PharmaSUG China 2018 - Paper 73

Best Practices for E2E DB build process and Efficiency on CDASH to SDTM data

Tao Yang, FMD K&L, Nanjing, China

Introduction of each phase of the trial

It is known to all that project management in clinical trials can be generally categorized into the following stages: study design, data management and data submission. The focus of different departments varied in each stage. For example, during project launch phase, DM, STST, Physician, MA and CO will all participate in the protocol and CRF design, whereas the focus of Medical is medical events, and Statistician will develop Statistical Analysis Plan. During project implementation, clinical trial data is interpreted from entirely different perspectives by different departments. For example, DM reviews data from data integrity perspective, while the SAS programming department analyzes data in standard format. In addition, in the implementation of the project, the support of various software is essential, to improve the efficiency of the process. The integrated application of SAS tools in the whole cycle of a project is described in the paper.

Procedure of each phase

From database design at the database development phase, to the data review during data management phase, and then standardization in the data submission process, SAS is always one of the most critical and useful tool, to ensure the high-quality result and high-efficiency operation from its powerful functions throughout the whole lifecycle of a clinical trial. Here is an introduction to our best practices:

Database Design Phase

Firstly, we need to understand the basic process of database design:

1. Specify the requirements

2. Design the database system according to the requirements

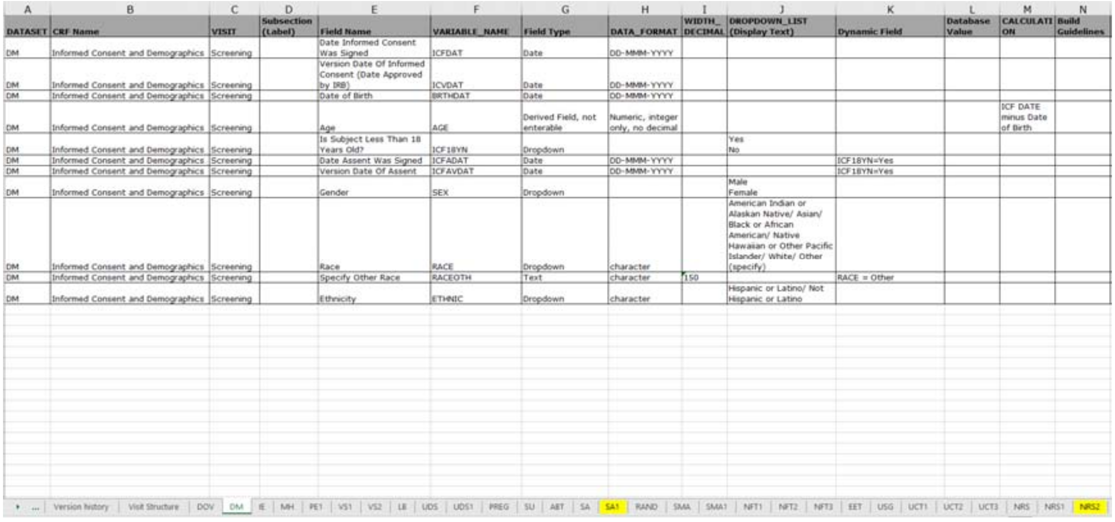
3. Determine consistency between the results of design and the original requirements.

In these processes, requirement changes may result in a series of other alternations.

In step1 and step 2, the operation of different database systems is different. Let’s take the Open source EDC as an example. The typical database design approach is to manually copy the eCRF requirements from raw aCRFs or eCRF specification workbook and manually paste to a machine-readable XML file template by Database Designers, and upload to the EDC. The manual process can be time-consuming with human mistakes. However, the manual transfer approach can be replaced by using SAS With SAS Macro 1, the requirements of Step 1 can be directly converted into the system file of Step 2, and then Step 3, i.e., consistency comparison is done with SAS Macro 2, and relevant prompts will be generated.

