

Linking Laboratory Data to Submission Documents Using SAS® Technologies

Dongmin Shen and Miu Ling Lau
Merck & Co, Inc., 556 Morris Ave, Summit, New Jersey, USA

Introduction

Merck is a global pharmaceutical company and so the sources of our data are global. Having the ability to link and transfer massive amounts of analytical data from various data sources into submission documents in an efficient and reproducible way is critical to producing successful regulatory submissions. SAS® technologies have been used to create various solutions ranging from data extractions, to data transformations, to documents generated in support of simultaneous worldwide new drug applications.

1. Connecting Laboratory Data to Submission Document

Capabilities have been developed using SAS® technologies, which enable the connection of laboratory data (across the globe) to our submission documents, generating standard outputs with the flexibility for customization within the different submissions/regions. Using electronic systems instead of manual documents allows us to gain efficiency, reproducibility and enhanced compliance (i.e. reduction in transcription errors). These capabilities include:

- Extracting data from various data sources, e.g. Laboratory Information Management System (LIMS) databases, using validated SAS macros
- Performing data transformations into standard formats that can be rendered into common document templates (CDT)
- Creating data tables in the CDT format, e.g. stability data tables, batch analysis tables
- Performing statistical analyses, e.g. regression analyses for shelf-life estimates
- Facilitating data requests for pre-approval inspection (PAI), auditing and specification setting.

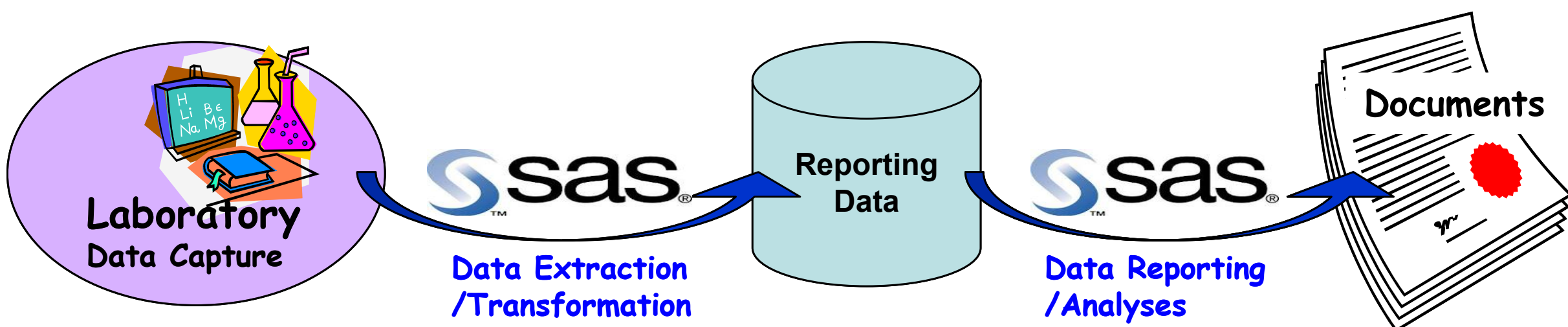


Figure 1. Connecting laboratory data to submission document using SAS technologies

2. Data Extraction and Transformation

SAS macro applications have been developed to perform data extractions and/or transformations. The high level program flow of one of the LIMS database extraction processes is illustrated in Figure 2. SAS pass-through SQL to Oracle databases is one of the key technologies applied for data extraction. Standard SAS macros have been developed to transform extracted laboratory data into standard data structures for data table generation.

