

Study Composer: a CRF design tool enabling the re-use of CDISC define.xml metadata

Dr. Philippe Verplancke, XClinical GmbH, Munich, Germany

ABSTRACT

define.xml is often created at the end of a clinical trial when delivering the SDTM datasets. However, with define.xml, data managers are just one click away from the automatic set-up of an EDC/CDM system at the beginning of a clinical trial. Define.xml and ODM XML are almost identical as far as the value-level metadata are concerned. They both contain the exact, unique and unambiguous definition of clinical items like systolic blood pressure, including a label, a data type, length, coded values/codelists, measurement units and question texts in different languages for the CRF. The use of such metadata for CRF design ensures all CRF data is collected such that it can be moved into SDTM datasets with no need for any difficult transformations or recoding. The tool Study Composer supports this independent of any EDC system. It can automatically generate an annotated CRF and can also be used to create a data validation plan directly pointing to the CDISC XML structure of the CRF.

INTRODUCTION

When submitting clinical trial data to the FDA, the metadata describing the content of the SAS datasets needs to be provided in CDISC define.xml format. This paper explains how such metadata can be readily re-used for other purposes and for CRF design in particular, thereby truly implementing a CDISC end-to-end approach. If all software systems supporting different activities along the clinical data management process would directly be based on one single set of metadata in XML format, the need for (legacy) data transformation would be eliminated. This would indeed create considerable time and cost savings and avoid quality and validation issues related to data transformations.

In contrast to what most data managers believe, it is not difficult to adopt CDISC XML standards in CRF design with existing EDC/CDM systems. In fact, all the necessary metadata is already available for immediate use: it is part of define.xml files that have been submitted for previous trials. Many EDC/CDM systems are capable and even certified to import such CDISC XML metadata. Those that are not can easily be enabled to do so via an import script.

THE VALUE OF RE-USING METADATA

Transforming data always causes time-delays, work effort (costs) and quality issues (transformation errors or even loss of information or precision). Therefore, clinical data managers will want to avoid any data transformation steps. Re-using metadata in different steps of the clinical data management process in fact means that data collected or processed during these steps do not need to be transformed between these steps. Re-using metadata in fact means that software systems (like CRF design tools, EDC/CDM systems, statistical analysis and reporting tools, etc.), which all need metadata to perform meaningful work with the data, share an identical set of metadata. Identical in the technical sense literally means they refer to exactly the same files or database tables containing the metadata.

DEFINE.XML AND ODM.XML (ALMOST) IDENTICAL

Within the context of this paper, we refer to the CDISC ODM and define.xml standards for storing metadata as XML files. In particular, we refer to value-level metadata which is in fact the collection of definitions for the clinical values (items) that are collected, stored, or submitted such as, e.g., the systolic blood pressure. As such, both within define.xml and odm.xml, there is an (almost) identical piece of XML that represents the definition of systolic blood pressure (an integer with length 3 with measurement unit mmHg).

Figure 1 illustrates the strong overlap between define.xml and odm.xml metadata. The latter is used to design CRF metadata, i.e. clinical data items, item groups, forms and visits. The former is used to describe the content of SDTM domains. The actual clinical items are just moved from the CRF pages into the SDTM tables. In SDTM, they are called topic variables and their original values in fact remain unchanged between data entry and submission.

