PharmaSUG 2017 - Paper AD03

Three Issues and Corresponding Work-Around Solution for Generating Define.xml 2.0 Using Pinnacle 21 Enterprise

Jeff Xia, Merck & Co., Inc., Rahway, NJ, USA

Lugang (Larry) Xie, Merck & Co., Inc., Rahway, NJ, USA

Shari Lisnov, Merck & Co., Inc., Rahway, NJ, USA

ABSTRACT

This paper discusses three technical issues that occur in define.xml 2.0 files generated by Pinnacle 21 Enterprise, explains their causes, and presents an integrated work-around solution to resolve them with a simple SAS® macro.

As per the latest "Standards Catalog" released by the FDA, the agency will stop accepting define.xml 1.0 in NDA/BLA submissions starting from March 2018, which will require the industry to move to define.xml 2.0. The web-based application Pinnacle 21 Enterprise, is emerging as a useful tool in generating define.xml 2.0 in a high quality format that meets the agency's expectations for submissions. However, we identified three technical issues with the define.xml 2.0 generated by Pinnacle 21 Enterprise, during our submission process: 1) certain hyperlinks to external files do not work in Microsoft Internet Explorer (IE) when the files are saved in a Windows NTFS system; 2) the order of datasets in the table of contents does not follow CDISC recommendations and industry conventions; and 3) the text of the define.xml does not wrap properly in some windows-based text editors. This improper formatting causes difficulty when reviewing the xml syntax or performing minor updates using these editors. These issues can be resolved by the solution presented in this paper.

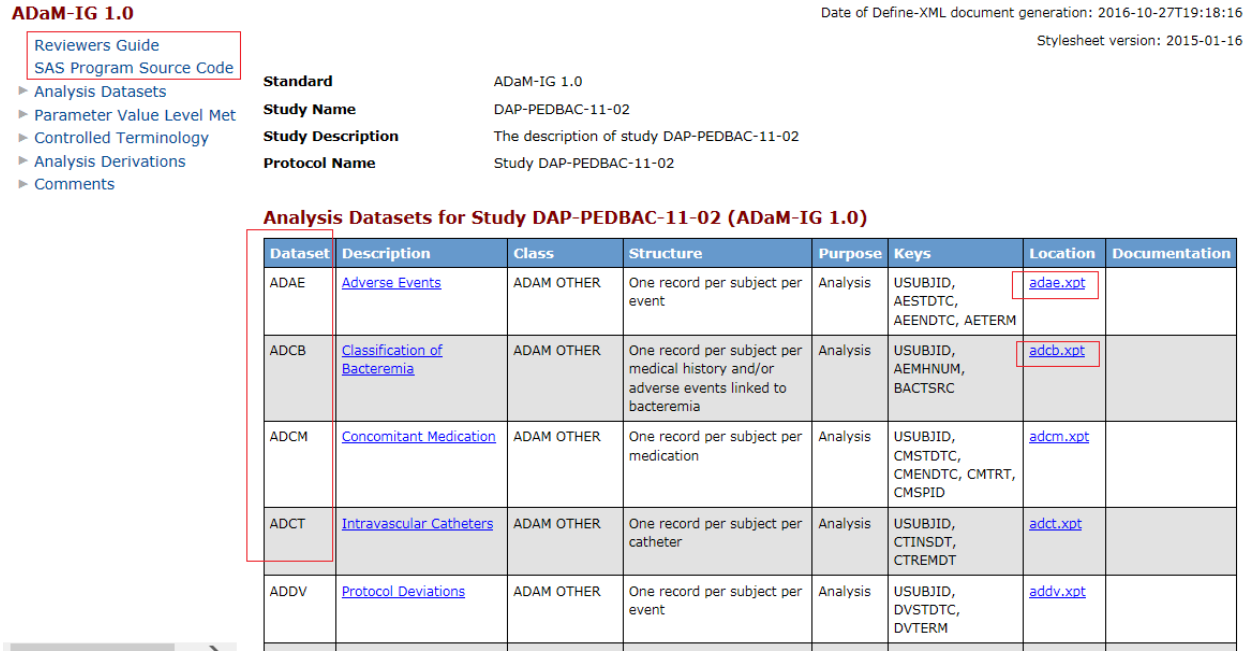
INTRODUCTION

Pinnacle 21 Enterprise is a web-based application, which can be used by sponsors to validate data packages in clinical trials. The FDA and PMDA also use Pinnacle 21 Enterprise to review submission data from sponsors. As a result, Pinnacle 21 Enterprise is becoming one of the most popular tools in the pharmaceutical industry for data review and technical conformance checking.

