PharmaSUG 2012 – AD14

Creating Define.pdf with SAS® Version 9.3 ODS RTF

Elizabeth Li, PharmaStat, LLC, Newark, CA

Carl Chesbrough, PharmaStat, LLC, Newark, CA

Abstract

It is becoming more common for regulatory submissions to include define.xml - the data definition documents for Study Data Tabulations Model (SDTM) data, Analysis Data Model (ADaM) data, and even legacy data. Although define.xml documents can help regulatory reviewers to navigate submission datasets, documents and variable derivations, they usually do not print out properly on paper. One solution is to generate a define.pdf document with the same content as the define.xml. The portable document format (PDF) file includes bookmarks and hyperlinks to facilitate online review, and it can also be printed out for hardcopy review. Providing define.pdf documents can help sponsors remove obstacles to the review of their regulatory submissions.

This paper presents tips for using SAS® Version 9.3 Output Delivery System (ODS) rich text format (RTF) to generate an RTF file, and then use Acrobat's PDF Maker to convert it to a define.pdf document. It discusses reasons for the use of ODS RTF instead of ODS PDF. It demonstrates how to create a user-defined style using SAS® PROC TEMPLATE. It shows how to use RTF code to set up bookmarks and hyperlinks to internal as well as external locations. It provides examples of using other RTF code to improve formatting of the RTF document. The features of bookmarks, hyperlinks, headings, and other formatting details are important to an online review of any document.

Key Words

Define.pdf, SAS® ODS, RTF, define.xml, bookmark, hyperlink

Introduction

In a regulated industry such as pharmaceutical and biotechnology industry, product submissions for marketing approval currently require define.xml (data definition) files for study data tabulation model (SDTM). It is a good idea to include define.xml files for analysis data model (ADaM), and/or legacy (Item 11) data, even though they are not required at this time. Although define.xml files can help the reviewers to navigate submission documents, datasets, and variable definitions, they usually do not print out properly on paper. One solution is to generate a define.pdf file, which contains the same contents as the define.xml file. The portable document format (PDF) file not only includes bookmarks and hyperlinks to allow online review, but also can be printed for hardcopy review. In the industry, every effort is made by sponsors to reduce

the review time of regulatory submissions. Generating define.pdf files may ease the review, hence may reduce the review time.

This paper presents tips for using SAS® Version 9.3 Output Delivery System (ODS) rich text format (RTF) to generate a PDF file (define.pdf) via conversion from RTF file by Acrobat's PDF Maker, for Study Data Tabulations Model (SDTM), Analysis Data Model (ADaM) or legacy (Item 11) data. In the following sections, detailed descriptions will be presented for the features of a typical define.pdf file, why to use ODS RTF instead of ODS PDF, how to use SAS® PROC TEMPLATE to create a user-defined style for RTF outputs, how to use RTF code to set up bookmarks and hyperlinks to internal/external locations, as well as other RTF formatting details.

Description of a Typical define.pdf File

There are three major features of a typical define.pdf file:

1. It should include the same contents as the define.xml, which is viewed through a stylesheet using a web browser. The contents include study information, references to additional documents (e.g., annotated case report form (aCRF) and/or data guides), dataset information, variable-level metadata, value-level metadata, computational methods, and controlled terminologies (i.e. code lists and external dictionaries).

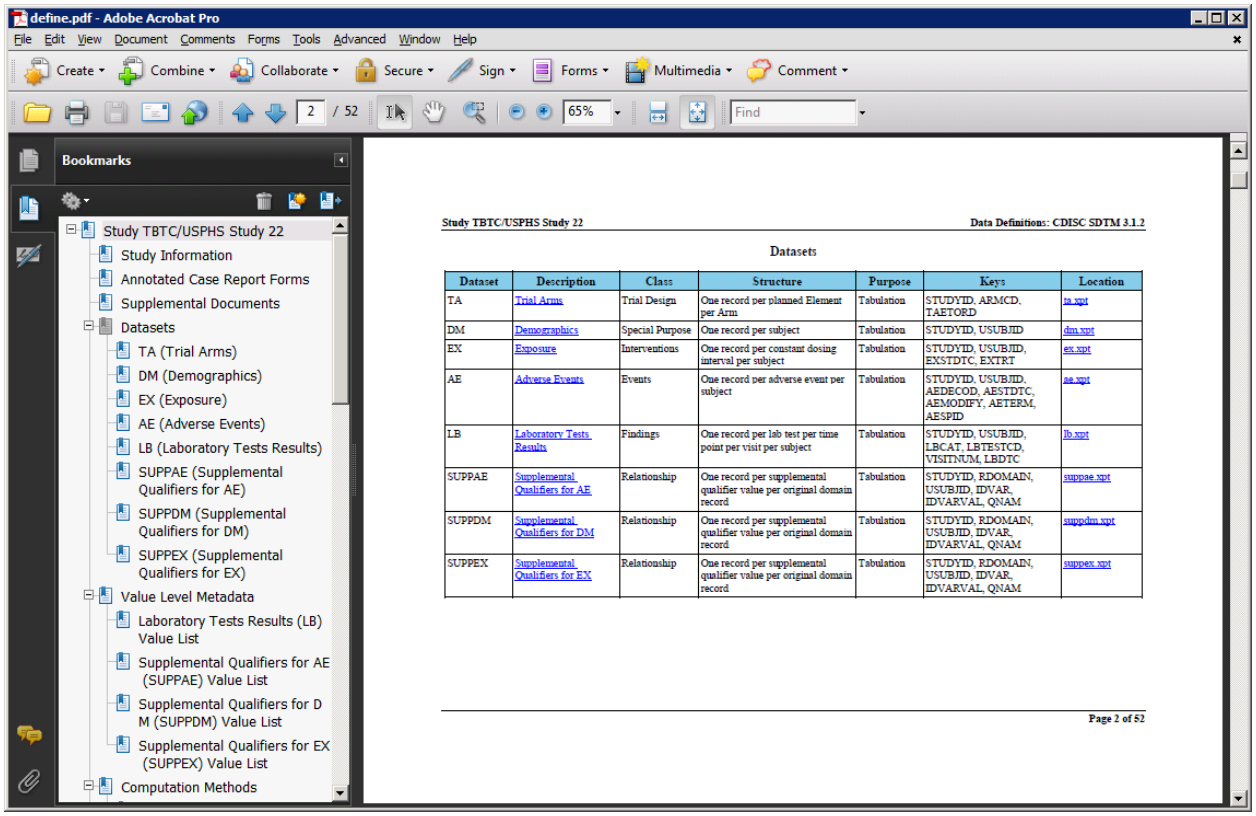
2. It should contain bookmarks and hyperlinks to access information during an online review. A bookmark contains a marker or address that identifies a document or a specific place within the document. A hyperlink links to another place in the same document or to an entirely different document. Bookmarks and hyperlinks provide regulatory reviewers the option to review online.

3. A define.pdf should incorporate the define.xml contents within printable space of a document. This enables the reviewer to print the define.pdf for a hardcopy review.

In addition to the above three major features of a define.pdf file, hierarchical bookmarks can help a viewer to navigate within the document efficiently. There are essential elements in hard copy review, such as header and/or footer information, page numbering, and formatting of the contents within the printable space.

Figure 1 shows an image of a page from a define.pdf document. In the image, bookmarks and hyperlinks are clearly shown. There are bookmarks for study information, annotated case report forms (aCRFs), supplemental documents, dataset metadata for each specific dataset, variable-level metadata, and specific value-level information. A dataset metadata table is displayed on the right hand side. There are two hyperlinks (in blue underline font) for each dataset. The hyperlink in the description column is connected to the define.pdf page for that dataset. The hyperlink in the location column is connected to the actual dataset outside the define.pdf document. For a more detailed example of the define.rtf document, please see Appendix A.

