

## Fast track to define.xml v2.0 – SDTM and ADaM

Jörn Bilow, Entimo AG, Berlin, Germany

### Abstract

With the FDA's announcement of ending support for Define-XML 1.0, version 2.0 is the focus for clinical research and development. Facilitating the generation of the define.xml for electronic submissions is an important and often time consuming part of standards and data management in the pharmaceutical industry. By utilizing an intelligent linking mechanism, the demonstrated define.xml tool enables not only the generation of the define.xml early in the development process as a means of communication, but easy and constant updates as information is collected with the progress of a study in various areas of the data and metadata repository – transport files, variables, programs, value level metadata, algorithms and codelists can be included easily along with the links to relevant documents. Automated system notifications provide information about changes to relevant areas, and the updated define.xml can be created with a few mouse clicks. The new version of the tool also provides support for Analysis Results Metadata tables.

### 1 Introduction

The Define-XML standard is currently supported in versions 1.0 and 2.0. The FDA has announced the end of support for Define-XML v1.0 for studies that start 12 months after March 15, 2017. Development of Entimo's Integrated Clinical Environment entimICE has therefore focused on the generation of define.xml v2.0 for the submission of clinical studies. All components for the define.xml can be stored in the entimICE repository. They are brought together by an intelligent linking mechanism. Unlike some other commercially available tools, the generation of define.xml is driven purely by metadata standards and not data or transport files. This allows utilization of the define.xml as a communication tool e.g. between sponsors and external partners even before data become available. As study data are generated and imported into the entimICE repository, the define.xml can be re-generated with more and more information content, thus becoming a "living" document. In its newest release 4.2 the entimICE define.xml generator supports generation of the define.xml v2.0 for SDTM and ADaM including Analysis Results Metadata tables.

### 2 Define-XML

Define-XML transmits metadata for SDTM, SEND and ADaM datasets; it is the metadata file sent with every study in each submission, which tells the FDA what datasets, variables, controlled terminologies, and other specified metadata were used. Here the focus will be on clinical submissions of SDTM and ADaM define.xml.

The update to Define-XML v2.0 was a result of experience with v1.0 and the continuing development of SDTM and ADaM standards. Key enhancements include:

- Support for Controlled Terminology and references to external terminologies/dictionaries
- Flexible definition of Value Level Metadata
- Enhanced documentation of data origin or source
- Improved support for ADaM metadata
- Improved handling of comments.

## PhUSE 2016

Define-XML v2.0 is the most current version of this standard and its use is recommended by CDISC.

### 2.1 Components

The define.xml provides a standardized representation for structure and content. It describes the study outcome and links all study outcome objects.

The central elements are structure and value level metadata, controlled terminology, algorithms and Analysis Metadata Tables. This is complemented by supplementing documents as aCRF, reviewers guide, comments and tables.

### 2.2 entimICE Folder Structure

The entimICE repository provides a highly configurable folder hierarchy to store all elements of the define.xml. The underlying technology enables storage of files, tables and documents under governance. Examples are

- **Metadata domains and variables with configurable attributes** as tables or as data elements on various levels like global, project, study
- **Controlled terminology** NCI, MedDRA, WHODrug, etc
- **Value level metadata**
- **Computational algorithms**
- **Datasets** as SAS datasets, CSV, XLS or database tables
- **Programs** from various statistical analysis tools as SAS®, R, S-PLUS, SQL
- **Documents** as PDF, MS-Office, images, tables, xml, html, etc
- **Output** of statistical programs
- **Log files** of program runs including complete traceability of execution

The system provides versioning of all objects, supports an audit trail and utilizes a role and group based access rights mechanism. The depth of the hierarchy as well as the folder structure, naming conventions and target folders for various content can be configured according to the individual need of the organization. The system supports configurable workflows and lifecycle management, e.g. for validation purposes. entimICE supports separated dev, val and prod environments and handles the transition of objects between these environments. Electronic signature, change management processes and configurable forms for data entry and editing are an integral part of the system.