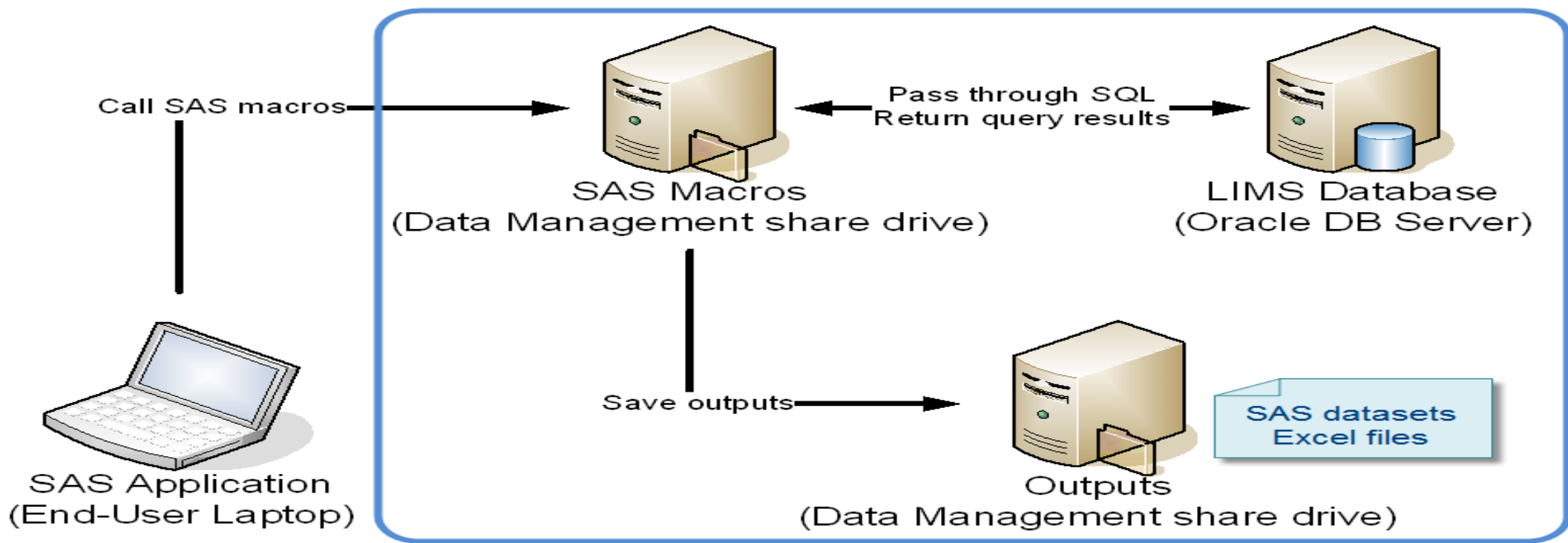


Figure 2. Validated SAS macros for LIMS data extraction

3. Document Generation

Figure 3 shows one example of the standard output generated by the SAS applications developed. The table title, batch information section, data section and footnotes are controlled by over 100 SAS macro parameters, some of which are highlighted in Figure 4. The table title is controlled by a few parameters, *tbl_no_auto* parameter (providing an option to turn the table number between auto and non-auto numbers), *tbl_no* parameter (linking input dataset variable *page* to table numbers), and *tbl_title* parameter (allowing one to assign table captions). Most of the macro parameters are designed with flexibility to allow mixed assignments of multiple dataset variables, e.g., *product*, *strength*, *batch* and *storage*, associated variable formats (e.g., *\$storage*) and character strings (e.g., 'Stability Data for Batch' in Figures 3 and 4). The batch information section can be customized via several parameters, each one of which is mapped to an associated cell position and can be assigned to various dataset variables, formats and strings. The number of columns under Time Point section is dynamically rendered by variables from the input dataset for each table, based on time points from the stability protocol for a given batch, condition and orientation. In addition, the time points can be formatted if needed (e.g. time zero is formatted to 'Initial'). The tests under Analysis column and test results in data section are automatically rendered from data in the input dataset and can be varied from table to table. The levels of identifiers for test results (listed under the analysis column) are controlled by data model parameters of the SAS macro in Figure 4. For example, 'Appearance' and 'Assay' tests assigned to parameter *model1* display one level of identifiers, 'Dissolution' associated with *model2* provides the second level of identifiers, e.g. '10 min, Avg(Max-Min)' and 'Degradates' using *model3* allows three levels of identifiers. The content of the footnote section is fed through an input dataset and can be varied from table to table.

Table 1 ACB 25 mg/tablet Stability Data for Batch M1234567, at 30°C/75% RH									
B	Batch Size	125,000 tablets		Package Container	75 mL White Opaque HDPE Bottle / Alkan		Drug Substance	ABC	
	Site of Manufacture	New Jersey, USA		Package Closure	White CR Plastic Cap, with Foil Induction Seal		Drug Substance Batch No	B12345	
	Date of Manufacture	25-JAN-2010		Packaging Size	60 Tablets		Site of Manufacture	New Jersey, USA	
	Study Start Date	31-MAR-2010							
	Time Point (Months)								
C		Initial	3	6	9	12	18	24	36
E	Analysis	CONFORMS	CONFORMS	CONFORMS	CONFORMS	CONFORMS	CONFORMS	Scheduled	Scheduled
	Assay (% Claim)	99.4	99.8	99.8	97.6	98.6	97.6	Scheduled	Scheduled
	Degradates (% Claim) Specified:	-	-	-	-	-	-	Scheduled	Scheduled
	Deg A	-	0.1	0.1	0.1	0.1	0.2	Scheduled	Scheduled
	Deg B	-	-	-	-	-	-	Scheduled	Scheduled
	Unspecified	-	-	-	-	-	-	Scheduled	Scheduled
	RRT+1.2	-	-	-	0.1	0.1	0.1	Scheduled	Scheduled
	Total Degradates (% Claim)	-	0.1	0.1	0.2	0.2	0.3	Scheduled	Scheduled
	Dissolution (%)	84 (89 – 79)	85 (90 – 81)	82 (88 – 76)	83 (88 – 78)	78 (81 – 76)	78 (81 – 76)	Scheduled	Scheduled
	10 min, Avg(Max-Min)	100 (104 – 96)	100 (103 – 98)	100 (103 – 98)	100 (102 – 96)	100 (104 – 98)	100 (101 – 97)	Scheduled	Scheduled
20 min, Avg(Max-Min)	100 (105 – 97)	99 (102 – 97)	100 (102 – 98)	100 (102 – 96)	100 (103 – 98)	100 (101 – 96)	Scheduled	Scheduled	
30 min, Avg(Max-Min)	100 (105 – 98)	100 (102 – 98)	100 (103 – 98)	100 (102 – 96)	100 (103 – 98)	100 (100 – 98)	Scheduled	Scheduled	
45 min, Avg(Max-Min)	100 (105 – 98)	100 (102 – 98)	100 (103 – 98)	100 (102 – 96)	100 (103 – 98)	100 (100 – 98)	Scheduled	Scheduled	
60 min, Avg(Max-Min)	99 (103 – 96)	99 (103 – 98)	100 (103 – 98)	100 (102 – 96)	100 (103 – 98)	100 (101 – 98)	Scheduled	Scheduled	
Moisture-KF (%)	0.8	0.8	0.8	0.7	0.9	1.1	Scheduled	Scheduled	
Microbial Limits	CONFORMS	NS	NS	NS	CONFORMS	NS	Scheduled	Scheduled	
NT = Not Tested, NS = Not Scheduled, - = Below Reporting Threshold (<0.1%)									

