PharmaSUG 2014 – PO16

Automation of ADAM Dataset Creation with a Retrospective, Prospective and Pragmatic Process

Karin LaPann, MSIS, PRA International, USA

Terek Peterson, MBA, PRA International, USA

ABSTRACT

In the CDISC Standards, analysis datasets (ADaM) hold a unique place in the end to end process and must be created with both a prospective and retrospective view of the entire clinical trial process. Analysis datasets must support the statistical analysis plan (SAP) by providing accurate efficacy and safety analyses. Companies must be pragmatic in deciding which processes can be automated by tools. Industry has tools to effectively transform data to the SDTM structure. These tools should be able build a large portion of the appropriate ADaM datasets with maximum efficiency. The burning question: Can ADaM datasets be built with a mapping tool just like SDTM?

INTRODUCTION

The theme of this paper is to describe how standards can aid in the process of automating analysis datasets, to give programmers and biostatistician more time to focus on the science and unique analyses for new novel indications and treatments. Automated processes require the proper governance by sponsors internally, and through collaboration between Clinical Research Organizations and the Sponsors. The decisions made interpreting the standards need to be collected and documented using a Meta Data Repository (MDR) so that rigorous and consistent implementation can be assured. The MDR can provide consistent input to the processing of the data from collection to Tables, Figures and Listings (TFL's) which then are used by medical writers to create the Clinical Study Report (CSR).

Although the CDISC standards have evolved greatly, there is still a lot of room for interpretation by users of these standards. Below is a table reviewing the standards.

Differentiating the CDISC Standards

Category	CDASH	SDTM	ADaM
Purpose	Collection of data from eCRF in a consistent manner	Submission in standardized tabular form	Analysis data for use in Tables, Listings and Figures
Structure	Flexible	Rigid	Flexible *
Part of Submission package	No	Yes	Yes
Output(s)	CRF collection data in electronic form	Transformed data from collection into SDTM tabular format in electronic form.	Copied data from SDTM for trace-ability. Derived data as per derivations from SAP.

* In order to automate, must decide on standard structure within set of studies.

Large Pharmaceuticals often have legacy tabulation standards which they then must update to SDTM. Often they hire Clinical Research Organizations (CRO's) to handle this conversion process. The ADaM portion, being newer in

acceptance, has more flexibility and standards are in the process broadened and improved based on industry usage..

Raw collection to SDTM is handled by many CRO's with in-house mapping tools. These tools, along with machine readable metadata, can be used with a MDR to automate the process of SDTM and ADaM dataset creation. This process must be exact in its interpretations of standards from the various CDISC documents, the collected data and the sponsor interpretations of the standards (in this case the pharmaceutical company sponsoring the study).

ADaM datasets also are candidates for automation, as long as the metadata is collected and stored in machine-readable form. The data comes not only from the SDTM domains which are the building blocks of the Analysis Datasets, but also from the final product which are the Tables, Figures and Listings mocks/shells providing the supporting analyses for the Statistical Analysis Plan. These must be annotated and documented so that a retrospective process can begin. ADaM is therefore the missing link between SDTM and TFL's.

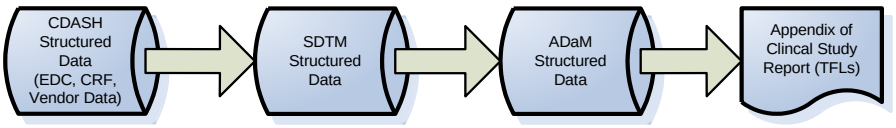
RETROSPECTIVE PROCESS

- **ret·ro·spec·tive**
- *adj.*
- **1.** Looking back on, contemplating, or directed to the past.
- **2.** Looking or directed backward.
- **3.** Applying to or influencing the past; retroactive.

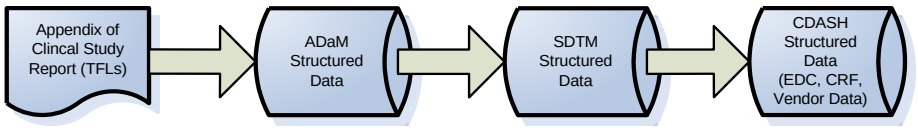
First and foremost, the analysis datasets must support the statistical analysis plan (SAP) by providing accurate efficacy and safety analyses. Part of the SAP is a set of mock tables. These can be annotated just like an eCRF with the ADaM-compliant variable names so that these can be derived or pulled in from the original collected data. This process is also being standardized by work groups within various organizations such as PhUse in developing standard analysis scripts by therapeutic area. These scripts intend to provide guidance on recommended Tables, Figures and Listings that are part of standard clinical submissions.