Pinnacle 21 Enterprise has many useful features, including a built-in function to generate define.xml 2.0 for the data in SDTM, ADaM, or SEND format. It has the capability to extract metadata information in different formats and from difference sources, such as data specifications in an Excel spreadsheet, datasets in SAS® transport format, or even another define.xml in a lower version. Furthermore, it can then generate a define.xml 2.0 file using the extracted metadata information, with minimal effort from end users. It is an effective and efficient tool, and the whole process is very straightforward.

We have, however, identified a few technical issues in the define.xml 2.0 generated by Pinnacle 21 Enterprise. Therefore, we have developed a SAS® macro-based work-around solution to resolve these issues. These technical issues are:

1. If the entire define package is saved in a Windows NTFS file system, some of the hyperlinks in the define.xml may not work as expected in IE. For example, hyperlinks to external PDF files, such as Analysis Data Reviewers Guide (ADRG), or to SAS® transport (xpt) files, such as adae.xpt, may be 'broken'; that is, when a reviewer clicks on these hyperlinks, IE does not open the associated files (see display 1).
2. The order of datasets in the Table of Contents (TOC) or Data Definition Table (DDT) does not follow industry conventions or recommendations in the Metadata Submission Guidance document published by CDISC. For example, the recommended order for SDTM datasets is "Trial Design Datasets", "Special-purpose Domains", "Interventions Domains", "Events Domains", "Findings Domains", and finally, "Relationship Datasets". For ADaM datasets, the industry convention is to include the ADSL dataset as the first entry in the define document, however, as you can see in display 1, Pinnacle 21 Enterprise arranges the datasets alphabetically.
3. As illustrated in display 2, the define.xml text lacks proper line-wrapping when opened in some MS Windows-based editors, like Notepad. This adds significant difficulty in reviewing/understanding the syntax of the define.xml and in performing certain edits when necessary.



ADaM-IG 1.0

Reviewers Guide
SAS Program Source Code
- Analysis Datasets
- Parameter Value Level Met
- Controlled Terminology
- Analysis Derivations
- Comments

Date of Define-XML document generation: 2016-10-27T19:18:16
Stylesheet version: 2015-01-16

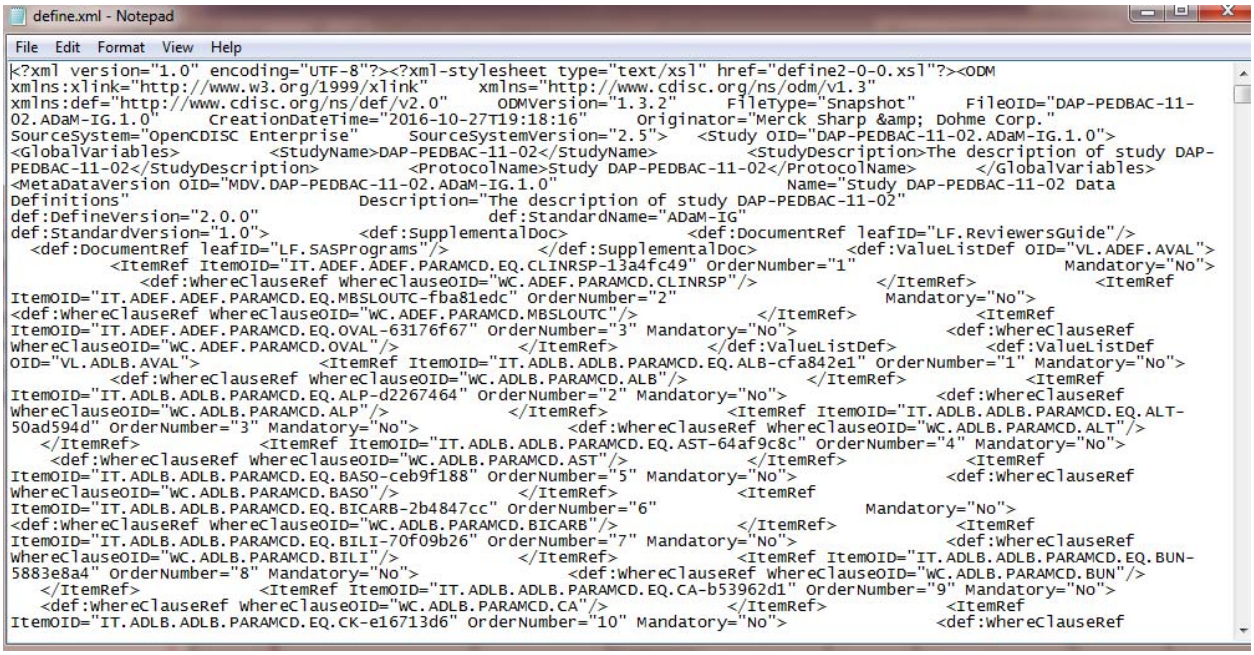
Standard	ADaM-IG 1.0
Study Name	DAP-PEDBAC-11-02
Study Description	The description of study DAP-PEDBAC-11-02
Protocol Name	Study DAP-PEDBAC-11-02