Figure 1 Sample Define.pdf (from ODS RTF) with Bookmarks and Hyperlinks.



Dataset	Description	Class	Structure	Purpose	Keys	Location
TA	Trial Arms	Trial Design	One record per planned Element per Arm	Tabulation	STUDYID, ARMCD, TAETORD	ta.xpt
DM	Demographics	Special Purpose	One record per subject	Tabulation	STUDYID, USUBJID	dm.xpt
EX	Exposure	Interventions	One record per constant dosing interval per subject	Tabulation	STUDYID, USUBJID, EXSTDTC, EXTRT	ex.xpt
AE	Adverse Events	Events	One record per adverse event per subject	Tabulation	STUDYID, USUBJID, AEDECOD, AESTDTC, AEMODIFY, AETERM, AESPID	ae.xpt
LB	Laboratory Tests Results	Findings	One record per lab test per time point per visit per subject	Tabulation	STUDYID, USUBJID, LBCAT, LBTESTCD, VISITNUM, LBDTC	lb.xpt
SUPPAE	Supplemental Qualifiers for AE	Relationship	One record per supplemental qualifier value per original domain record	Tabulation	STUDYID, RDOMAIN, USUBJID, IDVAR, IDVARVAL, QNAM	suppae.xpt
SUPPDM	Supplemental Qualifiers for DM	Relationship	One record per supplemental qualifier value per original domain record	Tabulation	STUDYID, RDOMAIN, USUBJID, IDVAR, IDVARVAL, QNAM	suppdm.xpt
SUPPEX	Supplemental Qualifiers for EX	Relationship	One record per supplemental qualifier value per original domain record	Tabulation	STUDYID, RDOMAIN, USUBJID, IDVAR, IDVARVAL, QNAM	suppex.xpt

Why Use ODS RTF Instead of ODS PDF

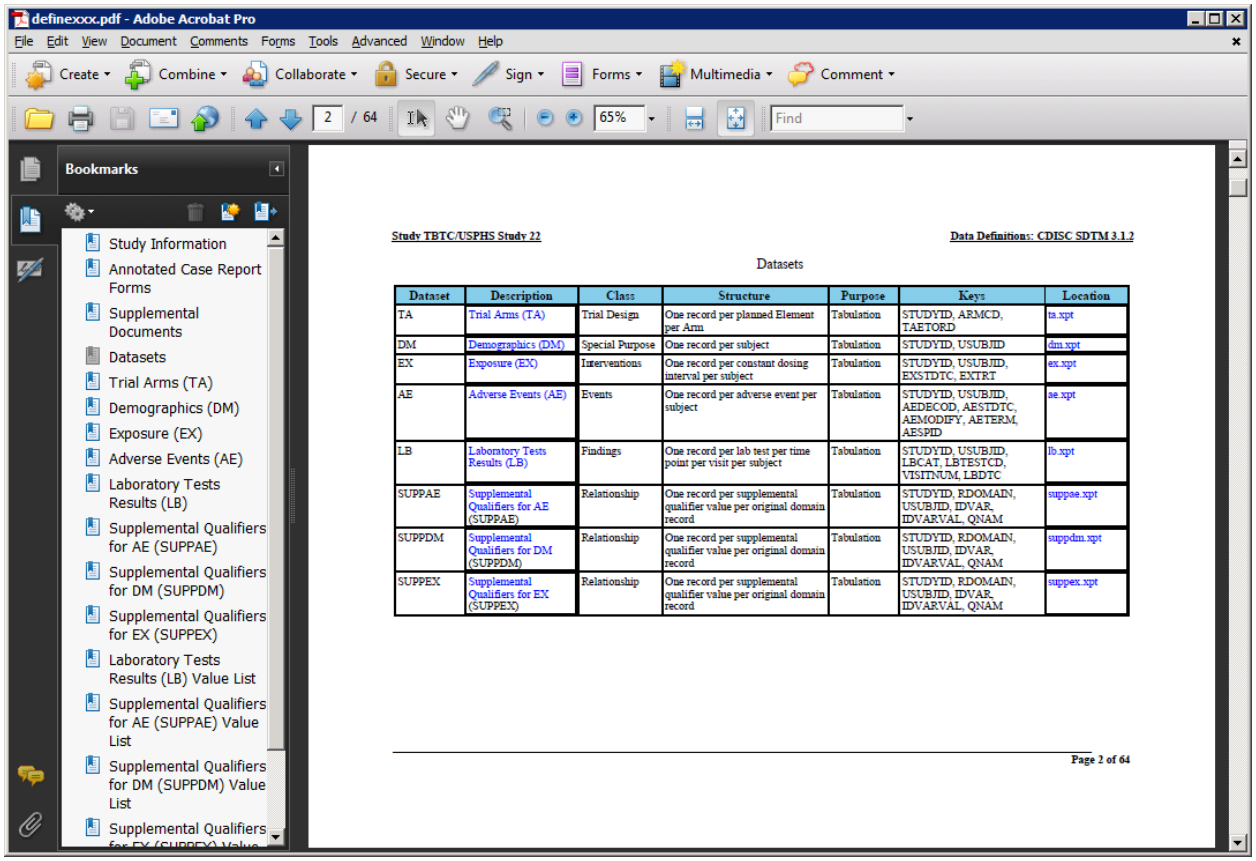
It would would be more straightforward if we could create a satisfactory define.pdf directly using SAS® ODS PDF. Figure 2 shows a document created using ODS PDF. We have found that ODS PDF can

- Generate define.xml content similar to that from ODS RTF
- Create hyperlinks, although lacking some desired features
- Create bookmarks, although lacking some desired features
- Generate reasonable header/ footer information
- Format acceptable contents

With ODS PDF, the document hyperlinks by default have a blue box around the link's display text. Using a SAS style, we can change the link color to be black, and change the color of the display text to blue with an underline. This gives the appearance of the text as a conventional browser hyperlink. The reality is that the link is the entire cell of the table. This also has the unintended effect of displaying thicker lines for the cells which contain links.

With the ODS PDF, we are able to suppress the extra two levels generated with the PROC REPORT, but we were not able to make the bookmarks hierarchical. In contrast to ODS RTF, we were unable to create multiple levels of bookmarks, i.e., insert a bookmark for the "Value Level Metadata" section, then insert an indented bookmark for each value-list within that section.

Figure 2 Sample Define.pdf (from ODS PDF) with Bookmarks and Hyperlinks.



Here is a summary table that compares the define.pdf files that generated using ODS PDF and ODS RTF.

Features of define.pdf	ODS PDF	ODS RTF
Contents of the define.xml	✓	✓
Hierarchical hyperlinks		✓
Bookmarks (blue underline text)	?	✓
Header/ footer information	?	✓
Page number	✓	✓

In order to generate a define.pdf with the features described in the previous section, and to have better control of the format of contents, using ODS RTF is a better choice than ODS PDF. There

are other ways to generate a define.pdf. One example is to create a postscript file first. We decided to create define.rtf first then use Acrobat's PDF Maker to convert the RTF file to PDF.

Creating a User-Defined Style with PROC TEMPLATE

Using PROC TEMPLATE, a user-defined style can be created for a RTF output file. Here is a list of style elements that are used in a define.rtf document.