Re-Use define.xml as Metadata for CRF design

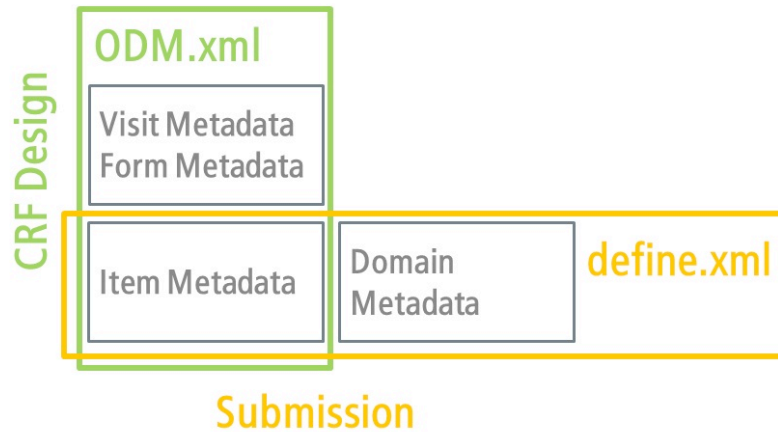


Figure 1: define.xml and odm.xml overlap

Figure 2 shows that value-level metadata (also called item metadata) are almost identical between ODM and define.xml. The differences are technically trivial and in fact could be easily harmonized in a future version of the standards.

Item Metadata in odm and define.xml (almost) identical



```
<ItemDef OID="VS.SYSBP"
  Name="Systolic Blood Pressure"   Name="SYSBP"
  def:Label="Systolic Blood Pressure" SDSVarName="SYSBP"
  DataType="integer"
  Length="3"
  SASFieldName="SYSBP">
  <Question>
    <TranslatedText xml:lang="en">Systolic Blood Pressure</TranslatedText>
    <TranslatedText xml:lang="de">Systolischer Blutdruck</TranslatedText>
  </Question>
  <MeasurementUnitRef MeasurementUnitOID="MU.mmHg"/>
</ItemDef>
```

define.xml
odm.xml

Figure 2 ODM and define.xml (almost) identical

STUDY COMPOSER

Study Composer is a CRF design tool that allows a non-technical user, e.g. a clinical study manager, to directly design the CRF and the data validation plan using a drag-and-drop and point-and-click graphical user interface. It is just as easy as any office application like Microsoft Word or Excel and in fact is meant to replace Word and Excel. Thus, the CRF specification process is clearly shortened. The Study Composer uses existing CDISC XML elements, such as those imported from existing define.xml files. Alternatively, the Study Composer can also be used to create new CDISC XML elements from scratch. The user actually does not need any XML knowledge since the user interface completely hides the underlying source code.

Study Composer is a desktop application that runs on any PC (Windows, Macintosh or Linux based). Since the CRF design is stored in an XML file, normal document management and versioning mechanisms can be used to control the CRF design process. Typically, a window on the left hand side shows the CRF structure (Visits, Forms, Item Groups, Items, Codelists and Measurement Units). A window on the right hand side shows the existing list of definitions available to be used in the CRF structure. Such elements are dragged from right to left with the mouse. This clearly supports the most important aspect of the CDISC ODM and define.xml standards, which is normalization: items should be defined only once and re-used in multiple locations of the CRF as needed.

Figure 3 shows an overview of the Study Composer application. In this example, the actual item definitions comply with SDTM naming conventions, data types and codelists. However, the tool allows to use any existing sponsor-defined items.