Analysis Datasets for Study DAP-PEDBAC-11-02 (ADaM-IG 1.0)

Dataset	Description	Class	Structure	Purpose	Keys	Location	Documentation
ADAE	Adverse Events	ADAM OTHER	One record per subject per event	Analysis	USUBJID, AESTDTC, AEENDTC, AETERM	adae.xpt	
ADCB	Classification of Bacteremia	ADAM OTHER	One record per subject per medical history and/or adverse events linked to bacteremia	Analysis	USUBJID, AEMHNUM, BACTSRC	adcb.xpt	
ADCM	Concomitant Medication	ADAM OTHER	One record per subject per medication	Analysis	USUBJID, CMSTDTC, CMENDTC, CMTRT, CMSPID	adcm.xpt	
ADCT	Intravascular Catheters	ADAM OTHER	One record per subject per catheter	Analysis	USUBJID, CTINSDT, CTREMDT	adct.xpt	
ADDV	Protocol Deviations	ADAM OTHER	One record per subject per event	Analysis	USUBJID, DVSTDTC, DVTERM	addv.xpt	

Display 1. Screen print of define.xml generated using Pinnacle 21 Enterprise.

(Note: 1. the hyperlink to Reviewers Guide does not respond to mouse click; 2. the hyperlink to adae.xpt does not respond to mouse click; 3. ADSL should be the first entry in above table.)