The retrospective process is described by the picture below. Traditionally, the ADaM datasets have been considered to be derived from the SDTM. The newer thinking is look through the Biostatistician's viewpoint and to look backwards from the end goal. The Biostatistician has input at the beginning as to what is collected as described in the protocol. Then the focus goes into the SAP, where the tables are mocked up and final hypotheses will be tested.

A Linear Approach to Standards



How we Should be Approaching the Standards



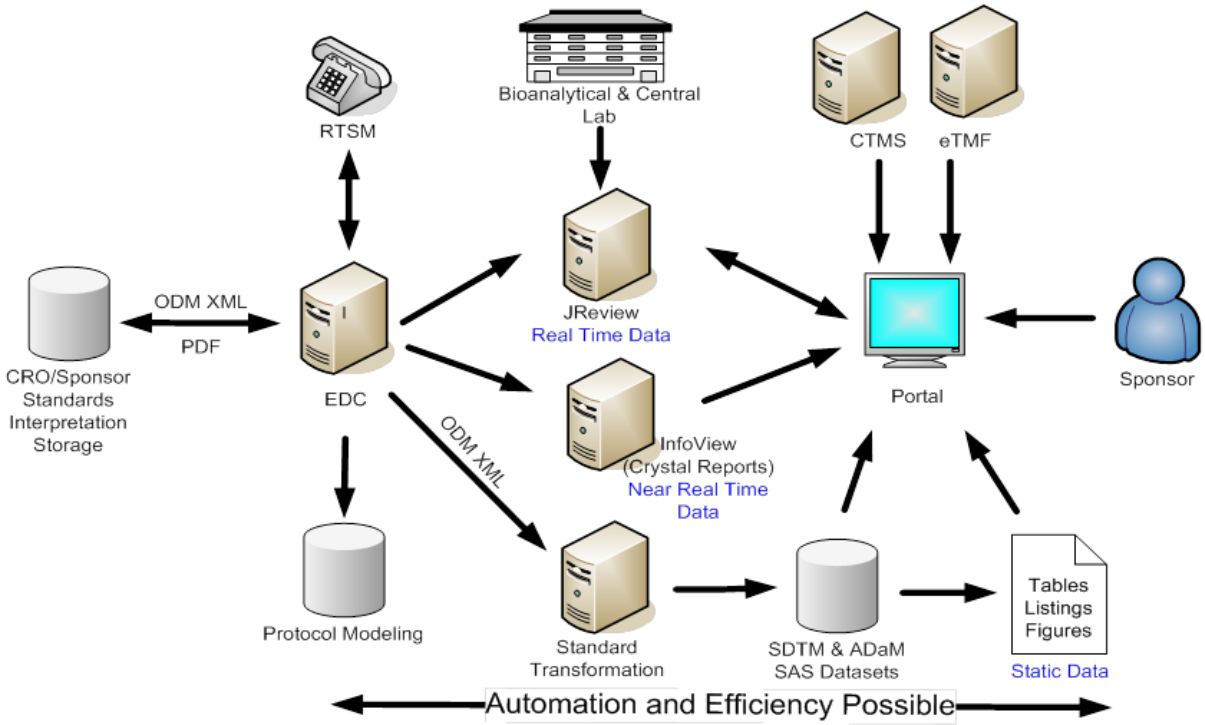
Building the MDR required for this approach includes the involvement of all the stakeholders along the process. Having the end in mind requires an integrated team approach from the following groups:

- Electronic Data Capture team (EDC)
- Data management (DM)
- Clinical Programming (CP)
- Biostatistics (Stats)

One of the ADaM premises, to be one 'proc' or step away from creating the tables and listings plays on this theme. Only by working backwards and literally 'annotating' the table with the required fields can we be sure to collect what is needed up front.

Metadata management is essential to automation. In the flowchart to the right we show various MDR stores that contribute to the automated process and reduce the amount of manual interaction to the key analyses and individualized endpoints that should always be a part of a scientific exploration.

In order for the process described to work we need the functional areas along the way to continuously ensure that standards are being adhered to and metadata is passed from one step to the next. Below is the big picture of the possibility with input and metadata:



PROSPECTIVE PROCESS

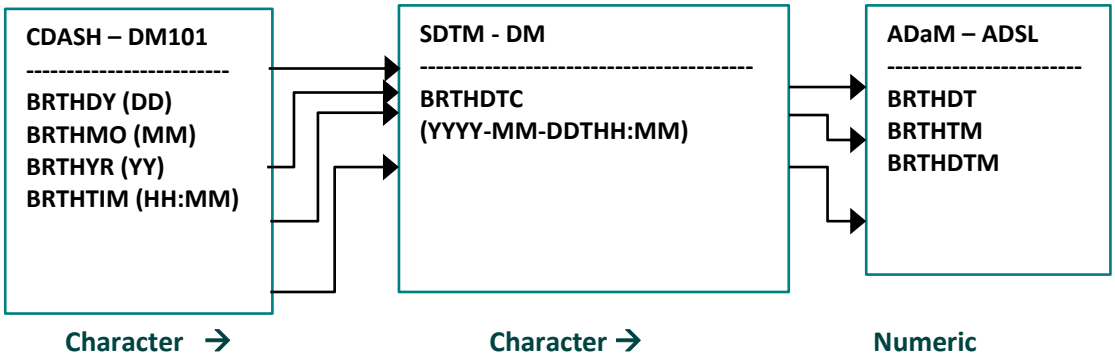
- **pro·spec·tive**
- *adj.*
- **1.** Likely or expected to happen.
- **2.** Likely to become or be