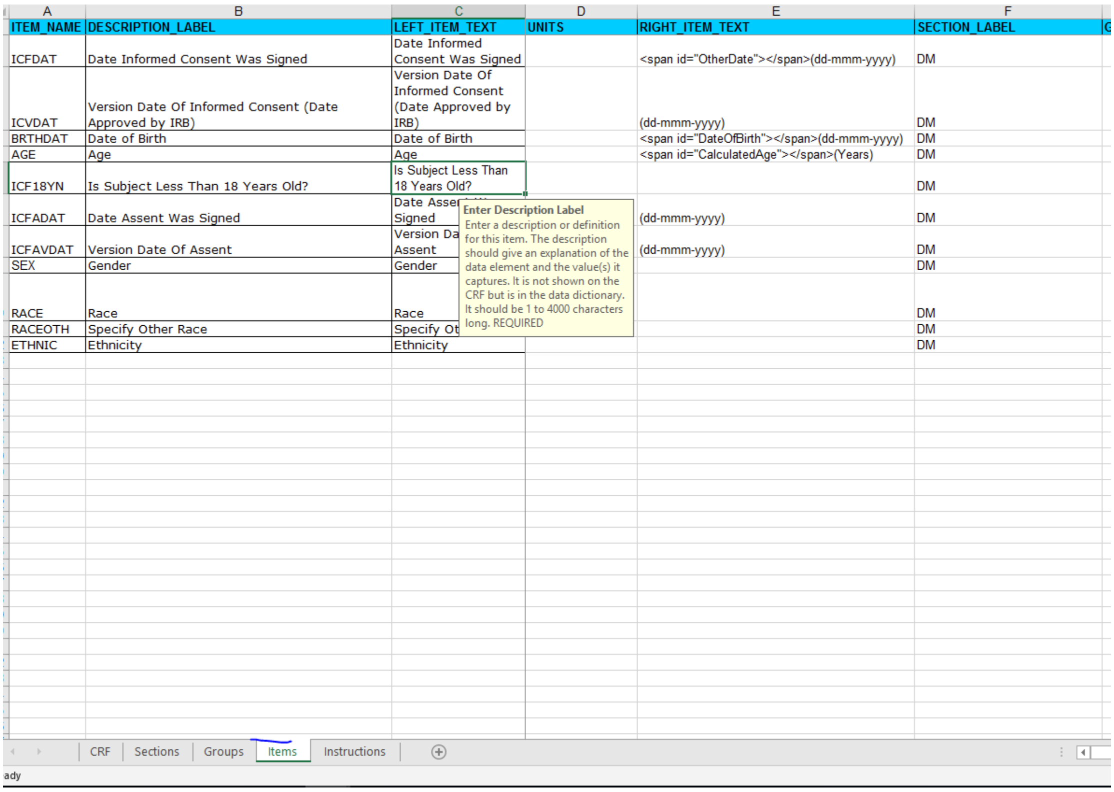
Specific steps of implementation:

Step 1, Original Request (Spec)



DATASET	CRF Name	VISIT	Subsection (Label)	Field Name	VARIABLE_NAME	Field Type	DATA_FORMAT	WIDTH_ DECIMAL	DROPDOWN_LIST (Display Text)	Dynamic Field	Database Value	CALCULATION	Build Guidelines
DM	Informed Consent and Demographics	Screening		Date Informed Consent Was Signed	ICFDAT	Date	DD-MMM-YYYY						
DM	Informed Consent and Demographics	Screening		Version Date Of Informed Consent (Date Approved by IRB)	ICVDAT	Date	DD-MMM-YYYY						
DM	Informed Consent and Demographics	Screening		Date of Birth	BRTHDAT	Date	DD-MMM-YYYY						
DM	Informed Consent and Demographics	Screening		Age	AGE	Derived Field, not enterable	Numeric, integer only, no decimal					ICF DATE minus Date of Birth	
DM	Informed Consent and Demographics	Screening		Is Subject Less Than 18 Years Old?	ICF18YN	Dropdown			Yes No				
DM	Informed Consent and Demographics	Screening		Date Assent Was Signed	ICFADAT	Date	DD-MMM-YYYY			ICF18YN=Yes			
DM	Informed Consent and Demographics	Screening		Version Date Of Assent	ICFAVDAT	Date	DD-MMM-YYYY			ICF18YN=Yes			
DM	Informed Consent and Demographics	Screening		Gender	SEX	Dropdown			Male Female				
DM	Informed Consent and Demographics	Screening		Race	RACE	Dropdown	character		American Indian or Alaskan Native/ Asian/ Black or African American/ Native Hawaiian or Other Pacific Islander/ White/ Other (specify)				
DM	Informed Consent and Demographics	Screening		Specify Other Race	RACEOTH	Text	character	150		RACE = Other			
DM	Informed Consent and Demographics	Screening		Ethnicity	ETHNIC	Dropdown	character		Hispanic or Latino/ Not Hispanic or Latino				