define.xml - Notepad

File Edit Format View Help

```
<?xml version="1.0" encoding="UTF-8"?><?xml-stylesheet type="text/xsl" href="define2-0-0.xsl"?><ODM
xmlns:xlink="http://www.w3.org/1999/xlink"     xmlns="http://www.cdisc.org/ns/odm/v1.3"
xmlns:def="http://www.cdisc.org/ns/def/v2.0"     ODMVersion="1.3.2"     FileType="Snapshot"     FileOID="DAP-PEDBAC-11-
02.ADaM-IG.1.0"     CreationDateTime="2016-10-27T19:18:16"     Originator="Merck Sharp &amp;amp; Dohme Corp."
SourceSystem="OpenCDISC Enterprise"     SourceSystemVersion="2.5">   <Study OID="DAP-PEDBAC-11-02.ADaM-IG.1.0">
<GlobalVariables>         <StudyName>DAP-PEDBAC-11-02</StudyName>         <StudyDescription>The description of study DAP-
PEDBAC-11-02</StudyDescription>         <ProtocolName>Study DAP-PEDBAC-11-02</ProtocolName>      </GlobalVariables>
<MetaDataVersion OID="MDV.DAP-PEDBAC-11-02.ADaM-IG.1.0"                                   Name="Study DAP-PEDBAC-11-02 Data
Definitions"                    Description="The description of study DAP-PEDBAC-11-02"
def:DefineVersion="2.0.0"                              def:StandardName="ADaM-IG"
def:StandardVersion="1.0">         <def:SupplementalDoc>            <def:DocumentRef leafID="LF.ReviewersGuide"/>
  <def:DocumentRef leafID="LF.SASPrograms"/>         </def:SupplementalDoc>         <def:ValueListDef OID="VL.ADEF.AVAL">
          <ItemRef ItemOID="IT.ADEF.ADEF.PARAMCD.EQ.CLINRSP-13a4fc49" OrderNumber="1"                    Mandatory="No">
             <def:WhereClauseRef WhereClauseOID="WC.ADEF.PARAMCD.CLINRSP"/>            </ItemRef>            <ItemRef
ItemOID="IT.ADEF.ADEF.PARAMCD.EQ.MBSLOUTC-fba81edc" OrderNumber="2"                    Mandatory="No">
<def:WhereClauseRef WhereClauseOID="WC.ADEF.PARAMCD.MBSLOUTC"/>            </ItemRef>            <ItemRef
ItemOID="IT.ADEF.ADEF.PARAMCD.EQ.OVAL-63176f67" OrderNumber="3" Mandatory="No">                <def:WhereClauseRef
WhereClauseOID="WC.ADEF.PARAMCD.OVAL"/>            </ItemRef>         </def:ValueListDef>         <def:ValueListDef
OID="VL.ADLB.AVAL">            <ItemRef ItemOID="IT.ADLB.ADLB.PARAMCD.EQ.ALB-cfa842e1" OrderNumber="1" Mandatory="No">
          <def:WhereClauseRef WhereClauseOID="WC.ADLB.PARAMCD.ALB"/>          </ItemRef>          <ItemRef
ItemOID="IT.ADLB.ADLB.PARAMCD.EQ.ALP-d2267464" OrderNumber="2" Mandatory="No">                <def:WhereClauseRef
WhereClauseOID="WC.ADLB.PARAMCD.ALP"/>            </ItemRef>            <ItemRef ItemOID="IT.ADLB.ADLB.PARAMCD.EQ.ALT-
50ad594d" OrderNumber="3" Mandatory="No">                <def:WhereClauseRef WhereClauseOID="WC.ADLB.PARAMCD.ALT"/>
   </ItemRef>            <ItemRef ItemOID="IT.ADLB.ADLB.PARAMCD.EQ.AST-64af9c8c" OrderNumber="4" Mandatory="No">
    <def:WhereClauseRef WhereClauseOID="WC.ADLB.PARAMCD.AST"/>            </ItemRef>            <ItemRef
ItemOID="IT.ADLB.ADLB.PARAMCD.EQ.BASO-ceb9f188" OrderNumber="5" Mandatory="No">                <def:WhereClauseRef
WhereClauseOID="WC.ADLB.PARAMCD.BASO"/>            </ItemRef>            <ItemRef
ItemOID="IT.ADLB.ADLB.PARAMCD.EQ.BICARB-2b4847cc" OrderNumber="6"                    Mandatory="No">
<def:WhereClauseRef WhereClauseOID="WC.ADLB.PARAMCD.BICARB"/>            </ItemRef>            <ItemRef
ItemOID="IT.ADLB.ADLB.PARAMCD.EQ.BILI-70f09b26" OrderNumber="7" Mandatory="No">                <def:WhereClauseRef
WhereClauseOID="WC.ADLB.PARAMCD.BILI"/>            </ItemRef>            <ItemRef ItemOID="IT.ADLB.ADLB.PARAMCD.EQ.BUN-
5883e8a4" OrderNumber="8" Mandatory="No">                <def:WhereClauseRef WhereClauseOID="WC.ADLB.PARAMCD.BUN"/>
   </ItemRef>            <ItemRef ItemOID="IT.ADLB.ADLB.PARAMCD.EQ.CA-b53962d1" OrderNumber="9" Mandatory="No">
   <def:WhereClauseRef WhereClauseOID="WC.ADLB.PARAMCD.CA"/>            </ItemRef>            <ItemRef
ItemOID="IT.ADLB.ADLB.PARAMCD.EQ.CK-e16713d6" OrderNumber="10" Mandatory="No">                <def:WhereClauseRef
```

Display 2. Screen print of define.xml when opened in Notepad in Win7. (Note: It lacks proper line wraps.)

MAIN DISCUSSION

1. CONFIGURATION OF PINNACLE 21 ENTERPRISE

The current version of Pinnacle 21 Enterprise we use at Merck is 3.1.1. This version has been installed on Amazon Web Service (AWS) and is offered to us as a SaaS (Software as a Service). SaaS is one of the three main categories of cloud computing; this implementation provides many benefits, such as unlimited scalability to the end users and

centralized administration of metadata updates, including new terminologies and new data standards. The underlying platform is usually not a Microsoft Windows-based system.