### 2.3 Linking Mechanism

Because of the high configurability and flexibility of entimICE, the folder structure and the location of all elements for the define.xml is different for each installed system. It is for example possible to store SDTM or ADaM structure metadata in a dedicated MDR part of the repository, or maintain that information in a folder hierarchy based on Therapeutic Areas, projects and studies. Therefore a hard coded or generic mechanism to bring all elements together into the define file is not possible.

Instead, entimICE provides an internal linking mechanism, which allows flexible and easily definable links between all objects and/or folders in the repository. Using these links, the entimICE define.xml Generator can pull information and files from the individually different locations of the hierarchical structure and build the define.xml structure and content from this information. entimICE supports the provision of default links as a starting point for the define.xml generation, which can be edited as needed.

### 2.4 Structure Metadata as Starting Point

The entimICE define.xml Generator uses SDTM and AdAM structure metadata from the entimICE MDR or 3<sup>rd</sup> party systems to generate the define.xml. This allows creation of the define.xml even before data become available. Transport files can be created as "planned

objects”, which allows users to easily recognize the study progress as these objects become physically available. The generation of define.xml at an early stage in the study progress allows utilization of the metadata based structure as a communication tool, e.g. between a sponsor and a CRO. As data become available, the define.xml can easily be recreated at various stages of the process. This way it can become a “living” document to get an overview of the study progress in a universal and easily exchangeable format.

### **2.5 aCRF Scanner**

A dedicated tool scans the linked annotated CRF, looking for occurrences of metadata attributes from domain definitions. The tool extracts page numbers for occurrences and stores them in metadata. Alternatively, users can manage page number references to the aCRF manually. The page numbers are used by the define.xml Generator to create hyperlinks from define.xml to the corresponding pages of the linked annotated CRF. The scanner supports regular expressions for attribute search. The patterns can be stored as templates and made available to other entimICE users.

## **3 Generating the define.xml**

### **3.1 Context Menu Action**

The entimICE repository allows the configuration of “actions” that are context sensitive for different folders and elements in the repository. The context menu “Generate define.xml” is accessible from SDTM and ADaM structure metadata. Users can select one or multiple domains for the creation of the define.xml. CDISC comments for variables are stored as attributes of the structure metadata and are included in the define.xml.

### **3.2 Define.xml Generator**

The entimICE define.xml Generator provides up to five sections for input. The “General Settings” control information about study name and description, structure metadata names, versions and descriptions and information about the format of domains and controlled terminology. The next two sections control the use of SDTM and ADaM domains in the define.xml. They determine which domains are going to be included in the creation of the define.xml, and whether transport files with data should be generated in the process. The order of domains is also determined here. “Analysis Results Metadata” includes the ADaM tables into the generated output. The section “Output Settings” determines the location of style sheets, supplementary documentation like aCRF and Reviewers Guide, and the output folder for the define.xml. It is also possible to switch between Define-XML v1.0 and v2.0 at this stage.

### **3.3 Output**

The define.xml is created in the configured output folder along with the transport files and style sheets. It can be viewed with any common current web browser. Hyperlinks guide the user or reviewer through the content. The entimICE repository also provides export mechanisms of the output to external locations like local or network drives for further distribution. A notification mechanism which can be connected to an organization’s email system will notify subscribed users about the creation of the define.xml.

## 4 Conclusion

The entimICE define.xml Generator provides a fast and easy to use way to generate the define.xml for SDTM and ADaM, including the Analysis Results Metadata tables. It is metadata driven, which allows creation of a define file even before data become available. Intelligent links provide for maximum flexibility in the organization of hierarchy, folder structures and storage locations of the various components. As data become available over the course of the study, the output can easily be recreated ensuring the content is always up to date.

### Contact Information

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

Jörn Bilow  
Entimo AG | Stralauer Platz 33-34 | 10243 Berlin | Germany  
Email: bil [at] entimo.de  
Work Phone: +49 30 520 024 100  
Web: www.entimo.com

Brand and product names are trademarks of their respective companies.