1) Set margins, font style, and font size of a document
2) Set borders, border width, cell padding, and cell space for a table
3) Set background color for the header row of a table
4) Set "pretext" and user text cell width and center justification
5) Set system footer to use RTF codes

Using the following SAS® PROC TEMPLATE code segments, a user can set the desired borders, define the font style, specify table style with a complete grid, assign the border spacing and cell spacing, as well as indicate the background color of the header row in a table. In addition, a user may specify styles for printed text before a procedure ("pretext") or user-defined text. Furthermore, a user may allow RTF codes to be rendered in the system footer. Here is the SAS® code segment used to generate a define.rtf output. The comments within the SAS code further explain the purposes.

```
proc template;
define style Styles.RtfMod;
   parent = styles.rtf;

   /* 1) set the margins to 1 inch all around */
   style Body from Body
      "Controls the Body file." /
      marginbottom = 1in  /* bottom margin */
      margintop = 1in     /* top margin */
      marginright = 1in   /* right margin */
      marginleft = 1in   /* left margin */
;...
      /* set Font and size for system titles and footers */
      'TitleFont2' = ("Times New Roman, <serif>", 11pt, bold)
                              /* Font and size for system titles and footers */
      'TitleFont' = ("Times New Roman, <serif>",13pt, bold)
                             /* Font and size for system titles and footers */
      'StrongFont' = ("Times New Roman, <serif>", 11pt, bold)
                              /* Font and size for row header */
      'headingEmphasisFont' = ("Times New Roman, <serif>", 11pt, bold)
                              /* Font and size for text from compute block */
      'headingFont' = ("Times New Roman, <serif>", 11pt, bold)
                              /* Font and size for column header */
      'docFont' = ("Times New Roman, <serif>", 10pt)
                              /* Font and size for data value in tables */
;...
   /* 2) set table with rule and updated cell padding */
   style table from output /
      rules = all
      frame = box          /* table frame */
```

```
      borderwidth = 1pt    /* table border */
      borderspacing = 1pt
      bordercolor = black
      background = white
      cellpadding = 2pt
      cellspacing = 1pt    /* gridline of table cells.  Default=0 pt makes lines */
                           /* inside lines of the table about 1/2 pt */
   ;
      /* 3) set background color for the header row of the table */
      style color_list from color_list
         "Colors used in the default style" /
         'link' = blue    /* hyperlink color */
         'bgH' = skyblue  /* background color for header. Was white by default */
         'fg' = black     /* foreground color */
         'bg' = white     /* background color */...
   ;
      /* 4) set pretext and user text cell width and center justified ***/
      class prepage /
         cellwidth=5in
              /* to avoid text wrap, provide ample space (5 in) for ods rtf prepage */
         just=center
   ;
      style UserText from Note/
         cellwidth=5in
                 /* to avoid text wrap, provide ample space (5 in) for ods rtf text */
         just=center
   ;
      /* 5) set system footer to allow RTF code ***/
      class systemfooter /
         protectspecialchars=off
   ;
   End;
   Run;
```

Using RTF Code to Generate Headings and Lines

In order to control the styles of a define.rtf document, RTF codes are used to enhance the features, such as line, headings, and keep with next text line. To make sure SAS recognizes the RTF code, the SAS ODS escapechar = statement is used. A set of SAS® macro variables is created to store those commonly-used RTF codes for the RTF output. SAS® recognizes any text string with single quote after ^R/RTF as a rich text format code. Here are examples of setting up SAS® macro variables that contain RTF codes. Comments are used to explain the meanings of the RTF codes.

```
ods escapechar='^';
%let line=^R/RTF'\brdrb\brdrs\brdrw10\brsp20'; /* add a line */
%let head1=^R/RTF'\s1\fs26\b\qc ';            /* Style 1, size=13pt, bold, center */
%let head2=^R/RTF'\s2\fs24\b\qc ';            /* Style 2, size=12pt, bold, center */
%let head22=^R/RTF'\s2\fs24\b\qc\page ';
                                  /* Style 2, size=12pt, bold, center, page break */
...
%let keepn=^R/RTF'\keepn ';         /* keep with next */
%let heading4=%str({\s1 Heading 1;\s2 Heading 2;\s3 Heading 3;\s4 Heading 4;});
                  /* define four headings in the RTF document as Styles 1 to 4 */
```

The *line* macro variable (see above SAS® code example) is used to add lines under the header for study names and above the page x of y in the footer. The *head1* and *head2* macro variables are used to add the RTF codes to define different heading styles. These heading styles are used for creating bookmarks in the PDF output. The *head1* macro variable defines style 1 (s1) as having the following characteristics: font size of 13pt (\fs26), bold (\b), and centered (\qc). See Appendix B for additional RTF codes.

To improve the visual effect of the tables, the *keepn* macro variable is used to add RTF codes to the first row and the next-to-last row of each table in order to avoid orphan rows. By adding the *keepn* macro variable to the first row of a table, the table will keep the following row on the same page as the first row. This keeps a table from starting at the bottom of a page with only one row before going to the next page. In addition, adding the *keepn* macro to the next-to-last row will prevent the last row of a table from starting a new page.

The following SAS® code puts the above-mentioned macro variables to use in a RTF output.

```
data study;
       set &UserDataLibname..study;
       length stdytitle $200;
       stdytitle="&head2.Study Information";
       call symput ('study',trim(studyname));
       call symput ('std',trim(StandardName));
       call symput ('stdv',trim(StandardVersion));
run;
%put study = &study;
...
title1 h=10pt j=l "Study &study.&line" j=r "&docver.&line";

footnote1 h=10pt j=c "&line" ;
footnote2 h=10pt j=r "Page ^{pageof}";

ods rtf Text = "&head1.Study &study";


ods rtf prepage = "&head2.Study^\~Information";
proc report data = study nowd split='~' ;
       column   StudyName StudyDescription ;
       define StudyName          / display "Study Name"
          style(column)=[cellwidth=2in just=left];
       define StudyDescription   / display "Study Description"
          style(column)=[cellwidth=6.5in just=left];
run;
```

Header Information with a line

Page number below a line in footer

Heading 1: Study Name and Number

Heading 2: Study Information

Generating Bookmarks, Hyperlinks, and Special Effects

Key features of the define.pdf are bookmarks and hyperlinks. They allow the reviewer to quickly navigate to the desired data. SAS® automatically creates bookmarks as each output table is created, called IDX, IDX1 through IDX<n>. These SAS-generated bookmarks are used to create hyperlinks for each dataset description within the RTF document. The RTF code "\\l" is used to indicate that the hyperlink is a local bookmark. The starting number of the IDX will depend on the number of output tables generated before the dataset variable-level description tables. The macro variable *_indx* contains the count of tables prior to the dataset variable-level

description tables, and is added to the macro variable *DatasetOrder* to give the correct IDX bookmark number to the hyperlink. The RTF code for the character style (\cs15), blue color (\cf2), and underline (\ul) is used to display the hyperlink.

The location of the SAS® xpt file is created with the RTF code "\\", without the l, as it is not a local bookmark. The *location* macro variable is used to create a relative link to the SAS® xpt file. Here is a sample SAS® code for generating bookmarks and hyperlinks:

```
...
/* Adding hyperlink dataset titles using SAS automatically generated
bookmarks IDXn in the same document - Dataset Order is very important */
/* There are 3 tables for SDTM data +3 (=&_indx), 2 for ADaM data +2 */
      titleh = '{\field\*{\*\fldinst\*HYPERLINK\*\\l\*' ||
           'IDX' || compress(put(DatasetOrder+%eval(&_indx),f8.)) ||
           '}{\fldrslt\*{\cs15\cf2\ul\*' || trim(left(title)) || '}}}';

/* adding external hyperlinks to *.xpt files which will be stored in the same
folder as the define.pdf */
      locationh = '{\field\*{\*\fldinst\*HYPERLINK\*"\\'  ||
               compress(location) ||
               '"}{\fldrslt\*{\cs15\cf2\ul\*' || compress(location) || '}}}';
...
```

Additional hyperlinks are needed for links to additional data, such as value-level metadata tables, computational algorithms, and code lists. To create these hyperlinks, a bookmark is created using the *\bkmkstart* and *\bkmkend* RTF code. The bookmark name is formed by adding a "BK" string to the name of the desired location.

