Importing Data Specifications from .RTF and .DOC files and producing Reports
Sandesh Jagadeesh, PPD, Austin, TX

ABSTRACT

Programming reports that are used in the process of data cleansing and patient level checks can be tedious. These reports are of great value to data management teams, providing in-depth information needed to identify adverse events, compare variables against database standards, and to reconcile data. The report specifications can be provided in .RTF or.DOC format, which would require manual programming to generate the desired reports. The number of variables in these reports can be quite extensive and time-consuming to review and program. This paper will introduce a method for importing the required data sets, variables and labels directly from the specification, reconciling them to the variables in the database, and producing reports in .PDF, .RTF, or .XLS format for review.

INTRODUCTION

Flow chart below describes the process. The conventional process of receiving the specification in .RTF, .DOC and manually programming the variables and labels in SAS and generating outputs
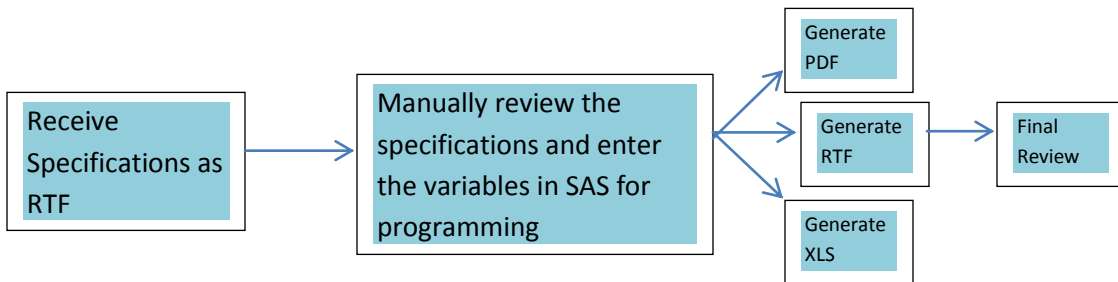
F.1 Flow chart of conventional data cleansing process

Flow chart below shows the additional step after importing the specifications and checking against the database, yielding high quality checks and data for review.
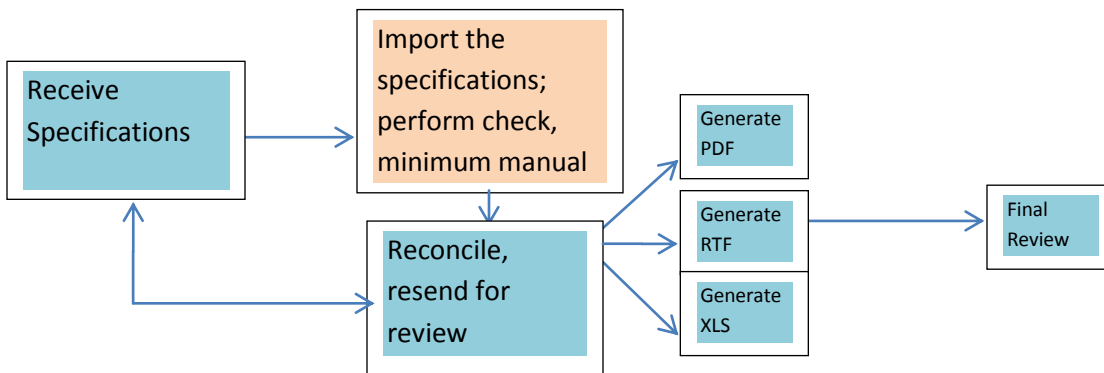
F.2 Flow chart with the automated macro minimizing manual process

The purpose of this paper is to convert a specification provided into a table so that it can be read as a data set, obtaining the variables required for querying. The following screenshots illustrates this

Order	Data Source	Variable Names	Special Column Headings	Criteria	Compare var (X)	Count key (X)	Key (X)	/Dataset Source	Additional Criteria (Programmer input)
1.	ANNNN	PT	Patient ID	**Required field**		N/A		AEAE	Not displayed
2.	ANNNN	CNTRY	Country	**Required field**	X	N/A		AEAE	
3.	ANNNN	CENTER	Site	**Required field**	X	N/A		AEAE	
4.	ANNNN	DDDDD	DREWEF	**Required field**	X	N/A		AEAE	

F.3 Sample of specifications in .RTF/.DOC format

col2	col3	col4	col5	col6	col7
DataSource	VariableNames	SpecialColumnHeadings	Criteria	Comparevar (X)	Countke
ANNNN	PT	Patient ID	Required field		N/A
ANNNN	CNTRY	Country	Required field	X	N/A
ANNNN	CENTER	Site	Required field	X	N/A
ANNNN	DDDDD	DREWEF	Required field	X	N/A

F.4 Sample of intermediate metadata converted from .RTF/.DOC

After we have the required domains and the variables for quering from the metadata, it is merged with the domains in the database for a quality check to make sure the specifications match the domains and variables in the database. This can be done with the help of PROC CONTENTS.