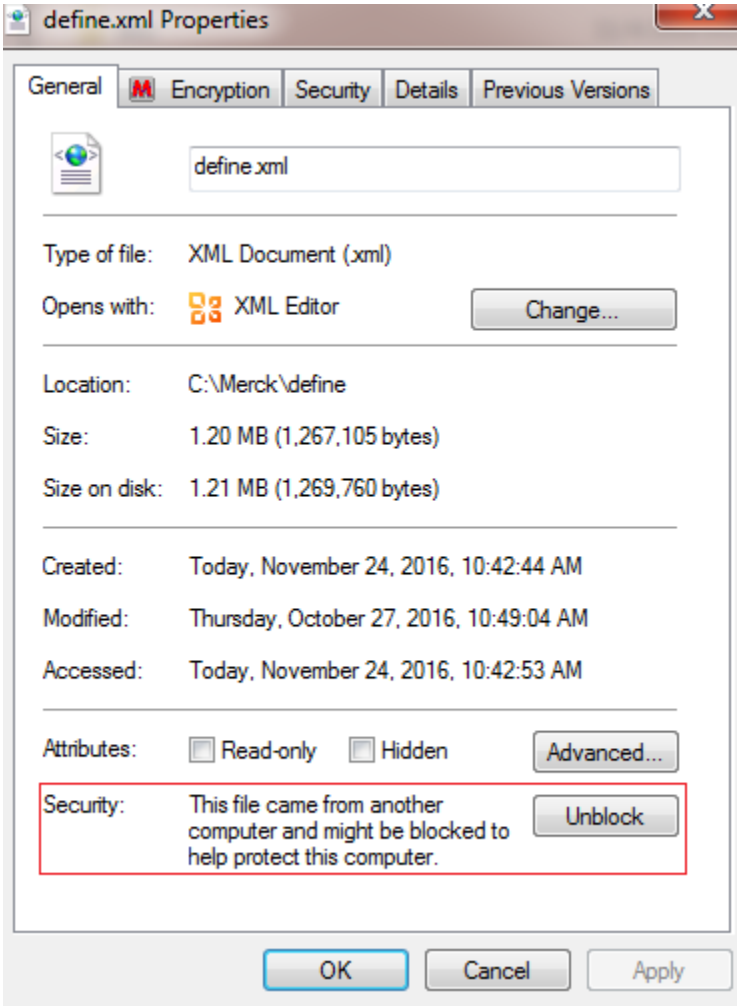
2. DIFFERENCE OF END OF LINE IN DEFINE.XML AND CORRESPONDING ISSUES

Extensible Markup Language (XML) is a simple, flexible text format used to exchange a wide variety of data on the Web and between applications. Define.xml is the data definition specification file in xml format; it is used extensively for data exchange or data submissions to agencies in the pharmaceutical industry. The web-based application Pinnacle 21 Enterprise is emerging as a useful tool in generating define.xml 2.0 in a high quality format that meets the agency's expectations for submissions.

Currently, due to differences in how End of Line (EOL) is represented in text files across platforms, there is a minor 'EOL-translation' problem when a define.xml file generated by Pinnacle 21 Enterprise (Unix-based) is used on/by a Windows platform. A Unix-based text file uses a single Line Feed <lf> to signify the end of a text line (EOL), whereas a Windows-based text file uses a Carriage Return/Line Feed <cr><lf>. The differences of the EOL among different platforms are summarized in table 1.

Platform	Description	End of Line
DOS (Windows)	two control-characters at the end of the line	<cr><lf>
Macintosh	one control-character at the end of the line	<cr>
Unix	one control-character at the end of the line	<lf>

Table 1: End of Line for text files in different platforms



Display 2. Screen print of define.xml 2.0 file properties. (Note: Windows Attachment Manager appends an