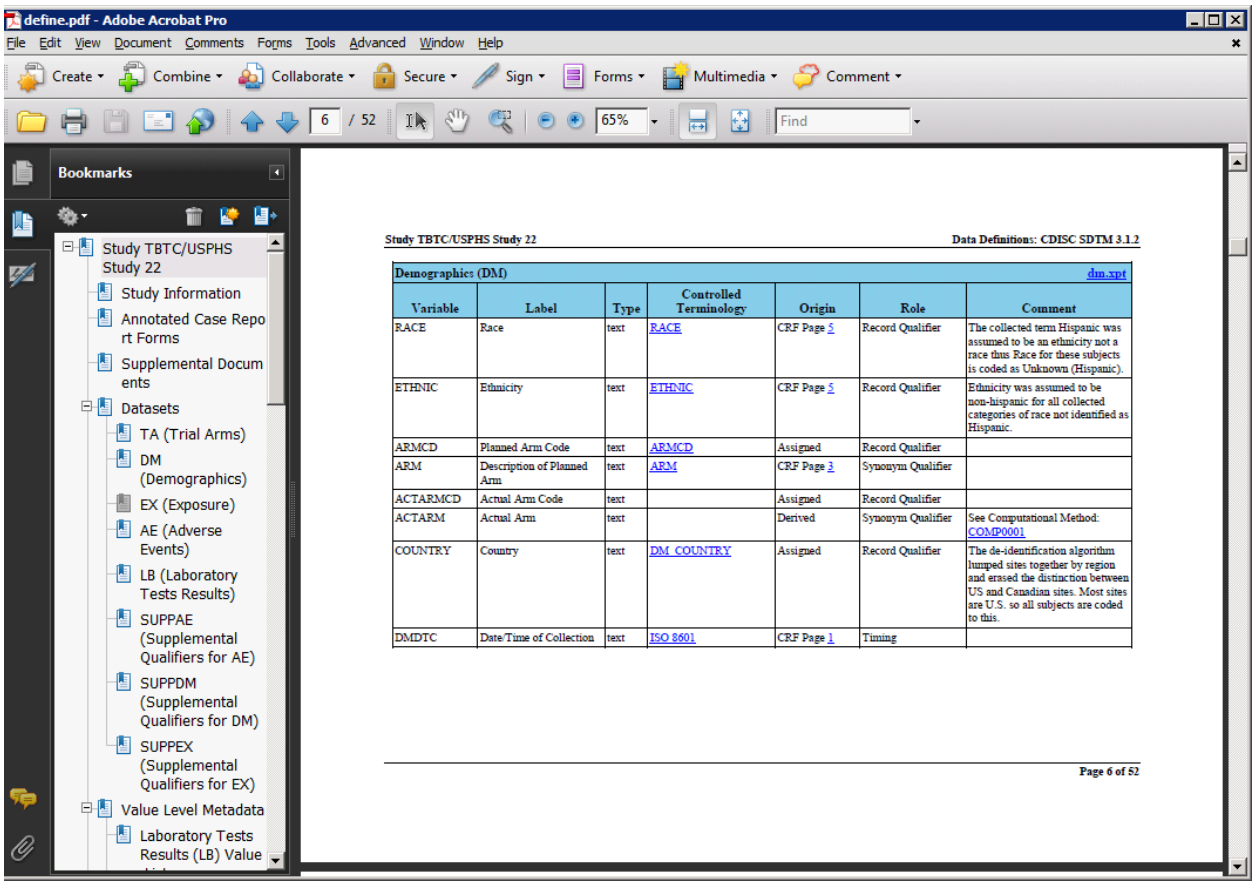
The following is an example of adding bookmarks to the computational method reference:

```
...
/* Identify the variables with computational method */
proc sort data=variable out=varcomp (keep=dataset variable comment ComputationMethod)
nodupkeys;
      by dataset variable;
      where ComputationMethod^=' ';
run;
data varcomp;
      set varcomp;
      by dataset variable;
      length compoid $40 compoidb $200;
      compoid='COMP'||put(_n_,z4.);

      /**** adding internal (within the document define.pdf) hyperlinks ****/
      /**** {\field {\*\fldinst HYPERLINK "filename_and_path"}
       {\fldrslt display_text}} ****/
      comment = trim(left(comment))||' See Computational Method: '
       ||"^R/RTF'{\field {\*\fldinst HYPERLINK \\l " ||
      'BK' || compress(compoid) || '}{\fldrslt {\cs15\cf2\ul '
       || trim(left(compoid)) || '}}}';

      /*** add bookmarks to the computational method reference number ***/
      /***{\*\bkmkstart bookmark_name}display_text
       {\*\bkmkend bookmark_name}****/
      compoidb="&head4"||'{\*\bkmkstart BK' || compress(compoid) ||
       '}' || compress(compoid)||'{\*\bkmkend BK'||compress(compoid)||'}';
```

```
      /*** count the total number of computational methods ***/
      %let ncomp=0;
      if _n_>0 then call symput ('ncomp',compress(put(_n_,4.)));
run;
```

Figure 3, shows the define.pdf image after it is converted from the define.rtf generated by the above SAS code.

Figure 3 Sample output with bookmarks and local hyperlinks (from SAS code above)



Putting define.rtf Together with ODS RTF and PROC REPORT Statements

Using multiple PROC REPORT procedures, we generated a complete RTF output. A COMPUTE statement in PROC REPORT is used to generate variable-level metadata for each dataset. The heading of the table contains the dataset name (left justified) and a link for the SAS transport (xpt) file (right justified). To have these two variables justified on opposite ends of the line, a right-justification tab is inserted using the RTF code \tqr\tx12550. This code inserts a tab at 8.7 inches which right justifies any text after the tab character. Inserting the RTF \tab code, a tab after the dataset name right justifies the xpt file name after the tab. In order to suppress trailing spaces in the dataset name, the Datasetb variable is printed with the $varying100. format.

```
ods rtf prepage = "&&&dst&i";
proc report data = &_indat nowd split='~';
      where dataset="&&&ds&i";
      column DatasetOrder datasetb locationh variable
             label datatype codelisth origin comment ;
      define DatasetOrder  / order order=internal noprint;
      define Datasetb      / order order=internal noprint;
      define locationh     / order order=internal noprint;
      define variableh     / display "Variable"
                             style(column)=[cellwidth=1in just=left];
      define label         / display "Label"
                             style(column)=[cellwidth=1.5in just=left];
      define datatype      / display "Type"
                             style(column)=[cellwidth=0.5in just=left];
      define codelisth     / display "Controlled~Terminology"
                             style(column)=[cellwidth=1in just=left];
      define origin        / display "Origin"
                             style(column)=[cellwidth=1.5in just=left];
      define comment       / display "Comment"
                             style(column)=[cellwidth=3.25in just=left];
      break after datasetb / page;
      compute before _page_ /style = [protectspecialchars = off
                                      background=cxc6dbf1 just=left];
            dslen = lengthn(Datasetb);
            line @1 '\tqr\tx12550' Datasetb $varying100. dslen '\tab'
            locationh $100.;
      endcomp;
run;
```

Discussion

Once the RTF output is generated, we use Adobe Acrobat Pro PDF Maker to convert it to a define.pdf document from Microsoft Word 2007, which is a reader of rich text format files. Handshakes between Microsoft Word and Acrobat should be carefully examined. For example, in define.xml a hyperlink to a specific page within an external PDF file contains a URL such as "blankcrf.pdf#21", where #21 is a pdf "named destination" for Page 21 in the blankcrf.pdf file. Using SAS, a user can create a hyperlink with this URL, but when the RTF file is converted into PDF, the "#21" part is ignored during the PDF conversion and the hyperlink will reference the entire document and only open the first (or most recently accessed) page, not the intended page.