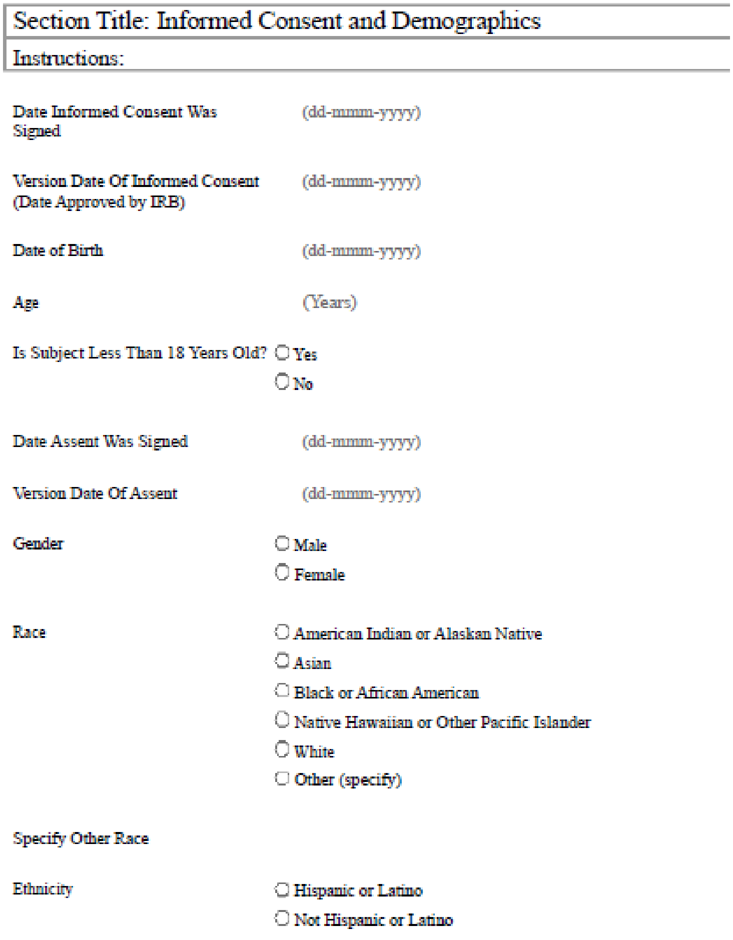
Step 2, Upload file which will be used in EDC system



ITEM_NAME	DESCRIPTION_LABEL	LEFT_ITEM_TEXT	UNITS	RIGHT_ITEM_TEXT	SECTION_LABEL
ICFDAT	Date Informed Consent Was Signed	Date Informed Consent Was Signed		<span id="OtherDate"></span>(dd-mmm-yyyy)	DM
ICVDAT	Version Date Of Informed Consent (Date Approved by IRB)	Version Date Of Informed Consent (Date Approved by IRB)		(dd-mmm-yyyy)	DM
BRTHDAT	Date of Birth	Date of Birth		<span id="DateOfBirth"></span>(dd-mmm-yyyy)	DM
AGE	Age	Age		<span id="CalculatedAge"></span>(Years)	DM
ICF18YN	Is Subject Less Than 18 Years Old?	Is Subject Less Than 18 Years Old?			DM
ICFADAT	Date Assent Was Signed	Date Asser Signed		(dd-mmm-yyyy)	DM
ICFAVDAT	Version Date Of Assent	Version Da Assent		(dd-mmm-yyyy)	DM
SEX	Gender	Gender			DM
RACE	Race	Race			DM
RACEOTH	Specify Other Race	Specify Ot			DM
ETHNIC	Ethnicity	Ethnicity			DM

Enter Description Label
Enter a description or definition for this item. The description should give an explanation of the data element and the value(s) it captures. It is not shown on the CRF but is in the data dictionary. It should be 1 to 4000 characters long. REQUIRED

