SDTM Domain Development on the Critical Path – Strategies for Success

David C. Izard, Octagon Research Solutions, Wayne PA

ABSTRACT

The design and development of SDTM domains from legacy source data and documents requires careful planning and execution in order to produce a high quality set of deliverables. The placement of this activity on the critical path introduces additional challenges that need to be addressed early and consistently revisited during study execution in order to ensure that business needs are met at the end of the day. This paper and presentation will explore the inherent challenges, risks and strategies that can be employed to ensure that the development of SDTM deliverables at this important time can effectively be managed and ultimately reduce overall time, energy & effort with on all Clinical activities related to study completion.

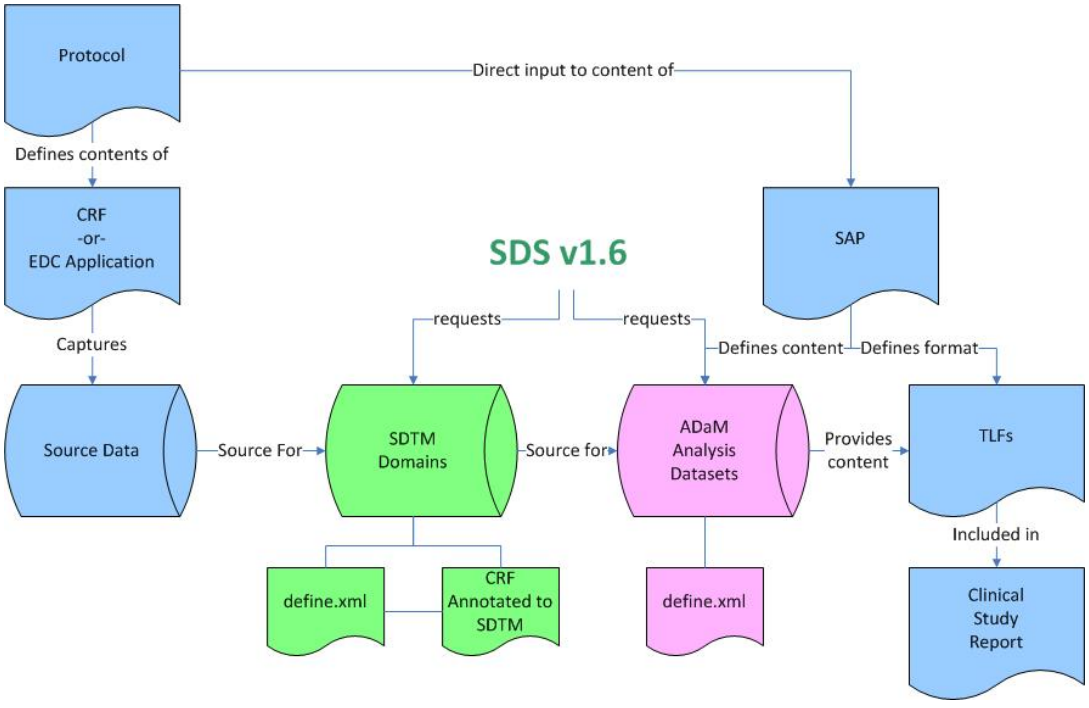
DEFINITIONS

In order to establish a framework for this discussion, first we must focus on definitions of "Legacy Conversion" and "Critical Path". A legacy conversion would be any conversion that involves source data being captured in a given format which requires a significant transformation effort to meet downstream business needs. In the context of this paper I will explore data captured in a format other than "CDASH conformant", "Near SDTM" or "SDTM like" with the goal of producing SDTM assets for use as input to analysis efforts as well as serving as the CRT format for study level submission requirements.

Literally, the critical path is defined as the longest sequence of activities in a project plan which must be completed on time for the project to complete on due date [1]. In this context I am focusing on the creation of SDTM domains in a variety of situations but the primary challenge being the time critical creation of analysis datasets and corresponding tables, figures and listings based on SDTM domains rather than produced straight from the source data. Coupling both of these definitions, the crux of the challenge is when this SDTM conversion is based on legacy data.

THE REGULATORY AND BUSINESS CASES FOR LEGACY SDTM CONVERSION

The following flow diagram represents the legacy conversion process for an individual study from conception through to submission in terms of both asset creation in compliant format and production of value added items in support of Clinical Study Report generation.