In conclusion, we were able to generate a robust define.rtf from the define.xml contents with bookmarks, hyperlinks, and proper pagination by using SAS® ODS RTF. The sample define.pdf in Appendix A was generated in SAS® version 9.3. The same define.pdf document was generated from SAS® version 9.2 (TS2M2) using the same program code.

References

Output Delivery System, Basics and Beyond, Lauren E. Haworth, Cynthia L. Zender, Michele M. Burlew, SAS Publication, 2009

To ODS RTF and Beyond, David Shannon, http://www2.sas.com/proceedings/sugi27/p001-27.pdf

Preproduction RTF Features in SAS 9.1, http://support.sas.com/rnd/base/ods/odsrtf/rtf901.html

RTF Pocket Guide, Sean M Burke, O'Reilly & Associates, Inc., 2003

Acknowledgement

Our special thanks go to John Brega of PharmaStat, LLC, for his review, suggestions, comments, and critiques on this paper. His contribution is greatly appreciated. We also like to thank Dr. Chad Heilig of the Centers for Disease Control for his support of this presentation. The USPH/TBTC Study 22 information is used in this presentation with permission.

Contact Information

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

Elizabeth Li
PharmaStat, LLC
39899 Balentine Drive, Suite 109
Newark, CA 94560
Work Phone: 510 656-2080
elizabethli@pharmastat.com

Carl Chesbrough
PharmaStat, LLC
39899 Balentine Drive, Suite 109
Newark, CA 94560
Work Phone: 510 656-2080
cchesbrough@pharmastat.com

The header uses SAS TITLE statement: title1 h=10pt j=l "Study &study.&line" j=r "&docver.&line";

Appendix A. Sample define.rtf document

Study TBTC/USPHS Study 22

from ODS RTF TEXT =

Study Information

from ODS RTF PREPAGE =

The shading is set in PROC TEMPLATE

Study Name	Study Description
TBTC/USPHS Study 22	Efficacy and Safety of Once-Weekly Rifapentine and Isoniazid Compared to Twice-Weekly Rifampin and Isoniazid in the Continuation Phase of Therapy for Pulmonary Tuberculosis

from PROC REPORT

Annotated Case Report Forms

Title	Document Name
Annotated Case Report Form	blankcrf.pdf

Supplemental Documents

Title	Document Name
Guide to SDTM Tabulations Datasets	tabulations-data-guide.pdf

Datasets

Dataset	Description	Class	Structure	Purpose	Keys	Location
TA	Trial Arms	Trial Design	One record per planned Element per Arm	Tabulation	STUDYID, ARMCD, TAETORD	ta.xpt
DM	Demographics	Special Purpose	One record per subject	Tabulation	STUDYID, USUBJID	dm.xpt
EX	Exposure	Interventions	One record per constant dosing interval per subject	Tabulation	STUDYID, USUBJID, EXSTDTC, EXTRT	ex.xpt
AE	Adverse Events	Events	One record per adverse event per subject	Tabulation	STUDYID, USUBJID, AEDECOD, AESTDTC, AEMODIFY, AETERM, AESPID	ae.xpt
LB	Laboratory Tests Results	Findings	One record per lab test per time point per visit per subject	Tabulation	STUDYID, USUBJID, LBCAT, LBTESTCD, VISITNUM, LBDTC	lb.xpt
SUPPEX	Supplemental Qualifiers for EX	Relationship	One record per supplemental qualifier value per original domain record	Tabulation	STUDYID, RDOMAIN, USUBJID, IDVAR, IDVARVAL, QNAM	suppex.xpt

<For illustration purposes, the metadata of datasets have been shortened.>

TA (Trial Arms)

Trial Arms (TA)						ta.xpt
Variable	**Label**	**Type**	**Controlled Terminology**	**Origin**	**Role**	**Comment**
STUDYID	Study Identifier	text	STUDYID	Assigned	Identifier	
DOMAIN	Domain Abbreviation	text	DOMAIN	Assigned	Identifier	
ARMCD	Planned Arm Code	text	ARMCD	Assigned	Topic	
ARM	Description of Planned Arm	text	ARM	Assigned	Synonym Qualifier	
TAETORD	Order of Element within Arm	integer		Assigned	Identifier	
ETCD	Element Code	text	ETCD	Assigned	Record Qualifier	
ELEMENT	Description of Element	text	ELEMENT	Assigned	Synonym Qualifier	
TABRANCH	Branch	text		Assigned	Rule	
TATRANS	Transition Rule	text		Assigned	Rule	
EPOCH	Epoch	text	EPOCH	Assigned	Timing	

DM (Demographics)

Demographics (DM)						dm.xpt
Variable	**Label**	**Type**	**Controlled Terminology**	**Origin**	**Role**	**Comment**
STUDYID	Study Identifier	text	STUDYID	Assigned	Identifier	
DOMAIN	Domain Abbreviation	text	DOMAIN	Assigned	Identifier	
USUBJID	Unique Subject Identifier	text		Assigned	Identifier	Protocol ID concatenated with subject ID
SUBJID	Subject Identifier for the Study	text		CRF Page 1	Topic	Site concatenated with collected patid ID for this study only
RFSTDTC	Subject Reference Start Date/Time	text	ISO 8601	Assigned	Record Qualifier	Date of first DOT dose collected on CRF (see EX) for treated subjects, otherwise date first dose scheduled for subjects not treated, otherwise date of enrollment.
RFENDTC	Subject Reference End Date/Time	text	ISO 8601	Assigned	Record Qualifier	Date of last treatment (see EX) for treated subjects, otherwise last date of participation (see DM.RFPENDTC).
RFXSTDTC	Date/Time First Study Treatment Exposure	text	ISO 8601	Derived	Record Qualifier	Date of first DOT dose collected on CRF (see EX).
RFXENDTC	Date/Time Last Study Treatment Exposure	text	ISO 8601	Derived	Record Qualifier	Date of last DOT dose collected on CRF (see EX).
RFICDTC	Date/Time of Informed Consent	text	ISO 8601	Assigned	Record Qualifier	Assumed to be date of enrollment.