Step 3, eCRF

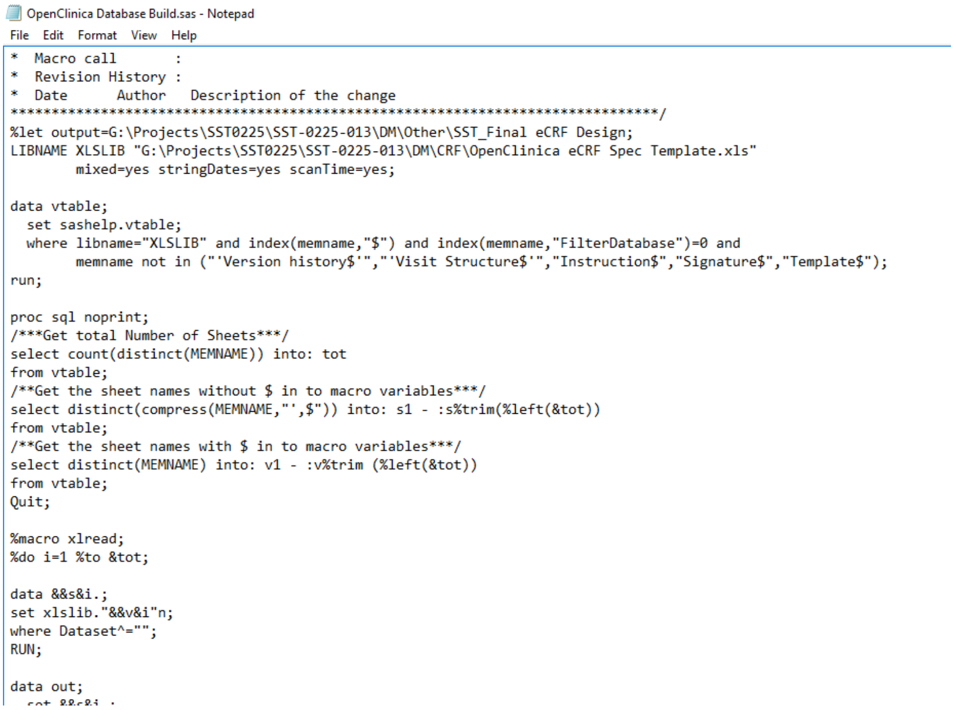


Section Title: Informed Consent and Demographics

Instructions:

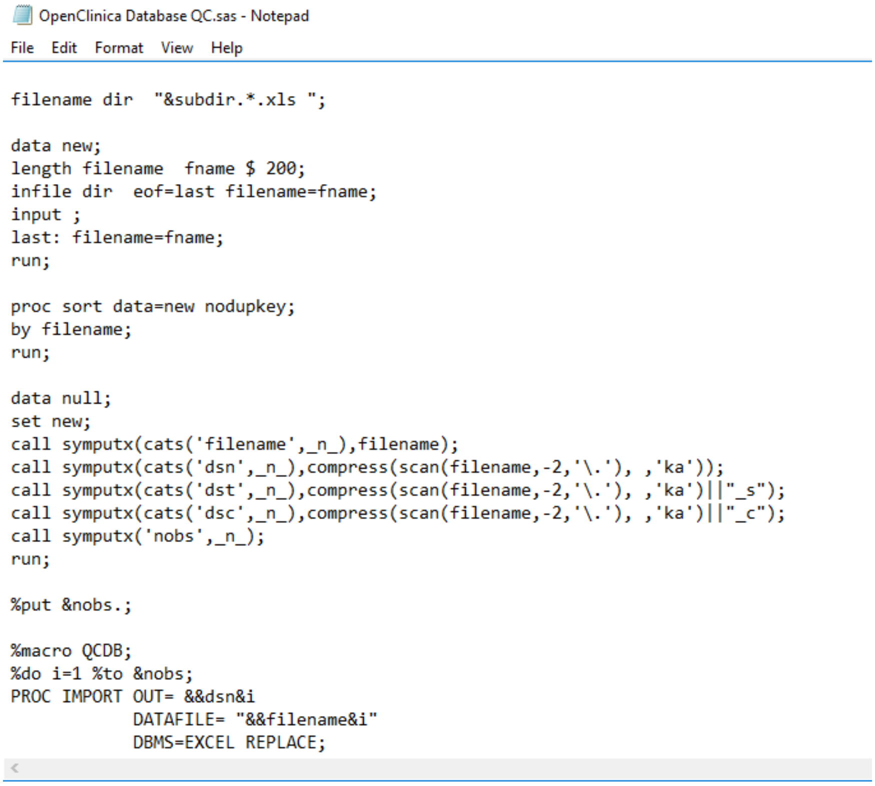
Date Informed Consent Was Signed	(dd-mmm-yyyy)
Version Date Of Informed Consent (Date Approved by IRB)	(dd-mmm-yyyy)
Date of Birth	(dd-mmm-yyyy)
Age	(Years)
Is Subject Less Than 18 Years Old?	○ Yes ○ No
Date Assent Was Signed	(dd-mmm-yyyy)
Version Date Of Assent	(dd-mmm-yyyy)
Gender	○ Male ○ Female
Race	○ American Indian or Alaskan Native ○ Asian ○ Black or African American ○ Native Hawaiian or Other Pacific Islander ○ White ○ Other (specify)
Specify Other Race	
Ethnicity	○ Hispanic or Latino ○ Not Hispanic or Latino

