Define.xml Content Validation – CRF Page Check

Shenglin Zhang, PAREXEL

PharmaSUG China 2016
Paper #77

CONTENTS

- Introduction
- Define.Xml Content Extraction
- Annotated CRF Comments Extraction
  - Parsing imported Comments With Perl Regular Expression
- Macro Invoke
- Validation Result
- References

INTRODUCTION

SDTM Submission Package

- Content
  - SAS xpt files(all expected datasets/variables)
  - define.xml (define.xsl and define.css)
  - blankcrf.pdf (SDTM annotated CRF)
  - Supporting documents, e.g., study-data-reviewers-guide.pdf

Importance of Define.xml

- A critical component of data submission is the define file. A properly functioning define.xml file is an important part of the submission of standardized electronic datasets and should not be considered optional
- An insufficiently documented define file is a common deficiency that reviewers have noted

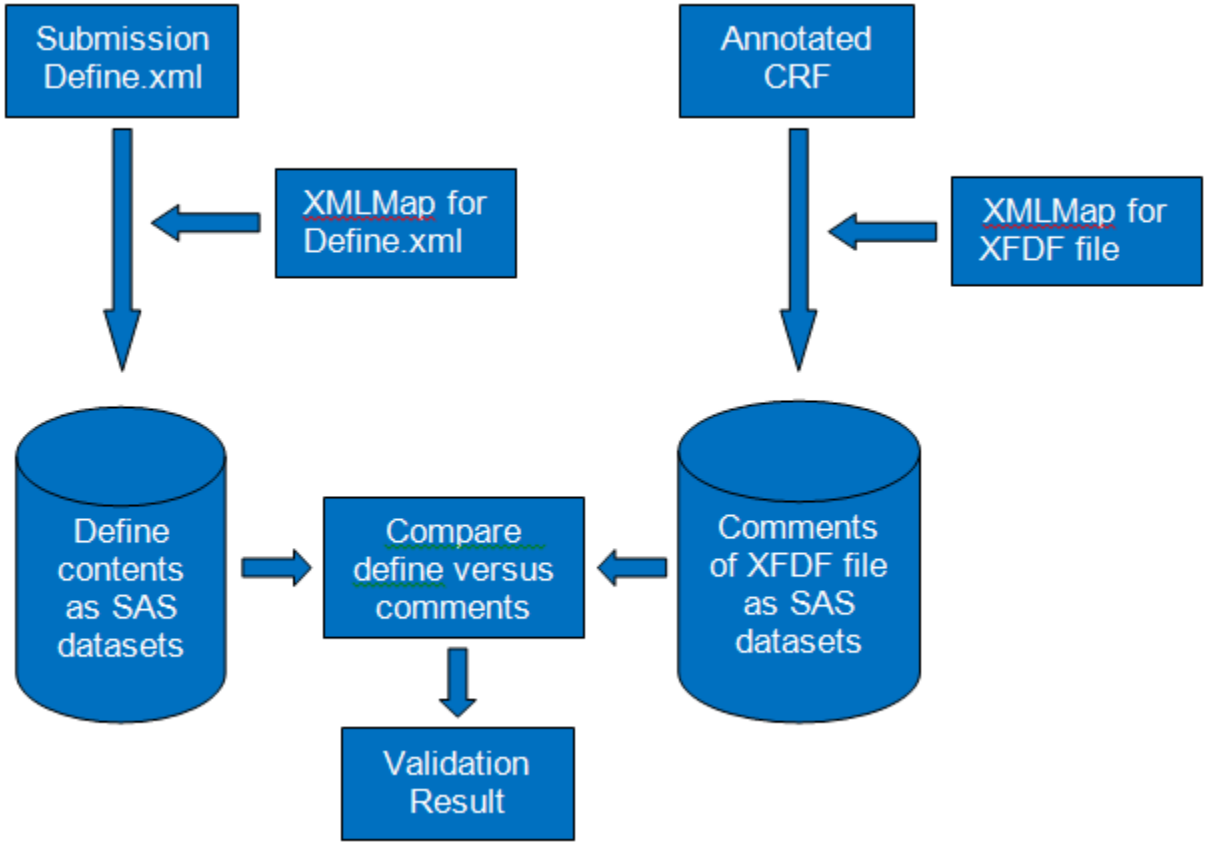
Origin Field in Define.xml

- Indicator of the origin of the variable.