Figure 3. Standard stability data table generated by SAS macros

```
%create_tables (  
  indsn = mydata, sortby = Page Product Strength Batch Orientation Storage Test_Seq  
  filename = &project.\Reports\Stability Tables &sysdate9.,  
  filetype = rtf, fontsize = 8, orient = L, spacing = 0, fnotes = Y,  
  tbl_title = product, strength 'Stability Data for Batch ' batch ', at ' storage Storage.,  
  tbl_no auto = Y, tbl_no = page,  
  b1c1 = 'Batch Size ' , b1c2 = BATCH_SIZE,  
  b2c1 = 'Site of Manufacture ' , b2c2 = MFG SITE $site.,  
  model1 = 'APPEARANCE ' ASSAY ' TOTAL_DRUG ' MOISTURE ' MLT',  
  model2 = 'DISSOLUTION', model3 = 'DEGRADATES',  
  h1c2 = 'Time Point (months)', no_of_cols = Scheduled,  
  h2c4 = T1 TIMEPMT., h2c5 = T2 TIMEPMT., h2c6 = T3 TIMEPMT., h2c7 = T4 TIMEPMT.,  
  h2c8 = T5 TIMEPMT., h2c13 = T10 TIMEPMT.,  
  datacla = Test Description, dataclaf = $STDESCR., dataclb = group, dataclbf = group.,  
  dataclc = Reported Name, dataclf = $RPTNAME.,  
  datacl4 = R1, datacl5 = R2, datacl6 = R3, datacl7 = R4, datacl13 = R10,  
  ) ;
```

Figure 4. Example of a SAS macro and its parameters used for generating data tables

4. Key Techniques

SAS data _null_ step is utilized to render information from input parameters and datasets to output documents in RTF format. The definition of a RTF file including various styles, formats and document template is stored in a SAS catalog as a SAS source file; which is then incorporated in an output RTF file using *infile* statement. This technique avoids writing thousands of *put* statements for RTF definitions and therefore simplifies SAS programs significantly.

```
data _null_;  
  file "&outfile" lrecl=30000;  
  infile "&mac_lib..sasmacr.rtfinit.source"  
  catalog end=__done missover length=len;  
  ...  
;
```

The data from input datasets are converted into RTF codes and output into RTF files using *put* statements. An example code for generating a table title is shown below (%*cellstr* and %*cellstp* macros define a table cell with a style of 'S34' and a given font size). The *put* statements are used to generate RTF codes for the table title with a bookmark and an auto table number field, as well as to render the macro parameters, *tbl_no* and *tbl_title*.

```
%cellstr(fontsize=fontsize, style=\s34 );  
  put "&outlinelevel0 {\*\bkmkstart Table" &tbl_no "}"  
  @;  
  
  put "{\b Table {\field{\*\fldinst { SEQ Table \\  
  ARABIC \* MERGEFORMAT }}" @;  
  
  put "{\fldrslt {\lang1024\langfel024\noproof"  
  &tbl_no "}}}\tab" &tbl_title @;  
  
  put "{\*\bkmkend Table" &tbl_no "}" @;  
%cellstp;
```

Conclusions

Various solutions/applications have successfully been developed using SAS® technologies to link laboratory data to submission documents in support of simultaneous worldwide submissions. The electronic processes increase efficiency and reproducibility of document generations, while avoiding manual data transcription errors, and reducing overall compliance risk.

Contact Information

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

Dongmin Shen
dong-min.shen@merck.com

Miu Ling Lau
miu-ling.lau@merck.com

Merck & Co. Inc
556 Morris Avenue, Summit, NJ 07922, USA

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.