Demographics (DM)						dm.xpt
Variable	**Label**	**Type**	**Controlled Terminology**	**Origin**	**Role**	**Comment**
RFPENDTC	Date/Time of End of Participation	text	ISO 8601	Derived	Record Qualifier	Last date of data collection looking across Treatment, Follow-up, AE, and Notification of Death. This takes the latest of dates in VS, MB, LB, and AE. It does not use Permanent Departure form because this only has the date the form was filled out.
DTHDTC	Date of Death	text	ISO 8601	Derived	Record Qualifier	This includes all the deaths known of in the study population. Derived as: AESTDTC where AE.AEOUT=FATAL.
DTHFL	Subject Death Flag	text		Derived	Record Qualifier	Subjects with one or more records with AE.AEOUT=FATAL
SITEID	Study Site Identifier	text		Assigned	Record Qualifier	Sites (CRF Page 1) were grouped into regions as part of the de-identification process; region is reported here.
BRTHDTC	Date/Time of Birth	text	ISO 8601	CRF Page 5	Record Qualifier	
AGE	Age	integer		Derived	Record Qualifier	Age = integer of (informed consent date - birth date + 1)/365.25
AGEU	Age Units	text	DM_AGEU	Assigned	Variable Qualifier	
SEX	Sex	text	SEX	CRF Page 5	Record Qualifier	

Demographics (DM)						**dm.xpt**
Variable	**Label**	**Type**	**Controlled Terminology**	**Origin**	**Role**	**Comment**
RACE	Race	text	RACE	CRF Page 5	Record Qualifier	The collected term Hispanic was assumed to be an ethnicity not a race thus Race for these subjects is coded as Unknown (Hispanic).
ETHNIC	Ethnicity	text	ETHNIC	CRF Page 5	Record Qualifier	Ethnicity was assumed to be non-hispanic for all collected categories of race not identified as Hispanic.
ARMCD	Planned Arm Code	text	ARMCD	Assigned	Record Qualifier	
ARM	Description of Planned Arm	text	ARM	CRF Page 3	Synonym Qualifier	
ACTARMCD	Actual Arm Code	text		Assigned	Record Qualifier	
ACTARM	Actual Arm	text		Derived	Synonym Qualifier	See Computational Method: COMP0001
COUNTRY	Country	text	DM_COUNTRY	Assigned	Record Qualifier	The de-identification algorithm lumped sites together by region and erased the distinction between US and Canadian sites. Most sites are U.S. so all subjects are coded to this.
DMDTC	Date/Time of Collection	text	ISO 8601	CRF Page 1	Timing	

EX (Exposure)

Exposure (EX)						ex.xpt
Variable	**Label**	**Type**	**Controlled Terminology**	**Origin**	**Role**	**Comment**
STUDYID	Study Identifier	text	STUDYID	Assigned	Identifier	
DOMAIN	Domain Abbreviation	text	DOMAIN	Assigned	Identifier	
USUBJID	Unique Subject Identifier	text		Assigned	Identifier	Protocol ID concatenated with subject ID
EXSEQ	Sequence Number	integer		Assigned	Identifier	
EXTRT	Name of Actual Treatment	text	EX_EXTRT	CRF Page 14	Topic	
EXDOSE	Dose per Administration	integer		Protocol	Record Qualifier	Assigned from information found in the protocol by treatment arm.
EXDOSU	Dose Units	text	EX_EXDOSU	Protocol	Variable Qualifier	Assigned from information found in the protocol.
EXDOSFRM	Dose Form	text	EX_EXDOSFRM	Protocol	Record Qualifier	
EXDOSFRQ	Dosing Frequency per Interval	text	EX_EXDOSFRQ	Assigned	Variable Qualifier	Assigned from information found in the protocol by treatment arm.
EXROUTE	Route of Administration	text	EX_EXROUTE	Assigned	Variable Qualifier	
VISITNUM	Visit Number	integer	VISITNUM	Assigned	Timing	
VISIT	Visit Name	text	VISIT	CRF Page 12	Timing	

Exposure (EX)						**ex.xpt**
Variable	**Label**	**Type**	**Controlled Terminology**	**Origin**	**Role**	**Comment**
EPOCH	Epoch	text	EPOCH	Assigned	Timing	Assigned based on planned timing using the data collection form name as a proxy, thus visits are not always nested as expected.
EXSTDTC	Start Date/Time of Treatment	text	ISO 8601	CRF Page 14	Timing	Dosing intervals are for a single DOT dose, thus start and end date are the same.
EXENDTC	End Date/Time of Treatment	text	ISO 8601	CRF Page 14	Timing	Dosing intervals are for a single DOT dose, thus start and end date are the same.

AE (Adverse Events)

Adverse Events (AE)						ae.xpt
Variable	**Label**	**Type**	**Controlled Terminology**	**Origin**	**Role**	**Comment**
STUDYID	Study Identifier	text	STUDYID	Assigned	Identifier	
DOMAIN	Domain Abbreviation	text	DOMAIN	Assigned	Identifier	
USUBJID	Unique Subject Identifier	text		Assigned	Identifier	Protocol ID concatenated with subject ID
AESEQ	Sequence Number	integer		Assigned	Identifier	
AESPID	Sponsor-Defined Identifier	text		Assigned	Identifier	Internal identifier used for tracing terms to source, including splits.
AETERM	Reported Term for the Adverse Event	text		CRF Pages 23, 24, 37	Topic	Adverse event-like data was collected into several data elements and these were mapped as different adverse events. Events on the Notification of Death report are included.
AEMODIFY	Modified Reported Term	text		Assigned	Synonym Qualifier	Used to separate concepts when more than one was included in the same verbatim term.
AELLT	Lowest Level Term	text	MedDRA	Dictionary	Synonym Qualifier	
AELLTCD	Lowest Level Term Code	integer	MedDRA	Dictionary	Synonym Qualifier	
AEDECOD	Dictionary-Derived Term	text	MedDRA	Dictionary	Synonym Qualifier	
AEPTCD	Preferred Term Code	integer	MedDRA	Dictionary	Synonym Qualifier	

Adverse Events (AE)						ae.xpt
Variable	**Label**	**Type**	**Controlled Terminology**	**Origin**	**Role**	**Comment**
AEHLT	High Level Term	text	MedDRA	Dictionary	Synonym Qualifier	
AEHLTCD	High Level Term Code	integer	MedDRA	Dictionary	Synonym Qualifier	
AEHLGT	High Level Group Term	text	MedDRA	Dictionary	Synonym Qualifier	
AEHLGTCD	High Level Group Term Code	integer	MedDRA	Dictionary	Synonym Qualifier	
AEBODSYS	Body System or Organ Class	text	MedDRA	Dictionary	Record Qualifier	
AESOCCD	Primary System Organ Class Code	integer	MedDRA	Dictionary	Synonym Qualifier	
AESER	Serious Event	text	NY	Derived	Record Qualifier	Set top Y if any of these collected qualifiers =Y: life-threatening, results in death, congenital abnormality in a pregnancy on study
AEACN	Action Taken with Study Treatment	text	ACN	CRF Page 26	Record Qualifier	
AEREL	Causality	text	AE_AEREL	CRF Page 26	Record Qualifier	
AEOUT	Outcome of Adverse Event	text	OUT	Assigned and CRF Pages 24, 26	Record Qualifier	Assigned the value of FATAL for events reported on the Notificaiton of Death CRF (Form 13)
AESCONG	Congenital Anomaly or Birth Defect	text	NY	CRF Page 26	Record Qualifier	

Adverse Events (AE)						ae.xpt
Variable	**Label**	**Type**	**Controlled Terminology**	**Origin**	**Role**	**Comment**
AESDTH	Results in Death	text	NY	CRF Pages 24, 26	Record Qualifier	
AESLIFE	Is Life Threatening	text	NY	CRF Page 24	Record Qualifier	
AETOXGR	Standard Toxicity Grade	text	AE_AETOXGR	CRF Pages 24, 26	Record Qualifier	
AETRTEM	Treatment Emergent Flag	text		Derived	Record Qualifier	Assign Y to events with a start date on or before the reference start date.
EPOCH	Epoch	text	EPOCH	Assigned	Timing	Assigned based on planned timing using the data collection form name as a proxy, thus visits are not always nested as expected.
AESTDTC	Start Date/Time of Adverse Event	text	ISO 8601	CRF Page 23	Timing	
AEENDTC	End Date/Time of Adverse Event	text	ISO 8601	Assigned	Timing	Not collected. All null.
AEENRF	End Relative to Reference Period	text	STENRF	Assigned	Timing	Not collected. Timing of resolution is unknown for all.