As a key, items in blue above are standard items produced as part of transforming the intentions of a clinical effort from concept to tangible proof of clinical effort per ICH guidelines. The green and pink shapes represent the assets generated

to support. Key to this figure is the label "SDS v1.6" in this picture. The FDA, via the eCTD guidance [n] states that data and related documentation should be produced per the FDA's Study Data Specifications, currently in Version 1.6 [2]. Starting on Page 3 this document defines the CDISC SDTM and SEND models as acceptable formats for provided Case Report Tabulation (CRT) data to the agency in support of a submission. The document goes on to articulate that CDISC ADaM is an accepted standard for analysis data and that you should clarify with your review division what formats are acceptable prior to sending in your submission.

Why go to all of this trouble? Why not continue to submit data in legacy formats? The FDA, via CDER's Common Data Standards Issues Document [3], reinforces the recommendation for use of the CDISC SDTM & ADaM models for tabulation and analysis data respectively but takes it farther by placing a significant amount of emphasis on traceability:

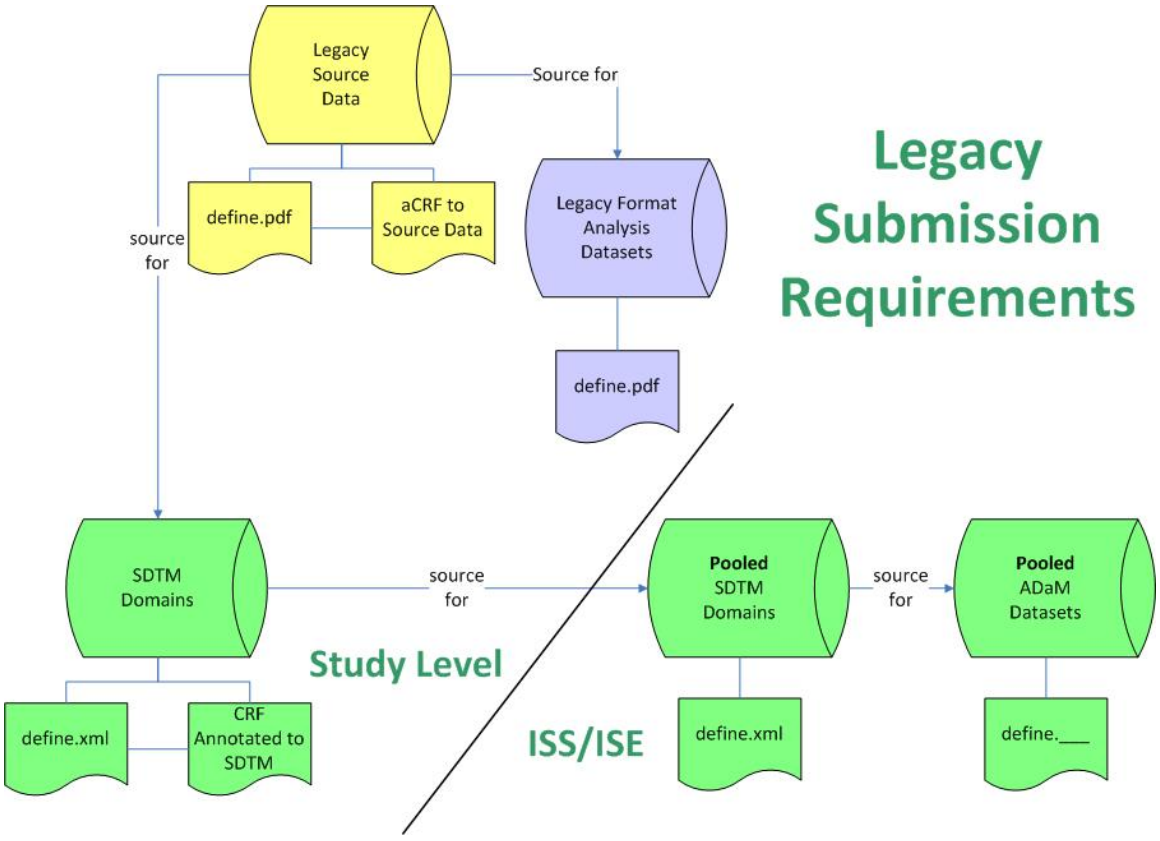
> "It is very important that the results presented in the Clinical Study Report be traceable back to the original data elements as they were collected in the case report form and represented in the SDTM datasets. The SDTM datasets must be able to support the results in the Clinical Study Report, either directly for some results, or, for other results, indirectly through analysis datasets that are derivable from the SDTM datasets." [[3], Page 5, 2nd paragraph]