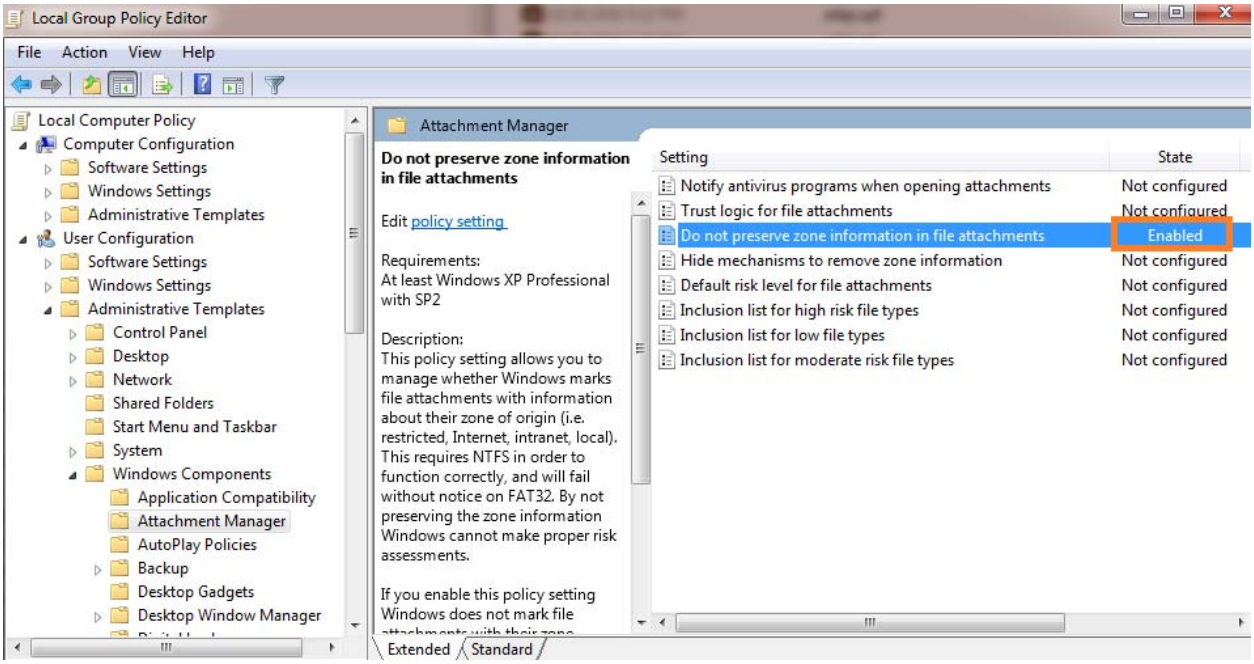
additional property for define.xml, which causes hyperlinks to PDF or XPT not to work as expected. Click the Unblock button to resolve the issue.)

In addition to non-working hyperlinks, the inconsistency in EOL across platforms can also cause problems with viewing/editing a define.xml 2.0 generated by Pinnacle 21 in a Windows-based text editor. This is because many Windows-based text editors do not recognize the single line feed EOL, and therefore, display the text with improper wrapping, as illustrated in display 2.

When the reviewers save the define.xml in an NTFS file system in MS Windows, the Attachment Manager in MS Windows appends additional file property information for security purposes, as seen in display 2. This additional security feature in MS Windows blocks any hyperlinks in the define.xml to external files, including PDF or XPT. There are two ways to address this issue:

1) Right click the define.xml in Windows File Explorer, choose “Properties” from the popup menu, and then click the “Unblock” button to release the blocking of external hyperlinks

2) Type “gpedit.msc” in the Windows command line, launch the Group Policy Object Editor, and change the status of “Do not preserve zone information in file attachments” from “Not configured” to “Enabled”. Saving this “Enabled” setting in the define.xml will prevent blocking of hyperlinks to external files including, PDF and XPT. However, we recommend consulting your system administrator before you make this change in group policy in MS Windows.



Display 3. Screen print of Group Policy Editor. (Note: Change the status of “Do not preserve zone information in file attachments” from “Not configured” to “Enabled”)

3. AN INTEGRATED WORK-AROUND SOLUTION

Xia and Xie (2016) presented a programming solution to re-order the datasets in the define.xml TOC, when up-versioning the define.xml from 1.0 to 2.0. The same technique can be used to re-order the define.xml in version 2.0.

There are many solutions on the market for converting the single line feed (LF) in a define.xml file to the carriage return/line feed (CRLF) recognized by Windows-based applications. A frequently used method is to go through an FTP server, which automatically converts the text file across different platforms. The other choice is to use some Unix-based utilities, such as unix2dos, to perform the conversion. These solutions either involve extra costs, such as an FTP server, or introduce extra processes and require expertise in the Unix environment. As SAS® programmers, developing a SAS® macro to make all necessary updates in a simple macro call is a very efficient solution.

