

# Define.xml Generator: A Metadata Driven Integrated Application Approach

Balakrishna Dandamudi, PharmaNet/i3, Kennett Square, PA

## ABSTRACT

FDA electronic submission (eSub) guidance and the electronic Common Technical Document (eCTD) specify that a document describing the content and structure of the included data should be provided within a submission. This document is commonly referred to as "define.xml", formally known as Case Report Tabulation Data Definition (CRT DD) specification.

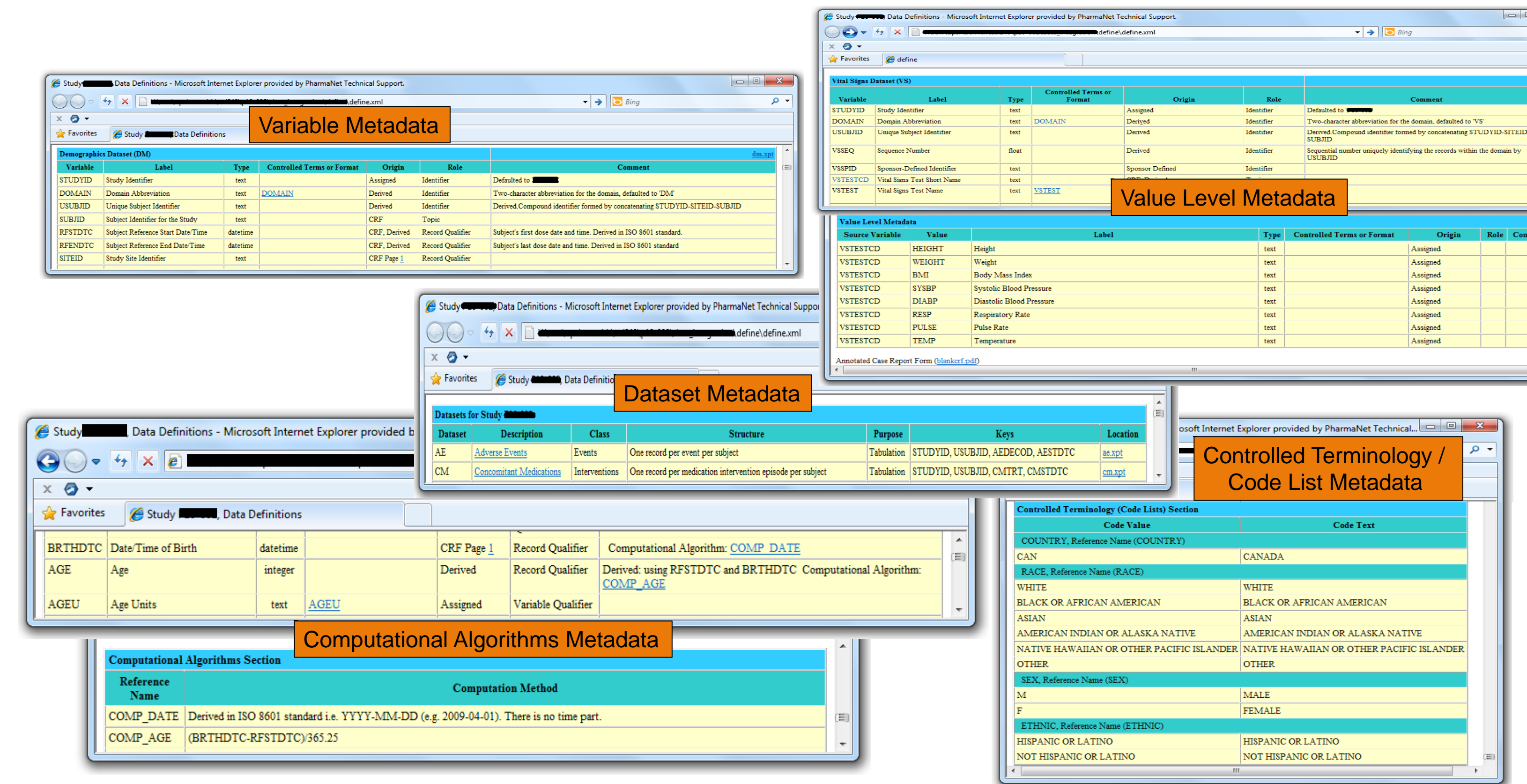
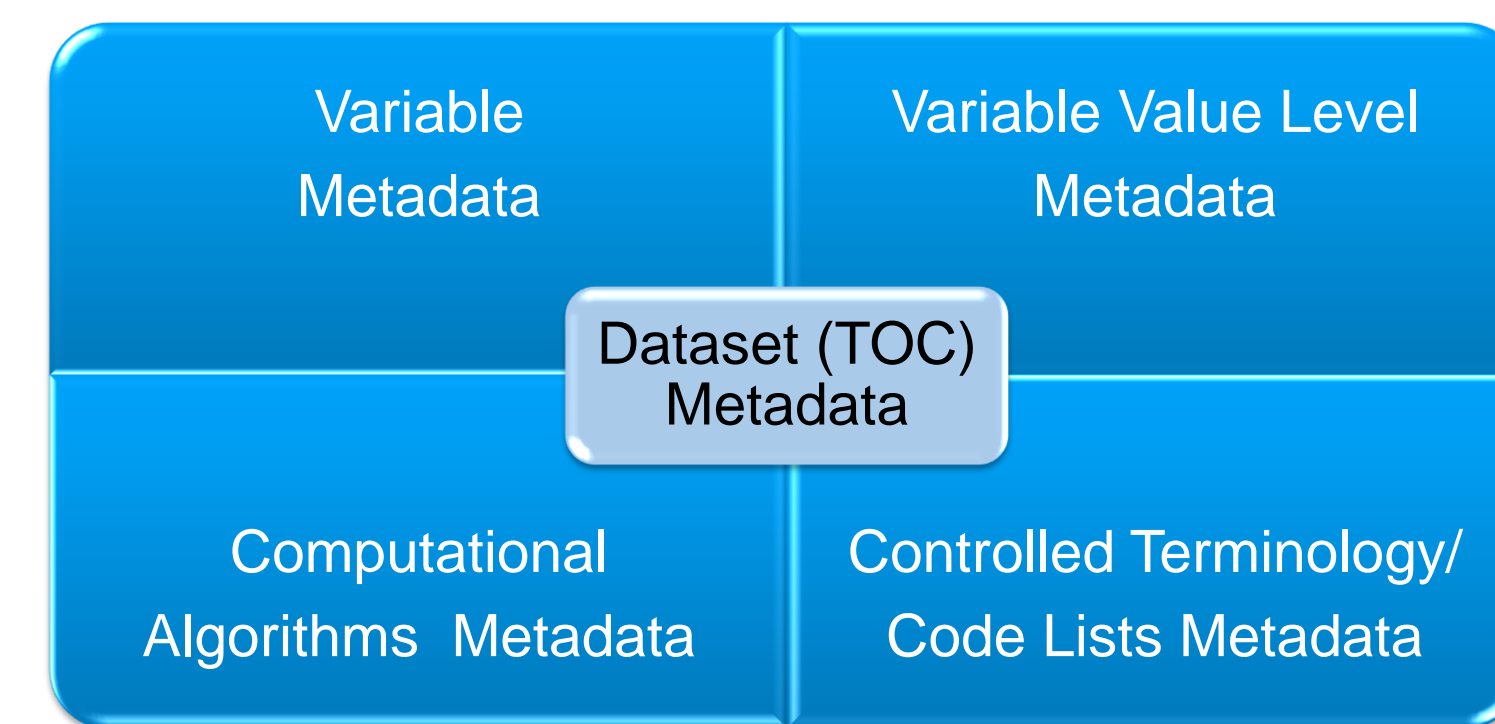
This paper discusses the current approach adopted by PharmaNet/i3 to create "define.xml". Possible ways to define.xml implementation are explored.

