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

Karin LaPann, MSIS, PRA International, Terek Peterson, MBA, PRA International

ABSTRACT

In the CDISC Standards, analysis datasets using the standards (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 poster is to describe how standards can aid in the process of automating analysis datasets, to give programmers and biostatisticians more time to focus on the science and unique analyses for new 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 assimilated and documented using a Metadata Repository (MDR) so that rigorous and consistent implementation can be assured. The MDR provides consistent input to the processing of the data from collection to Tables, Figures and Listings (TFL's) which are then 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. (see below:)

Category	CDASH	SDTM	ADaM
Purpose	**Collection of data from eCRF in a consistent manner**	**Submission data 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.

Pharmaceutical Companies often have legacy tabulation standards which they then update to SDTM. Often they hire Clinical Research Organizations (CRO's) to handle this conversion process. The ADaM standards, being newer in acceptance, have more flexibility.

Raw collection to SDTM is handled by many CRO's with in-house mapping tools. These tools, along with machine readable metadata, are 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.

ADaM datasets also are candidates for automation using an MDR to provide mappings from SDTM to ADaM. Additional metadata can be created from the TFLs and mock shells of the supporting analyses defined in the Statistical Analysis Plan. ADaM is therefore the missing link between SDTM and TFL's.

•ret·ro·spec·tive

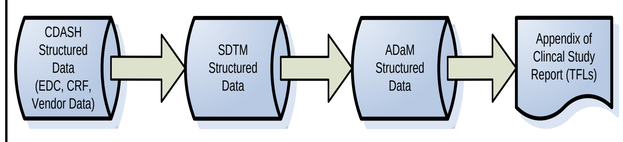
•*adj.*
•**1.** Looking back on, contemplating, or directed to the past.
•**2.** Looking or directed backward.
•**3.** Applying to or influencing the past; retroactive.

Retrospective Process

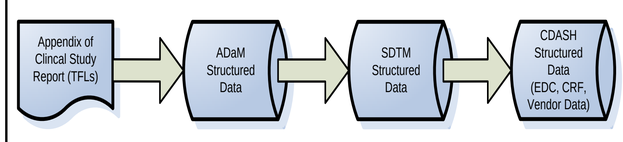
First and foremost, the analysis datasets must support the statistical analysis plan (SAP) by providing accurate efficacy and safety analyses. The Biostatistician creates the SAP for the study and this becomes a detailed roadmap for the clinical programmer to create any derived fields. Part of the SAP is a set of mock tables. These can be annotated just like an eCRF with ADaM-compliant variable names so that these can be derived or mapped from the SDTM 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 (TFLs) 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 domains without influence from the analysis. The newer method is to think from the biostatistician's viewpoint and 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



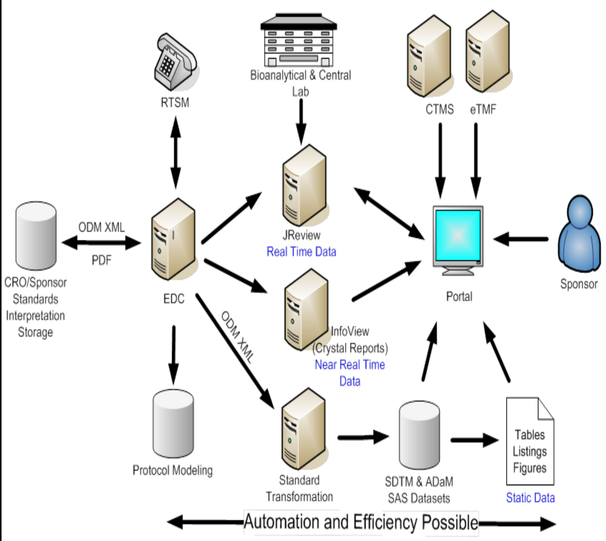
Building the MDR required for this approach includes the involvement of all the stakeholders along the process. Having the end in mind requires an integrated team approach from the following groups:

- 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 see how the MDR stores metadata that contribute to the automated processes from collection to conversion to analysis.