Macro 1: Read Spec file to datasets, and then transfer to load format


OpenClinica Database Build.sas - Notepad
File Edit Format View Help

```
*  Macro call       :
*  Revision History :
*  Date      Author   Description of the change
*******************************************************************************/
%let output=G:\Projects\SST0225\SST-0225-013\DM\Other\SST_Final eCRF Design;
LIBNAME XLSLIB "G:\Projects\SST0225\SST-0225-013\DM\CRF\OpenClinica eCRF Spec Template.xls"
        mixed=yes stringDates=yes scanTime=yes;

data vtable;
  set sashelp.vtable;
  where libname="XLSLIB" and index(memname,"$") and index(memname,"FilterDatabase")=0 and
        memname not in ("'Version history$'","'Visit Structure$'","Instruction$","Signature$","Template$");
run;

proc sql noprint;
/***Get total Number of Sheets***/
select count(distinct(MEMNAME)) into: tot
from vtable;
/**Get the sheet names without $ in to macro variables***/
select distinct(compress(MEMNAME,"',$")) into: s1 - :s%trim(%left(&tot))
from vtable;
/**Get the sheet names with $ in to macro variables***/
select distinct(MEMNAME) into: v1 - :v%trim (%left(&tot))
from vtable;
Quit;

%macro xlread;
%do i=1 %to &tot;

data &&s&i.;
set xlslib."&&v&i"n;
where Dataset^="";
RUN;

data out;
  set &&s&i ;
```

Macro 2: DB QC without eSpec


OpenClinica Database QC.sas - Notepad
File Edit Format View Help

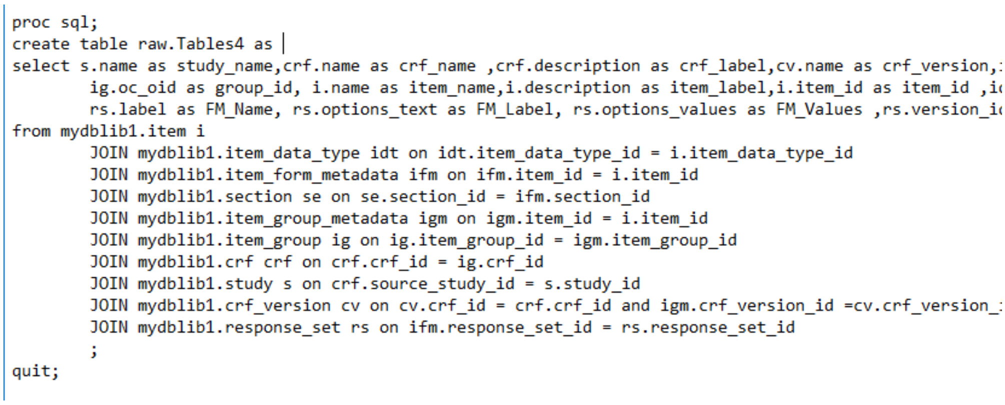
```
filename dir  "&subdir.*.xls ";

data new;
length filename  fname $ 200;
infile dir  eof=last filename=fname;
input ;
last: filename=fname;
run;

proc sort data=new nodupkey;
by filename;
run;

data null;
set new;
call symputx(cats('filename',_n_),filename);
call symputx(cats('dsn',_n_),compress(scan(filename,-2,'\.'), ,'ka'));
call symputx(cats('dst',_n_),compress(scan(filename,-2,'\.'), ,'ka')||"_s");
call symputx(cats('dsc',_n_),compress(scan(filename,-2,'\.'), ,'ka')||"_c");
call symputx('nobs',_n_);
run;

%put &nobs.;

%macro QCDB;
%do i=1 %to &nobs;
PROC IMPORT OUT= &&dsn&i
            DATAFILE= "&&filename&i"
            DBMS=EXCEL REPLACE;
```

Clinical Operations Phase

To achieve high-efficiency, we should design the database on the basis of CDASH as much as possible in database design, so that we can customize different data viewing models during the implementation. For example, for the DM department, they can generate project reports such as project progress status report, query processing report of sites, actual enrollment, and data reports at any time. For the Statistics department, data status maps and individual proportion data models can be generated periodically. For medical personnel, medical history, the relationship between adverse events and medications can be checked regularly. In FMD, these are realized by using SAS.