An integrated Windows application developed using Visual Basic for Applications to generate "define.xml" is presented. This application has a single GUI, where user can select domains to be included in define.xml. There is also an option to make changes to value level metadata, standard comments, and computations for standard variables provided in SDTM mapping excel spec document.

## DEFINE.XML STRUCTURE

The define.xml has two main parts, a Table of Contents (TOC) and a collection of Data Definition Tables. The TOC lists all of the datasets (also known as domains) included in the submission per study or collection of studies (e.g., in support of the Summary of Clinical Safety or Summary of Clinical Efficacy). Whereas the Data Definition Tables provide detailed information about the variables contained within each of the domains.

These two main parts can be further classified as five major sections :



define.xml, generated using Define.xml Generator

## DEFINE.XML GENERATOR

Although, metadata information is available within SAS datasets, this is not sufficient to automate the entire define.xml generation process. Thus, we adopted an integrated approach, which keeps all metadata in one document, mapping spec, which can be used throughout the life cycle of SDTM mapping process.

The application takes the data from individual tabs of excel and creates staging tables in metadata repository and this data gets processed to generate define.xml. The main advantage of the tool is that it encapsulates all the technicality of xml generation and it provides a simple click and run approach. User can generate the define.xml by simply making content updates within excel, mapping spec, document.

## CONCLUSION

There are different ways to generate define.xml. The most effective and efficient way is to have a tool which is user friendly, adaptive to the changes in CDISC standards as they evolve and less resource dependent. The tool we developed internally meets these requirements. The main advantage of having an in-house tool is that the new standards can be implemented as they made available.

## INTRODUCTION

The FDA reviewer requires a document which navigates through, how study data is presented in individual datasets. Define.xml is the one such document which describes the structure and content of clinical data in a machine readable format.

The XML schema used in define.xml is an extension to the CDISC Operational Data Model (ODM). ODM is designed to facilitate the archive and interchange of the metadata and data for clinical research.

This paper describes a methodology to generate define.xml from a mapping specification document, which is a central document, used in producing SDTM datasets.

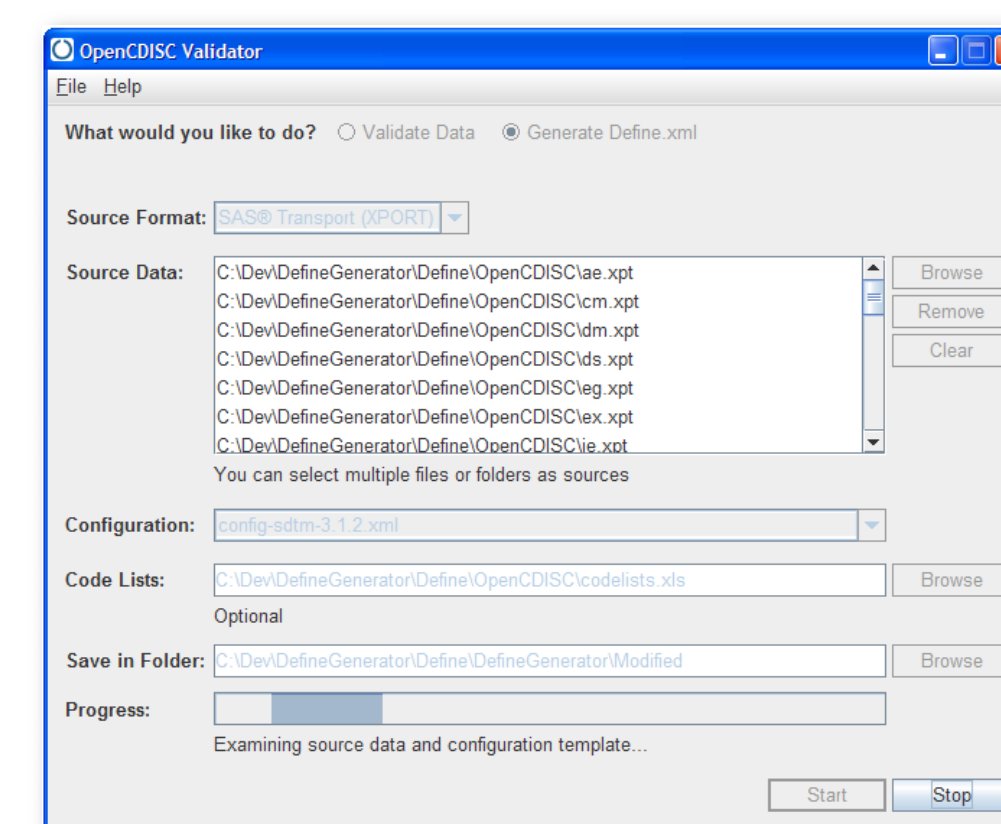
The combination of ODM and data definition schema structures used in the define.xml are intended to meet or exceed the minimum requirements for electronic transmission of metadata of a wide range of standards including, but not limited to the CDISC Study Data Tabulation Model (SDTM), CDISC Analysis Dataset Model (ADaM), and Standard for Exchange of Non-clinical Data (SEND).

We limit scope of this poster to CDISC Study Data Tabulation Model (SDTM).