CDISC standards are the first component in the metadata build. These are available in machine readable form and read into the CRO/Sponsor Global Metadata Repository (GMDR). Any interpretations for individual custom forms, domains, and ADaM datasets are added to the GMDR. Since CDASH standards offer some flexibility, the Sponsor must have their interpretations clearly described in the metadata that is stored in the GMDR.

Data in the above model can be accessed at all levels of processing, whether it is at the collection point, the data reviewing by data management teams, or by programs that process and push the data to the next data store.

Machine consumable metadata in the form of detailed specifications, control terminology and requirements forms the basis of automation. Further automation is possible through ODM. In the future through ODM, we should be able to

automatically push additional data from CDASH to SDTM to ADaM. See example below. (here we collect time, although not usually collected unless a pediatric study).



When the system sees these fields in CDASH ending in DY, MO, YR and TIM it can then appropriately 'push' the data to SDTM ISO8601 format, and to ADaM numeric format as per information stored in the GMDR on all these types of fields. Due to standardization of naming conventions, the process is further automated for potential new date elements. The character date kept for traceability from SDTM to ADaM.

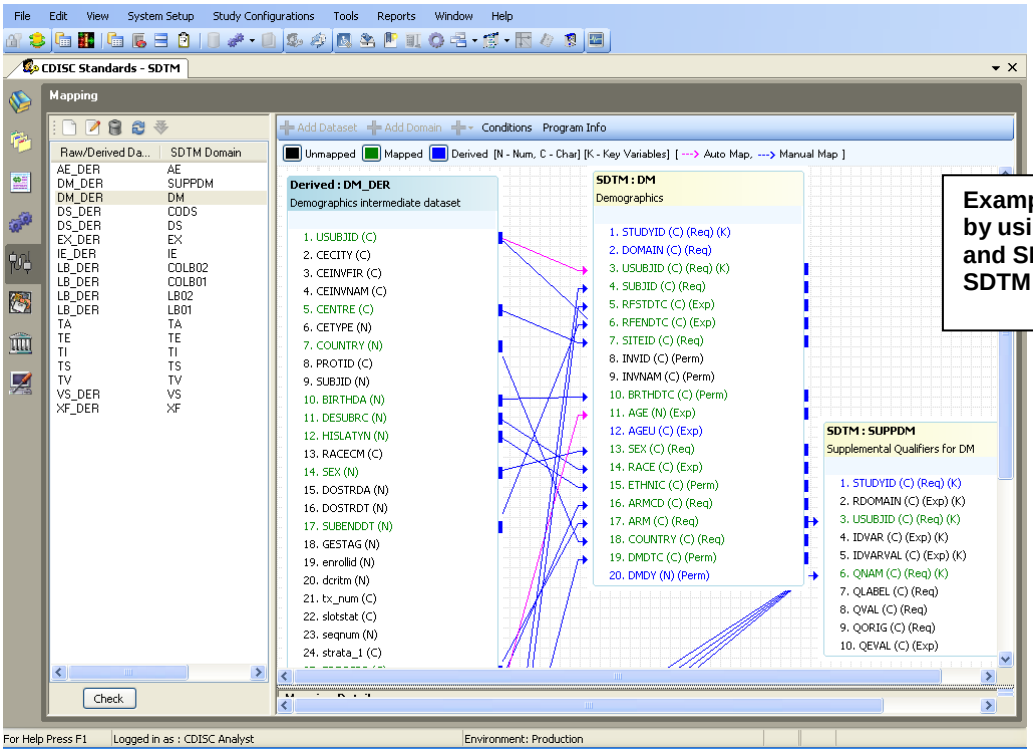
PRAGMATIC

- **prag·ma·tic**
- *adj.*
- **1.** Dealing or concerned with facts or actual occurrences; practical.
- **2.** *Philosophy* of or relating to pragmatism.

