PhUSE US Connect 2019

Paper DH14

Organize and report your clinical data today!

Michael Litzsinger, Project Data Sphere, LLC, Cary, NC, U.S.A.

ABSTRACT

Don't wait for the perfect standard or software package!

Create open-ended data acquisition and reporting processes, to build interoperable metadata and set-up flexible methods that aggregate and summarize data for clinical data research.

Start today!

The idea, in 3 bullets:

- Use existing business assets.
- Define metadata for “data” and “reporting”.
- Plan, implement in steps, and track progress.

Examples of metadata for “data” and “reporting”:

- ♦1 List of Study Data.
- ♦2 Schedule of Snapshots.
- ♦3 List of Standard Reports.
- ♦4 List of Reporting Requests.

INTRODUCTION

Designing metadata to represent your data and reporting creates clear inputs/outputs for reuse by other researchers and automated tools. This clarity benefits research by not re-inventing the wheel and encouraging interoperability.

Examples provided here are intended to spur metadata usage within research programming code, reducing effort and enabling transparency. Standards are a good thing. A great thing, in fact. But it can be overwhelming to incorporate metadata standards before they have converged.

Implementation of the metadata mentioned in this paper is not controversial and can be implemented quickly now. This metadata also facilitates reporting on your department webpage or to your department management to transparently show your progress in utilizing metadata!

METADATA FOR DATA: AUTOMATE SAS® LIBNAME STATEMENTS

Let’s begin with one type of metadata, the location of each clinical study database. You (and other researchers) want to find each clinical study quickly and transparently so you build a ♦1 List of Study Data.

♦1 LIST OF STUDY DATA

This data table has one row for each clinical study database, with columns for the identifier, path, and any other useful descriptive information that you want to maintain. There must be a column containing a unique identifier for

each study (either a protocol number or a registry identifier) as this is used in other linked metadata. Ideally there is a field to define the `libname` in program code as well. There also must be a column containing the directory path. You could also choose to store other descriptive information here about the study (phase) or the data type (format of the data, especially to stack raw and analysis data records). I do not recommend to add redundant descriptions if this can be linked (perhaps from a protocol data table). This example above depicts paths to SDTM data for each clinical study and its associated compound pooled dataset.

The ♦1 List of Study Data metadata should enable automatic aggregation of studies. Each path specified should contain data that is write protected and reliably "ready to use". No exceptions allowed. Then reviewers, programmers, and automated programs can be confident that this is always the latest and best source to use, whether to be read directly into a program or for manual browsing. There must be also a defined process in place to update this source data, and any replacement data must be checked before overwriting the current data. If the replacement data is incomplete or fails validation it is not allowed to overwrite the current data in this location.

♦1 LIST OF STUDY DATA

Product	Study	Status	Study SDTM Path	Pool SDTM Path
<name1>	<study1>	FINAL_CSR	...\<name1>\<study1>\sdtm	...\<name1>\pool\fin_sdtm
<name1>	<study2>	FINAL_CSR	...\<name1>\<study2>\sdtm	...\<name1>\pool\fin_sdtm
<name1>	<study3>	FINAL_DATA	...\<name1>\<study3>\sdtm	...\<name1>\pool\ong_sdtm
<name1>	<study4>	ONGOING	...\<name1>\<study4>\sdtm	...\<name1>\pool\ong_sdtm
<name2>	<study11>	FINAL_DATA	...\<name2>\<study11>\sdtm	...\<name2>\pool\ong_sdtm
<name2>	<study12>	ONGOING	...\<name2>\<study12>\sdtm	...\<name2>\pool\ong_sdtm
<name2>	<study13>	ONGOING	...\<name2>\<study13>\sdtm	...\<name2>\pool\ong_sdtm
<name2>	<study14>	NO_DATA		