## OPENCDISC VALIDATOR

There are certainly many ways to create define.xml. One approach is using OpenCDISC Validator to generate Define.XML. This is a very simple easy to use application. With a nice one screen GUI feature.

With couple of mouse clicks, one can easily generate define.xml. This application is a good start for basic define.xml generator. However, this version needs



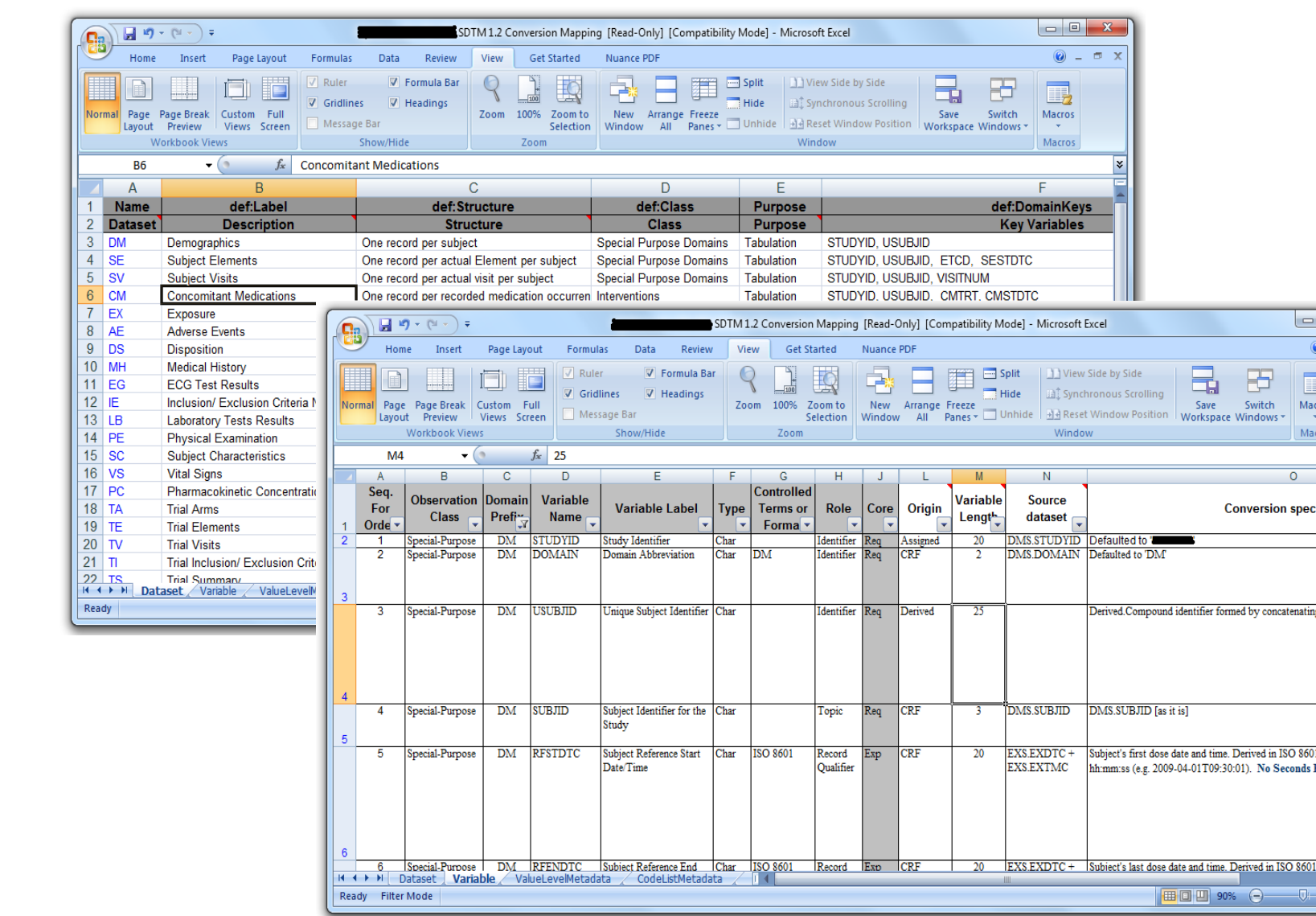
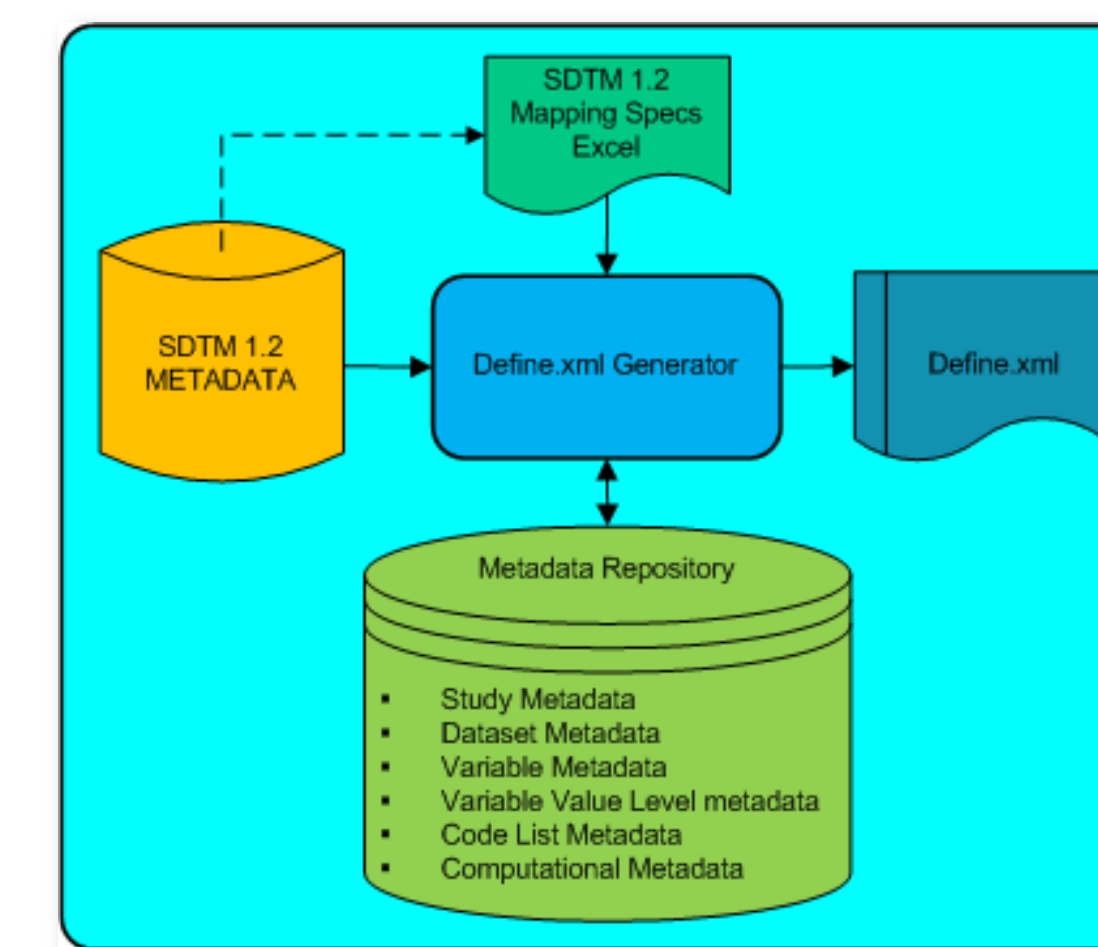
additional manual updates to define.xml. This could be error prone, time intensive and requires familiarity with xml and define.xml structure.

Some of the missing metadata that requires manual input includes:

- Study Metadata
- Keys in Dataset Level Metadata
- Origin and Comment in Variable Level Metadata
- Variable Value Level Metadata
- Computation Method Metadata

To fill this gap, we developed an in-house application to generate define.xml using VBA and MS Access Database.

## METADATA APPROACH



Define.xml Generator

## REFERENCES

1. CDISC Case Report Tabulation Data Definition Specification (define.xml), Version 1.0, February 9, 2005 (<http://www.cdisc.org/models/def/v1.0/index.html>)
2. <http://www.cdisc.org>
3. <http://www.opencdisc.org>

## CONTACT INFORMATION

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

Name: Krishna (Balakrishna) Dandamudi  
 Address: 415 McFarlan Rd # 201  
 City, State ZIP: Kennett Square, PA 19348-2454  
 Work Phone: 484-738-7247  
 E-mail: bdandamudi@pharmanet.com

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. © indicates USA registration.

Other brand and product names are trademarks of their respective companies.