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 of automation using metadata:



•pro·spec·tive

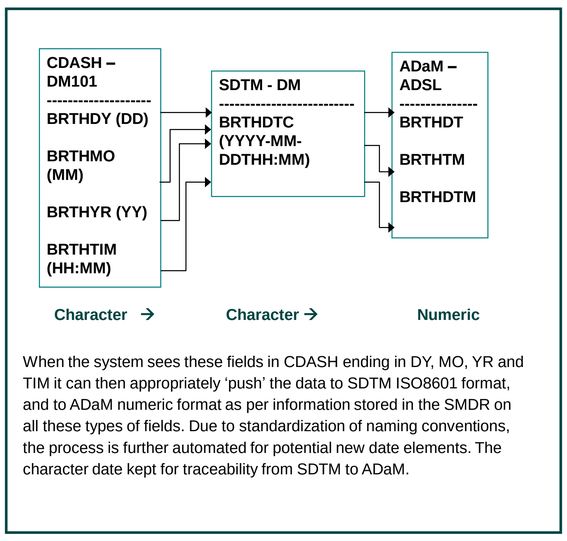
•*adj.*
•**1.** Likely or expected to happen.
•**2.** Likely to become or be.

Prospective Process

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 Study/Sponsor Metadata Repository (SMDR). Since CDASH standards offer some flexibility, the Sponsor must have their interpretations clearly described in the metadata that is stored in the SMDR.

Data in the above model can be accessed at all levels of processing, whether it is at the collection point, the data reviewed 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 XML. In the future through ODM XML, we should be able to automatically push additional metadata 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 SMDR 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.

•prag·mat·ic

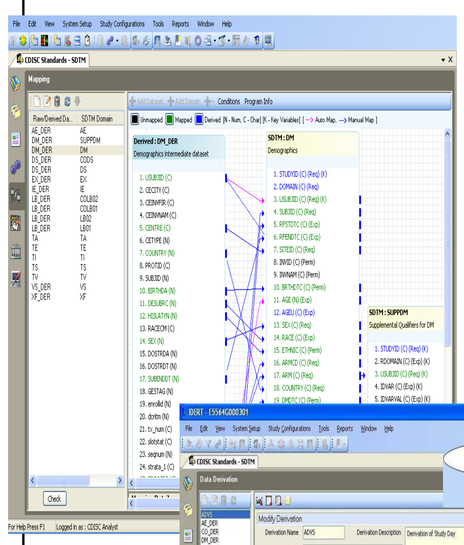
•*adj.*
•**1.** Dealing or concerned with facts or actual occurrences; practical.
•**2.** *Philosophy* Of or relating to pragmatism.

Pragmatic

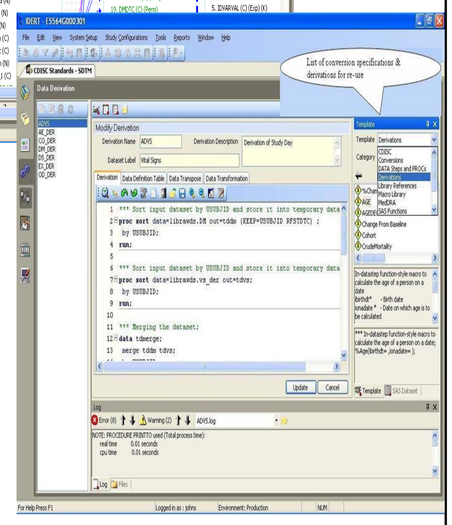
The third element is a pragmatic one, where one decides what portion of the process can be automated by tools. This part requires extensive governance by the sponsor and the CRO. It needs:

- Company-wide standards implementation
- Sponsorship
- Standard content alone is not a solution without a tool or tools
- Perseverance
- 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

CONCLUSION

To automate the creation of ADaM datasets no single approach will get the desired result. A combination of retrospective, prospective, and pragmatic processes must be used to successfully 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. 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.

For additional information please contact:

Karin LaPann
Principal Clinical Programmer, PRA International
lapannkarin@PRAIntl.com

Terek Peterson, MBA
Senior Director of Clinical Programming, PRA International
petersonterek@PRAIntl.com