The third element is a pragmatic one, where one decides what portion of the process can be automated by tools. This part requires extensive governance by the sponsor and the CRO. It needs:

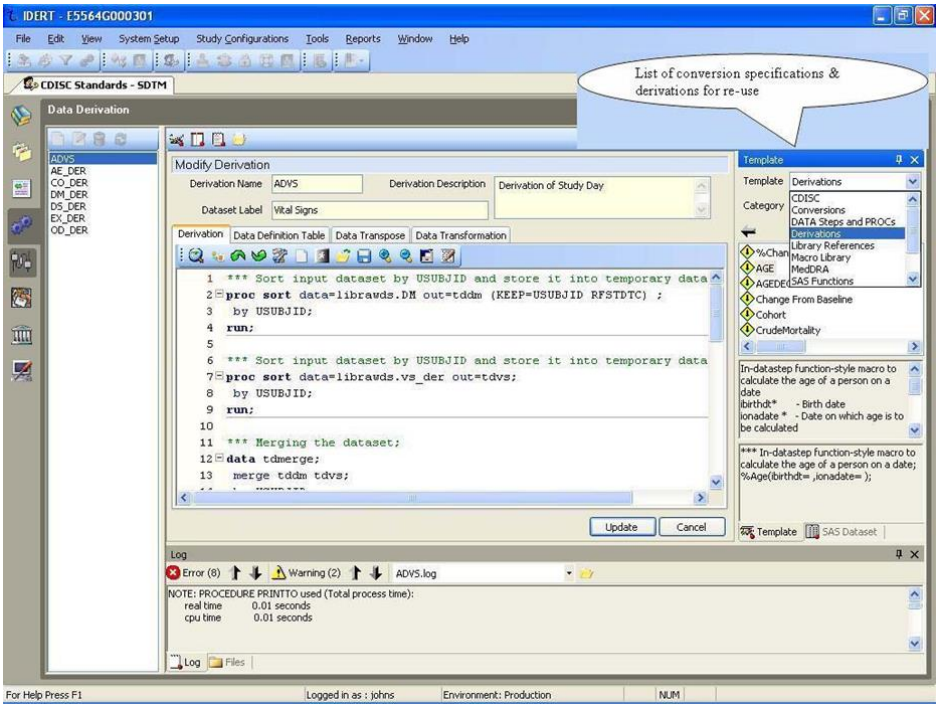
- company-wide standards implementation
- Sponsorship
- Standard content alone is not a solution without a tool or tools
- Jointly living the vision

An example of an actual implementation is given below. The tool below shows how SDTM and ADaM creation are being automated.



Example of Mapping by using an GMDR and SMDR to build SDTM Domains

Example of Implementing Standard Derivations to ADaM



The above two slides show a real live implementation of a tool built to take advantage of standards to facilitate mapping and conversion efforts. The idea of metadata driven systems is becoming more and more a reality, across both individual companies and also at CDISC with the current development of the SHARE metadata repository.

CONCLUSION

To automate the creation of ADaM datasets no single approach will get the desired result. A retrospective,

prospective, and pragmatic process must be used to automate the creation of ADaM datasets. ADaM datasets are the keys between SDTM domains and TFLs, which must support the statistical analysis plan by providing accurate efficacy and safety analyses.

Retrospectively, keeping the end in mind from the beginning with an integrated team approach will better influence systems earlier during data collection and SDTM conversion. For this to work, solid draft SAPs including annotated TFL shells must be completed before EDC systems are built. Strict adherence to CDASH and SDTM interpretations sets the building blocks for ADaM automation.

Prospectively for the process to be automated we need to produce machine consumable metadata in the form of an MDR and other tools that are able to push metadata form one standard to the other in a linear process. The use of ODM XML does have features that allow for information system interoperability so that hardware devices and software routines work harmoniously together.

Organizations must be pragmatic in not taking on too much at one time. For success, clear commitment and sponsorship from the leadership in the company is essential with funding for tool implementation and governance. Single groups cannot work in siloes; it takes company-wide standards implementation with longevity and perseverance over potentially many years to get to the point of automation. By adhering to the standards developed by the CDISC organization, we can get closer to the goal of automating as much of the process as possible, giving programmers and biostatisticians more time to focus on the science and unique analyses for new novel indications and treatments.

REFERENCES

www.Phuse.eu

www.CDISC.org

ACKNOWLEDGMENTS

The authors would like to thank PRA International for participating in the PhUSE initiatives in 2013.

CONTACT INFORMATION

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

Name:	Karin LaPann
Title:	Principal CDISC Standards Consultant
Enterprise:	PRA International
Address:	630 Dresher Road
City, State ZIP:	Horsham, PA 19044
Work Phone:	860-4348613
E-mail:	LaPannKarin@PRAIntl.com
Web:	www.praintl.com

Name:	Terek Peterson, MBA
Title:	Senior Director, Global Standards Strategies
Enterprise:	PRA International
Address:	630 Dresher Road
City, State ZIP:	Horsham, PA 19044
Work Phone:	215-444-8613
E-mail:	PetersonTerek@PRAIntl.com
Web:	www.praintl.com