/♦1 List of Study Data screenshot/

```
%macro Win_LIBNAMES;
%* Import LIBNAMES worksheet to Win_LIBNAMES dataset
 ********************************************************************************;
proc import datafile="C:\sddshared\SASWorkspaces\Files\Metadata\Win_LIBNAMES.csv"
DBMS=CSV OUT=Win_LIBNAMES;
run;

%* Create macro variable &NObs (number of observations)
 * Create macro variable &&study&i (libref for observation&i)
 * Create macro variable &&path&i (path for observation&i)
 ********************************************************************************;
proc sql noprint;
  select count(*) into :Nobs from Win_LIBNAMES;
  select study into :study1-:study%left(&NObs) from Win_LIBNAMES;
  select path into :path1-:path%left(&NObs) from Win_LIBNAMES;
quit;

%* Write LIBNAME statement for each observation
 ********************************************************************************;
%do i=1 %to &NObs;
LIBNAME &&study&i "&&path&i";
%end;
run;
%mend Win_LIBNAMES;

%Win_LIBNAMES
```

/Sample SAS macro code screenshot to read in metadata for libnames…/

```
MPRINT:   proc import datafile="C:\sddshared\SASWorkspaces\Files\Metadata\Win_LIBNAMES.csv" DBMS=CSV
OUT=Win_LIBNAMES;
MPRINT:   ADLM;
MPRINT:   run;
NOTE: 132 records were read from the infile 'C:\sddshared\SASWorkspaces\Files\Metadata\Win_LIBNAMES.csv'.
NOTE: The data set WORK.WIN_LIBNAMES has 132 observations and 5 variables.

MPRINT:   proc sql noprint;
MPRINT:   select count(*) into :NObs from Win_LIBNAMES;
MPRINT:   select study into :study1-:study132 from Win_LIBNAMES;
MPRINT:   select path into :path1-:path132 from Win_LIBNAMES;
MPRINT:   quit;
NOTE: PROCEDURE SQL used 0.01 seconds

MPRINT:   LIBNAME PDS79 "C:\sddshared\SASWorkspaces\Files\PDS79\SDTM";
NOTE: Libref PDS79 was successfully assigned as follows:
      Physical Name: C:\sddshared\SASWorkspaces\Files\PDS79\SDTM

MPRINT:   LIBNAME PDS80 "C:\sddshared\SASWorkspaces\Files\PDS80\SDTM";
NOTE: Libref PDS80 was successfully assigned as follows:
      Physical Name: C:\sddshared\SASWorkspaces\Files\PDS80\SDTM

MPRINT:   LIBNAME PDS81 "C:\sddshared\SASWorkspaces\Files\PDS81\SDTM";
NOTE: Libref PDS81 was successfully assigned as follows:
      Physical Name: C:\sddshared\SASWorkspaces\Files\PDS81\SDTM
```

/…resolves into these SAS program statements/

The ♦1 List of Study Data metadata status likely will need to reflect versioning of the data. The /♦1 List of Study Data screenshot/ depicts metadata status as NO_DATA, ONGOING data, FINAL_DATA (locked but not fully reported), and FINAL_CSR (locked and fully reported).

At Project Data Sphere, LLC there is another category for status: FINAL_DE-ID (final and de-identified).

Ongoing versions might also track the accrual of domains or track changes in standards.

Tumor_Type	Study	Status	Study Path
Liver	PDS1	FINAL_DE-ID	C:\sddshared\SASWorkspaces\Files\PDS1
Liver	PDS2	FINAL_DE-ID	C:\sddshared\SASWorkspaces\Files\PDS2\final
Prostate	PDS79	FINAL_DE-ID	C:\sddshared\SASWorkspaces\Files\PDS79\SDTM
Prostate	PDS80	FINAL_DE-ID	C:\sddshared\SASWorkspaces\Files\PDS80\SDTM
Prostate	PDS81	FINAL_DE-ID	C:\sddshared\SASWorkspaces\Files\PDS81\SDTM
Prostate	PDS127	FINAL_DE-ID	C:\sddshared\SASWorkspaces\Files\PDS127\ADaM
Prostate	PDS132	FINAL_DE-ID	C:\sddshared\SASWorkspaces\Files\PDS132\ADaM

/Sample PDS metadata /

METADATA FOR REPORTING: TRACK WHAT YOU REPORT

Types of metadata about reporting are the ♦2 Schedule of Snapshots, the ♦3 List of Standard Reports, and the ♦4 List of Reporting Requests. These are the reports you build and execute on the ♦1 List of Study Data.