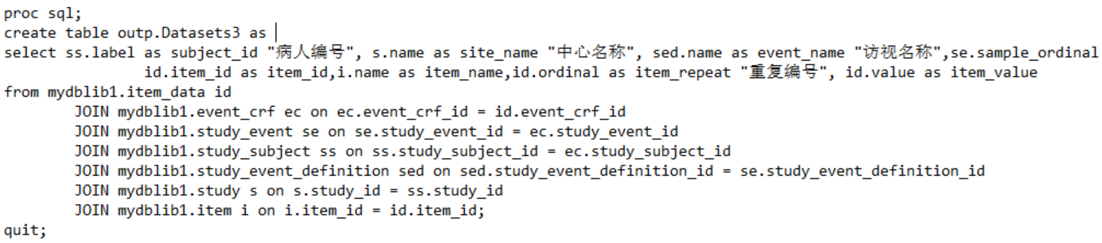
Specific steps of Macro:

Step 1：Read raw data from DB



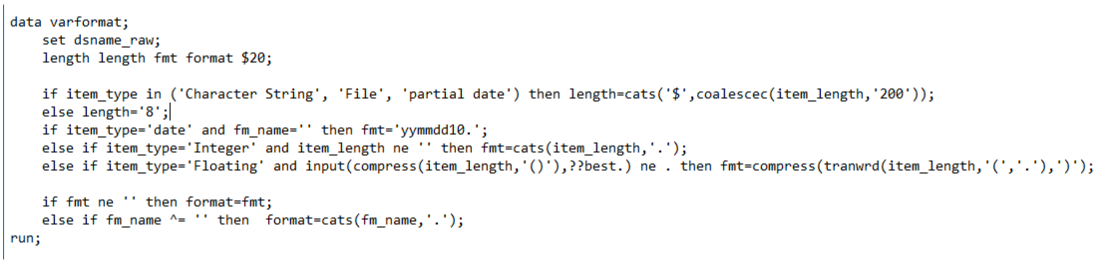
```
proc sql;
create table raw.Tables4 as
select s.name as study_name,crf.name as crf_name ,crf.description as crf_label,cv.name as crf_version,
        ig.oc_oid as group_id, i.name as item_name,i.description as item_label,i.item_id as item_id ,i
        rs.label as FM_Name, rs.options_text as FM_Label, rs.options_values as FM_Values ,rs.version_i
from mydblib1.item i
        JOIN mydblib1.item_data_type idt on idt.item_data_type_id = i.item_data_type_id
        JOIN mydblib1.item_form_metadata ifm on ifm.item_id = i.item_id
        JOIN mydblib1.section se on se.section_id = ifm.section_id
        JOIN mydblib1.item_group_metadata igm on igm.item_id = i.item_id
        JOIN mydblib1.item_group ig on ig.item_group_id = igm.item_group_id
        JOIN mydblib1.crf crf on crf.crf_id = ig.crf_id
        JOIN mydblib1.study s on crf.source_study_id = s.study_id
        JOIN mydblib1.crf_version cv on cv.crf_id = crf.crf_id and igm.crf_version_id =cv.crf_version_
        JOIN mydblib1.response_set rs on ifm.response_set_id = rs.response_set_id
        ;
quit;
```

Step 2: Read All tables from DB



```
proc sql;
create table outp.Datasets3 as
select ss.label as subject_id "病人编号", s.name as site_name "中心名称", sed.name as event_name "访视名称",se.sample_ordinal
                id.item_id as item_id,i.name as item_name,id.ordinal as item_repeat "重复编号", id.value as item_value
from mydblib1.item_data id
        JOIN mydblib1.event_crf ec on ec.event_crf_id = id.event_crf_id
        JOIN mydblib1.study_event se on se.study_event_id = ec.study_event_id
        JOIN mydblib1.study_subject ss on ss.study_subject_id = ec.study_subject_id
        JOIN mydblib1.study_event_definition sed on sed.study_event_definition_id = se.study_event_definition_id
        JOIN mydblib1.study s on s.study_id = ss.study_id
        JOIN mydblib1.item i on i.item_id = id.item_id;
quit;
```