  - CRF
  - Derived
  - eDT
  - Assigned
  - Protocol
- Hyperlink is provided to display the specific CRF Pages
  - If the reviewer clicks the hyperlink of a CRF Page, then there should be a annotation in corresponding CRF Page

Validity of Origin Field

- All the CRF Page with the same annotation should be listed in the Origin field for that variable in the define.xml
- To do this validation manually can be cumbersome and time consuming

A Programming Way to Check

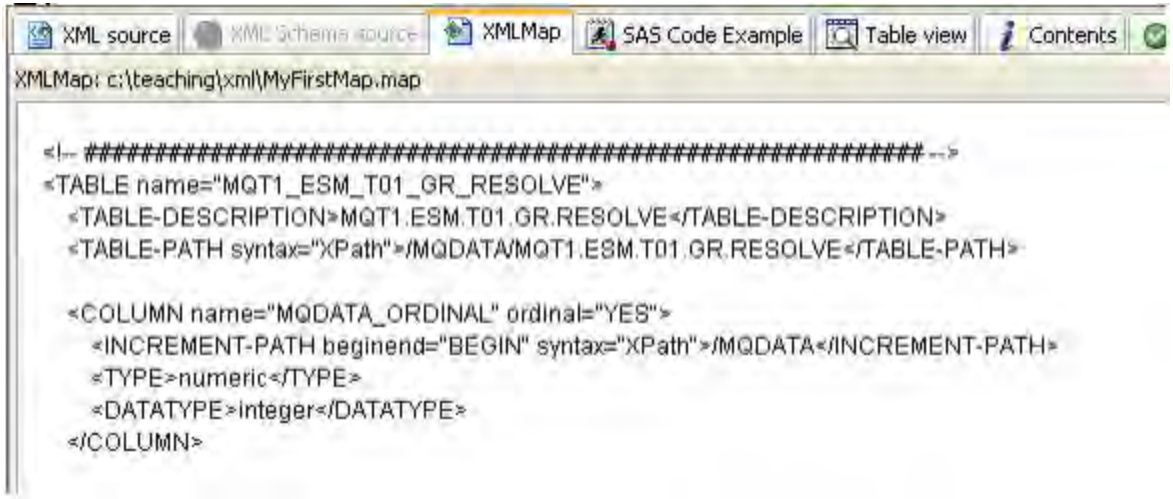


XML Mapper

- Download SAS XML mapper
  - http://support.sas.com/downloads/browse.htm?cat=12

How to Customize a XML Mapper

- Wendi L. Wright. 2010. How to Create an XML Map with the XML Mapper. NESUG 2010.



Define.Xml Content Extraction

Syntax for Reading define.xml into SAS Dataset

- filename define "./define.xml";
- filename defmap "./define2sas.map";
- libname define xmlv2 xmlmap=defmap automap=replace access=readonly;

- proc copy in=define out=work;
- run;