LB (Laboratory Tests Results)

Laboratory Tests Results (LB)						lb.xpt
Variable	**Label**	**Type**	**Controlled Terminology**	**Origin**	**Role**	**Comment**
STUDYID	Study Identifier	text	STUDYID	Assigned	Identifier	
DOMAIN	Domain Abbreviation	text	DOMAIN	Assigned	Identifier	
USUBJID	Unique Subject Identifier	text		Assigned	Identifier	Protocol ID concatenated with subject ID
LBSEQ	Sequence Number	integer		Assigned	Identifier	
LBTESTCD	Lab Test or Examination Short Name	text	LB_LBTESTCD	Assigned	Topic	
LBTEST	Lab Test or Examination Name	text	LB_LBTEST	Assigned	Synonym Qualifier	Assigned according to the test name associated with a result. See --ORRES for pages with results annotated.
LBCAT	Category for Lab Test	text	LB_LBCAT	Assigned	Grouping Qualifier	
LBORRES	Result or Finding in Original Units	text		CRF Pages 2, 8	Result Qualifier	
LBORRESU	Original Units	text	LB_LBORRESU	CRF Pages 2, 8	Variable Qualifier	
LBORNRLO	Reference Range Lower Limit in Orig Unit	text		Assigned	Variable Qualifier	Textbook normal ranges were applied. Details are available from the sponor upon request.

Laboratory Tests Results (LB)						lb.xpt
Variable	**Label**	**Type**	**Controlled Terminology**	**Origin**	**Role**	**Comment**
LBORNRHI	Reference Range Upper Limit in Orig Unit	text		Assigned	Variable Qualifier	Textbook normal ranges were applied. Details are available from the sponor upon request.
LBSTRESC	Character Result/Finding in Std Format	text		Assigned	Result Qualifier	See value-level metadata. Conversion details are available from the sponsor upon request.
LBSTRESN	Numeric Result/Finding in Standard Units	float		Assigned	Result Qualifier	
LBSTRESU	Standard Units	text	LB_LBSTRESU	Assigned	Variable Qualifier	
LBSTNRLO	Reference Range Lower Limit-Std Units	float		Assigned	Variable Qualifier	Textbook normal ranges were applied. Details are available from the sponor upon request.
LBSTNRHI	Reference Range Upper Limit-Std Units	float		Assigned	Variable Qualifier	Textbook normal ranges were applied. Details are available from the sponor upon request.
LBSTNRC	Reference Range for Char Rslt-Std Units	text		Assigned	Variable Qualifier	
LBNRIND	Reference Range Indicator	text	LB_LBNRIND	Derived	Variable Qualifier	Derived by comparing the standardized result to the assigned normal range.
LBBLFL	Baseline Flag	text	NY	Derived	Record Qualifier	Last nonmissing observation prior to the reference start date.
VISITNUM	Visit Number	integer	VISITNUM	Assigned	Timing	
VISIT	Visit Name	text	VISIT	CRF Pages 2, 8	Timing	

Laboratory Tests Results (LB)						**lb.xpt**
Variable	**Label**	**Type**	**Controlled Terminology**	**Origin**	**Role**	**Comment**
EPOCH	Epoch	text	EPOCH	Assigned	Timing	Assigned based on planned timing using the data collection form name as a proxy, thus visits are not always nested as expected.
LBDTC	Date/Time of Specimen Collection	text	ISO 8601	CRF Pages 2, 8	Timing	

SUPPEX (Supplemental Qualifiers for EX)

Supplemental Qualifiers for EX (SUPPEX)						suppex.xpt
Variable	**Label**	**Type**	**Controlled Terminology**	**Origin**	**Role**	**Comment**
STUDYID	Study Identifier	text	STUDYID	Assigned	Identifier	
RDOMAIN	Related Domain Abbreviation	text	DOMAIN	Assigned	Identifier	
USUBJID	Unique Subject Identifier	text		Assigned	Identifier	Protocol ID concatenated with subject ID
IDVAR	Identifying Variable	text	IDVAR	Assigned	Identifier	
IDVARVAL	Identifying Variable Value	text		Assigned	Identifier	
QNAM	Qualifier Variable Name	text	SUPPEX_QNAM	Assigned	Topic	
QLABEL	Qualifier Variable Label	text	SUPPEX_QLABEL	Assigned	Synonym Qualifier	
QVAL	Data Value	text		Assigned and CRF Pages	Topic Value	See value-level metadata
QORIG	Origin	text		Assigned	Variable Qualifier	
QEVAL	Evaluator	text	EVAL	Assigned	Record Qualifier	

Value Level Metadata

Laboratory Tests Results (LB) Value List						
Variable	**Value**	**Label**	**Type**	**Controlled Terminology**	**Origin**	**Comment**
LBTESTCD	ALP	Alkaline Phosphatase	float		CRF Page 8	
LBTESTCD	ALT	Alanine Aminotransferase	integer		CRF Page 8	
LBTESTCD	AST	Aspartate Aminotransferase	integer		CRF Page 8	
LBTESTCD	BILI	Bilirubin	float		CRF Page 8	
LBTESTCD	CD4	CD4	integer		CRF Page 8	
LBTESTCD	CREAT	Creatinine	integer		CRF Page 8	
LBTESTCD	HCG	Choriogonadotropin Beta	text	LB_LBTESTCD_HCG	CRF Page 8	
LBTESTCD	HCT	Hematocrit	float		CRF Page 8	
LBTESTCD	HGB	Hemoglobin	float		CRF Page 8	
LBTESTCD	HIV12AB	Hiv-1/2 Antibody	text	LB_LBTESTCD_HIV12AB	CRF Page 8	
LBTESTCD	LYM	Lymphocytes	integer		CRF Page 8	
LBTESTCD	LYMLE	Lymphocytes/Leukocytes	float		CRF Page 8	
LBTESTCD	PLAT	Platelet	integer		CRF Page 8	
LBTESTCD	WBC	Leukocytes	float		CRF Page 8	

Supplemental Qualifiers for EX (SUPPEX) Value List						
Variable	**Value**	**Label**	**Type**	**Controlled Terminology**	**Origin**	**Comment**
QNAM	DOS_MODI	Dosage Modified for Weight	text	NY	CRF Page 14	
QNAM	MISSEDRX	Patient Missed DOT Visit	text	NY	CRF Page 14	
QNAM	MISSOTSP	Other Missed/Modified Doses Specify	text		CRF Page 14	
QNAM	NEWDRUG	Specify New Drug Dosage	text	SUPPEX_QNAM_NEW DRUG	CRF Page 14	
QNAM	REDU_TOX	Dose Held or Reduced for Toxicity	text	NY	CRF Page 14	
QNAM	STUDY_RX	Study Drug Status	text	NY	CRF Page 14	
QNAM	UNAVLDOT	Not Available for DOT, Self Admin	text	NY	CRF Page 14	

Computation Methods

Computational Method	
Reference Name	**Computational Method**
COMP0001	Derived from dosing information on CRF Page 14. If received only Rifampin or received only Rifapentine then assign one of those treatments and assume it was given as planned. If received both Rifampin and Rifapentine then assign both. If no doses were recorded then assign NOT TREATED.

