end approach, saving time in the transition between activities. The attribution of the data programming to the data manager is facilitated by the use of the SDTM format. This standard is now commonly used and understood by both, data managers and data analysts, and facilitates the communication between the two functions.

The XClinical's solution was found to be ideal since it supports the data managers in all their activities within a single environment. Based on the ODM-XML structure issued by CDISC, the data manager defines an eCRF taking into account the requirements of the SDTM data sets to be produced, anticipating any mapping difficulties one might face especially due to the handling of clinical data not covered by the regular SDTM implementation guidance. In parallel to the study conduct, using the same ODM-XML structure as defined for the eCRF, the data manager specifies the tabulation that will generate the SDTM data sets. Both, the management of the study data as well as the design of the ODM-XML, brings to the data manager a deep insight of the actual clinical data and of the ODM-XML structure to be mapped and facilitates the programming of SDTM data sets.

In addition, the XClinical's solution allows isolating and storing each element of the ODM-XML and of the SDTM tabulation as part of libraries such that modules can be reused from one clinical study to another, improving the consistency of the outputs and saving programming time.

This CDM model enables Clinical Pharmacology Data Management at Actelion to deliver an eCRF within four weeks (on average) after the clinical protocol is released and to deliver a first set of SDTM datasets two weeks after first data entry. Thus, data management activities have a limited impact on the study timelines particularly with respect to clinical data analysis.

REFERENCES

- [1] Operational Data Model, <http://www.cdisc.org/odm>
- [2] Standard Data Tabulation Model, <http://www.cdisc.org/sdtm>
- [3] SAS Institute Inc. (2006) Base SAS 9.1.3 Procedures Guide, vol 1-4. SAS Institute Inc., Cary, NC
- [4] CDISC End-to-End using ODM, LAB and SDTM, Kurt Hellstern, PhUSE SDE in Basel, March 18, 2009
- [5] Define-XML, <http://www.cdisc.org/define-xml>

CONTACT INFORMATION

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

Alexandre Mathis, MSc

Department of Clinical Pharmacology

Actelion Pharmaceuticals Ltd.

Gewerbestrasse 16

CH-4123 Allschwil

Switzerland

Work Phone: +41 61 565 57 33

Fax: +41 61 565 62 00

Email: alexandre.mathis@actelion.com