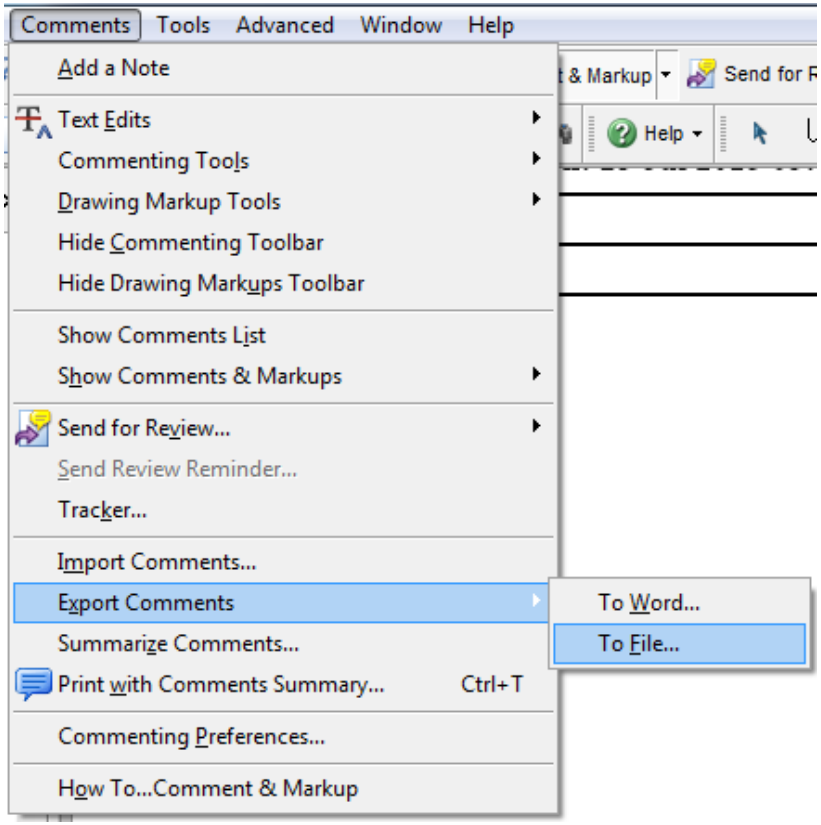
Annotated CRF Comments Extraction

Read Comments in PDF to SAS Dataset

- Export Annotation in blankcrf to Data file (XFDF, a xml version of form data in PDF)
- Version: Acrobat version 7.0

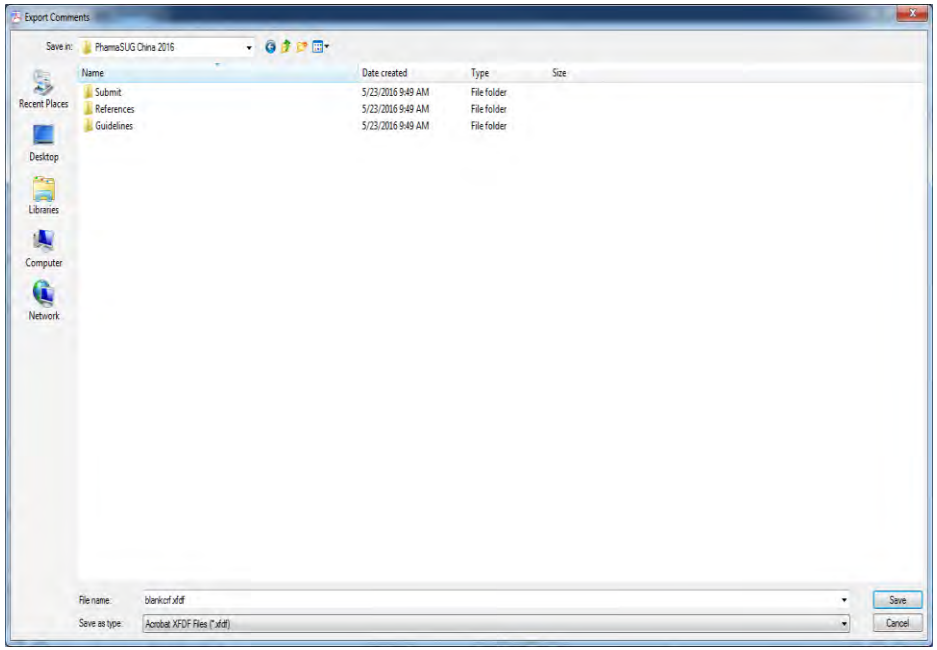
Read Comments in PDF to SAS Dataset

- Select “Comments → Export Comments → To File” from the main menu



Read Comments in PDF to SAS Dataset

- A new box with the title “Export Comments” appears. Save this resulting PDF as XFDF file at your location. Select save as type to be “Acrobat XFDF Files (*.xfdf)”