The macro published by Xia and Xie (2016) has been revised for the purpose of EOL conversion. The macro reads in the entire define.xml file generated by Pinnacle 21 Enterprise, re-arranges the dataset order to follow the industry convention, and outputs the updated text to a separate external file using the SAS® data null step and PUT statement. NOTE: It is necessary to specify the option of encoding = “utf-8” in the file statement within the data step

to overwrite the default encoding value of “WLatin1” in a given SAS® session. Otherwise, the stylesheet will have problems in syntax parsing.

```
File  Edit  Format  View  Help
<?xml version="1.0" encoding="UTF-8"?>
<?xml-stylesheet type="text/xsl" href="define2-0-0.xsl"?>
<ODM
    xmlns:xlink="http://www.w3.org/1999/xlink"
    xmlns="http://www.cdisc.org/ns/odm/v1.3"
    xmlns:def="http://www.cdisc.org/ns/def/v2.0"
    ODMVersion="1.3.2"
    FileType="Snapshot"
    FileOID="DAP-PEDBAC-11-02.ADaM-IG.1.0"
    CreationDateTime="2016-10-27T19:18:16"
    Originator="Merck Sharp &amp;amp; Dohme Corp."
    SourceSystem="OpenCDISC Enterprise"
    SourceSystemVersion="2.5">
  <Study OID="DAP-PEDBAC-11-02.ADaM-IG.1.0">
    <GlobalVariables>
       <StudyName>DAP-PEDBAC-11-02</StudyName>
       <StudyDescription>The description of study DAP-PEDBAC-11-02</StudyDescription>
       <ProtocolName>Study DAP-PEDBAC-11-02</ProtocolName>
    </GlobalVariables>
    <MetaDataVersion OID="MDV.DAP-PEDBAC-11-02.ADaM-IG.1.0"
                     Name="Study DAP-PEDBAC-11-02 Data Definitions"
                     Description="The description of study DAP-PEDBAC-11-02"
                     def:DefineVersion="2.0.0"
                     def:StandardName="ADaM-IG"
                     def:StandardVersion="1.0">
```

Display 4. Screen print of define.xml with proper alignments after conversion using the SAS macro.

Because the PUT statement, by default, strips all leading spaces before SAS® writes the text to the external file, the proper indentation and alignment of each line in the define.xml is difficult to maintain. Two extra steps have been added to address this issue:

1. Counting the number of leading spaces in each line of the original text;
2. Instructing SAS® to output the text at certain locations, depending on how many spaces precede the original text.

```
** Output to XML **;
data _null_;
    file "&outfile" lrecl=8196 nopad encoding="utf-8";
    set xml2(keep=xmlcode);
    x =length(xmlcode)- length(left(xmlcode));
    put @x xmlcode    ;
run;
```

Display 5. SAS code to address the indentation and alignment. Refer to Xia and Xie (2016) for the entire logic.

CONCLUSION

The Pinnacle 21 Enterprise application has been an advantageous tool in generating define.xml 2.0 for regulatory submissions. Minor issues such as broken hyperlinks to external files, improper line wrapping, and alignment issues could be worked around by post-processing the define.xml generated by Pinnacle 21 Enterprise.

RECOMMENDED READING

1. Sergiy Sirichenko, Michael DiGiantomasso, Travis Collopy. 2015. Usage of OpenCDISC Community Toolset 2.0 for Clinical Programmers. PharmaSUG 2015.
2. Jeff Xia and Larry Xie, 2016. Up-Versioning Define.xml from 1.0 to 2.0. PharmaSUG 2016.

ACKNOWLEDGMENTS

The authors would like to thank Cynthia He for her great support and valuable input of this paper; and Travis Collopy from Pinnacle 21 for his great support.

CONTACT INFORMATION

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

Name: Jeff Xia
Enterprise: Merck
Address: 126 E. Lincoln Avenue
City, State ZIP: Rahway, NJ 07065-4607
Work Phone: 732-594-6439
Fax:
E-mail: jeff.xia@merck.com
Web: www.merck.com
Twitter:

Name: Lugang (Larry) Xie
Enterprise: Merck
Address: 126 E. Lincoln Avenue
City, State ZIP: Rahway, NJ 07065-4607
Work Phone: 732-594-1527
Fax:
E-mail: larry.xie@merck.com
Web: www.merck.com
Twitter:

Name: Shari Lisnov
Enterprise: Merck
Address: 126 E. Lincoln Avenue
City, State ZIP: Rahway, NJ 07065-4607
Work Phone: 732-594-6102
Fax:
E-mail: shari.lisnov@merck.com
Web: www.merck.com
Twitter:

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.