This document goes farther to clarify expectations regarding performing legacy conversions:

> "If a sponsor decides to convert trial data to SDTM that was originally collected in non-SDTM format, it is important to note that the resulting SDTM data should support the accompanying analysis data sets and sponsor's reports (study reports, etc.). CDER has received applications in which the converted SDTM data sets were not consistent with the submitted analysis datasets and study report analyses, thus causing confusion during application review." [[3], Page 5, 3rd paragraph]

The agency is looking for sponsors to be accountable for providing a clear path, a chain of custody from initial collection through to analysis, reporting and ultimately submission in regulatory compliant form. It is an arduous process yet sponsors are willing to go down that pathway. Why? In order for a sponsor organization to maximize profitability they need to balance and control time, quality and resources effectively. As Sponsors experience with the CDISC standards continues to mature it is becoming more evident that legacy conversion has an upfront cost but the positive impact on downstream cost far outweighs the initial expense.

Consider the following diagram describing the situation where SDTM domains are created at the study level in support of integration and standardization efforts in support of Integrated Summaries of Safety (ISS) and Efficacy (ISE):



The yellow shapes represent the source data which would be published in the legacy format defined by the FDA's Submission Data Standards. The study level green shapes represent a SDTM assets produced as part of a legacy conversion effort in support of integrated SDTM domains that will serve as the foundation for ISS & ISE analysis. The quandary is the blue shapes which represent the analysis datasets developed to support the tables, figures & listings present in the Clinical Study Report written for this study.

Based on the FDA's statements on traceability quoted above, you have a few choices. One option would be to redo your analysis activities which would involve creating ADaM analysis datasets from the SDTM domains, reproducing your tables, figures and listings from the ADaM analysis datasets and highlighting where results differed between the original analysis datasets / TFLs and the standards based results as part of the Reviewer's Guide. A second option could be to rewrite your legacy analysis dataset data definition table to reflect that you generated your analysis datasets from SDTM domains (even though you didn't) in order to theoretically maintain traceability from SDTM through to analysis. In this scenario you must be very careful to ensure that all items can be derived; I have seen many cases where analysis decisions were based on source data that did not make it to the SDTM domains, such as system variables and CRF form names, making full traceability impossible.

The third option would be to fully publish and submit your source data, your analysis data and the fruits of your legacy SDTM conversion efforts at the study level when developing your submission. While this may seem like overkill, it is fully permissible per the eCTD. You may get agency questions during the review cycle as the analysis at the study level will not reflect the look and feel of the analysis data at the integrated analysis level but it is fully defensible. As with any submission, whatever path you choose, ensure that you have the conversation with your review division early and often. The key message is that all three options involve performing additional time consuming steps in order to meet agency requirements for traceability and transparency and at the same time leave your study level analysis assets in a non-standard format. Already you have spent additional time and effort working around the agency's expectation that you provide SDTM domains in support of CRT requirements and ADaM analysis datasets derived from SDTM domains in support of study level analysis, and you still have the risk that review and approval activities could be delayed by having to deal with analysis data presented in disparate formats across the submission.

LEGACY SDTM CONVERSION – KNOW WHEN TO SAY WHEN

Now that the case for performing legacy SDTM conversion work has been established, you should consider when you would want to undertake these efforts and when you might not want to. The key question to ask is, "Will the results of the work that I will do based on the current representation of the data directly contribute to the content of a regulatory submission?" Any work related to developing analyses that will contribute to the authoring of study level or integrated documents that will be part of a regulatory submission are clearly the sweet spot of legacy SDTM conversion work.

There are areas of clinical development, however, that fall outside of these activities. Usually the driving force behind these reporting requirements is timely delivery of analysis following availability of input data. Some examples of these activities include analysis to support dose escalation decisions, medical publication deadlines, press release support or data and analysis in support of data safety monitoring boards. In these cases you will likely produce the same results based on standards study assets at the time of submission but the timeliness of the initial reporting cannot support the rigor required to produce submission level assets.