♦2 SCHEDULE OF SNAPSHOTS

The ♦2 Schedule of Snapshots metadata has one row per scheduled snapshot. Each snapshot can be planned in advance to alert providers of obligatory reporting so that resources are in place for just-in-time refreshes. This could be to enable same day turnaround of datasets for a database lock, for quarterly safety reporting, or for ad hoc purposes. Clarifying a calendar of data needs up to a year in advance enables efficient planning. Aligning transfers for quarterly reporting to coincide with annual reporting (e.g., DSUR reporting) and a publication could mean only 4 refreshes rather than 6. Providing the ♦2 Schedule of Snapshots also as a calendar helps data providers and recipients to see the impact of weekends and local holidays and adjust accordingly.

Year	Product	Report Type	SDTM snap	SDTM delivery	Safety reporting	Annual reporting
2015	<name1>	Safety	2-Jan-15	9-Jan-15	12-Jan-15 to 23-Jan-15	
2015	<name1>	Safety	2-Apr-15	9-Apr-15	13-Apr-15 to 24-Apr-15	
2015	<name1>	Safety & Annual	2-Jul-15	9-Jul-15	13-Jul-15 to 24-Jul-15	13-Jul-15 to 17-Jul-15
2015	<name1>	Safety	2-Oct-15	9-Oct-15	12-Oct-15 to 23-Oct-15	

/♦2 Schedule of Snapshots screenshot/

♦3 LIST OF STANDARD REPORTS

The ♦3 List of Standard Reports metadata describes the standard reports available and it advertises best options for fast reporting. Clickable links to screenshots of the canned reports can be provided to view a sample of the output format (very helpful to those unfamiliar with the layouts already programmed). A secondary use of the ♦3 List of Standard Reports metadata is for departmental metric reporting about the progression of reporting standards.

Report Type	Report Name	Products; Studies	Data	Methods
Patient Profile	Graphical Patient Profile	<name2>; All	Study SDTMs	<tool3>
Patient Profile	Formatted Patient Profile (By-Patient Listing)	<name1>, <name2>; All	Study SDTMs	<tool3>
Patient Profile	Formatted Patient Profile (By-Patient Listing)	<name1>; <study1>, <study2>	Study ADaMs	<tool1>
Cleaning/ Review	SAE Reconciliation Differences	All; All	Study OC, Safety DB	<tool2>
Annual	Cumulative Exposure (3 reports)	All; All	Study SDTMs	<tool1>, <tool3>
Safety	Disposition and Discontinuation Reasons	All; All	Study SDTMs	<tool1>, <tool3>
Safety	Demographics	All; All	Study SDTMs	<tool1>, <tool3>
Safety	Study Medication Exposure	All; All	Study SDTMs	<tool1>, <tool3>
Safety	Incidence of AEs	All; All	Study SDTMs	<tool1>, <tool3>

/♦3 List of Standard Reports screenshot/

♦4 LIST OF REPORTING REQUESTS

You (and other researchers) want to find each report quickly and transparently so you build and maintain a ♦4 List of Reporting Requests. A primarily use of this metadata is for programmers to find a past report in order to have a starting point for creating a new request or for revising a past report. A secondary use of this is for departmental metric reporting about task counts, data sources used, and active customers. Don't stress out to build a ♦4 List of Reporting Requests retrospectively; it is still useful to build it from the present.