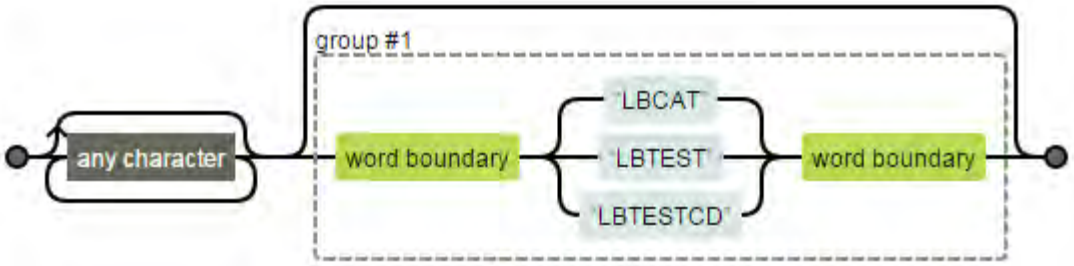
Syntax for Reading XFDF file into SAS Dataset

- Same methodology used to extract define.xml
  - filename blankcrf "./blankcrf.xfdf";
  - filename acrfmap "./acrf2sas.map";
  - libname blankcrf xmlv2 xmlmap=acrfmap automap=replace access=readonly;

  - proc copy in=blankcrf out=work;
  - run;

Parsing Imported Comments With Perl Regular Expression

- PRXCHANGE Function:
  - prxchange("s/.*?(\b(?:LBCAT|LBTEST|LBTESTCD)\b)?/\1 /o", -1, cats(COMMENTS));
- Regular expression visualization:



Macro Invoke

Parameter	Description
DEFCRFPATH	Path where define.xml to be validated and annotated CRF are located
XMLMAPPATH	Path where XMLMap is located
OUTPATH	Path where check result is located

- Using example:
  - %ChkDefCrfPage(defcrfpath = &demo_path
  - , xmlmappath = &demo_path
  - , outpath = &demo_path
  -);

Validation Result

Domain Abbreviation	SDTM Variable	CRF Page in Define.xml	CRF Page in Annotated CRF
DD	DDDTC		115
DD	DDTESTCD	115	115, 116
LB	LBTEST	32, 33, 34, 37	32, 33, 34, 37, 39, 41, 43, 45, 47, 49
LB	LBTESTCD	30, 32, 33, 34, 35, 36, 37, 55, 56	30, 32, 33, 34, 35, 36, 37, 39, 41, 43, 45, 47, 49, 55, 56
MH	MHENTPT	13	
PC	PCTEST	138, 139	
TU	TUTEST	119, 123, 124, 127, 128, 132	
XZ	XZTESTCD	54, 117	54, 117, 118

Reference

- Jeff Xia, Sangeetha Mahalingam. “Cross Check between Define.xml and blankcrf.pdf”. NJ CDISC User Group Meeting 2014.
- FDA CDER, 2011. “Common Data Standards Issue Document.”
- Prafulla Girase, Robert Agostinelli. “Automating Validation of Define.xml using SAS”. PharmaSUG 2013.
- Wendi L. Wright. 2010. “How to Create an XML Map with the XML Mapper”. NESUG 2010.
- Joel Campbell, Ryan Wilkins. “Importing and Parsing Comments From a PDF Document With Help From Perl Regular Expressions”. PharmaSUG 2011.

Name: Shenglin Zhang
Organization: PAREXEL International
Address: 20F, Taiping Finance Tower, No. 488, Middle YinCheng Road, Pudong
City, State ZIP: Shanghai, 200120
Work Phone: +86 21-51118092
Fax: +86 21 61609196
E-mail: Shenglin.Zhang@PAREXLE.com
Web: http://www.parexel.com/

Thanks!
Any questions?