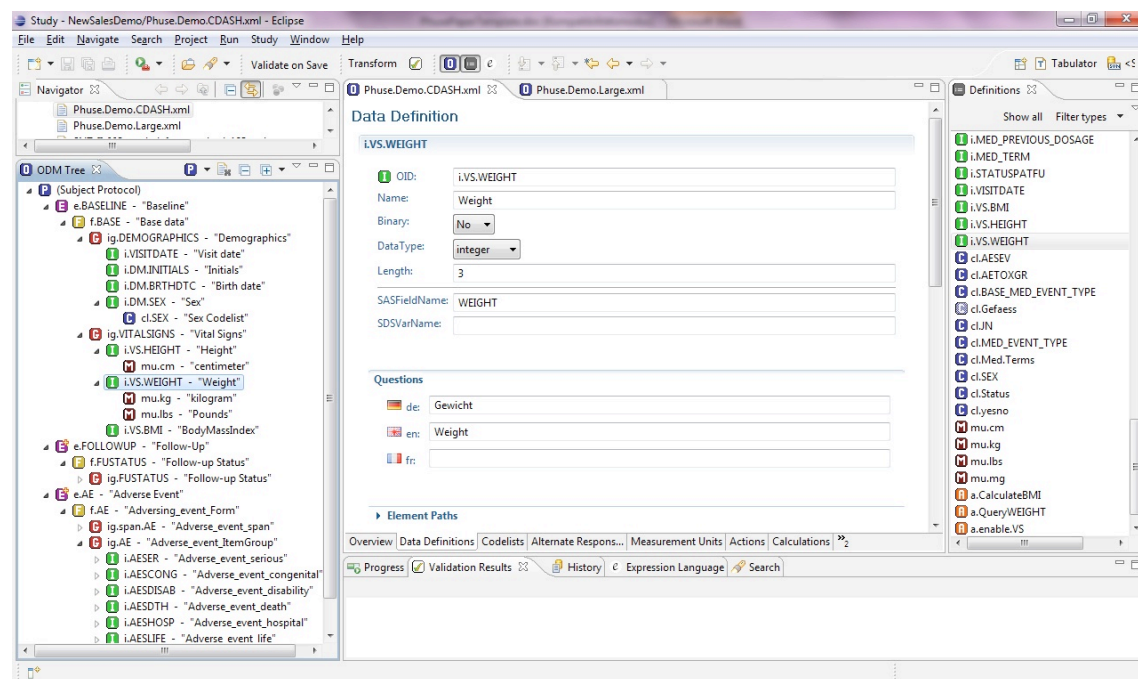


Figure 3 Overview of Study Composer

AUTOMATIC GENERATION OF ANNOTATED CRFS AND OTHER DOCUMENTATION

The real power of a tool like Study Composer based on CDISC XML standards is its capability to automatically generate all kinds of documentation from a single description of the CRF. Such documents include the annotated CRF (in all variations), the visit matrix (schedule of events), the list of items, list of codelists, etc. Figure 4 shows an extract of an annotated CRF automatically generated from Study Composer. Its design can be customized according to sponsor requirements. Most importantly, in case of last minute changes to the CRF, all documentation can be regenerated automatically at the push of a button. So clearly, Study Composer accelerates the CRF design process and guarantees consistency between all CRF-related documentation.

Base data^{*f.BASE}

Demographics ^{*ig.DEMOGRAPHICS; SAS:DM}	
Date ^{*i.VISITDATE; date10}	___/___/___ mm/dd/yyyy
Initials ^{*i.DM.INITIALS; text2}	___
Date of birth ^{*i.DM.BRTHDTC; date10}	___/___/___ mm/dd/yyyy
Sex ^{*i.DM.SEX; text2; display.style radio}	☐ Male ☐ Female ☐ Undifferentiated ☐ Unknown ^{cl.SEX; text; SAS:\$SEX}
Vital Signs ^{*ig.VITALSIGNS; SAS:VS}	
Height ^{VS.HEIGHT; integer3; SAS:HEIGHT}	___ cm ^{mu.cm}
Weight ^{VS.WEIGHT; integer3; SAS:WEIGHT}	___ []kg ^{mu.kg} []lbs ^{mu.lbs}
BMI ^{i.VS.BMI; float3,1; SAS:BMI}	___

Figure 4 Extract of an annotated CRF generated by Study Composer**REDUCING VALIDATION EFFORTS**

Study Composer can be used independent of any EDC system and the CRF design produced by the tool can be used by many downstream stakeholders and tools. E.g., the XML-based CRF design can be sent to different CROs to guarantee they will all send back data in exactly the same format.