Status	Due Date	Type	Report Name	Data	Method
Completed	2015-10-09	New	<name1> <study4> Missing Pages Report	Study SDTMs	<tool1>
Ongoing	2015-10-16	Change	<name1> Subject narrative list for 120DSU - ongoing LTFU studies - <study3>	Study ADaMs	<tool1>
Ongoing	2015-10-23	New	<name1> Subject narrative list for 120DSU - ongoing LTFU studies - <study4>	Study ADaMs	<tool1>
Ongoing	2015-10-28	Change	Standard Side-by-Side-Reconciliation-Listing-OC-Pt-vs-SafetyDB-Pt: CR6 Update	Study OC, Safety DB	<tool2>
Not_started	2015-11-06	Refresh	<name2> 2015 Annual Report	Study SDTMs	<tool1>
Not_started	2015-11-13	Refresh	<name2> 2015 Q4 Safety Report	Study SDTMs	<tool1>

/♦4 List of Reporting Requests screenshot/

A reporting team at my previous job compiled a spreadsheet of 3300 rows of reporting requests over a 9-year period.

LINK TO OTHER METADATA

I touched on this earlier in the ♦1 List of Study Data section when I recommended to not duplicate descriptive information about the study. Instead, add only a unique identifier to link to this other metadata (same unique identifier must be used in both places).

This linked metadata could store planned data elements or actual data elements. Planned data elements could link to protocol metadata while actual data elements could link to data summary counts.

PROTOCOL METADATA (SDTM.TS)
Link to the clinical study phase, planned subject counts, planned treatments, planned age range, planned countries.

DEMOGRAPHIC METADATA (SUMMARY COUNTS OF SDTM.EX, SDTM.DM, …)
Link to the actual subject counts, actual treatments, actual age range, actual countries, et cetera.

CONCLUSION

Interoperable metadata can be used by a variety of tools to aggregate and summarize data for clinical data research.

- The ♦1 List of Study Data metadata can be built in a day or so, and then you have a useful map of all clinical studies in one place.
- The ♦2 Schedule of Snapshots metadata can be built and agreed with partners within 2 weeks, and achieves transparency of expectations for data refreshes. It significantly raises the likelihood that data is exchanged efficiently in the timeframe needed.
- The ♦3 List of Standard Reports metadata transparently tracks standards created (and planned). Perhaps you have a list already. However, maintaining this as metadata enables posting on a departmental webpage and encourages a broader awareness.
- The ♦4 List of Reporting Requests metadata is essential to track the status (and history) of your tasks and deliverables. Maintain it to manage day-to-day accomplishments, but also think of this as an intermediate step towards automation. Such metadata is invaluable to remember years back and reduce the knowledge drain from employee turnover. Both customers and employees will use this metadata for queries if you keep it simple and searchable.

Don't wait to create simple metadata to enable automated processes later. A sound clinical data accrual and reporting process has to be designed, implemented, and maintained first. Start today!

RECOMMENDED READING

1. PhUSE CSS Emerging Technology Working Group's *Metadata Definitions* (Version 1.0, June 6, 2014): https://www.phusewiki.org/wiki/images/c/cc/ET-project-_Metadata_definitions-V1-2014_07_15.pdf
2. Tucker, Katherine. *Data sharing 3 year on – from baby to toddler…* PhUSE 2016, paper DH12. https://www.lexjansen.com/phuse/2016/dh/DH12.pdf

CONTACT INFORMATION

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

Michael Litzsinger
Project Data Sphere, LLC
100 SAS Campus Drive
Cary, NC U.S.A. 27513
mike.litzsinger@projectdatasphere.org
linkedin.com/in/michael-litzsinger
projectdatasphere.org

Project Data Sphere, LLC is an independent not-for-profit initiative of the *CEO Roundtable on Cancer's* Life Sciences Consortium. The initiative provides a free digital library-laboratory that gives the research community one place to broadly share, integrate and analyze historical, patient-level data from academic and industry phase III cancer clinical trials.