STRATEGIES FOR SUCCESS

Setting Realistic Expectations

Your company traditionally goes from Database Lock (DBL) or an equivalent milestone reflecting that the source data has been captured, reviewed and certified as final and available for the next steps in the clinical development cycle, to draft headline tables in something on the order of two weeks and final analysis datasets and tables, figures and listing in the neighborhood of five weeks. You now introduce the production of SDTM domains based on proprietary source data collection standards, perhaps ADaM analysis datasets produced from the SDTM domains and then the required tables, figures and listings based on the ADaM analysis datasets. Using rough numbers based on a typical Phase III study, where you typically had approximately 30 source datasets for your study and you produced approximately 12 to 15 analysis datasets from the source datasets, you now introduce the generation of approximately 30 SDTM domains plus trial design and special purpose domains dependent on trial design domains into the timeline. Using these numbers you are producing 71% more datasets / domains when developing SDTM domains than if you weren't. Put simply, the more you produce, the more time it will take you to (initially) produce it.

Controlling your SDTM Development Process

Leaning back on an earlier PharmaSUG paper [4] you should look at defining a consistent, repeatable process for legacy SDTM conversion that will produce predictable results based on a wide variety of input analogous to Level 3 of the Software Engineering Institute's Capability Maturity Model. Some characteristics of a process that is consistent and repeatable would involve developing and following SOPs specifically related to legacy SDTM conversion work, using a

consistent method and/or tool to document domain mapping strategies prospectively and consistently performing high quality QC steps at multiple points in the process that address both structural integrity and well as clinical relevance of the SDTM assets developed.

Interim Conversions and End Game Planning

One strategy to minimizing time required once database lock is achieved is to establish as much of the conversion activities as you can prior to database lock. Performing initial conversion work on an interim data cut is one strategy that can work in your favor but you have to manage it carefully. Some things to consider are the degree to which your programs can accommodate changes in parameterized source data, such as code lists that are controlled based on CRF permissible values versus open ended items (e.g., unexpected lab tests, ECG interpretations) as well as tools that are available to detect differences in this parameterized source data from one delivery to the next. Another is structural checks on input data to ensure no changes have occurred to the input datasets and, if they have, they are quickly identified and either challenged or accommodated.

What can you Influence Upstream?

It takes considerable effort to implement a legacy SDTM conversion for a study, particularly when using interim / non-final data. Two key additional questions to ask, beyond the strategies of considering the transition from interim to final data when performing end game planning, are (1) at what point in the lifecycle of the study are input sources such as the source data, case report form and external data going to achieve a stable form and (2) in the event that they do change, to what degree are these changes communicated throughout the organization in a timely manner to allow for affected processes to react? Knowing this information along with taking a proactive role to influence these input sources to remain stable allows you to make an educated, risk-based decision on when to embark on interim conversion efforts in order to maximize the impact of interim conversion work while minimizing rework.

CONCLUSION

Legacy SDTM conversion is a complicated task that requires careful forethought and planning, a well documented process and detailed accountability for implementation and verification steps in order to execute effectively. While there are many challenges to developing SDTM assets from proprietary source dataset formats, the downstream benefits of clear traceability and transparent presentation of data to the agency for the purposes of review are vast. Specific activities can be undertaken at appropriate points during study execution which support performing these legacy conversion tasks on the critical path that minimize the disruption introduced by the required additional effort in order to make it as painless as possible.

REFERENCES

[1] http://www.businessdictionary.com/definition/critical-path.html
Web location verified April 1, 2012.

[2] FDA Submission Data Standards – (Version 1.6 / June 26, 2011)
http://www.fda.gov/downloads/Drugs/DevelopmentApprovalProcess/FormsSubmissionRequirements/ElectronicSubmissions/UCM199759.pdf
Web location verified April 1, 2012.

[3] CDER Common Data Standards Issues Document - (Version 1.1/December 2011)
http://www.fda.gov/downloads/Drugs/DevelopmentApprovalProcess/FormsSubmissionRequirements/ElectronicSubmissions/UCM199759.pdf
Web location verified April 1, 2012.

[4] Izard, D., *Pre-requisite to Effective Validation: Evaluating Your Clinical Trials Software Development Environment Using SEI's Capability Maturity Model for Software,* Conference Proceedings, PharmaSUG 2001. Available at www.lexjansen.com via http://www.lexjansen.com/pharmasug/2001/proceed/fdacomp/fda08_izard.pdf .
Web location verified April 1, 2012.

CONTACT INFORMATION

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

Name:	David C. Izard
Enterprise:	Octagon Research Solutions
Address:	585 East Swedesford Road
City, State ZIP:	Wayne, PA 19087
Work Phone:	610.535.6500, x5541
Fax:	610.535.6515
E-mail:	dizard@octagonresearch.com
Web:	www.octagonresearch.com

TRADEMARK CITATION