Controlled Terminology

CodeLists

Reference Name	Code Value	Code Text
ACN	DOSE INCREASED	DOSE INCREASED
	DOSE NOT CHANGED	DOSE NOT CHANGED
	DOSE REDUCED	DOSE REDUCED
	DRUG INTERRUPTED	DRUG INTERRUPTED
	DRUG WITHDRAWN	DRUG WITHDRAWN
	NOT APPLICABLE	NOT APPLICABLE
	UNKNOWN	UNKNOWN
AE_AEREL	NOT RELATED	NOT RELATED
	POSSIBLY RELATED	POSSIBLY RELATED
	PROBABLY RELATED	PROBABLY RELATED
	UNKNOWN	UNKNOWN
AE_AETOXGR	1	1
	2	2
	3	3

Reference Name	Code Value	Code Text
	4	4
	5	5
	U	U
ARM	RIFAMPIN + INH BIS	RIFBIS
	RIFAPENTINE + INH QS	RPTQS
ARMCD	RIFBIS	RIFAMPIN + INH BIS
	RPTQS	RIFAPENTINE + INH QS
DM_AGEU	YEARS	YEARS
DM_COUNTRY	USA	USA

< For illustration purposes more codelist have been removed. >

External Dictionaries

Reference Name	Dictionary	Version
MedDRA	MedDRA	12.0

Appendix B. List of Useful RTF code

Category	RTF code	Explanation
Command	\ab4, example: \fs22 (font size 11 point.)	This is a command. It starts with backslash, some lowercase letter, maybe an integer (may have a negative sign)
Command	{\pard … par}	Paragraph
Document Prolog	{\rtf1	RTF version 1
Document Prolog	\ansi	document is in ANSI character set.
Document Prolog	\deffN	declares font number N is default font for this document
Font table	\fmodern	monospace font
Font table	\froman	proportionally spaced serif font
Font table	\fswiss	proportionally spaced sans serif font
Font table	\fnil	Unknown/other
Color table	{colortbl; … declaration…}	
Paragraph	\ql	Left justified
Paragraph	\qr	Right justified
Paragraph	\qj	Full justified
Paragraph	\qc	Center justified
Paragraph	\txN, example: \tx6120	Tab stop at location N (4.25 inches)
Paragraph	\tqc	Center around tab
Paragraph	\tql	Left align at tab
Paragraph	\tqr	Right align at tab
Paragraph	\sbN, example: \sb180	Add N twips of vertical space before the paragraph
Paragraph	\saN, example: \sa180	Add Nwips of vertical space after the paragraph
Paragraph	\fiN, example \fi720	Indent the first line of this paragraph by 720 twips (0.5 inch).
Paragraph	\liN	block indentation from left margin.
Paragraph	\riN	block indentation from right margin.
Paragraph	\pagebb	Put a page break before the paragraph
Paragraph	\keep	keep the paragraph in one piece.
Paragraph	\keepn	keep together with the next (following) paragraph.
Paragraph	\widctlpar	widow-and-orphans control for this paragraph (antonym: \nowidctlpar)
Paragraph	\nowidctlpar	No widow-and-orphans control for this paragraph
Paragraph	\hyphpar	

Appendix B. List of Useful RTF code

Category	RTF code	Explanation
Paragraph	\hyphpar0	
Paragraph	\sl360\slmult1	1.5-spacing between lines of text
Paragraph	\sl480\slmult1	double=space between lines of text
Character command	\i	italics
Character command	\b	bold
Character command	\ul, example: \ul TEST (underlining TEST)	Underlining a string that follows \ul.
Character command	\super	superscript
Character command	\sub	subscript
Character command	\scaps	smallcaps
Character command	\strike	strike through
Character command	\\	a backslash, (same as \'5c)
Character command	\{	an open-brace (same as \'7b)
Character command	\}	an close-brace (same as \'7d)
Character command	\bullet	bullet character (same as Latin-1 character 149)
Character command	\endash	n-width dash
Character command	\emdash	m-width dash
Character command	\enspace	n-width non-breaking space
Character command	\emspace	m-width non-breaking space
Character command	\lquote	single openquote (6)
Character command	\rquote	single closequote (9)
Character command	\ldblquote	double openquote (66)
Character command	\rdblquote	double closequote (99)

Appendix B. List of Useful RTF code

Category	RTF code	Explanation
Character command	\sect\sectd	new section. (Resets header and columnation.)
Character command	\ftnbj\ftnrestart	initialize footnote numbering
Document formatting	\widowctrl	turn on widows-and-orphans control for the document
Document formatting	\pgnstartN	for page numbering, set first page to N
Document formatting	\landscape	document is in landscape format
escapes	\'xy, example: \'ea (an ê character)	backslash, two hex digits
escapes	\'f1 (a ñ character)	unicode 241 = f1 ([15(f)*16+1] hexdecimal), between unicode 1(01) to 255 (ff = 15(f)*16+15(f)).
escapes	\uc1\unumber*, example: \uc1\u21487*(可)	between 256 to 32767
escapes	\uc1\u-number*, example: \uc1\u-28589* (道)	between 32768 (=8*16*16*16) to 65535 (=16*16*16*16 - 1), express 65536 - number
escapes	\~	non-breaking space
escapes	\-	hyphenation point
escapes	_	non-breaking hyphen
groups	{ anything}, example: {\i Hi There!} [*Hi There! (Italicized)*]	Whatever between {}.
Info group	{\title XX}	In the document property title as XX
Info group	{\author XX}	Author's name in the document property
Info group	{company YY}	Company name in the document property.
Info group	{doccomm MMMMM	comments in the document property.
Preliminaries	\deflangN, example: \deflang1033 (US English)	set document default language
Preliminaries	\colN	set N columns per page
Preliminaries	\linebetcol	draw a line between columns
Preliminaries	\colsx1440	make columns 1 inch apart.
Preliminaries	\fntbj	footnote bottom justified
plaintext	{\cb5 yow!}	Highlighting for text "yow!" in light yellow, if color number 5 is defined as light yellow in the color table.
Change text color	{\cf2 yow!}	Color text "yow!" in blue, if color number 2 is defined as blue in the color table.

Appendix B. List of Useful RTF code

Category	RTF code	Explanation
Hyperlink	{\field{*\fldinst HYPERLINK "link_name"}{fldrslt{\ul\cs15\cf2 link display name}}}	{\field {*\fldinst HYPERLINK \\l "hyperlink name"}{\fldrslt {\ul\cs15\cf2 link display name}}} Go to bookmark "hyperlink name" in this document (\\l) when mouse over the text "link display name", which is in blue color and underlined (color style number 15, foreground color number 2).
Bookmark	{*\bkmkstart book-mark-name} bookmarktext {*\bkmkend book-mark-name}	Bookmark the *bookmarktext* in a RTF document
Margin	\margtN, example: \margt1440 (1 inch top margin)	top margin N twips
Margin	\margbN, example: \margb1440 (1 inch bottom margin)	bottom margin N twips
Margin	\marglN, example: \margl1440 (1 inch left margin)	left margin N twips
Margin	\margrN, example: \margr1440 (1 inch right margin)	right margin N twips
Line drawing	{\pard \brdrb \brdrs \brdrw10 \brsp20 \par} , example: {\pard \li2268 \ri567 \brdrb \brdrs \brdrw10 \brsp20 \par} (draw a horizontal line that starts 4 cm (= 2268 twips) from the left margin and ends 1 cm (= 567 twips) from the right margin.)	draw a horizontal line between left and right margins
underline	\brdrb\brdrs\brdrw1	
Newline	\line	not a real paragraph break
tab	\tab	better than using a literal tab character
character formatting	\plain, example: {\header\pard\qr\plain\f0\chpgn\par} (This turns on page numbering in header area.)	turn off all formatting
Pagebreak	\page	pagebreak