Step 3: Read ALL formats from system

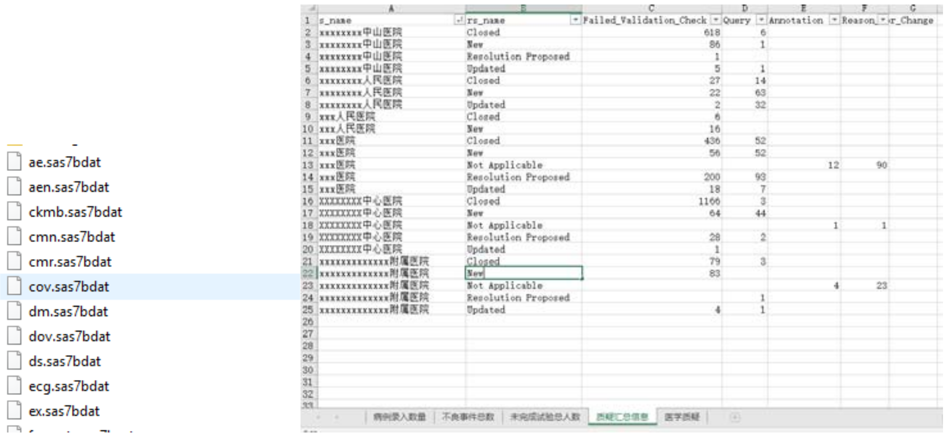


```
data varformat;
    set dsname_raw;
    length length fmt format $20;

    if item_type in ('Character String', 'File', 'partial date') then length=cats('$',coalescec(item_length,'200'));
    else length='8';
    if item_type='date' and fm_name='' then fmt='yymmdd10.';
    else if item_type='Integer' and item_length ne '' then fmt=cats(item_length,'.');
    else if item_type='Floating' and input(compress(item_length,'()'),??best.) ne . then fmt=compress(tranwrd(item_length,'(','.'),')');

    if fmt ne '' then format=fmt;
    else if fm_name ^= '' then  format=cats(fm_name,'.');
run;
```

Step 4: Generate all dataset during the generate macro

```
%macro transpose;
%do i=1 %to &ntot;

    %put dataset = &&ds&i;
    data _var;
        set varformat;
        where group_name="&&ds&i";
    run;

    %let varid=;
    proc sql noprint;
        select distinct item_id into :varid separated by ',' from _var
    quit;
    %put &varid;
    proc sort data=raw.datasets3 out=_data;
        where item_id in (&varid);
        by subject_id site_name event_name event_repeat item_repeat;
    run;

    data _null_;
    %let _nobs=0;
        set _data nobs=nobs;
        call symputx('_nobs',nobs);
        stop;
    run;

    %if &_nobs>0 %then %do;
    proc transpose data=_data out=&&ds&i(drop=_name_ _label_);
        by subject_id site_name event_name event_repeat item_repeat;
        id item_name;
        var item_value;
    run;

    %let nvar=;
    proc sql noprint;
        create table _var0 as
        select item_name, item_label,length, fmt, format
        from _var
        where item_id in (select distinct item_id from _data)
        order by item_name
        ;
    quit;
```

Step 5: Datasets and listing generated



ae.sas7bdat
aen.sas7bdat
ckmb.sas7bdat
cmn.sas7bdat
cmr.sas7bdat
cov.sas7bdat
dm.sas7bdat
dov.sas7bdat
ds.sas7bdat
ecg.sas7bdat
ex.sas7bdat

s_name	rs_name	Failed_Validation_Check	Query	Annotation	Reason_	r_Change
xxxxxxxx中山医院	Closed	618	6			
xxxxxxxx中山医院	New	86	1			
xxxxxxxx中山医院	Resolution Proposed	1				
xxxxxxxx中山医院	Updated	5	1			
xxxxxxxx人民医院	Closed	27	14			
xxxxxxxx人民医院	New	22	63			
xxxxxxxx人民医院	Updated	2	32			
xxx人民医院	Closed	6				
xxx人民医院	New	16				
xxx医院	Closed	436	52			
xxx医院	New	56	52			
xxx医院	Not Applicable			12	90	
xxx医院	Resolution Proposed	200	93			
xxx医院	Updated	18	7			
XXXXXXXX中心医院	Closed	1166	3			
XXXXXXXX中心医院	New	64	44			
XXXXXXXX中心医院	Not Applicable			1	1	
XXXXXXXX中心医院	Resolution Proposed	28	2			
XXXXXXXX中心医院	Updated	1				
xxxxxxxxxxxxx附属医院	Closed	79	3			
xxxxxxxxxxxxx附属医院	New	83				
xxxxxxxxxxxxx附属医院	Not Applicable			4	23	
xxxxxxxxxxxxx附属医院	Resolution Proposed		1			
xxxxxxxxxxxxx附属医院	Updated	4	1			

Delivery Phase

After standardized data is available, fast delivery can be realized. (We will only do a brief introduction here)

With Macro, you can submit data with one click. In FMD, there is a department that is specialized in preparing data for submission.