The next step involves generating the report.

Record Status
Patient ID   Document Number   CRF #   Change Status

Subject Characteristics
Country   Site   Birth Date   Gender

F.6 Sample of the output generated

DETAILS OF THE PROCESS

The aim of this paper is to automate the process of report creation with minimum manual process. Earlier, the process of report generation heavily involved manual inputting of the specification into the SAS code. The following points describe how exactly this can be achieved.

1. The Specifications is received in .DOC/.RTF format as shown in illustration F.3

Order	Data Source	Variable Names	Special Column Headings	Criteria	Compare var (X)	Count key (X)	Key (X)	/Dataset Source	Additional Criteria (Programmer input)
1.	ANNNN	PT	Patient ID	**Required field**		N/A		AEAE	Not displayed
2.	ANNNN	CNTRY	Country	**Required field**	X	N/A		AEAE	
3.	ANNNN	CENTER	Site	**Required field**	X	N/A		AEAE	
4.	ANNNN	DDDDD	DREWEF	**Required field**	X	N/A		AEAE	

2. The .RTF/.DOC is read using a simple infile statement[1]:

```
%let basefil="                                    \Order.rtf";
data readrtf_a;
        infile &basefil missover length = l end = lastobs lrecl = 5000;
        input string $varying2000. l;
        rownum = _n_;
        run;
```

3. This would generate the following data in SAS when parsing .RTF/.DOC

4. After extracting the required data, we get the intermediate output in SAS which mimics the specification[2].

col2	col3	col4	col5	col6	col7
DataSource	VariableNames	SpecialColumnHeadings	Criteria	Comparevar (X)	Countke
ANNNN	PT	Patient ID	Required field		N/A
ANNNN	CNTRY	Country	Required field	X	N/A
ANNNN	CENTER	Site	Required field	X	N/A
ANNNN	DDDDD	DREWEF	Required field	X	N/A

5. Further macro processing involves merging the above variables with the database for a quality check and labeling them based on specifications, producing the output below

Record Status

Patient ID	Document Number	CRF #	Change Status

Subject Characteristics

Country	Site	Birth Date	Gender

OTHER TECHNIQUES AND FUTURE WORK

Besides the above method, .RTF/.DOC can be imported using one of the following [3]:
1. Microsoft Libname Engines
2. SAS PROC IMPORT and EXPORT
3. ODBC (Oracle Database Connection)
4. ODS (Output Delivery System)
5. DDE (Dynamic Data Exchange)
6. Microsoft Office Plug In
7. Third Party Software

- Future work involves using the specifications for creating dummy text, which can be merged to get the reports.
- Dynamically extracting SAS statements from the comments section of specifications will further add to the automation.

CONCLUSION

"We live in a time when automation is ushering in a second industrial revolution"[4]. The report specifications which are provided in .RTF or.DOC can be automated to be imported into SAS and converted to an intermediate metadata which mimics the specifications and reconciling them with the variables in the database, and producing reports in .PDF, .RTF, or .XLS format for review. This would reduce errors, improve efficiency and time required for reviews.

REFERENCES

[1] PharmaSUG 2014 Paper BB14. A SAS® Macro Utility to Modify and Validate RTF Outputs for Regional Analyses Jagan Mohan Achi, PPD, Austin, TX, Joshua N. Winters, PPD Rochester, NY. http://www.pharmasug.org/proceedings/2014/BB/PharmaSUG-2014-BB14.pdf

[2] PharmaSUG2010 - Paper TT13 %rtf2data : A utility macro to convert RTF Table to SAS dataset Neeral Beladia, MaxisIT Inc., Metuchen, NJ . Scott Osowski. Thomas Fritchey (2006): Hyperlinks and Bookmarks with ODS RTF Microsoft® Rich Text Format(RTF) Specification, RTF v1.7. http://www.pharmasug.org/proceedings/2014/BB/PharmaSUG-2014-BB14.pdf

[3] Linking SAS and Microsoft Products Presented to The Virginia SAS® User's Group Nat Wooding, Dominion Virginia Power June 17, 2009

[4] Adlai E. Stevenson. (American Politician, Governor of Illinois (1949-53) and Ambassador to the United Nations).

ACKNOWLEDGMENTS

Thanks to Ken Borowiak, David Gray, Rama Boorlagadda, Heather Sullivan for their reviews and comments.

DISCLAIMER

The content of this paper are the works of the authors and do not necessarily represent the opinions, recommendations, or practices of PPD.

CONTACT INFORMATION

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

Name: Sandesh Jagadeesh
Enterprise: PPD
Address: 7551 Metro Center Drive, Suite 300
City, State ZIP: Austin, Texas 78744
Work Phone: 512-747-5923
E-mail: Sandesh.Jagadeesh@ppdi.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.