If Study Composer is used in combination with an EDC system that is ready to import CDISC XML metadata, e.g. XClinical's MARVIN system, maximum time and cost savings can be achieved. In fact, if the CRF is directly designed in Study Composer (skipping Microsoft Word or Excel), the design can be printed, reviewed and signed by the sponsor to confirm that it actually matches the requirements of the trial. Thus, the CRF design within Study Composer serves as a formal CRF specification document. The EDC system can then import the design and automatically set-up the e-CRFs. Since the Study Composer as well as the EDC system are GCP system validated software products, the EDC system is guaranteed to execute the specification and to exactly match the CRF design as defined by the Study Composer. In other words, the implementation is identical to the specification. Therefore, the effort of user acceptance testing could be completely eliminated. Since most auditors will still require some form of user acceptance testing (e.g., entering test patient data), the effort will in fact not be zero but indeed be greatly reduced.

BUILDING A DATA VALIDATION PLAN WITHOUT EXCEL

Another strong advantage of Study Composer is its capability of attaching data validation rules directly to the annotated CRF. The tool is just as easy to use as Microsoft Excel to define the checks and to attach them to the CRF structure. In fact, the specification of the data validation plan in Excel can be completely eliminated and replaced by the specification of the data validation plan directly in Study Composer. The tool will attach the rules to the XML references of the relevant items in an unambiguous way. Creating such unambiguous references does not require any XML knowledge since the tool references the items which are dragged into the formula editor with the mouse.

Study Composer can automatically generate a list of checks that can be readily opened and reviewed in Excel. Thus, the Excel document is an output instead of an input of the process. Figure 5 shows an example of a few data validation rules created by Study Composer. Please note how easy it is to create cross-page, cross-visit edit checks, e.g. "a.query.FUDate", which compares the date of the follow-up visit with the date of the baseline visit.

	A	B	C	D	E	F
1	Data Validation Plan					
2	OID	Type	Context Affected	Condition	Message / Calculation	Variables
3	a.CalculateBMI	calculate	1111.VS.BMI		(\$L_WEIGHT / (\$L_HEIGHT * \$L_HEIGHT)) * 10000	\$L_HEIGHT = i.VS.HEIGHT \$L_WEIGHT = i.VS.WEIGHT
4	a.query.FUDate	query	e.FOLLOWUPi.FUSTATUSig.FUSTATUSi	\$L_VISITDATE_FU <= \$L_VISITDATE_BL		\$L_VISITDATE_FU = i.VISITDATE \$L_VISITDATE_BL = e.BASELINEi.BASELINE.DEMOGRAPHICSi.VISITDATE
5	a.query.Weight	query	e.BASELINEi.BASELINE.VITALSIGNSi	\$L_VS_WEIGHT < 40 \$L_VS_WEIGHT > 150	Please confirm this abnormal weight value.	\$L_VS_WEIGHT = i.VS.WEIGHT

Figure 5 Data Validation Rules created by Study Composer

PhUSE 2011

The data validation rule format is not yet standardized by CDISC. Study Composer stores these rules as extensions to the CDISC standard. These cannot immediately be interpreted by other EDC systems. However, the exact and unambiguous references to the CRF structure make the data validation plan in Study Composer superior to a specification in an Excel sheet. Therefore, such a plan is valuable independent of any EDC system.

Maximum time and cost savings are achieved when importing the data validation plan from Study Composer into the MARVIN EDC system, because MARVIN can actually interpret and immediately execute the rules thereby eliminating the need for any programming. Since MARVIN and Study Composer are validated, the need for user acceptance testing of the data validation plan is strongly reduced.

CONCLUSION

Software tools using CDISC XML based metadata, whether define.xml or odm.xml, create considerable time and cost savings in the clinical data management process by re-using exactly the same metadata for different purposes. Re-using value-level metadata contained in existing define.xml files for CRF design automatically ensures the CRF data will match the CDISC SDTM and CDASH requirements.

REFERENCES

The latest versions of CDISC standard specifications can be downloaded from www.cdisc.org.

CONTACT INFORMATION

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

Philippe Verplancke
XClinical GmbH
Arnulfstrasse 19
80803 Munich
Germany
Work Phone: +49 89 4522775100
Email: philippe.verplancke@xclinical.com
Web: www.xclinical.com

Brand and product names are trademarks of their respective companies.