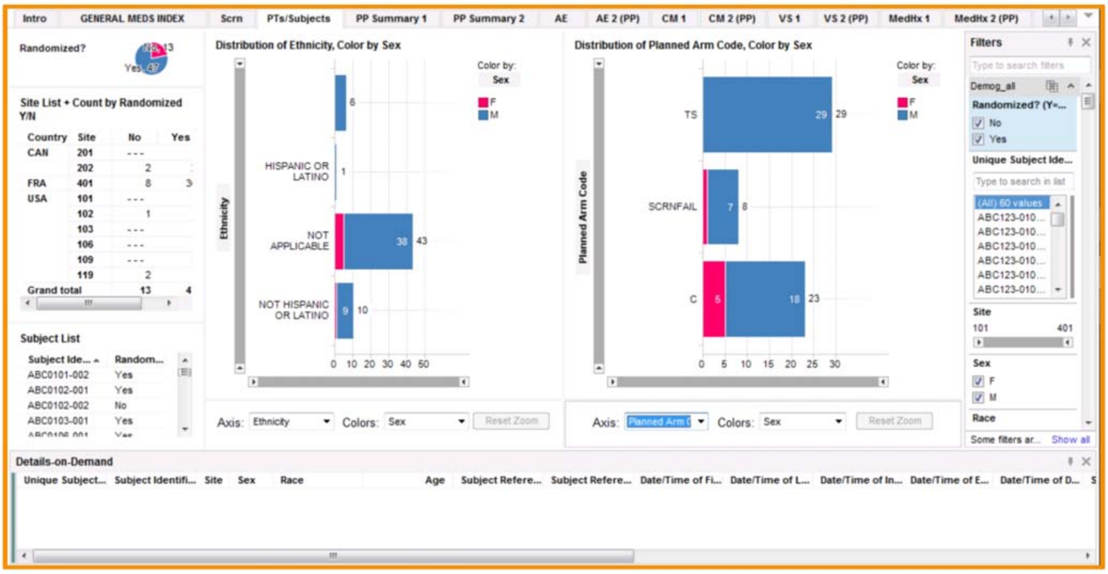
Such as: SDTM Data

Both CDASH and SDTM follow the CDISC standard for data collection and naming. The two are almost 80% same in Following CDASH from database design phase could dramatically reduce the data standardization programming efforts at SDTM programming stage. The remaining different or derived parts, such as xxSEQ and xxDY, can be automatically assigned via macro.

Define file

Define displays that serve as data elements. You can extract the attributes of CRF and data to achieve automatic filling of each module. In the later stage, only simple adjustment and review are needed.

Data visualization base on SDTM/CDISH datasets(End user will review it base on these format)



Site List + Count by Randomized Y/N

Country	Site	No	Yes
CAN	201	---	
	202	2	
FRA	401	8	3
USA	101	---	
	102	1	
	103	---	
	106	---	
	109	---	
	119	2	
Grand total		13	4

Subject List

Subject Ide...	Random...
ABC0101-002	Yes
ABC0102-001	Yes
ABC0102-002	No
ABC0103-001	Yes
ABC0106-001	Yes

Axis: Ethnicity Colors: Sex

Axis: Planned Arm C Colors: Sex

Filters

Type to search filters

Demog_all

Randomized? (Y=...

☑ No

☑ Yes

Unique Subject Ide...

Type to search in list

(All) 60 values
ABC123-010...
ABC123-010...
ABC123-010...
ABC123-010...
ABC123-010...
ABC123-010...

Site

101 401

Sex

☑ F

☑ M

Race

Some filters ar... Show all

Details-on-Demand

Unique Subject...	Subject Identifi...	Site	Sex	Race	Age	Subject Refere...	Subject Refere...	Date/Time of Fi...	Date/Time of L...	Date/Time of In...	Date/Time of E...	Date/Time of D...	S

Thanks for the teamwork

Xuhua.deng@klserv.com;

Tiantian.zhao@klserv.com;